A Model of the Data Economy

Maryam Farboodi* and Laura Veldkamp†

January 14, 2022

Abstract

Are the economics of the data economy new? There are surely increasing returns: Data helps firms grow; larger firms, generate more data. At the same time, data has diminishing returns: using more data to inform already precise estimates has low value. To explore how increasing and decreasing returns play out, we construct a dynamic, stochastic economy, where data is a long-lived asset. We learn that in the long run, diminishing returns dominate and dynamics are like a classical economy: Comparative advantage dictates who produces what, and, without additional data externalities, allocations are efficient. But in the short run, increasing returns dominate and new forces arise: Data-intensive goods or services are given away for free; many new firms are unprofitable; and some of the biggest firms profit primarily from selling data.

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

†Columbia Graduate School of Business, NBER, and CEPR, 3022 Broadway, New York, NY 10027; lv2405@columbia.edu.
Thanks to Rebekah Dix and Ran Liu for invaluable research assistance and to participants and discussants at the 2019 SED plenary as well as numerious research seminars and conferences for helpful comments and suggestions. Keywords: Data, growth, digital economy, data barter.

Does the data economy have new economics? In the information age, production increasingly revolves around information and specifically, data. Many firms, particularly the most valuable U.S. firms, are valued primarily for the data they have accumulated. Collection and use of data is as old as book-keeping. But recent innovations in data-intensive prediction technologies allow firms to use more data more efficiently. We have known since Wilson (1975) that ideas, information, data and other non-rival inputs have returns to scale. Because large firms benefit more from data, produce more data and grow bigger, data typically has increasing returns. At the same time, any data scientist will tell you that data has decreasing returns: Most of the predictive value comes from the first few observations. Understanding these opposing forces and what they mean for an economy requires constructing a new, dynamic equilibrium framework, with data as a state variable. Our model of the data economy teaches us that the long-run dynamics and welfare resemble an economy with capital accumulation and decreasing returns, with the twist that firms may specialize in selling goods or selling data. However, the short-run features new dynamics, like increasing returns, S-shaped growth, negative profits, and the barter of data for goods.

Most of the discussion about big data is about a particular type of digitized information: transaction-generated data, used by firms to optimize their business processes, by accurately predicting future outcomes. The hype about data economics arose because of breakthroughs in machine learning and artificial intelligence. These are prediction algorithms. They require enormous amounts of data, which are naturally generated by transactions: Buyer characteristics, traffic images, textual analysis of user reviews, click through data, and other evidence of economic activity. Predictions made with this big data are typically used for business process optimization, such as forecasting demand, costs, earnings, labor needs, advertising or selecting investments or product lines. While data can also be an input in the innovation process, big data in business does not typically mean research inputs, just like capital investment does not usually mean research facilities.

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

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

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

The use and sale of data is governed by a mostly familiar force – comparative advantage – but with non-rivalry playing a central role. Section 2.3 shows that in the long run, a data economy will feature specialization, if data is sufficiently non-rival. In such an economy, large firms that have a comparative advantage in data production derive most of their profit from data sales. The business model of these large firms is to do lots of transactions at a low price for goods and earn more revenue from data sales. While we know that many firms execute a strategy like this and the idea of comparative advantage is familiar, this is different from a capital accumulation economy: Unlike traditional goods production, the extent of specialization in a firm's comparative advantage is not complete. The extent of specialization also depends on non-rivalry and on the concentration of data processing talent.

Section 3 explores transition dynamics before a firm gets to steady state (the short run). When data is scarce, it may have increasing returns, because of a "data feedback loop." More data makes a firm more productive, which results in more production and transactions, which generate more data, further increasing productivity and data generation. This is the dominant force when data is scarce. Increasing returns also generate poverty traps. Firms with low levels of data earn low profits, which makes little production optimal. But little production generates little data, which keeps the firm poor. Firms may even choose to produce with negative profits. This rationalizes the commonly-observed practice of data barter, as a form of costly investment in data (Section 3.3).

Many digital services, like apps, which were costly to develop, are given away to customers at zero price. The exchange of customer data for a service, at a zero monetary price, is a classic barter trade.

To answer many regulatory questions about data, one needs a fully micro-founded model that speaks to social welfare. Section 4 adds microfoundations and finds that, despite the non-rivalry, the increasing returns, and the production of data as a by-product, equilibrium choices are efficient. However, this answer depends crucially on the assumption that all data is used for product quality improvement. In reality, many firms use data for marketing, to steal business from rivals. Privacy concerns are another source of externalities. Such negative externalities obviously incentivize excess data production, which inflates good production, in order to generate the data. Because of the non-rivalry of data, externalities prompt excessive data trade, a novel feature of data economies. By mapping the extent of data inefficiency to excess capital investment, this framework can guide more targeted regulation.

The primary contribution of the paper is not the particular predictions we explore. Some of those predictions are more obvious, some more specific to data economies. The larger contribution is a tool to think clearly about the economics of aggregate data accumulation. Because our tool is a simple one, many applications and extensions are possible. We explore some in the paper; others, such as imperfect competition or firm size dispersion, are discussed in the conclusion. While adding features to the main model could allow it to better address one question or another, keeping the main model simple allows the tool to be used in many different ways.

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

Related Literature. Work on information frictions in business cycles (Caplin and Leahy, 1994; Veldkamp, 2005; Lorenzoni, 2009; Ordonez, 2013; Ilut and Schneider, 2014; Fajgelbaum et al., 2017) have versions of a data-feedback loop that operate at the level of the aggregate economy: More data enables more aggregate production, which in turn, produces more data. The key difference is

that in those papers information is a public good, not a private asset. The private asset assumption in the current paper changes firms' incentives to produce data, allows data markets to exist and is what raises welfare concerns.

Work on media in the macroeconomy (Chahrour et al., 2019; Nimark and Pitschner, 2019) shares our focus on markets where information is bought and sold. Work on screening incentives in lending (Asriyan et al., 2021) also views data as a durable asset with value. However, in a data economy, transactions create more data. This feedback is absent in these other literatures and is essential for increasing returns.

Compared to the existing literature on data and growth, the key difference in our model is that data is information, used to forecast a random variable. In Jones and Tonetti (2018), Cong et al. (2021) and Cong et al. (2020), the focus is on exploring growth versus privacy. Their data contributes directly to productivity. It is not information with diminishing returns. Without diminishing returns, these models cannot explore the tension between diminishing and increasing returns that is at the heart of our analysis.

In models of learning-by-doing (Jovanovic and Nyarko, 1996; Oberfield and Venkateswaran, 2018) and organizational capital (Atkeson and Kehoe, 2005; Aghion et al., 2019), firms also accumulate a form of knowledge. But the economics differ. Unlike prediction data, this knowledge need not have long-run diminishing returns. Also, it is not a tradeable asset. Our increasing returns to data differs from growth models with increasing returns (Farmer and Benhabib, 1994), because those are based on positive spillovers between firms. Ours is a feedback loop within a firm.

Work exploring the interactions of data and innovation complements ours. For example, in Garicano and Rossi-Hansberg (2012), IT allows agents to accumulate more knowledge, which facilitates innovation. Agrawal et al. (2018) develop a combinatorial-based knowledge production function to explore how breakthroughs in AI could enhance discovery rates and economic growth.[1] Acemoglu and Restrepo (2018) explore similar questions to ours, but about the growth potential from robots. Although robots require data, they are rival capital goods, distinct from the data itself. Our work analyzes big data and new prediction algorithms, in the absence of technological change. Once we understand this foundation, one can layer these insights about data, innovation and automation on

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

top.

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

Finally, the insight that the stock of knowledge can serve as a state variable comes from the five-equation toy model sketched in Farboodi et al. (2019). That was a partial-equilibrium numerical exercise, designed to explore the size of firms with heterogeneous data. This paper builds an aggregate equilibrium model, with goods markets, data markets, data non-rivalry, analytical solutions and welfare analysis. These new dimensions to this model fundamentally shape the answers to our main questions about aggregate dynamics and long-run outcomes.

1 A Data Economy

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

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

1.1 Model

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

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

The quality of a good depends on a firm's choice of a production technique $a_{i,t}$. Each period firm i has one optimal technique, with a persistent and a transitory components: $\theta_t + \epsilon_{a,i,t}$. Neither component is separately observed. The persistent component θ_t follows an AR(1) process: $\theta_t = \bar{\theta} + \rho(\theta_{t-1} - \bar{\theta}) + \eta_t$. The AR(1) innovation $\eta_t \sim N(0, \sigma_\theta^2)$ is $i.i.d.$ across time.[2] Firms have a noisy prior about the realization of θ_0. The transitory shock $\epsilon_{a,i,t} \sim N(0, \sigma_a^2)$ is $i.i.d.$ across time and firms and is unlearnable.

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

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

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

Data The role of data is that it helps firms to choose better production techniques. One interpretation is that data can inform a firm whether blue or green cars or white or brown kitchens will be more valued by their consumers, and produce or advertise accordingly. In other words, a technique could represent a resource allocation. Transactions help to reveal customers' marginal values, but these values are constantly changing. Firms must continually learn to catch up. Another interpretation is that the technique is inventory management, or other cost-saving activities. Ob-

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

serving production and sales processes at work provides useful information for optimizing business practices. For now, we model data as welfare-enhancing. Section 4 relaxes that assumption.

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

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

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

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

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

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

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

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

Of course, data is non-rival. Some firms use data and also sell that same data to others. If there

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

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

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

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

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

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

Data adjustment costs capture the idea that a firm that does not store or analyze any data cannot freely transform itself to a big-data machine learning powerhouse. That transformation

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

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.

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

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

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

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

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

$$\begin{aligned} P_t &= \bar{P} Y_t^{-\gamma}, \\ Y_t &= \int_i A_{i,t} k_{i,t}^{\alpha} di. \end{aligned} \tag{7}$$

At each time t, the output is deterministic. Appendix A shows that, because there are infinitely many firms with independent signals and a noisy prior, independent forecast errors imply independence in $A_{i,t}$'s. But the central limit theorem, the aggregate or average A converges to a known value. Therefore, price is also deterministic.

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

1.2 Solution

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

Optimal Technique and Expected Quality Taking a first order condition with respect to the technique choice, we find that the optimal technique is $a^*_{i,t} = \mathbb{E}_i[\theta_t|\mathcal{I}_{i,t}]$. Thus, expected quality of firm i's good at time t in (2) can be rewritten as $\mathbb{E}[A_{i,t}] = E\left[g\left((\mathbb{E}_i[\theta_t|\mathcal{I}_{i,t}] - \theta_t - \epsilon_{a,i,t})^2\right)\right]$. The squared term is a squared forecast error. It's expected value is a conditional variance, of $\theta_t + \epsilon_{a,i,t}$. That conditional variance is denoted $\Omega^{-1}_{i,t} + \sigma^2_a$.

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

$$\mathbb{E}[A_{i,t}|\mathcal{I}_{i,t}] \approx g\left(\Omega^{-1}_{i,t} + \sigma^2_a\right) + g''\left(\Omega^{-1}_{i,t} + \sigma^2_a\right) \cdot \left(\Omega^{-1}_{i,t} + \sigma^2_a\right). \tag{8}$$

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

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

The Stock of Knowledge Since the stock of knowledge $\Omega_{i,t}$ is the sufficient statistic to keep track of information and its expected utility, we need a way to update or keep track of how much

of this stock there is. Lemma 1 is just an application of Bayes' law, or equivalently, a modified Kalman filter, that tell us how the stock of knowledge evolves from one period to the next.

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

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

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

The inflows of data are new pieces of data that are generated by economic activity. The number of new data points $n_{i,t}$ is assumed to be data mining ability times end of period physical output: $z_i k_{i,t}^{\alpha}$. By Bayes' law for normal variables, the total precision of that information is the sum of the precisions of all the data points: $n_{i,t}\sigma_\epsilon^{-2}$. The term σ_a^{-2} in (9) is the additional information learned from seeing one's own realization of quality $A_{i,t}$, at the end of period t. That information also gets added to the stock of knowledge. At the firm level, we need to keep track of whether a firm buys or sells data. Thus the newly added stock of data $n_{i,t}$ has to be adjustd for data trade. That is the role of the indicator functions at the end of (9).

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

How does data flow out or depreciate? Data depreciates because data generated at time t is about next period's optimal technique θ_{t+1}. That means that data generated s periods ago is about θ_{t-s+1}. Since θ is an AR(1) process, it is constantly evolving. Data from many periods ago, about a θ realized many periods ago, is not as relevant as more recent data. So, just like capital, data

depreciates. Mathematically, the depreciated amount of data carried forward from period t is the first term of (9): $\left[(\rho^2(\Omega_{i,t}+\sigma_a^{-2}))^{-1}+\sigma_\theta^2\right]^{-1}$. The $\Omega_{i,t}+\sigma_a^{-2}$ term represents the stock of knowledge at the start of time t plus the information about period t technique revealed to a firm by observing its own output. This stock of knowledge is multiplied by the persistence of the AR(1) process squared, ρ^2. If the process for optimal technique θ_t was perfectly persistent then $\rho = 1$ and this term would not discount old data. If the θ process is i.i.d. $\rho = 0$, then old data is irrelevant for the future. Next, the formula says to invert the precision, to get a variance and add the variance of the AR(1) process innovation σ_θ^2. This represents the idea that volatile θ innovations make knowledge about past θ's less relevant. Finally, the whole expression is inverted again so that the variance is transformed back into a precision. This precision represents a (discounted) stock of knowledge. The depreciation of knowledge is the period-t stock of knowledge, minus the discounted stock.

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

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

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

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

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

$$V(\Omega_{i,t}) = \max_{k_{i,t},\delta_{i,t}} P_t\mathbb{E}[A_{i,t}|\mathcal{I}_{i,t}]k_{i,t}^\alpha - \Psi(\Delta\Omega_{i,t+1}) - \pi_t\delta_{i,t} - rk_{i,t} + \left(\frac{1}{1+r}\right)V(\Omega_{i,t+1}) \tag{12}$$

where $\mathbb{E}[A_{i,t}|\mathcal{I}_{i,t}]$ is an increasing function of $\Omega_{i,t}$, given by (8), $n_{i,t} = z_i k_{i,t}^\alpha$, and the law of motion for $\Omega_{i,t}$ is given by (9).

