NBER WORKING PAPER SERIES

A MODEL OF THE DATA ECONOMY

Maryam Farboodi
Laura Veldkamp

Working Paper 28427
http://www.nber.org/papers/w28427

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2021, Revised June 2022

Thanks to Rebekah Dix and Ran Liu for invaluable research assistance and to participants and discussants at the 2019 SED plenary as well as numerous research seminars and conferences for helpful comments and suggestions. Maryam Farboodi receives research support from MIT Sloan school of management. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

A Model of the Data Economy
Maryam Farboodi and Laura Veldkamp
NBER Working Paper No. 28427
February 2021, Revised June 2022
JEL No. O3,O4

ABSTRACT

In a data economy, transactions of goods and services generate information, which is stored, traded and depreciates. How are the economics of this economy different from traditional production or innovation economies? How do these differences matter for measurement of GDP, firm values, depreciation rates, welfare and externalities? Despite incorporating active experimentation and data as an intangible asset, we devise a tractable recursive representation. Because the resulting model maps to many observable macro and finance measures, it can be calibrated and estimated like its old-economy DSGE counterpart. The model also delivers insights: It rationalizes why apps are often "free," why firm size is diverging and why even non-digital economic activity might be greater than GDP suggests.

Maryam Farboodi
MIT Sloan School of Management
100 Main Street, E62-627
Cambridge, MA 02142
and NBER
m.farboodi@gmail.com

Laura Veldkamp
Columbia Business School
725 Kravis Hall
665 W 130th St
New York, NY 10027
and NBER
lv2405@columbia.edu

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. 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. However, the short-run features new dynamics, like increasing returns, negative profits, and the barter of data for goods.

The primary contribution of the paper is not the particular predictions we explore, but our model as a tool. Some of our predictions are unsurprising, given the model assumptions. But the realism of the predictions supports the notion that the framework is a relevant and useful one. The larger contribution is a tool to value data, measure its effects and to think clearly about the aggregate economic consequences of data accumulation. Measuring and valuing data is complicated by the fact that frequently, customers provide their data, in exchange for a free digital service. Our value function assigns a positive value to goods and to data, even if they have a zero transaction price. In so doing, it moves aggregate models beyond price-weighted valuation and toward an updated way of thinking about economic value in a data economy.

Modeling the data economy is a challenge. A key feature is that firms/customer actions produce data, which is a form of information. When actions are chosen, taking into account the information those actions will generate, this is active experimentation. Micro models of active experimentation are typically challenging to solve (Bergemann and Välimäki, 2000), even without the complicating equilibrium forces. As an additional challenge, a useful model of the data economy should feature data as a long-lived, depreciating and tradeable asset. That calls for a recursive Bellman approach, with a data state variable. Tractably valuing data that a) comes from active experimentation, b) generates value for many periods, c) is traded in markets with equilibrium prices and d) eventually depreciates, calls for a new set of tools. While the resulting model looks like a standard DSGE

framework, achieving this degree of simplicity requires care.

The model in Section 1 describes a particular type of digitized information: transaction-generated data, used by firms to optimize their business processes, by accurately predicting future outcomes. The data economy blossomed with breakthroughs in machine learning and artificial intelligence. These are prediction algorithms. They require troves 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. Such data is typically used to help firms optimize by forecasting demand, costs, earnings, labor needs, targeting advertising or selecting investments or product lines.

Because of its simple structure, the model can be applied and extended in many ways. 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 model streamlined allows it to be used flexibly.

Section 2 shows how to value and depreciate data. Both are tough to observe directly. However, our model offers a way to estimate how quickly a particular type of data loses value. Bayes' Law and its cousin, the Kalman filter dictate the rate at which information precision depreciates and point us to a simple estimation procedure. Knowing how data depreciates allows us to build up a recursive value function structure, that looks similar to ones used to value capital, but embodies the value of production as active experimentation and the unique way in which data depreciates.

Section 3 explores the path a given firm takes when growing to its 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 generates 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 data-poor. Firms may even choose to produce with negative profits, as a form of costly investment in data and may still have high equity market valuations, despite having minimal book value. This rationalizes observed data barter. 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 barter trade.

Section 4 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. The long-run data economy looks just like a long-run capital economy, but for different reasons: First, prediction errors can only be reduced to zero. That places a natural bound on how much prediction error data can possibly resolve. Second, unforecastable randomness limits how accurate firms' forecasts can possibly be. Either one of these forces ensures that when data is abundant, it has diminishing returns and cannot sustain growth.

Of course, if we change the model to make data an input into research and development (R&D), it can sustain growth (Section 4.3). The message is not that anything is possible. The point is to inform measurement: We should measure data used for R&D separately, just like macroeconomists typically distinguish between regular capital investment and R&D investments.

Some of the most heated policy debates today revolve around firms' use of data. Thinking about regulation and welfare requires building out the household side of the model that micro-founds the demand curve. Section 5 does this and finds that, despite the non-rivalry, the increasing returns, and the production of data as a by-product, equilibrium choices are efficient. That doesn't mean that data cannot cause harm. It just means that the simple forces our model describes do not compromise welfare, by themselves.