This result greatly simplifies the problem by collapsing it to a deterministic problem with choice variables k and δ and one state variable, $\Omega_{i,t}$. In expressing the problem this way, we have already substituted in the optimal choice of production technique. The quality $A_{i,t}$ that results from the

optimal technique depends on the conditional variance of θ_t. Because the information structure is similar to that of a Kalman filter, that sequence of conditional variances is deterministic.

The non-rivalry of data acts like a kinked price of data, or a negative transactions cost in (9).[5]

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

2 Long-Run Features of a Data Economy

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

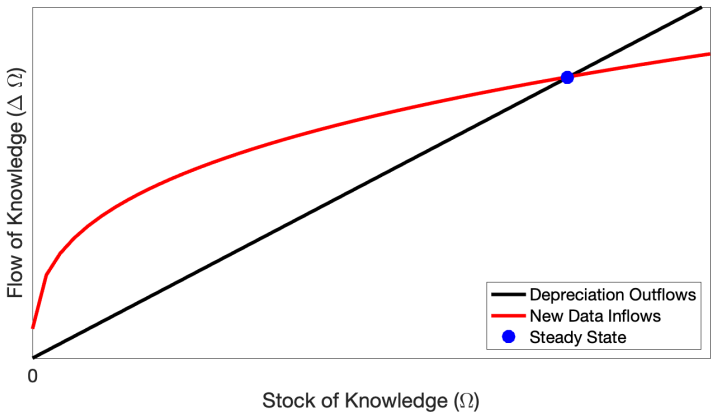

Figure 1: Economy converges to a data steady state: Aggregate inflows and outflows of data.
Line labeled inflows plots the quantity in (10) for the aggregate economy, for different levels of initial data stock and optimal capital choices. Line labeled outflows plots the quantity in (11). This is equivalent to the outflow and inflow for a representative firm i who operates in an economy populated with identical firms with no trade. In all examples, we adopt a linear quality function $g(z) = g(0) - z$ and a quadratic data adjustment cost: $\Psi(\Delta\Omega_{i,t+1}) = \psi(\Delta\Omega_{i,t+1})^2$, where ψ is a constant parameter and Δ represents the percentage change: $\Delta\Omega_{i,t+1} = (\Omega_{i,t+1} - \Omega_{i,t})/\Omega_{i,t}$.

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 the distance between inflows and outflows diminishing, which represents growth slowing. Eventually, inflows and outflows cross at the steady state. If the stock of knowledge ever reached its steady state level, it would no longer change, as inflows and outflows just balance each other. Likewise, quality and GDP would stop growing.

One difference between data and capital accumulation is the nature and form of depreciation. In the Solow model of capital accumulation, depreciation is a fixed fraction of the capital stock, always linear. In the data accumulation model, depreciation is not linear, but is very close to linear. Lemma 5 in the appendix shows that depreciation is approximately linear in the stock of

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

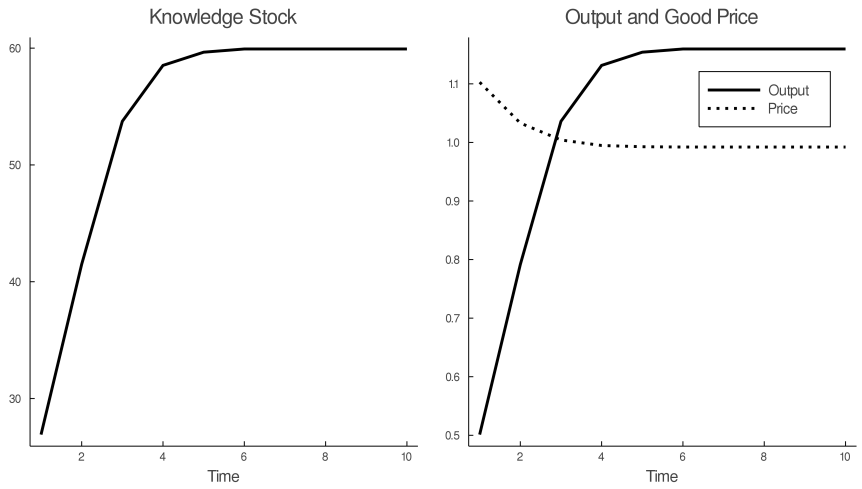

Figure 2: Aggregate growth dynamics: Data accumulation grows knowledge and output over time, with diminishing returns. Parameters: $\rho = 1, r = 0.2, \beta = 0.97, \alpha = 0.3, \psi = 0.4, \gamma = 0.1, \overline{A} = 1, \overline{P} = 1, \sigma_a^2 = 0.05, \sigma_\theta^2 = 0.5, \sigma_\epsilon^2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

knowledge, with an error bound that depends primarily on the variance of the innovation in θ.

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

Conceptually, diminishing returns arise because we model data as information, not as an addition to productivity. Information has diminishing returns because its ability to reduce variance gets smaller and smaller as beliefs become more precise. Forecast errors can, at best, be zero. Mathematically, diminishing returns comes from two distinct and independent sources: the finite-valued quality function and unlearnable risk. The next set of results explain why either feature bounds the growth from data.

Long Run Growth Impossibility Results Can data accumulation sustain growth in an economy without innovation? For sustained growth to be possible, two things must both be true: 1) Perfect one-period-ahead foresight implies infinite real output; and 2) the future is a deterministic function of today's observable data.[6] While empirical studies support the idea of decreasing

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

returns to data (Bajari et al., 2018), it is not possible to prove the nature of a production function theoretically. What theory can do is tell us that if we believe data can sustain long-run growth, this logically implies other economic properties. In this case, long-run growth implies two conditions that are at odds with most theories. If a researcher does not believe either property to be true, they must then believe data-induced growth, without innovation, cannot be sustained.

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

Definition 1 (Sustainable Growth) *Let $Y_t = \int_i \mathbb{E}[A_{i,t}]k_{i,t}^{\alpha}di$, such that $ln(Y_{t+1}) - ln(Y_t)$ is the aggregate growth rate of expected output. A data economy can sustain a minimum growth rate $\underline{g} > 0$ if $\exists\, T$ such that in each period $t > T$, $ln(Y_{t+1}) - ln(Y_t) > \underline{g}$.*

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

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

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

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

Fundamental randomness means future events that are not deterministic functions of observable events today. If they are not deterministic functions of something that can be observed today, then no signal can perfectly predict these future events. In other words, fundamentally random variables are not perfectly learnable. In our model, fundamental randomness or unlearnable risk is present when $\sigma_a^2 > 0$.

Proposition 2 ***Data-Driven Growth Implies No Fundamental Randomness*** *Suppose the quality function g is finite almost everywhere, except $g(0) \to \infty$. Sustainable growth requires that productivity-relevant variables (θ_t and $\varepsilon_{a,i,t}$) have no time-$(t-1)$ fundamental randomness.*

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

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

Thus, if one believes some events tomorrow are fundamentally random, then even if perfect precision can potentially generate infinite output, data will still have diminishing returns. Conversely, even if one believes that nothing is truly random, but they believe that with one-period ahead knowledge, an economy can only produce the finite amount today that they would otherwise produce tomorrow, then data must also have diminishing returns. For process-optimizing data, without technological innovation, to produced sustained growth, one must embrace both the infinite output

and no-fundamental-randomness assumptions. Keep in mind that this analysis holds technology fixed. This technology includes advances in prediction technology. So to sustain growth, without technological advance, what is required is that both perfect one-period-ahead forecasts enable infinite production, with current production technology, and that current prediction technology can, with sufficient data, achieve one-period-ahead forecasts with zero prediction error.

2.2 Endogenous Growth

Every result hinges on its assumptions. In this case, long-run stagnation comes from the assumption that transactions data is used for process optimization, as opposed to technological innovation. If we relax this assumption, we find that data can sustain growth, under the same circumstances as those in which capital can sustain growth. If data (or capital) is used for research and development, the constant addition to the stock of new ideas can sustain growth. This extension connects our main framework, which is a data economy version of Solow (1956), to a data economy version of endogenous growth with quality ladders, in the spirit of Grossman and Helpman (1991), Aghion and Howitt (1992) or Garicano and Rossi-Hansberg (2012). Of course, for this extension to make sense, one needs to believe that information about who buys what can be used to discover growth-sustaining technologies.

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

$$A_{i,t} = A_{i,t-1} + \max\{0, \Delta A_{i,t}\}, \tag{13}$$

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

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

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

$\mathbb{E}_i[\Delta A_{i,t}] = \bar{A} - \Omega_{i,t}^{-1} - \sigma_a^2$.

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

But in a model where physical capital is an essential ingredient in research, the same would be true of capital. The conclusion then is that, in the long run, the economic forces of the data economy are not very new at all.

2.3 Specialization in Data Production

There are two possible ways a high data-productivity firm might profit. First, it could retain the data, to make high-quality goods, to sell at a high price. Such firms are specialized in the production of high-quality goods. Alternatively, they could sell off most of their data and produce low-quality goods. Their goods would earn little or even no revenue. But their data sales would earn profits. We say that such a firm specializes in data production or data services.

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

From here on, we return to the quality Equation (2), without endogenous growth and adopt a linear quality function $g(z) = g(0) - z$, for simplicity.

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

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

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

When there is a single, high data-productivity (H) firm in a market populated by L-firms ($\lambda = 1$), the steady state outcome is what is intuitively expected. Lemma 3 in the appendix shows that the data-productive H firm accumulates more knowledge ($\Upsilon^{ss} > 0$). The positive knowledge gap implies that the data-productive firm is larger and produces high-quality goods.

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

Proposition 3 ***High Data-Productivity Firms Accumulate Less Knowledge*** *Suppose that there is a strictly positive measure of high data-productivity firms, $\lambda < 1$. If $\alpha < \frac{1}{2}$ and γ is sufficiently small, then there exist $\bar{\iota}$ and $\bar{\iota}_2$ such that*

a. *if data is sufficiently non-rival, $\iota < \bar{\iota}$, the steady state knowledge gap is negative: $\Upsilon^{ss} < 0$;*

b. *if high data-productivity firms are sufficiently productive $z_H > \underline{z}_H$, and if data is sufficiently non-rival $\iota < \bar{\iota}_2$, then greater efficiency widens the knowledge gap: $\frac{d\Upsilon^{ss}}{dz_H} < 0$.*

Proposition 3 is a specific illustration of specialization due to comparative advantage when firms operate in two different markets. The familiar intuition is that when a firm, in this case an H-firm, has a comparative advantage in a market, the data market, it will specialize in that market. This intuition holds true in the model independent of ι: H-firms are more data-productive. They are always the sellers in the data market and make profits there. The counter-intuitive outcome is that if and only if data is sufficiently non-rival, the optimal degree of specialization in the data market is at the cost of producing low quality goods in the product market ($\Upsilon^{ss} < 0$).

This is only possible when non-rivalry of data allows H-firms to sell a lot of data and still keep enough of it to achieve a reasonable quality, albeit low, on the good market, which in turn enables

them to produce a lot of goods and generate a lot of data to sell on the data market going forward. On the other hand, when data is rival, $\iota = 1$, this strategy is not optimal for H-firms anymore, and they choose to be the higher quality producers in the good market as well ($\Upsilon^{ss} > 0$). This outcome is more similar to what one expects in models of learning-by-doing, or intangible capital, which are deemed as rival factors of production.

Data non-rivalry acts like a negative bid-ask spread in the data market. It drives a wedge between the value of the data sold and the opportunity cost, the amount of data lost through the act of selling. While a bid-ask spread typically involves some loss from exchange of an asset, with non-rivalry, exchanging data results in more total data being owned. If the buyer pays a price π per unit of data gained, the seller earns more than π per unit of data forfeited, because they forfeit only a fraction of the data sold. This negative spread, or subsidy, on transactions incentives data producers to be prolific sellers of data. The incentive to sell data can be so great that these data producers are left with little data for themselves.

Comparing Proposition 3 and the single high data-productivity firm case raises the question: How does a positive mass of high-data-productivity firms cause the result to change sign? The key is that the single, measure-zero H-firm cannot influence the amount of data held by the continuum of L-firms. The knowledge gap falls in Proposition 3, not because H-firms lose knowledge but because L-firms gain knowledge. That gain cannot happen when there is a single, measure-zero H-firm because that one firm is simply not large enough to provide data for all L-firms.

Since many economists and policy makers are concerned about concentration in data markets, we also explore what happens to data specialization when the the data market is more concentrated. We interpret λ close to 1, where there is a small measure of high data-productivity firms, as being data market concentration. The numerical example in Figure 3 illustrates realistic features of data specialization. Since data is multi-use (non-rival), the knowledge gap is negative. As a result, high data-productivity firms earn more of their profits from data sales. Low data-productivity firms earn negative data profits because they are data purchasers. The divergence of profit lines in Figure 3 illustrates how data market concentration amplifies the specialization of high data-productivity firms. While different relationships are possible, which one arises hinges largely on the question of rivalry: How much value does one firm's data lose when it is sold to another firm?

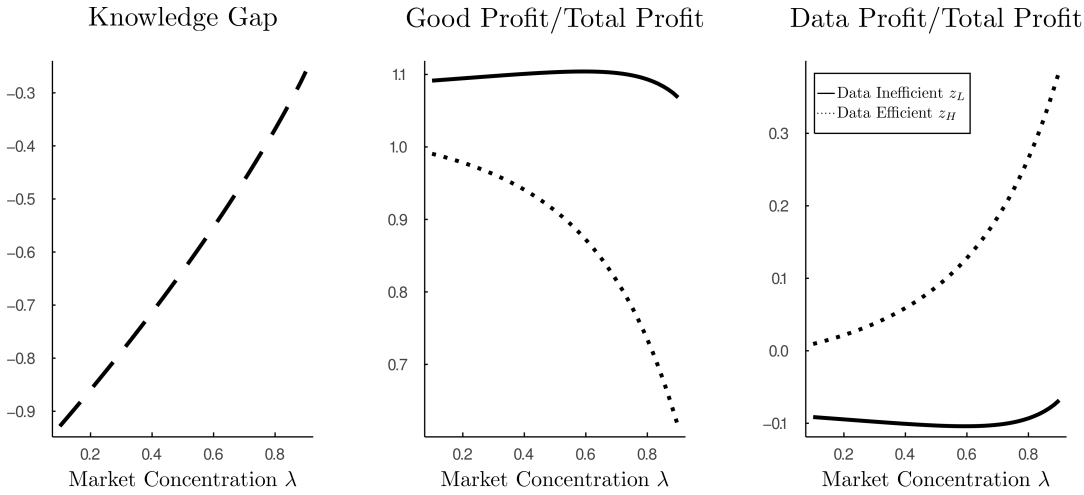

Figure 3: Data market concentration (λ) causes large (H) firms to derive most profits from data. Data market concentration is one minus the fraction of high data-productivity (H) firms. Parameters: $\rho = 1, r = 0.2, \beta = 0.97, \alpha = 0.3, \gamma = 0.09, \overline{A} = 1, \overline{P} = 0.5, \sigma_a^2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

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

3 Short Run Features of a Data Economy

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

3.1 Increasing Returns in the Short Run

While the previous results focused on diminishing returns, the other force at work is increasing returns. Increasing returns arise from the data feedback loop: A firm with more data produces higher-quality goods, which induces them to invest more, produce more, and sell more. This, in turn, generates more data for them. That feedback causes aggregate knowledge accumulation to accelerate. The feedback loop competes against diminishing returns. Diminishing returns always dominate when data is abundant; the previous results about the long run were unambiguous. But when firms are young, or data is scarce, increasing returns can be strong enough to create an increasing rate of growth. While that sounds positive, it also creates the possibility of a firm growth trap, with very slow growth, early on the in the lifecycle of a new firm.

While we have been talking about symmetric firms, we now relax the symmetry assumption. The next set of results is about one firm growing, while all others are in steady state. Then, we drop in one, atomless, low-data (low $\Omega_{i,t}$) firm and observe its transition. From this exercise, we learn about barriers to new firm entrants.

Before stating the formal result, we need to define net data flow. Recall that aggregate data inflows Ω_t^+ are the total precision of all new data points at t (eq. 10). Aggregate data outflows Ω_t^- are the end-of-period-t stock of knowledge minus the discounted stock (eq. 11). Aggregate data flows are the difference $d\Omega_t = d\Omega_t^+ - d\Omega_t^-$. At the individual level, data flows are defined using the firm-i specific version of Equations (10)-(11), which incorporates data trade: $d\Omega_{i,t} = d\Omega_{i,t}^+ - d\Omega_{i,t}^-$. From here on, we also adopt a linear quality function, for simplicity: $g(x) = \bar{A} - x$.

Proposition 4 ***S-Shaped Accumulation of Knowledge*** *When all firms are in steady state, except for one firm i, then the firm's net data flow $d\Omega_{i,t}$*

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

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

The difference between one firm entering when all other firms are in steady state (Proposition 4), and all firms growing together (Propositions 1 and 2), is prices. When all firms are data-poor,

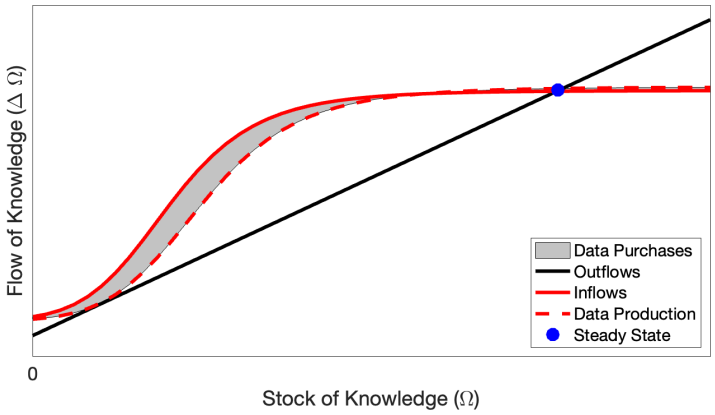

Figure 4: A single new firm grows slowly: Inflows and outflows of one firm's data.
Line labeled inflows plots the individual firm i version of the data inflows in Equation (10). Line labeled outflows plots the individual firm i version of the quantity in (11). Firm i is in an economy where all other firms are in steady state.

all goods are low quality. Since quality units are scarce, prices are high. The high price of goods induces these firms to produce goods, creating data. When the single firm enters, others are already data-rich. Quality goods are abundant, so prices are low. This makes it costlier and slower for the single firm to grow. What works in the opposite direction is that data may also be abundant, keeping the price of data low.

Figure 4 illustrates the inflows, outflows and dynamics of a single firm. This differs from Figure 1 because it represents a single firm's transition, not the transition of a whole economy of symmtric firms, growing together. This figure illustrates one possible economy. Data production may lie above or below the data outflow line. For some parameter values, the diminishing returns to data is always stronger than the data feedback loop. Proposition 9 in the appendix shows that, when learnable risk is abundant, knowledge accumulation is concave. In such cases, each firm's trajectory looks like the concave aggregate path in Figure 2. But when parameters make the equilibrium price force sufficiently strong, the increasing returns of the data feedback loop overwhelms diminishing returns, to make data inflows convex at low levels of knowledge.

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

The quality-adjusted production path of a single, growing firm mimics the path of its stock of knowledge. The difference between the S-shaped inflows and nearly linear outflows in Figure 4 traces out the S-shaped output path of a new entrant firm in this environment. We explore

characteristics of new data firms next.

3.2 New Firms' Profits, Book Value and Market Value

In a data economy, the trajectory of a single firm's profits, book value and market value are quite different from those in an economy driven by capital accumulation. Since empirical evidence on profits, book value and market value are easily available, it is useful to explore the model's predictions along these dimensions. In doing so, we relate to the literature on using Tobin's Q to measure intangible capital.

In a standard model, a young, capital-poor firm has a high marginal productivity of capital. The firm offers high returns to its owners and has a book and market value that differ only by the capital adjustment cost. In a data economy, data scarcity makes a young firm's quality and profits low. In fact, there is a range of parameters for which young firms cannot possibly make positive initial profits. Start by defining a firm's profit:

$$\text{Profit}_t = P_t A_{i,t} k_{i,t}^{\alpha} - \Psi(\Delta\Omega_{i,t+1}) - \pi_t \delta_{i,t} - r k_{i,t}. \tag{14}$$

Proposition 5 ***A Single New Firm Loses Money*** *Assume that $g(\sigma_a^2 + \sigma_\theta^2) < 0$. Then for a firm entering with zero data, $\Omega_{i,0} = \sigma_\theta^{-2}$, the firm cannot make positive expected profit at any period t unless it has made strictly negative expected profit at some $t' < t$.*

The reason such a firm produces, even though producing loses money, is that production generates data, which has future value to the firm. This firm is doing costly experimentation. This is like a bandit problem. There is value in taking risky, negative expected value actions because they generate information. Such behavior is also called active experimentation. Production at time t is like paying to generate information, which will allow the firm to be profitable in the future. The reason that the firm's production looses money is that if $g(\sigma_a^2 + \sigma_\theta^2) < 0$, the initial expected quality of the firm's good is too low to earn a profit. But production in one period generates information for the next, which raises the average quality of the firm's goods, and enables future profits.

The idea that data unlocks future firm value implies that in order to increase its stock of knowledge, a new firm both produces low quality goods to self-produce data, and buys some data on the data market, as depicted in Figure 4. The two mechanisms of building stock of knowledge

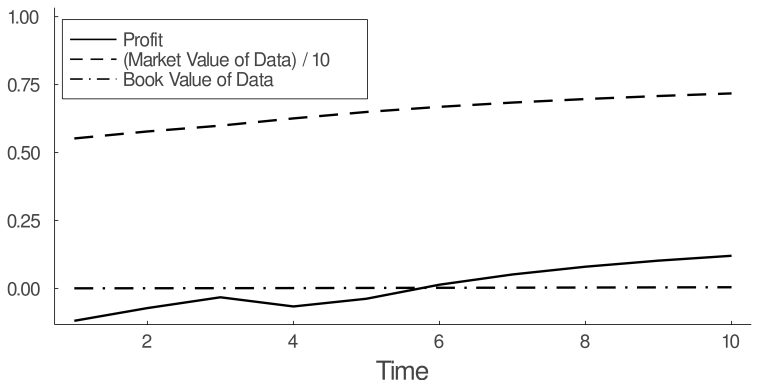

Figure 5: S-shaped growth can create initial profit losses and high market-to-book value of data. Knowledge stock defined in Lemma 1. Book value defined in (15). Parameters: $\rho = 1, r = 0.2, \beta = 0.97, \alpha = 0.3, \psi = 4, \overline{A} = 0.5, \sigma_a^2 = 0.05, \sigma_\theta^2 = 0.5, \sigma_\epsilon^2 = 0.1, z = 0.01, \pi = 0.002, P = 1, \iota = 1$

lead to a discrepancy between a firm's book value and market value. It is so because accounting rules do not allow a firm's book value to include data, unless that data was purchased. In the context of our simple model, the firm rents but does not own any capital, and data is the firm's only asset. Therefore, we define the firm book value to be the discounted value of all purchased data. The indicator function $\mathbf{1}_{\delta_{i,t}>0}$ captures only data purchases, not self-produced data. If we equate the book value depreciation rate to the household's rate of time preference β, then[7]

$$\text{Book Value}_t = \sum_{\tau=0}^{t} \beta^{t-\tau} \pi_t \delta_t \mathbf{1}_{\delta_{i,\tau}>0}. \tag{15}$$

The market value of the data is the Bellman equation value function $V(\Omega)$ in (12).

Figure 5 plots the market value (divided by 10 to facilitate visualization), book value and profits of a young firm, over time. The difference between the market and book value of a firm is used to measure intangible assets. The high market value and low book value, given firm characteristics, are typically interpreted as a large intangible asset stock. Our firms exhibit this hallmark. The negative profits described in Proposition 5, representing costly experimentation, also show up here. This result connects our model to work measuring intangible capital as a gap between market and book values, as well as to work exploring financial barriers to firm entry.

[7]Accounting rules depreciate data and software assets, by amortizing them over three years. That amounts to accounting depreciation or 30% or $\beta = 0.7$ per year.

3.3 Data Barter

Data barter arises when goods are exchanged for customer data, at a zero price. While this is a knife-edge possibility in this model, it is an interesting outcome because it illustrates a phenomenon we see in reality. In many cases, digital products, like apps, are being developed at great cost to a company and then given away "for free." Free here means zero monetary price. But obtaining the app does involve giving one's data in return. That sort of exchange, with no monetary price attached, is a classic barter trade. The possibility of barter is not shocking, given the assumptions. But the result demonstrates the plausibility of the framework, by showing how it speaks to data-specific phenomena we see.

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

At $P_t = 0$, the marginal benefit of investment is additional data that can be sold tomorrow, at price π_{t+1}. If the price of data is sufficiently high, and/or the firm is a sufficiently productive data producer (high z_i), then the firm should engage in costly production, even at a zero goods price, to generate the accompanying data.

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

Our framework allows us to assign a value to such barter trades, despite their zero monetary price. In practice, a whole segment of the economy is not being captured by traditional GDP measures because the transactions price misses the value of data being paid. A framework that captures this value is a first step toward better measurement of aggregate economic activity.

4 Welfare and Data Externalities

Discussions of data regulation abound. Optimal policy depends on what aspects of a data economy are efficient or inefficient. Our framework can be used for welfare analysis, but lacks an important

consideration: Data is not always used for a socially productive purpose. Firms can use data to steal customers away from other firms. This section models the household side of the economy and add a business stealing externality, to examine the welfare properties of a data economy.

4.1 A Micro-founded Model for Welfare Analysis

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

Households There is a continuum of homogeneous infinitely lived households, with quasi-linear preferences over consumption of the retail good c_t and the numeraire good m_t. The representative household's optimization problem is

$$\begin{aligned} &\max_{c_t, m_t} \sum_{t=0}^{+\infty} \frac{u(c_t) + m_t}{(1+r)^t} \\ &\text{s.t.} \quad P_t c_t + m_t = \Phi_t \qquad \forall t \end{aligned} \tag{16}$$

Households have CRRA utility for retail good consumption: $u(c_t) = \bar{P}\frac{c_t^{1-\gamma}}{1-\gamma}$. The household budget constraint equates the expenditure on the two consumption goods to household income, which is aggregate firm profits Φ_t. Since aggregate output is non-random, as argued earlier, aggregate profits and the household's optimization problem are also not random in each period t.

Retail Good Production The producers of the retail goods live forever. They use capital, rented at a constant exogenous cost r, trade data, and produce the retail good using their capital and data. There are two types of retail firms. They are identical, except for their, z_i, the efficiency with which they convert produced units into data. We consider a measure λ of low data-productivity firms with $z_i = z_L$, and a measure $(1-\lambda)$ of high data-productivity firms with $z_i = z_H$, where $z_L < z_H$.

Profit is revenue minus adjustment costs, minus data costs (if $\delta > 0$) or plus revenue from data sales (if $\delta < 0$), minus the cost of capital, $\Phi_{it} := P_t A_{i,t} k_{i,t}^{\alpha} - \Psi(\Delta\Omega_{i,t+1}) - \pi\delta_{i,t} - rk_{i,t}$. The profit

the households get is the aggregate firm profit,

$$\Phi_t = \int \Phi_{it} di = P_t \int_i A_{i,t} k_{i,t}^{\alpha} di - \int_i -\Psi(\Delta\Omega_{i,t+1}) di - r \int_i k_{i,t} di,$$

Firms maximize the expected present discounted value of their profit:

$$\max_{\{k_{i,t},\delta_{i,t}\}_{t=0}^{\infty}} V(\Omega_{i,0}) = \sum_{t=0}^{+\infty} \frac{1}{(1+r)^t} \left(P_t \mathbb{E}[A_{i,t}|\mathcal{I}_{i,t}] k_{i,t}^{\alpha} - \Psi(\Delta\Omega_{i,t+1}) - \pi\delta_{i,t} - r k_{i,t} \right). \tag{17}$$

Data governs the expected quality of goods, $\mathbb{E}[A_{i,t}]$, described by Equations (5) and (8). Law of motion for data is expressed in Equation (9).