We augment the model to capture data externalities, such as lost privacy or stealing business from rivals with data-targeted marketing. Such negative externalities obviously incentivize excess data production, which in turn, inflates goods production, in order to generate the extra data. One way in which the data economy reacts differently to externalities is that they prompt excessive data trade. This finding offers new perspective on where to look for evidence on the strength of data externalities.

Related Literature. This work builds on insights from many literatures, each of which has some, but not all, of the features of this model. 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.

Choosing to acquire data is technically similar to other forms of information choice like the information acquisition in Broer et al. (2021) or rational attention choices in Maćkowiak and Wiederholt (2009), Matějka and McKay (2015) or Reis (2008). Our work borrows modeling strategies directly from Morris and Shin (2002) and Angeletos et al. (2006) and shares a focus on the social value of information. 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 data that can be sold. This feedback is absent in these other literatures and is central to the data measurement challenges we address.

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), data contributes directly to productivity. This is okay for their objective – exploring growth versus privacy. But without modeling data as information, they cannot explore the tension between diminishing and increasing returns that is central to data valuation.

Work exploring the interactions of data and innovation sounds similar, but has essential differences. For example, in Garicano and Rossi-Hansberg (2012), IT allows agents to accumulate more knowledge, which facilitates innovation. Agrawal et al. (2018) explore how breakthroughs in AI could enhance discovery rates and economic growth. 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 unlike prediction data, this knowledge is not a tradeable asset. Acemoglu and Restrepo (2018) consider the growth potential from robots. Although robots require data, they are rival capital goods, distinct from the data itself. Our work analyzes data accumulation, in the absence of technological change. Once we understand this foundation, one can layer these insights about innovation and automation on top.

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. This paper adds an aggregate equilibrium model,

with goods markets, data markets, data non-rivalry, analytical solutions and welfare analysis. The new dimensions to this model fundamentally shape the answers to our main questions about the value of data and the functioning of a data economy.

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 component: $\theta_t + \epsilon_{a,i,t}$. Neither component is separately observed. The persistent component θ_t follows an AR(1) process: $\theta_t =$

$\bar{\theta} + \rho(\theta_{t-1} - \bar{\theta}) + \eta_t$. The AR(1) innovation $\eta_t \sim N(0, \sigma_\theta^2)$ is *i.i.d.* across time.[1] 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. Observing production and sales processes at work provides useful information for optimizing business practices. For now, we model data as welfare-enhancing. Section 5 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 $\epsilon_{i,t,m} \sim N(0, \sigma_\epsilon^2)$.

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

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 transitory shock $\epsilon_{a,i,t}$ 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, after producing in date t. If $\delta_{i,t} < 0$, the firm is selling data. If $\delta_{it} > 0$, the firm purchased data.We restrict $\delta_{i,t} \geq \underline{\delta}$, where $\underline{\delta} \leq 0$. This does not prohibit selling or even selling multiple copies of data. But it does bound sales so that a firm cannot sell so much data that it is left with a negative stock of knowledge. If the firm buys $\delta_{i,t} > 0$ units of data, it adds the data it produced and the data it purchased, $n_{i,t} + \delta_{i,t}$ units of data, to its stock of data.

Let the price of one piece of data be denoted π_t.

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

regulations that prevent multiple organizations from using some types of personal data. Another interpretation of this assumption is that there is a transaction cost of trading data, proportional to the data value.

Data Adjustment and the Stock of Knowledge The information set of firm i when it chooses its technique $a_{i,t}$ is[2] $\mathcal{I}_{i,t} = \{\mathcal{I}_{i,t-1}, \{s_{i,t-1,m}\}_{m=1}^{\omega_{i,t-1}}, A_{i,t-1}\}$, where $\omega_{i,t-1}$ is the net number of data points added (or subtracted if ω is negative), after accounting for data purchases or sales. 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, $\mathbb{E}[\theta_t|\mathcal{I}_{i,t}] - \theta_t$, 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.

Our data adjustment cost Ψ captures the idea that if a firm that does not store or analyze any data wants to transform itself to a machine learning powerhouse, it would require new computer systems, workers with different skills, and learning by the management team. As a practical matter, if there is no data adjustment cost, a firm would immediately purchase the optimal amount of data, just as in models of capital investment without capital adjustment costs. Data adjustment costs are important because they make dynamics gradual.

[2]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.

Equilibrium definition. 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 $\mathcal{I}_{i,t}$.

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. By assumption, the inverse demand function and the industry quality-adjusted supply are:

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

Firms take the industry price P_t and the parameter $\bar{P}$ as given. Price is not random because, by the central limit theorem, the aggregate or average A converges to a known value.[3] The data price π_t equates data demand and supply. As in Solow (1956), we take the rental rate of capital as given. This reveals the data-relevant mechanisms as clearly as possible. This could be an industry or a small open economy, facing a world rate of interest r.