The retail sector represents an industry where consumption and data are industry-specific, but capital is rented from an inter-industry market, at rate r, paid in units of numeraire.[8]

Market Clearing conditions are also the resource constraints in the planner problem:

$$\begin{aligned}
&\text{retail good :} && c_t = \lambda A_{L,t} k_{L,t}^{\alpha} + (1-\lambda) A_{H,t} k_{H,t}^{\alpha}, \\
&\text{numeraire good :} && m_t + r\left(\lambda k_{L,t} + (1-\lambda) k_{H,t}\right) + \Big(\lambda\Psi(\Delta\Omega_{L,t+1}) + (1-\lambda)\Psi(\Delta\Omega_{H,t+1})\Big) = 0 \\
&\text{data :} && \lambda\delta_{L,t} + (1-\lambda)\delta_{H,t} = 0.
\end{aligned}$$

The micro foundations make very little difference for our conclusions so far. They deliver the same inverse demand as in (7). But these foundations allow us to compare the decentralized equilibrium and optimal social planner outcomes. That comparison reveals if inefficiencies in this economy arise.

Proposition 7 ***Welfare*** *The steady state allocation is socially efficient.*

Equilibrium capital investment and data production are efficient because there are no externalities. The constraint, that data may only be produced through the production of goods, is a constraint that is faced both by the planner and the firm. Prices of goods and data reflect their

[8] Equivalently, we can interpret this as a small, open economy where capital and numeraire goods are tradeable and retail goods are non-tradeable. The world rental rate of capital is r. This simplification puts the focus on data. An endogenously determined rental rate of capital would increase when firms are more productive. This would create a wealth effect for capital owners. These equilibrium effects are well-studied in previous frameworks, but are not related to economics of data.

marginal social value. This aligns the private and social incentives for production. Consistent with the previous results, we find that the steady state is similar to a capital-accumulation economy.

4.2 Data for Business Stealing

Policy analysis needs to consider potential data externalities. For example, when data can be used for marketing, advertising or other forms of business stealing, firms' use of data harms others. Privacy costs can also take the form of a non-pecuniary externality. In the presence of such an externality, firms' choices will obviously be socially inefficient. By incorporating such an externality in a tractable way, our framework can provide guidance about measuring the extent of the data inefficiency.

Using data for business stealing can be represented through quality:

$$A_{i,t} = \bar{A} - \left[\left(a_{i,t} - \theta_t - \epsilon_{a,i,t} \right)^2 - b \int_{j=0}^{1} \left(a_{j,t} - \theta_t - \epsilon_{a,j,t} \right)^2 dj \right] \qquad \text{for } b \in [0,1] \tag{18}$$

The baseline model is represented by $b = 0$, i.e. Equations (18) and (2) are identical. In this case, firm's quality solely depends on the precision of its own prediction of the optimal technique, and there is no externality as shown in Proposition 7.

Alternatively, if $b = 1$, firm's quality depends on the different between the precision of its own prediction and the average precision of the predictions of all other firms. In other words, each firm's maximum quality can be attained when there is maximum difference between the optimal technique and the chosen technique by every other firm. This captures the idea that when one firm uses data to reduce the distance between their chosen technique $a_{i,t}$ and the optimal technique $\theta + \epsilon$, that firm benefits, but all other firms lose.

$b = 1$ captures the extreme case where data does not have any social value, the losses from business stealing entirely cancel out the productivity gains from data: $\int A_{i,t} di = \bar{A}$. Moving b between 0 and 1 regulates the extent to which data enhances welfare.

Proposition 8 ***Welfare with Business Stealing*** *There is over-investment in the steady state level of capital and excessive trade in the data market in equilibrium.*

Proposition 8 points out to two distinct inefficiencies: there is excessive production, and there

is excessive data trade. First, notice that the business stealing externality does not change firms' choices because it does not enter in a firm's first order condition.[9] Therefore, it does not change data inflows, outflows, data sales or accumulation, and capital choices, at a given set of prices. However, it does influence the aggregate qualities, which leads to two externalities.

The first externality is a result of too much data sales. When selling data to another firm j, firm i does not internalize that selling data increases j's quality, which decreases quality of i's good and all other goods. Because each firm is measure-zero, they neglect their effect on aggregates and only consider the positive profits generated by selling their data. This leads to too much data trade.

The second externality is a result ofover-invest in capital, production generates data. While firms internalize the benefits of this additional data, they neglect the external effect, through which high data reduces the quality of other firms' goods. Thus, in equilibrium, too much output is produced and too much data is traded.

5 Conclusions

The economics of transactions data bears some resemblance to technology and some to capital. It is not identical to either. Data has the diminishing returns of capital, in the long run. But it has the increasing returns of ideas and technologies, early in the transition path to steady state. 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.

We conclude with future research possibilities that our framework could enable.

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

Firm competition. Instead of the price taking behavior that we assume, one can model a finite number of firms that consider the price impact of their production decisions. In such a setting, the

[9] To see why this is the case, note that firm i's actions have a negligible effect on the average productivity term $\int_{j=0}^{1} \left(a_{j,t} - \theta_t - \epsilon_{a,j,t}\right)^2 dj$. So the derivative of that new externality term with respect to i's choice variables is zero. If the term is zero in the first order condition, it means it has no effect on choices of the firm. This formulation of the externality is inspired by Morris and Shin (2002).

degree of imperfect competition will modulate firms' use of data. At the same time, firms' data will affect the extent to which they compete. Working this out in a dynamic, recursive setting like this one, could give us insights about how data changes firms' competitive strategies.

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

Data and product portfolio choice. We examined a model about how much data a firm produces and accumulates. Just as important a question is what type of data that is. Appendix B.2 sketches a model where different goods can be informative about different risks. In such a world, firms could invest in a portfolio of products to diversify and learn some about each or could specialize to become expert in producing one good with high quality. The choices, in turn, would shape the forces of market competition.

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

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

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

References

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

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

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

_ , Antonin Bergeaud, Timo Boppart, Peter J Klenow, and Huiyu Li, "A theory of falling growth and rising rents," Technical Report, National Bureau of Economic Research 2019.

_ , 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.

Asriyan, Vladimir, Luc Laeven, and Alberto Martin, "Collateral Booms and Information Depletion," *Review of Economic Studies*, 2021, *forthcoming.*

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

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

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

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

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

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

Chahrour, Ryan, Kristoffer Nimark, and Stefan Pitschner, "Sectoral media focus and aggregate fluctuations," 2019. Working Paper, Boston College.

Cong, Lin William, Danxia Xie, and Longtian Zhang, "Knowledge Accumulation, Privacy, and Growth in a Data Economy," *Management Science*, 2021, *67* (10), 5969–6627.

Cong, Lin William, Wenshi Wei, Danxia Xie, and Longtian Zhang, "Endogenous Growth Under Multiple Uses of Data," 2020.

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

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

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

Garicano, Luis and Esteban Rossi-Hansberg, "Organizing growth," *Journal of Economic Theory*, 2012, *147* (2), 623–656.

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

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

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

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

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

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

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

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

Nimark, Kristoffer P. and Stefan Pitschner, "News media and delegated information choice," *Journal of Economic Theory*, 2019, *181*, 160–196.

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

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

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

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

Wilson, Robert, "Informational Economies of Scale," *Bell Journal of Economics*, 1975, *6*, 184–95.

A Appendix: Derivations and Proofs. Not For Publication.

A.1 Proof of Lemma 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.

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

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

$$s^{a}_{i,t-1} = \theta_{t-1} + \epsilon_{a,t-1}, \tag{19}$$

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

The second type of signal the firm observes is data points. They are a by-product of economic activity. Here, we introduce a new piece of notation for brevity. It is the number of new data points added to the firm's data set. $\omega_{i,t}$. For firms that do not trade data, this is $\omega_{i,t} = n_{i,t} = zk^{\alpha}_{i,t}$. More generally, for all firms, the number of new data points depends on the amount of data traded:

$$\omega_{i,t} := n_{i,t} + \delta_{i,t}(\mathbf{1}_{\delta_{i,t}>0} + \iota \mathbf{1}_{\delta_{i,t}<0}).$$

The set of signals $\{s_{t,m}\}_{m\in[1:\omega_{i,t}]}$ are informationally equivalent to a single average signal $\bar{s}_t$ such that:

$$\bar{s}_t = \theta_t + \epsilon_{s,t}, \tag{20}$$

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

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

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

At time, t, the firm takes as given:

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

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

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

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

Then, use the equations above as prior beliefs and use Bayes law to update them with the new signals $\bar{s}_t$:

$$\widehat{\mu}_t = \mathbb{E}\big[\theta_t \mid \mathcal{I}_t\big] = \frac{\Big[\rho^2 \big[\Omega_{t-1} + \sigma_a^{-2}\big]^{-1} + \sigma_\theta^2\Big]^{-1} \cdot \mathbb{E}\big[\theta_t \mid \mathcal{I}_{t-1}, s^a_{i,t-1}\big] + \omega_t \sigma_\epsilon^{-2} \bar{s}_t}{\Big[\rho^2 \big[\Omega_{t-1} + \sigma_a^{-2}\big]^{-1} + \sigma_\theta^2\Big]^{-1} + \omega_t \sigma_\epsilon^{-2}} \tag{21}$$

$$\Omega_t^{-1} = Var\big[\theta_t \mid \mathcal{I}_t\big] = \Big\{\Big[\rho^2 \big[\Omega_{t-1} + \sigma_a^{-2}\big]^{-1} + \sigma_\theta^2\Big]^{-1} + \omega_t \sigma_\epsilon^{-2}\Big\}^{-1} \tag{22}$$

Multiply and divide Equation (21) by Ω_t^{-1} as defined in Equation (22) to get

$$\widehat{\mu}_{i,t} = (1 - \omega_t \sigma_\epsilon^{-2} \Omega_t^{-1}) \left[\bar{\theta}(1-\rho) + \rho\left((1 - M_t)\mu_{t-1} + M_t s^a_{i,t-1}\right)\right] + \omega_t \sigma_\epsilon^{-2} \Omega_t^{-1} \bar{s}_t, \tag{23}$$

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

Equations (22) and (23) constitute the Kalman filter describing the firm dynamic information problem. Importantly, note that Ω_t is deterministic.

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

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:

$$\max \sum_{t=0}^{\infty} \left(\frac{1}{1+r}\right)^t \mathbb{E}\left[P_t A_t k_t^\alpha - \Psi(\Delta\Omega_{i,t+1}) - \pi_t \delta_{i,t} - r k_t | \mathcal{I}_{i,t}\right]$$

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

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

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

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

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:

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

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

$$V(\{s_{t,m}\}_{m\in[1:\omega_t]}, y_{t-1}, \hat{\mu}_{t-1}, \Omega_{t-1}) = \max_{k_t} \mathbb{E}\Big[P_t g\big((\mathbb{E}[\theta_t|I_{i,t}] - \theta_t - \epsilon_{a,t})^2\big) k_t^{\alpha} - \Psi(\Delta\Omega_{i,t+1}) - \pi_t \delta_{i,t} - r k_t + \left(\frac{1}{1+r}\right) V(\{s_{t+1,m}\}_{m\in[1:\omega_{t+1}]}, y_t, \hat{\mu}_t, \Omega_t) | \mathcal{I}_{i,t}\Big].$$

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

$$\mathbb{E}\Big[\mathbb{E}[\theta_t|I_t] - \theta_t - \epsilon_{a,t})^2\Big] = \mathbb{E}\Big[(\mathbb{E}[\theta_t|I_{i,t}] - \theta_t)^2\Big] + \sigma_a^2 = \Omega_{i,t}^{-1} + \sigma_a^2$$

$\mathbb{E}[(\mathbb{E}[\theta_t|I_t] - \theta_t)^2 | I_{i,t}]$ is the time-t conditional (posterior) variance of θ_t, and the posterior variance of beliefs is $\mathbb{E}[(\mathbb{E}[\theta_t|I_t] - \theta_t)^2] := \Omega_t^{-1}$. Expected productivity determines the within period expected payoff, which using Equation (8) depends on posterior variance. The posterior variance Ω_t^{-1} is given by the Kalman system Equation (22), which depends only on Ω_{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:

$$\begin{aligned} V(\Omega_t) &= \max_{k_t} P_t \mathbb{E}[A_{i,t}|\mathcal{I}_{i,t}] k_t^{\alpha} - \Psi(\Delta\Omega_{i,t+1}) - \pi_t \delta_{i,t} - r k_t + \left(\frac{1}{1+r}\right) V(\Omega_{t+1})\Big] \\ \text{s.t.} \quad \Omega_{t+1} &= \left[\rho^2(\Omega_t + \sigma_a^2)^{-1} + \sigma_\theta^2\right]^{-1} + \omega_{i,t}\sigma_\epsilon^{-2} \end{aligned}$$

Data use is incorporated in the stock of knowledge through (9), which still represents one state variable.

A.3 Equilibrium and Steady State Without Trade in Data

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

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

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

$$k_{i,t} = \left[\frac{\alpha}{r}\left(P_t A_{i,t} + z_i \sigma_\epsilon^{-2}\left(\left(\frac{1}{1+r}\right) V'(\cdot) - 2\psi(\cdot)\right)\right)\right]^{1/(1-\alpha)} \tag{24}$$

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

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

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

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

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

$$\omega_{ss}\sigma_\epsilon^{-2} + \left[\rho^2(\Omega_{ss} + \sigma_a^{-2})^{-1} + \sigma_\theta^2\right]^{-1} = \Omega_{ss} \tag{25}$$

This is equating the inflows of data $\omega_{i,t}\sigma_\epsilon^{-2}$ with the outflows of data $\left[\rho^2(\Omega_{i,t} + \sigma_a^{-2})^{-1} + \sigma_\theta^2\right]^{-1} - \Omega_{i,t}$. Given a number of new data points ω_{ss}, this pins down the steady state stock of data. The number of data points depends on the steady state level of capital. The steady state level of capital is given by (24) for A_{ss} depending on Ω_{ss} and a steady state level of $V'(\Omega_{ss})$. We use the term V'_{ss} to refer to the partial derivative $\partial V/\partial \Omega$, evaluated at the steady state value of Ω. We solve for that steady state marginal value of data next.

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

$$V'_{ss} = \left[1 - \left(\frac{1}{1+r}\right) \frac{\partial \Omega_{t+1}}{\partial \Omega_t}|_{ss}\right]^{-1} P_t k_{ss}^{\alpha} \Omega_{ss}^{-2} \tag{26}$$

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

A.4 Equilibrium With Trade in Data

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

$$\begin{aligned} V(\Omega_{i,t}) = \max_{k_{i,t},\omega_{i,t}} \; & P_t\left(\bar{A} - \Omega_{i,t}^{-1} - \sigma_a^2\right) k_{i,t}^{\alpha} - \pi\left(\frac{\omega_{i,t} - z_i k_{i,t}^{\alpha}}{\mathbf{1}_{\omega_{i,t} > n_{i,t}} + \iota \mathbf{1}_{\omega_{i,t} < n_{i,t}}}\right) - r k_{i,t} \\ & - \Psi\left(\Delta\Omega_{i,t+1}\right) + \left(\frac{1}{1+r}\right) V(\Omega_{i,t+1}) \end{aligned} \tag{27}$$

$$\text{where} \quad \Omega_{i,t+1} = \left[\rho^2(\Omega_{i,t} + \sigma_a^{-2})^{-1} + \sigma_\theta^2\right]^{-1} + \omega_{i,t}\sigma_\epsilon^{-2} \tag{28}$$

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

$$FOC[k_{i,t}]: \quad \alpha P_t A_{i,t} k_{i,t}^{\alpha-1} + \frac{\pi \alpha z_i k_{i,t}^{\alpha-1}}{\mathbf{1}_{\omega_{i,t}>n_{i,t}} + \iota \mathbf{1}_{\omega_{i,t}<n_{i,t}}} - r = 0 \tag{29}$$

Solving for $k_{i,t}$ gives

$$k_{i,t} = \left(\frac{1}{r} \left(\alpha P_t A_{i,t} + \tilde{\pi} \alpha z_i \right) \right)^{\frac{1}{1-\alpha}} \tag{30}$$

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

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

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

$$FOC[\omega_{i,t}]: \quad -\Psi'(\cdot) \frac{\partial \Delta \Omega_{i,t+1}}{\partial \omega_{i,t}} - \tilde{\pi} + \left(\frac{1}{1+r} \right) V'(\Omega_{i,t+1}) \frac{\partial \Omega_{i,t+1}}{\partial \omega_{i,t}} = 0 \tag{31}$$

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

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

$$\omega_{ss} \sigma_\epsilon^{-2} + \left[\rho^2 (\Omega_{ss} + \sigma_a^{-2})^{-1} + \sigma_\theta^2 \right]^{-1} = \Omega_{ss} \tag{32}$$

This is equating the inflows of data $\omega_{i,t} \sigma_\epsilon^{-2}$ with the outflows of data $\left[\rho^2 (\Omega_{i,t} + \sigma_a^{-2})^{-1} + \sigma_\theta^2 \right]^{-1} - \Omega_{i,t}$. Given a number of new data points ω_{ss}, this pins down the steady state stock of data. The number of data points depends on the steady state level of capital. The steady state level of capital is given by Equation 30 for A_{ss} depending on Ω_{ss} and a steady state level of V'_{ss}. We solve for that steady state marginal value of data next.

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

$$V'_{ss} = \left[1 - \left(\frac{1}{1+r} \right) \frac{\partial \Omega_{t+1}}{\partial \Omega_t} |_{ss} \right]^{-1} P_{ss} k_{ss}^{\alpha} \Omega_{ss}^{-2} \tag{33}$$

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

From the first order condition for $\omega_{i,t}$ (eq 31), the steady state marginal value is given by

$$V'_{ss} = (1+r) \tilde{\pi} \sigma_\epsilon^2 \tag{34}$$

The Equations (30), (31), (32) and‘(33) form a system of 4 equations in 4 unknowns. The solution to this system delivers the steady state levels of capital, knowledge, data, and marginal value data.

A.4.1 Characterization of Firm Optimization Problem in Steady State

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

Individual Optimization Problem:

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

where i denotes the firm data productivity.

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

$$\Omega_L = \left(\rho^2(\Omega_L+\sigma_a^{-2})^{-1}+\sigma_\theta^2\right)^{-1} + \left(z_L k_L^{\alpha} + \delta_L\right)\sigma_\epsilon^{-2} \tag{35}$$

$$\Omega_H = \left(\rho^2(\Omega_H+\sigma_a^{-2})^{-1}+\sigma_\theta^2\right)^{-1} + \left(z_H k_H^{\alpha} + \iota\delta_H\right)\sigma_\epsilon^{-2} \tag{36}$$

$$\alpha P(\bar{A} - \Omega_L^{-1} - \sigma_a^2)k_L^{\alpha-1} + \pi\alpha z_L k_L^{\alpha-1} = r \tag{37}$$

$$\alpha P(\bar{A} - \Omega_H^{-1} - \sigma_a^2)k_H^{\alpha-1} + \frac{\pi\alpha z_H k_H^{\alpha-1}}{\iota} = r \tag{38}$$

$$P\sigma_\epsilon^{-2}k_L^{\alpha} = \pi\Omega_L^2\left(1 + r - \frac{\rho^2}{\left(\rho^2+\sigma_\theta^2(\Omega_L+\sigma_a^{-2})\right)^2}\right) \tag{39}$$

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

$$P = \bar{P}\left(\lambda(\bar{A} - \Omega_L^{-1} - \sigma_a^2)k_L^{\alpha} + (1-\lambda)(\bar{A} - \Omega_H^{-1} - \sigma_a^2)k_H^{\alpha}\right)^{-\gamma} \tag{41}$$

$$\lambda\delta_L + (1-\lambda)\delta_H = 0 \tag{42}$$

A.5 Proof of Proposition 1: Data Must Enable Infinite Output to Sustain Growth

Suppose not. Then, for every firm $i \in I$, with $\int_{i\notin I} di = 0$, producing infinite data $n_{i,t} \to \infty$ implies finite firm output $y_{i,t} < \infty$. Thus $M_y \equiv \sup_i\{y_{i,t}\} + 1$ exists and is finite. By definition, $y_{i,t} < M_y$, $\forall i$. If the measure of all firms is also finite, that is $\exists 0 < N < \infty$ such that $\int_i di < N$. As a result, the aggregate output is also finite in any period $t+s$, $\forall s > 0$:

$$Y_{t+s} = \int_i y_{i,t} di < M_y \int_i di < M_y N < \infty. \tag{43}$$

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

$$\ln(Y_{t+s}) - \ln(Y_t) = [\ln(Y_{t+s}) - \ln(Y_{t+s-1})] + \cdots + [\ln(Y_{t+1}) - \ln(Y_t)] > \underline{g}s, \tag{44}$$

which implies

$$Y_{t+s} > Y_t e^{\underline{g}s}. \tag{45}$$

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

$$Y_{t+s} > Y_t e^{\underline{g}s} > Y_t e^{\underline{g}\underline{s}} > Y_t e^{\underline{g}\frac{\ln(M_y N)-\ln(Y_t)}{\underline{g}}} = M_y N, \tag{46}$$

which contradicts (43).

A.6 Proof of Proposition 2: Data-Driven Growth Implies a Deterministic Future

We break this result into two parts. Part (a) of the result is that in order to have infinite output in the limit, an economy will need $(a_{i,t} - \theta_t - \epsilon_{a,i,t})^2$ to approach zero.

Part (b) says: For $(a_{i,t} - \theta_t - \epsilon_{a,i,t})^2$ to approach zero, marginal utility relevant variables θ_t and $\epsilon_{a,i,t}$ must be in the set Ξ_{t-1}.

Proof part a: From proposition 1, we know that sustaining aggregate growth above any lower bound $\underline{g} > 0$ arises only if a data economy achieves infinite output $Y_t \to \infty$ when some firm has infinite data $n_{i,t} \to \infty$. Since Y_t is a finite-valued function, except at 0, infinite output requires that the argument of g, which is $(a_{i,t} - \theta_t - \epsilon_{a,i,t})^2$ becomes arbitrarilty close to zero.

Proof of part b. Suppose not. The optimal action that can achieve infinite output when g is not finite-valued is $a_t^* = \theta_t + \epsilon_{a,i,t}$. If the optimal action is not in Ξ_{t-1}, then it is not a t-measurable action. There is some unforecastable error such that $\mathbb{E}[(a_{i,t} - \theta_t - \epsilon_{a,i,t})^2] > \underline{z} > 0$.

If it is not a measurable action, it cannot be chosen with strictly positive probability in a continuous action space. Since the optimal action must be in Ξ_{t-1}, then $\theta_t + \epsilon_{a,i,t}$ must be in Ξ_{t-1} as well. Since θ_t and $\epsilon_{a,i,t}$ are unconditionally and conditionally independent, for the sum to be perfectly predictable, each element must also be perfectly predictable. Thus, θ_t and $\epsilon_{a,i,t}$ must be in Ξ_{t-1}.

A.7 Proof of Lemma 3: Knowledge Gap When High Data Productivity is Scarce

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

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

$$\Omega_L = \left(\rho^2(\Omega_L + \sigma_a^{-2})^{-1} + \sigma_\theta^2\right)^{-1} + z_L k_L^\alpha \sigma_\epsilon^{-2} \tag{47}$$

$$\alpha P(\bar{A} - \Omega_L^{-1} - \sigma_a^2)k_L^{\alpha-1} + \pi\alpha z_L k_L^{\alpha-1} = r \tag{48}$$

$$P\sigma_\epsilon^{-2} k_L^\alpha = \pi\Omega_L^2 \left(1 + r - \frac{\rho^2}{\left(\rho^2 + \sigma_\theta^2(\Omega_L + \sigma_a^{-2})\right)^2}\right) \tag{49}$$

$$P = \bar{P}\left((\bar{A} - \Omega_L^{-1} - \sigma_a^2)k_L^\alpha\right)^{-\gamma} \tag{50}$$

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

$$\alpha P(\bar{A} - \Omega_H^{-1} - \sigma_a^2)k_H^{\alpha-1} + \frac{\pi\alpha z_H k_H^{\alpha-1}}{\iota} = r \tag{51}$$

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

$$\Omega_H = \left(\rho^2(\Omega_H + \sigma_a^{-2})^{-1} + \sigma_\theta^2\right)^{-1} + \left(z_H k_H^\alpha + \iota\delta_H\right)\sigma_\epsilon^{-2} \tag{53}$$

Manipulate to get

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

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

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

Next we show three steps:

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

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

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

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

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

The 3 steps deliver the results.

Part a. Take the total derivative of Equation (56) wrt to z_H and simplify. It implies

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

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

$$A(i) = \frac{\alpha^2\iota_i^{2-\frac{1}{1-\alpha}}\left(\iota_i r k_i^{1-\alpha}\right)^{\frac{1}{1-\alpha}-2}}{(1-\alpha)\Omega_i^2} = \frac{\alpha^2 r^{\frac{2\alpha-1}{1-\alpha}}k_i^{2\alpha-1}}{(1-\alpha)\Omega_i^2} \tag{57}$$

$$B(i) = \frac{2\pi\sigma_\epsilon^2\Omega_i\left(1+r-\frac{\rho^2(\rho^2+\sigma_a^2\sigma_\theta^2)}{\left(\rho^2+\sigma_\theta^2(\sigma_a^2+\Omega_i)\right)^3}\right)}{P} = \pi\frac{dC(i)}{d\Omega_i} \tag{58}$$

$$C(i) = \frac{\sigma_\epsilon^2\Omega_i^2\left(1+r-\frac{\rho^2}{\left(\rho^2+\sigma_\theta^2(\sigma_a^2+\Omega_i)\right)^2}\right)}{P} = \frac{\iota_i k_i^\alpha}{\pi} \tag{59}$$

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

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

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

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

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

Furthermore, consider Equation (40). Keeping the prices constant, the left hand side is increasing in k_H. Alternatively, the derivative of the right hand side with respect to Ω_H is given by

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

$\frac{(\rho^2+\sigma_\theta^2\sigma_a^{-2})}{\left(\rho^2+\sigma_\theta^2(\Omega_H+\sigma_a^{-2})\right)} < 1$, thus Equation (40) 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 (55) implies that fixing ι, k_H moves in the same direction as Ω_H.

Next consider $\hat{z}_H = z_L + \epsilon$. As $\epsilon \to 0$, $\hat{z}_H \to z_L$, $\hat{\Omega}_H \to \Omega_L$, and $\hat{\Upsilon} \to 0$. The above inequality shows that if $\iota < \bar{\iota}$, $\frac{d\Upsilon}{dz_H} > 0$. Thus for $z_H > \hat{z}_H = z_L$, $\Upsilon > 0$.

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

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

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

A.8 Proof of Proposition 3

a. Negative Knowledge Gap with Non-rival Data When High Data-Productivity is Abundant

The proof proceeds in a few steps. We will do the proof for $\gamma = 0$, which implies $P = \bar{P}$. Then, by continuity, the same result holds for γ sufficiently small. Part b of the proposition about the widening knowledge gap is proven separately in the following corrollary.

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

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

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

Part c. Negative knowledge gap: $\exists \bar{\iota} \mid \iota \leq \bar{\iota} \Rightarrow \mathbf{\Upsilon^{ss}} < \mathbf{0}$. Merge Equations (37)-(40) and use $P = \bar{P}$ to write Ω_H

and Ω_L:

$$\frac{\pi}{\bar{P}\sigma_\epsilon^{-2}}\Omega_L^2\left(1+r-\frac{\rho^2}{\left(\rho^2+\sigma_\theta^2(\Omega_L+\sigma_a^{-2})\right)^2}\right)=\frac{1}{r^{\frac{\alpha}{1-\alpha}}}\left(\alpha\bar{P}(\bar{A}-\Omega_L^{-1}-\sigma_a^2)+\pi\alpha z_L\right)^{\frac{\alpha}{1-\alpha}} \tag{60}$$

$$\frac{\pi}{\bar{P}\sigma_\epsilon^{-2}}\Omega_H^2\left(1+r-\frac{\rho^2}{\left(\rho^2+\sigma_\theta^2(\Omega_H+\sigma_a^{-2})\right)^2}\right)=\frac{\iota^{\frac{1-2\alpha}{1-\alpha}}}{r^{\frac{\alpha}{1-\alpha}}}\left(\iota\alpha\bar{P}(\bar{A}-\Omega_H^{-1}-\sigma_a^2)+\pi\alpha z_H\right)^{\frac{\alpha}{1-\alpha}}. \tag{61}$$

Consider Equation (61). Since $\alpha<\frac{1}{2}$, $\iota\to 0$ implies that the first term on the right hand since goes to zero. Every other term in the left and right hand side of the equation is finite and bounded away from zero, except Ω_H^2, so $\Omega_H\to 0$. By continuity, as ι gets small, keeping everything else constant Ω_H has to decline while there is no effect in Equation (60) on Ω_L. Thus $\exists\bar{\iota}$ such that $\iota\leq\bar{\iota}\Rightarrow\Upsilon^{ss}<0$.

b. Widening Knowledge Gap with Non-rival Data When High Data-Productivity is Abundant and z_H is High

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

Equations (37) and (39) can be solved to get (k_L,Ω_L) in terms of data price π:

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

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

The second equation implies

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

The same argument as in Proposition 3 shows that using Equation (39), the derivative of the left hand side with respect to Ω_L is positive. Next, using implicit function theorem on both sides of Equation (62) 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 (37), (39), (38), and (40) to get:

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

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

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

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

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

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

Using Definition (57)-(59) the above expressions simplify to:

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

Thus the derivative of the knowledge gap is given by

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

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

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

Note that

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

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

change

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

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

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

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

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

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

A.9 Proof of Proposition 4: S-shaped Accumulation of Knowledge

We proceed in two parts: convexity and then concavity.

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

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

Recall that data inflow is $d\Omega_{i,t}^+ = z_{i,t}k_{i,t}^{\alpha}\sigma_\epsilon^{-2}$ and its first derivative is $\frac{\partial d\Omega_{i,t}^+}{\partial\Omega_{i,t}} = \alpha z_{i,t}k_{i,t}^{\alpha-1}\sigma_\epsilon^{-2}\frac{\partial k_{i,t}}{\partial\Omega_{i,t}}$. 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 (24)) drops out and it reduces to $k_{i,t} = \left[\frac{\alpha}{r}\left(P_t A_{i,t} + z_i\sigma_\epsilon^{-2}\frac{1}{1+r}V'(\Omega_{i,t+1})\right)\right]^{1/(1-\alpha)}$, which implies

$$k_{i,t}^{1-\alpha} = \frac{\alpha}{r}\left(P_t A_{i,t} + z_i\sigma_\epsilon^{-2}\frac{1}{1+r}V'(\Omega_{i,t+1})\right). \tag{65}$$

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

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

Differentiating (65) with respect to $\Omega_{i,t}$ and using the relationships $\frac{\partial A_{i,t}}{\partial\Omega_{i,t}} = {\Omega_{i,t}}^{-2}$ and $\frac{\partial\Omega_{i,t+1}}{\partial\Omega_{i,t}} = \rho^2[\rho^2 + \sigma_\theta^2(\Omega_{i,t} +$

$\sigma_a^{-2})]^{-2}$, yields

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

Therefore,

$$\begin{aligned}\frac{\partial d\Omega_{i,t}^{+}}{\partial \Omega_{i,t}} &= z_{i,t}k_{i,t}^{2\alpha-1}\sigma_\epsilon^{-2}\frac{\alpha^2}{(1-\alpha)r}\left(P_t{\Omega_{i,t}}^{-2} + z_i\sigma_\epsilon^{-2}\frac{1}{1+r}V''(\Omega_{i,t+1})\rho^2[\rho^2+\sigma_\theta^2(\Omega_{i,t}+\sigma_a^{-2})]^{-2}\right)\\ &= z_{i,t}k_{i,t}^{2\alpha-1}\sigma_\epsilon^{-2}\frac{\alpha^2}{(1-\alpha)r}P_t{\Omega_{i,t}}^{-2} + z_{i,t}^2k_{i,t}^{2\alpha-1}\sigma_\epsilon^{-4}\frac{\alpha^2}{1-\alpha}\frac{1}{r(1+r)}V''(\Omega_{i,t+1})\rho^2[\rho^2+\sigma_\theta^2(\Omega_{i,t}+\sigma_a^{-2})]^{-2}.\end{aligned} \tag{66}$$

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

$$\frac{\partial d\Omega_{i,t}^{-}}{\partial \Omega_{i,t}} = 1 - \frac{1}{\rho^2(\Omega_{i,t}+\sigma_a^{-2})^2(\sigma_\theta^2+\rho^{-2}(\Omega_{i,t}+\sigma_a^{-2})^{-1})^2}. \tag{67}$$

The derivatives of net data flow is then

$$\begin{aligned}\frac{\partial d\Omega_{i,t}^{+}}{\partial \Omega_{i,t}} - \frac{\partial d\Omega_{i,t}^{-}}{\partial \Omega_{i,t}} &= z_{i,t}k_{i,t}^{2\alpha-1}\sigma_\epsilon^{-2}\frac{\alpha^2}{(1-\alpha)r}P_t{\Omega_{i,t}}^{-2} + z_{i,t}^2k_{i,t}^{2\alpha-1}\sigma_\epsilon^{-4}\frac{\alpha^2}{1-\alpha}\frac{1}{r(1+r)}V''(\Omega_{i,t+1})\rho^2[\rho^2+\sigma_\theta^2(\Omega_{i,t}+\sigma_a^{-2})]^{-2}\\ &\quad + \frac{1}{\rho^2(\Omega_{i,t}+\sigma_a^{-2})^2(\sigma_\theta^2+\rho^{-2}(\Omega_{i,t}+\sigma_a^{-2})^{-1})^2} - 1.\end{aligned} \tag{68}$$

For notational convenience, denote the first term in (68) as $M_1 = z_{i,t}k_{i,t}^{2\alpha-1}\sigma_\epsilon^{-2}\frac{\alpha^2}{(1-\alpha)r}P_t{\Omega_{i,t}}^{-2} > 0$, the second term as $M_2 = z_{i,t}^2k_{i,t}^{2\alpha-1}\sigma_\epsilon^{-4}\frac{\alpha^2}{1-\alpha}\frac{1}{r(1+r)}V''(\Omega_{i,t+1})\rho^2[\rho^2+\sigma_\theta^2(\Omega_{i,t}+\sigma_a^{-2})]^{-2} \leq 0$ and the third term as $M_3 = \frac{1}{\rho^2(\Omega_{i,t}+\sigma_a^{-2})^2(\sigma_\theta^2+\rho^{-2}(\Omega_{i,t}+\sigma_a^{-2})^{-1})^2} > 0$. Notice that $M_3 - 1 < 0$ always holds, and thus $M_2 + M_3 - 1 < 0$. $\frac{\partial d\Omega_{i,t}^{+}}{\partial \Omega_{i,t}} - \frac{\partial d\Omega_{i,t}^{-}}{\partial \Omega_{i,t}} > 0$ only holds when P_t is sufficiently large so that M_1 dominates. P_t is sufficiently large when $\bar{P}$ is sufficiently large.

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

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

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

$$z_{it}k_{it}^{2\alpha-1}\frac{\alpha^2}{(1-\alpha)r}\sigma_\epsilon^{-2}\left[-2\bar{P}\Omega_{it}^{-3} - z_{it}\sigma_\epsilon^{-2}\frac{1}{1+r}\nu\rho^2\frac{2\sigma_\theta^2}{\left(\rho^2+\sigma_\theta^2\left(\Omega_{it}+\sigma_a^{-2}\right)\right)^3}\right] > 0 \tag{69}$$

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

$$f(\Omega_{i,t}) := -z_{it}\sigma_\epsilon^{-2}\frac{1}{1+r}\nu\rho^2\Omega_{it}^3\frac{\sigma_\theta^2}{\left(\rho^2+\sigma_\theta^2\left(\Omega_{it}+\sigma_a^{-2}\right)\right)^3} \tag{70}$$

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

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

$$h(\Omega_{i,t}) = 1 - M_3, \tag{71}$$

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

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

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

For the derivative of data inflow (66), note that $\frac{\partial d\Omega^+_{i,t}}{\partial \Omega_{i,t}} \leq z_{i,t} k_{i,t}^{2\alpha-1} \sigma_\epsilon^{-2} \frac{\alpha^2}{(1-\alpha)r} P_t {\Omega_{i,t}}^{-2}$ because $0 < \alpha < 1$ and $V'' < 0$. Thus $\lim_{\Omega_{i,t}\to\infty} \frac{\partial d\Omega^+_{i,t}}{\partial \Omega_{i,t}} \leq 0$.

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

A.10 Proof of Proposition 5: New Firms Earn Negative Profits

Without any production or any data purchased, $\Omega_0 = \sigma_a^{-2}$, because this is the prior variance of the state θ. This is the case when the firm is entering.

Consider the approximation in Equation (9): $\mathbb{E}_i[A_{i,t}] \approx g\left(\Omega_{i,t}^{-1} + \sigma_a^2\right) + g''\left(\Omega_{i,t}^{-1} + \sigma_a^2\right) \cdot \left(\Omega_{i,t}^{-1} + \sigma_a^2\right)$. $g(v)$ is decreasing. When $g''(.) = 0$ (the standing assumption of this part of the paper), then the second term is zero. Thus $\mathbb{E}[A_{i,0}] = g(\sigma_a^2 + \sigma_\theta^2) < 0$. The inequality is the assumption stated in the proposition.

If expected quality $\mathbb{E}[A_{i,0}]$ is less than zero, then expected profit is negative, for any positive level of production, because the steady state price level for goods is positive $P^{ss} > 0$. This can be seen in (14), noting that adjustment cost Ψ, capital rental r and data prices π are all non-negative, by assumption or by free disposal.

Of course, a firm can always choose zero production $k_{i,t} = 0$ and zero data to achieve zero profit. A firm that chose this every period, would have no profit ever and thus zero firm value.

Thus, the only way to get to positive firm value is to produce. Either the firm first buys data and then produces, first produces, or does both together. If the firm first buys data, then profit is negative in the period when the firm buys the data and is not yet producing. If the firm produces first, profit is negative because expected quality is negative, as per the argument above. If the firm produces and buys data at the same time, then profit is more negative because of negative expected quality and the cost of the data purchase. In every scenario, the firm must incur some negative profit to achieve positive production and positive firm value.

A.11 Proof of Proposition 6: Firms Sell Goods at Zero Price (Data Barter)

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

A.12 Proposition 9: Accumulation Can be Purely Concave

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

Proposition 9 ***Concavity of Data Inflow*** $\exists \epsilon > 0$ *such that* $\forall \Omega_{i,t} \in B_\epsilon(0)$*, the net data flow decreases with* $\Omega_{i,t}$ *if* $\sigma_\theta^2 > \sigma_a^2$.

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

Step 1: Data outflows are approximately linear when $\Omega_{i,t}$ *is small.*
This is proven separately in Lemma 4.

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

Recall that the derivative of data inflows with respect to the current stock of knowledge Ω_t is $\frac{\partial d\Omega^+_{i,t}}{\partial \Omega_{i,t}} = \rho^2 \left[\rho^2 + \sigma_\theta^2(\Omega_{i,t} + \sigma_a^{-2})\right]^{-2} > 0$ (see the Proof of Proposition 4 for details). Thus

$$\frac{\partial d\Omega^+_{i,t}}{\partial \Omega_{i,t}} - \frac{\partial d\Omega^-_{i,t}}{\partial \Omega_{i,t}} \approx \rho^2 \left[\rho^2 + \sigma_\theta^2(\Omega_{i,t} + \sigma_a^{-2})\right]^{-2} - 1 + \rho^2(1 + \rho^2 \sigma_\theta^2 \sigma_a^{-2})^{-2}. \tag{72}$$

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

A.13 Lemma 4, 5, 6: Linearity of Data Depreciation

One property of the model that comes up in a few different places is that the depreciation of knowledge (outflows) is approximately a linear function of the stock of knowledge $\Omega_{i,t}$. There are a few different ways to establish this

approximation formally. The three results that follow show that the approximation error from a linear function is small i) when the stock of knowledge is small; ii) when the state is not very volatile; and iii) when the stock of knowledge is large.

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

Proof:

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

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

Proof:

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

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

Proof:

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

A.14 Welfare: Proof of Propositions 7 and 8

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

Household Problem Let Γ_t denote the Lagrangian multiplier of the individual problem on his budget constraint. Individual problem can be written as:

$$\max_{c_t, m_t} \sum_{t=0}^{+\infty} \frac{1}{(1+r)^t} \left(u(c_t) + m_t\right) \qquad \text{with } u(c_t) = \bar{P}\frac{c_t^{1-\gamma}}{1-\gamma}$$

$$\text{s.t.} \quad P_t c_t + m_t = \Phi_t \qquad \forall t$$

where Φ_t is the aggregate profit of all firms:

$$\Phi_t = \int \Phi_{it} di = P_t \int_i A_{i,t} k_{i,t}^{\alpha} di - \int_i -\Psi(\Delta\Omega_{i,t+1}) di - r \int_i k_{i,t} di.$$

The first order conditions for optimal household choices of consumption of c_t and the numeraire good m_t are

$$c_t : \qquad \frac{1}{(1+r)^t} u'(c_t) = P_t \Gamma_t,$$
$$m_t : \qquad \Gamma_t = \frac{1}{(1+r)^t},$$

The first order conditions imply that agents equate their marginal utility of c to its price: $P_t = u'(c_t)$.

Deterministic Aggregate Productivity. We take a short digression here to explain why there is no expectation operator around aggregate profits. Φ_t is not random at date t because aggregate quality $\int A_{i,t} di$ converges to a non-random value, even though each $A_{i,t}$ for each firm i is a random variable. The reason is that the random shocks to $A_{i,t}$'s are independent and converge, by the central limit theorem.

Recall that quality is $A_{i,t} = g((a_{it} - \theta_t - \epsilon_{a,i,t})^2)$. The ϵ_a shocks are obviously idiosyncratic and independent. That is not a cause for concern so we set those aside. However, one might think that shocks to θ_t would cause $A_{i,t}$ to covary across firms and create aggregate shocks to quality and output. The reason this does not happen is that the action choice a_{it} is firm i's conditional expectation of θ_t. So, $a_{it} - \theta_t$ is a forecast error. The forecast errors are what are independent. What ensures this is the noisy prior assumption made in the model setup. When the prior is noisy, beliefs about θ_t are the true θ_t, plus idiosyncratic signal noise. Thus, forecast errors are idiosyncratic, or independent. Since any function of an independent random variable or variables is independent, $A_{i,t} = g((a_{it} - \theta_t - \epsilon_{a,i,t})^2)$ is independent across firms. Since the random component of $A_{i,t}$ is independent, its integral over an infinite number of firms, its mean, converges to a constant, by the central limit theorem.

Since we have a continuum of firms, then for any finite types of firms, like the H and L firms later, the quality of each type of firm also has independent noise. Therefore, the type-specific quality averages $A_{L,t}$ and $A_{H,t}$, that we make use of later, will also be non-random variables.

Firm Problem Firms' sequential optimization problem is

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

Equivalently, in recursive form

$$V(\Omega_{i,t}) = \max_{k_{i,t},\delta_{i,t}} P_t \mathbb{E}[A_{i,t}] k_{i,t}^{\alpha} - \Psi(\Delta\Omega_{i,t+1}) - \pi_t\delta_{i,t} - r k_{i,t} + \frac{V(\Omega_{i,t+1})}{1+r} \tag{73}$$

$$\text{s.t. } \Omega_{i,t+1} = \left(\rho^2(\Omega_{i,t} + \sigma_a^{-2})^{-1} + \sigma_\theta^2\right)^{-1} + \left(z_i k_{i,t}^{\alpha} + \left(\mathbf{1}_{\delta_{i,t}>0} + \iota \mathbf{1}_{\delta_{i,t}<0}\right)\delta_{i,t}\right)\sigma_\epsilon^{-2} \tag{74}$$

The profits of the firm at time t are $\Phi_{i,t} = P_t A_{i,t} k_{i,t}^{\alpha} - \Psi(\Delta\Omega_{i,t+1}) - \pi\delta_{i,t} - r k_{i,t}$.

Market Clearing (Resource Constraint)

$$\begin{aligned}
&\text{retail good :} && c_t = \int_i A_{i,t} k_{i,t}^{\alpha} di, \\
&\text{numeraire good :} && m_t + \int_i \left(r k_{i,t} + \Psi(\Delta\Omega_{i,t+1})\right) di = 0 \\
&\text{data :} && \int_i \delta_{i,t} di = 0.
\end{aligned}$$

The adjustment cost terms are incorporated in the market clearing/resource constraint for the numeraire good so that they show up in the planner's objective function.

A.14.1 Competitive Equilibrium in Steady State

In equilibrium, households (HHs, hereafter) maximize utility by choosing c_t and m_t, firms maximize profits by choosing $\{k_{i,t}, \delta_{i,t}\}_{i=L,H}$, and markets clear.

In this section we focus on steady state equilibrium outcomes with two types of firms, $i = L, H$. HH budget constraint simplifies to

$$\begin{aligned}
&P^{eq} c^{eq} + m^{eq} = \Phi^{eq} \\
&\Phi^{eq} = P^{eq}\Big(\lambda \mathbb{E}[A_L^{eq}](k_L^{eq})^{\alpha} + (1-\lambda)\mathbb{E}[A_H^{eq}](k_H^{eq})^{\alpha}\Big) - r\Big(\lambda k_L^{eq} + (1-\lambda)k_H^{eq}\Big)
\end{aligned}$$

where HH optimization implies $P^{eq} = u'(c^{eq})$. Furthermore, the market clearing conditions simplify to

$$\begin{aligned}
&\text{retail good :} && c^{eq} = \lambda \mathbb{E}[A_L^{eq}](k_L^{eq})^{\alpha} + (1-\lambda)\mathbb{E}[A_H^{eq}](k_H^{eq})^{\alpha}, \\
&\text{numeraire good :} && m^{eq} + r\Big(\lambda k_L^{eq} + (1-\lambda)k_H^{eq}\Big) = 0 \\
&\text{data :} && \lambda\delta_L^{eq} + (1-\lambda)\delta_H^{eq} = 0.
\end{aligned}$$

There are 6 real equilibrium steady state variables. They are: $(\Omega_L^{eq}, \Omega_H^{eq}, k_L^{eq}, k_H^{eq}, \delta_L^{eq}, \delta_H^{eq})$. Thus we need 6 equations to determine them. 3 of the equations are straightforward

1. Two equations for the the dynamic evolution of firm stock of knowledge, $i = L, H$, Equation (80)

2. One equation is the resource constraint for traded data, Equation (83).

Firms' optimal capital choices. There are two equations for first order condition (FOC) with respect to k_i, $i = L, H$. We will use the sequential problem to get this first order condition. Consider FOC of firm i with respect to $k_{i,t}$:

$$\frac{1}{(1+r)^t}\left(\alpha P_t \mathbb{E}[A_{i,t}]k_{i,t}^{\alpha-1} - \frac{\partial\Psi(\Delta\Omega_{i,t+1})}{\partial k_{i,t}} - r\right) + \frac{1}{(1+r)^{t+1}}\left(P_{t+1}\frac{\partial\mathbb{E}[A_{i,t+1}]}{\partial k_{i,t}}k_{i,t+1}^{\alpha} - \frac{\partial\Psi(\Delta\Omega_{i,t+1})}{\partial k_{i,t}}\right) = 0.$$

Substitute

$$\frac{\partial\mathbb{E}[A_{i,t+1}]}{\partial k_{i,t}} = \alpha z_i \sigma_\epsilon^{-2} k_{i,t}^{\alpha-1}\Omega_{i,t+1}^{-2}.$$

Multiply both sides by $\frac{1}{(1+r)^t}$. Steady state implies a stable level of knowledge ($\Delta\Omega = 0$). With a quadratic adjustment cost function that is 0 at 0, $\Psi'(0) = 0$. Thus, in the steady state $\frac{\partial\Psi(\Delta\Omega_{i,t+2})}{\partial k_{i,t}} = \frac{\partial\Psi(\Delta\Omega_{i,t+1})}{\partial k_{i,t}} = 0$. Imposing this condition simplifies the firm's FOC:

$$\alpha P k_i^{\alpha-1}\Big(\mathbb{E}[A_i] + \frac{z_i\sigma_\epsilon^{-2}}{1+r}\Omega_i^{-2}k_i^{\alpha}\Big) = r. \tag{75}$$

which implies Equation (81) with $P = u'(c)$.

Firm's optimal data choices. In the steady state, where the adjustment cost is zero, the firm's FOC with respect to data purchases/sales is

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

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

$$V'(\Omega_{i,t}) = P_t k_{i,t}^{\alpha}\Omega_{i,t}^{-2} + \frac{1}{1+r}V'(\Omega_{i,t+1})\frac{\partial\Omega_{i,t+1}}{\Omega_{i,t}}, \tag{77}$$

Differentiating Equation (74) with respect to $\Omega_{i,t}$,

$$\frac{\partial\Omega_{i,t+1}}{\partial\Omega_{i,t}} = \frac{\rho^2}{\left(\rho^2 + \sigma_\theta^2(\Omega_{i,t} + \sigma_a^{-2})\right)^2}. \tag{78}$$

Substitute Equation (76) for $V'(\Omega_{i,t}) = V'(\Omega_{i,t+1})$ (in steady state) in (77):

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

Next substitute for $V'(\Omega_{i,t})$ in Equation (76), using the expression for $\frac{\partial\Omega_{i,t+1}}{\partial\Omega_{i,t}}$ from Equation (78). Then, multiply through by $1 + r$, and re-arrange. This yields one condition for the optimal capital-knowledge ratio for L firms and

one for H firms:

$$\left(1+r-\frac{\rho^2}{(\rho^2+\sigma_\theta^2(\Omega_i+\sigma_a^{-2}))^2}\right)\frac{\pi}{P\sigma_\epsilon^{-2}(\mathbb{1}_{\delta_i>0}+\iota\mathbb{1}_{\delta_i<0})}=k_i^\alpha\Omega_i^{-2}\qquad i=L,H \tag{79}$$

If we guess and verify that H firms will sell data and L firms will buy it, then we can simplify $(\mathbb{1}_{\delta_i>0}+\iota\mathbb{1}_{\delta_i<0})$, by equating it to 1 for L firms and ι for H firms. Taking the ratio of the L and H optimality conditions allows us to cancel out P_t, which delivers Equation (82).

Thus the 6 equilibrium steady state real variables, $(\Omega_L^{eq},\Omega_H^{eq},k_L^{eq},k_H^{eq},\delta_L^{eq},\delta_H^{eq})$ are determined by the following system of 6 equations. Note that (80) and (81) represent two equations each.

$$\Omega_i^{eq}=\left[\rho^2(\Omega_i^{eq}+\sigma_a^{-2})^{-1}+\sigma_\theta^2\right]^{-1}+\left(z_i(k_i^{eq})^\alpha+\delta_i^{eq}(\mathbf{1}_{\delta_i^{eq}>0}+\iota\mathbf{1}_{\delta_i^{eq}<0})\right)\sigma_\epsilon^{-2}\qquad i=L,H \tag{80}$$

$$r=\alpha u'(c^{eq})(k_i^{eq})^{(\alpha-1)}\left[\mathbb{E}[A_i^{eq}]+\frac{z_i\sigma_\epsilon^{-2}}{1+r}(k_i^{eq})^\alpha(\Omega_i^{eq})^{-2}\right]\qquad i=L,H \tag{81}$$

$$\frac{\frac{(k_L^{eq})^\alpha}{(\Omega_L^{eq})^2}}{\iota\frac{(k_H^{eq})^\alpha}{(\Omega_H^{eq})^2}}=\frac{\left(1+r-\frac{\rho^2}{(\rho^2+\sigma_\theta^2(\Omega_L^{eq}+\sigma_a^{-2}))^2}\right)}{\left(1+r-\frac{\rho^2}{(\rho^2+\sigma_\theta^2(\Omega_H^{eq}+\sigma_a^{-2}))^2}\right)} \tag{82}$$

$$\lambda\delta_L^{eq}+(1-\lambda)\delta_H^{eq}=0. \tag{83}$$

Equation (80) represents the two law of motions for stock of knowledge, one for each type of firm $i=L,H$. Equation (81) represents the two first order conditions for capital choice, one for each type of firm $i=L,H$. (82) is a single equation, the ratio of first order conditions for the data choice for the two types of firm. Taking the ratio enables us to eliminate the steady state data price π^{eq} from the system of equations. Finally, Equation (83) is the resource constraint for the total traded data, which should be zero.

We also have $c^{eq}=\lambda\mathbb{E}[A_L^{eq}](k_L^{eq})^\alpha+(1-\lambda)\mathbb{E}[A_H^{eq}](k_H^{eq})^\alpha$, and we adopt the quality function $g(a_{i,t}-\theta_t-\epsilon_{a,i,t})=\bar{A}-(a_{i,t}-\theta_t-\epsilon_{a,i,t})^2$. Recall that the optimal technique is $a_{i,t}^*=\mathbb{E}[\theta_i|\mathcal{I}_{i,t}]$, which implies that in steady state,

$$\begin{aligned}A_i^{eq}&=\bar{A}-\left(\mathbb{E}[\theta_t|\mathcal{I}_{i,t}]-\theta_t-\epsilon_{a,i,t}\right)^2, &&i=L,H\\ \mathbb{E}[A_i^{eq}]&=\bar{A}-(\Omega_i^{eq})^{-1}-\sigma_a^2. &&i=L,H\end{aligned}$$

Solving for Equilibrium Prices Goods price P^{eq} is simply determined by the HH FOC with respect to c_t. Data price π^{eq} is derived from combining the three Equations (76), (77), and (78) in one equation:

$$P^{eq}=u'(c^{eq}) \tag{84}$$

$$\pi^{eq}=\frac{P^{eq}\sigma_\epsilon^{-2}(K_L^{eq})^\alpha}{(\Omega_L^{eq})^2\left(1+r-\frac{\rho^2}{(\rho^2+\sigma_\theta^2(\Omega_L^{eq}+\sigma_a^{-2}))^2}\right)}=\frac{\iota P^{eq}\sigma_\epsilon^{-2}(K_H^{eq})^\alpha}{(\Omega_H^{eq})^2\left(1+r-\frac{\rho^2}{(\rho^2+\sigma_\theta^2(\Omega_H^{eq}+\sigma_a^{-2}))^2}\right)}. \tag{85}$$

Recall that household optimization problem in each period is deterministic. Thus using the functional form for

$u(c)$, Equation (81) can be written as

$$r = \alpha \bar{P}(c^{eq})^{-\gamma}(k_i^{eq})^{\alpha-1}\left[\mathbb{E}[A_i^{eq}] + \frac{z_i\sigma_\epsilon^{-2}}{1+r}(k_i^{eq})^{\alpha}(\Omega_i^{eq})^{-2}\right] \qquad i = L, H \tag{86}$$

Next, use the recourse constraint for the retail good to write Equation (84) as

$$P^{eq} = \bar{P}\Big(\lambda\mathbb{E}[A_L^{eq}](k_L^{eq})^{\alpha} + (1-\lambda)\mathbb{E}[A_H^{eq}](k_H^{eq})^{\alpha}\Big)^{-\gamma}. \tag{87}$$

A.14.2 Social Planner Problem

The planner maximizes HH total discounted utility, taking the resource constraints into account. Thus planner's problem can be written as

$$\max_{\{k_{i,t},\delta_{i,t}\}_{i=L,H}} \sum_{t=0}^{\infty} \frac{1}{(1+r)^t}\left(u(c_t) - r\int_i k_{i,t}di - \int_i \Psi(\Delta\Omega_{i,t+1})di\right)$$

or in recursive form

$$\begin{aligned}
V^P(\{\Omega_{i,t}\}_i) = \max_{\{k_{i,t},\delta_{i,t}\}_i} \; & u(c_t) - r\int_i k_{i,t}di - \int_i \Psi(\Delta\Omega_{i,t+1})di + \frac{1}{1+r}V^P(\{\Omega_{i,t+1}\}_i) \\
\text{s.t.} \quad & c_t = \int_i A_{i,t}k_{i,t}^{\alpha}di \qquad (\Xi_t) \qquad \forall t \\
& \int_i \delta_{i,t}di = 0 \qquad (\eta_t) \qquad \forall t \\
& \Omega_{i,t+1} = \left[\rho^2(\Omega_{i,t} + \sigma_a^{-2})^{-1} + \sigma_\theta^2\right]^{-1} + \left(z_i(k_{i,t})^{\alpha} + \delta_{i,t}(\mathbf{1}_{\delta_{i,t}>0} + \iota\mathbf{1}_{\delta_{i,t}<0})\right)\sigma_\epsilon^{-2} \quad \forall i,t \\
& \mathbb{E}[A_{i,t}] = \bar{A} - \Omega_{i,t}^{-1} - \sigma_a^2 \qquad \forall i,t.
\end{aligned}$$

Similar to equilibrium, as the household consumption equal the aggregate production of a continuum of firms, it is deterministic at each time t.

Social Planner's optimal capital choice. The planner's first order condition with respect to $k_{i,t}$ is

$$r\lambda_i = \frac{\partial u(c_t)}{\partial k_{i,t}} + \frac{1}{1+r}\frac{\partial u(c_{t+1})}{\partial k_{i,t}} \qquad \text{for } \; i = L, H \tag{88}$$

Again, focus on two types of firm $i = L, H$ where the firms in each group are identical. Then $\lambda_i = \lambda$ when $i = L$ and $\lambda_i = 1-\lambda$ when $i = H$. The planner objective simplifies to

$$\begin{aligned}
V^P(\Omega_{L,t}, \Omega_{H,t}) = \max_{\{k_{i,t},\delta_{i,t}\}_{i=L,H}} \; & u(c_t) - r\Big(\lambda k_{L,t} + (1-\lambda)k_{H,t}\Big) - \Big(\lambda\Psi(\Delta\Omega_{L,t+1}) + (1-\lambda)\Psi(\Delta\Omega_{H,t+1})\Big) \\
& + \frac{1}{1+r}V^P(\Omega_{L,t+1}, \Omega_{H,t+1})
\end{aligned}$$

Furthermore, $c_t = \lambda \mathbb{E}[A_{L,t}]k_{L,t}^{\alpha} + (1-\lambda)\mathbb{E}[A_{H,t}]k_{H,t}^{\alpha}$ and $\mathbb{E}[A_{i,t}] = \bar{A} - \Omega_{i,t}^{-1} - \sigma_a^2$. Thus

$$\begin{aligned}\frac{\partial c_t}{\partial k_{i,t}} &= \alpha\lambda_i \mathbb{E}[A_{i,t}]k_{i,t}^{\alpha-1} \\ \frac{\partial c_t}{\partial \Omega_{i,t}} &= \lambda_i \frac{\partial \mathbb{E}[A_{i,t}]}{\partial \Omega_{i,t}} k_{i,t}^{\alpha} = \lambda_i \Omega_{i,t}^{-2} k_{i,t}^{\alpha}\end{aligned} \tag{89}$$

In steady state, substitute in the expressions above into (88),

$$r = \alpha \bar{P}(c^{opt})^{-\gamma}(k_i^{opt})^{\alpha-1}\left[\mathbb{E}[A_i^{opt}] + \frac{z_i \sigma_\epsilon^{-2}}{1+r}(k_i^{opt})^{\alpha}(\Omega_i^{opt})^{-2}\right] \qquad i = L, H \tag{90}$$

This is the same as Equation (86). Thus the capital FOCs are the same between optimum and equilibrium.

Social Planner's optimal data choice. Let V_i^P denote the derivative of the social planner value function with respect to $\Omega_{i,t}$, $i = L, H$. The data first order condition reveals that the Lagrange multiplier on the data constraint is

$$\lambda_i \eta_t = \frac{1}{1+r} V_i^P(\Omega_{L,t+1}, \Omega_{H,t+1})\sigma_\epsilon^{-2}(\mathbb{1}_{\delta_{i,t}>0} + \iota \mathbb{1}_{\delta_{i,t}<0}). \tag{91}$$

To solve for V_i^P in steady state, differentiate the value function and apply the envelope condition to get:

$$V_i^P(\Omega_{i,t}, \Omega_{-i,t}) = \frac{\partial u(c_t)}{\partial \Omega_{i,t}} + \frac{1}{1+r} V_i^{P\prime}(\Omega_{i,t+1}, \Omega_{-i,t+1}) \frac{\partial \Omega_{i,t+1}}{\partial \Omega_{i,t}}$$

In steady state, equate $V_i^p(\Omega_{i,t}, \Omega_{-i,t})$ and $V_i^P(\Omega_{i,t+1}, \Omega_{-i,t+1})$ in the previous equation, and use Equations (78) and (91) to replace for $\frac{\partial \Omega_{i,t+1}}{\partial \Omega_{i,t}}$ and $V_i^P(\Omega_{i,t+1}, \Omega_{-i,t+1})$ to get

$$\left(1 + r - \frac{\rho^2}{(\rho^2 + \sigma_\theta^2(\Omega_{i,t} + \sigma_a^{-2}))^2}\right) \frac{\eta_t \lambda_i}{\sigma_\epsilon^{-2}(\mathbb{1}_{\delta_{i,t}>0} + \iota \mathbb{1}_{\delta_{i,t}<0})} = \frac{\partial u(c_t)}{\partial \Omega_{i,t}} \qquad i = L, H \tag{92}$$

which in steady state can be written as

$$\left(1 + r - \frac{\rho^2}{(\rho^2 + \sigma_\theta^2(\Omega_i^{opt} + \sigma_a^{-2}))^2}\right) \frac{\eta \lambda_i}{\bar{P}(c^{opt})^{-\gamma}\sigma_\epsilon^{-2}(\mathbb{1}_{\delta_i^{opt}>0} + \iota \mathbb{1}_{\delta_i^{opt}<0})} = \lambda_i (k_i^{opt})^{\alpha}(\Omega_i^{opt})^{-2} \qquad i = L, H \tag{93}$$

In steady state, H firms sell data and L firms buy data. As with the decentralized problem, take the ratio of the H and L conditions. $(c^{opt})^{-\gamma}$ and the Lagrange multiplier η both drop out of the resulting equation, thus we have

$$\frac{\frac{(k_L^{opt})^{\alpha}}{(\Omega_L^{opt})^2}}{\iota \frac{(k_H^{opt})^{\alpha}}{(\Omega_H^{opt})^2}} = \frac{1 + r - \frac{\rho^2}{(\rho^2 + \sigma_\theta^2(\Omega_L^{opt} + \sigma_a^{-2}))^2}}{1 + r - \frac{\rho^2}{(\rho^2 + \sigma_\theta^2(\Omega_H^{opt} + \sigma_a^{-2}))^2}}, \tag{94}$$

which is the same as Equation (82).

Finally, the planner's first order conditions with respect to consumption choice tells us that the Lagrange multiplier on the consumption resource constraint is $\Xi_t = u'(c_t)$.

A.14.3 Proof of Proposition 7: Efficient Equilibrium

The decentralized equilibrium is characterized by Equations (80) for $i = L, H$, (82), (83), and (86) for $i = L, H$.

The social planner's optimum is characterized by Equations (80) for $i = L, H$ and (83) (all for optimum variables), Equation (90) for $i = L, H$, and Equation (94).

The resulting capital first order conditions for each form $i = L, H$, as well as the ratio of the data first order conditions across two types of firms, for both problems are the same. Thus, the equilibrium is efficient because the decentralized economy and the social planner end up making the same choices.

A.14.4 Proof of Proposition 8: Inefficiency with Business Stealing

With business stealing externality, i.e. when $b = 1$, the only difference is that A_i is determined via Equation (18) to be

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

Thus in a symmetric allocation where all firms of type i are the same, in equilibrium we have

$$\mathbb{E}[A_{i,t}] = \left(\bar{A} - \left(\Omega_{i,t}\right)^{-1} - \sigma_a^2\right) + \left(\lambda_i \left(\Omega_{i,t}^{-1} + \sigma_a^2\right) + (1-\lambda_i)\left(\Omega_{-i,t}^{-1} + \sigma_a^2\right)\right) = \bar{A} - (1-\lambda_i)(\Omega_{i,t}^{-1} - \Omega_{-i,t}^{-1}).$$

By construction, aside from the change in the equilibrium steady state value of $\mathbb{E}[A_i^{eq}]$, the business stealing externality does not change the firm optimization problem. In particular, it does not affect any of the first order condition, such as $\frac{\partial \mathbb{E}[A_{i,t+1}]}{\partial k_{i,t}}$. Thus the equilibrium is still characterized by Equations (80) for $i = L, H$, (82), (83), and (86) for $i = L, H$.

For the optimum, Equations (80) for $i = L, H$ and (83) clearly remains the same. The other optimum equations change as the quality of every firm is affected by the capital and data choices of each individual firm i.

Planner FOC for Data with Business Stealing Observe that the amount of data traded by firm i at time t, $\delta_{i,t}$ does not affect the stock of knowledge of firm j at $t+1$, $\Omega_{j,t+1}$ conditional on $\delta_{j,t}$. Furthermore, $\Omega_{i,t}$ does not affect $\Omega_{j,t+1}$, $j \neq i$. However, $\frac{\partial c_t}{\partial \Omega_{i,t}}$ is adjusted to reflect data used for business stealing:

$$\begin{aligned}\frac{\partial c_t}{\partial \Omega_{i,t}} &= \lambda_i k_{i,t}^{\alpha} \frac{\partial \mathbb{E}[A_{i,t}]}{\partial \Omega_{i,t}} + (1-\lambda_i) k_{-i,t}^{\alpha} \frac{\partial \mathbb{E}[A_{-i,t}]}{\partial \Omega_{i,t}} = \lambda_i (1-\lambda_i) k_{i,t}^{\alpha} \Omega_{i,t}^{-2} - (1-\lambda_i)^2 k_{-i,t}^{\alpha} \Omega_{i,t}^{-2} \\ &= (1-\lambda_i)\Omega_{i,t}^{-2}\left(\lambda_i k_{i,t}^{\alpha} - (1-\lambda_i) k_{-i,t}^{\alpha}\right). \end{aligned} \tag{95}$$

Comparing Equations (89) and (95) clarifies that data with business stealing, data is less useful to increase the consumption level. The firms do not internalize that selling data ot others decreases their quality. Thus, there is an over-supply of data on the data market, and too much data trade. With business stealing, Equations (93) and (94)

change to

$$\left(1+r-\frac{\rho^2}{(\rho^2+\sigma_\theta^2(\Omega_i^{opt}+\sigma_a^{-2}))^2}\right)\frac{\eta\lambda_i}{\bar{P}(c^{opt})^{-\gamma}\sigma_\epsilon^{-2}(\mathbb{1}_{\delta_i^{opt}>0}+\iota\mathbb{1}_{\delta_i^{opt}<0})}$$

$$=(1-\lambda_i)(\Omega_i^{opt})^{-2}\left(\lambda_i(k_i^{opt})^\alpha-(1-\lambda_i)(k_{-i}^{opt})^\alpha\right)\quad\forall i \tag{96}$$

$$\left(\frac{1-\lambda}{\lambda}\right)^2\frac{\frac{\lambda(k_L^{opt})^\alpha+(1-\lambda)(k_H^{opt})^\alpha}{(\Omega_L^{opt})^2}}{\iota\frac{(1-\lambda)(k_H^{opt})^\alpha+\lambda(k_L^{opt})^\alpha}{(\Omega_H^{opt})^2}}=\frac{1+r-\frac{\rho^2}{(\rho^2+\sigma_\theta^2(\Omega_L^{opt}+\sigma_a^{-2}))^2}}{1+r-\frac{\rho^2}{(\rho^2+\sigma_\theta^2(\Omega_H^{opt}+\sigma_a^{-2}))^2}}, \tag{97}$$

Equation (97) is different from equilibrium Equation (82).

This is the first externality. With business stealing, the planner internalizes that the data that a firms sells on the data market, decreases its own quality. Since firms do to internalize this effect, they sell more data on the data market than what is efficient. There is excessive data trade.

Planner FOC for Capital with Business Stealing

$$r\lambda_i=\frac{\partial u(c_t)}{\partial k_{i,t}}+\frac{1}{1+r}\frac{\partial u(c_{t+1})}{\partial k_{i,t}}\qquad i=L,H$$

Thus we have

$$r=\alpha\bar{P}(k_i^{opt})^{\alpha-1}(c^{opt})^{-\gamma}\left[\mathbb{E}[A_i^{opt}]+\frac{z_i\sigma_\epsilon^{-2}(1-\lambda_i)}{1+r}\left((k_i^{opt})^\alpha-\frac{1-\lambda_i}{\lambda_i}(k_{-i}^{opt})^\alpha\right)(\Omega_i^{opt})^{-2}\right].\qquad i=L,H \tag{98}$$

Equation (98) is different from Equation (86).

This is the second externality. With business stealing, the planner internalizes that an increase in capital of firm i, increases data production, which decreases the quality of every other firm in the sector. Since firms do to internalize this effect, they over-invest in capital to get more data than what is efficient. There is excessive production.

B Numerical Examples and Model Extensions

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

Parameter Selection The results below are not calibrated.[10] However, the share of aggregate income paid to capital is commonly thought to be about 0.4. Since this is governed by the exponent α, we set $\alpha = 0.4$. For the rental rate on capital, we use a riskless rate of 3% , which is an average 3-month treasury rate over the last 40 years. The inverse demand curve parameters determine the price elasticity of demand. We take γ and $\bar{P}$ from the literature. Finally, we model the adjustment cost for data ψ in the same was as others have the adjust cost of

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

capital.This approach makes sense because adusting one's process to use more data typically involves the purchase of new capital, like new computing and recording equipment and involves disruptive changes in firm practice, similar to the disruption of working with new physical machinery.

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

B.1 Computational Procedure

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

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

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

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

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

B.2 Data Portfolio Choice

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

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

$$V(\Omega_{i,t}) = \max_{k_{i,t},\delta_{i,t}} P_t' \left(\mathbb{1}'(\bar{A} - \sigma_a^2)\mathbb{1} - \Omega_{i,t}^{-1}\right) k_{i,t}^{\alpha} - \Psi(\Delta\Omega_{i,t+1}) - \pi_t'\delta_{i,t} - rk_{i,t}'\mathbb{1} + \left(\frac{1}{1+r}\right) V(\Omega_{i,t+1}) \tag{99}$$

where $k_{i,t}^{\alpha}$ means that each element is raised to the power α, $\mathbb{1}$ is an $n \times 1$ vector of ones, and the law of motion for $\Omega_{i,t}$ is given by (9).

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