1.2 Interpreting Model Assumptions

Alternatives to data as a forecasting tool. In this model, the defining feature of data is that it is a tool to forecast a future state θ_{t+1}. This is not the only way to represent data. As mentioned before, some papers model more data as a direct contribution to TFP, which may well be a useful shorthand for data that is an input into R&D. Another approach to modeling data is as an improved matching technology. It could improve the match between customers and goods or between workers and tasks. Matching and noisy information are not separate phenomena. They are two ways of

[3] 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 and that this implies a deterministic price and aggregate output.

representing an information friction. So, this could be a matching model. In this case, the noisy signal model was a more tractable formulation.

Can data be sold multiple times? Our setting allows this. Whether a firm sells d data points or sells 1 data points d times makes no difference, as long as ι of knowledge is lost, each time a firm sells a data point.

Investing in data quality. If a firm can pay for a higher z data processing ability, then this will further accentuate the data feedback loop and increasing returns. Larger firms with more transactions to process will get a higher marginal benefit from better data technology and will acquire even more knowledge than small firms. While that additional channel is interesting and may be quantitatively important, it doesn't change any of the ideas we develop in this paper. Therefore, we hold z fixed for simplicity.

Why this formulation of quality? The function g is an arbitrary function. It makes sense to assume it is decreasing because otherwise, worse forecasts improve quality. But the argument of the g function is quadratic in the difference between actions and optimal actions. This quadratic form is an approximation to many relationships. It has a long history in tracking problems like this one where agents use information to forecast a changing state. In economics, this quadratic loss formulation is used extensively in the global games literature, following Morris and Shin (2002). Most importantly, this formulation simplifies the solution because it ensures that conditional variance is a sufficient statistic for mapping what a firm knows to their value function.

1.3 Solution: Optimal Technique and Expected Quality

A key to simplifying the problem to a one-state variable problem lies in understanding the expected quality that results from the optimal choice of technique.

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 $\Omega_{i,t}^{-1} + \sigma_a^2$ and variance $2(\Omega_{i,t}^{-1} + \sigma_a^2)$, the expected quality of firm i's good at time t in (2) can be approximated as

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

If the g function is not too convex, then quality is a deceasing function of expected forecast errors. Or put simply, more data precision increases the quality of a firm's good. We will return to the question of highly convex, unbounded g functions in the next section.

Importantly, only the variance (or precision) of signals enters in expected utility, 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.

2 Valuing and Depreciating Data

Before exploring predictions of the model, we work out what this model structure teaches us about how data should be depreciated and valued.

2.1 Depreciating Data

Solving our dynamic model requires taking a stand on the depreciation rate of data. This depreciation rate estimation is of independent interest. For the most valuable firms in the world, data is arguably their most valuable asset. Yet, data valuation and data accounting are in their infancy. A key question for valuing data is assessing how quickly data depreciates.

Luckily, our model also points us to a method for quantifying depreciation. It teaches us that the depreciation rate of data is a particular function of the persistence and volatility of the environment that data is used to forecast. We derive and explain this depreciation formula, which can be used in this model, or in any environment where data is used for forecasting and where a linear and

normal stochastic environment is a reasonable approximation.

To derive this depreciation formula, we start from the state evolution equation. Recall that it is an AR(1): $\theta_{t+1} = \bar{\theta} + \rho(\theta_t - \bar{\theta}) + \epsilon_{t+1}$. Consider the beliefs about the time-t state and how they change when the same information is used to forecast the $t+1$ state. At the start of date t, the conditional variance of beliefs about the state θ_t is $V[\theta_t|\mathcal{I}_t] := \Omega_t^{-1}$, where Ω_t is what we've called the "stock of knowledge" and is the object we want to depreciate.

Next, we simply apply the same conditional variance operator, with the same information set, to the AR(1) equation above: $V[\theta_{t+1}|\mathcal{I}_t] = \rho^2 V[\theta_t|\mathcal{I}_t] + \sigma_\epsilon^2 = \rho^2 \Omega_t^{-1} + \sigma_\epsilon^2$. This holds in the absence of learning any additional information about the state during all of period t. In this no date-t learning case, we invert the variance and rearrange $V[\theta_{t+1}|\mathcal{I}_t]^{-1}$ to get:

$$\Omega_{t+1}^{no\ learning} = \frac{\Omega_t}{\rho^2 + \sigma_\epsilon^2 \Omega_t}.$$

To be clear, this is not the correct law of motion for the state Ω in this model because firms learn new information every period. But examining the no-learning case is instructive because the only thing changing the stock of knowledge from one period to the next is depreciation. While typically, one would depreciate a capital stock by multiplying capital k_t times a term like $(1-\delta^k)$. The equivalent multiplicative term here is $(\rho^2 + \sigma_\epsilon^2 \Omega_t)^{-1}$, which multiplies Ω_t. Thus, the depreciation rate, the equivalent of δ^k in a capital accumulation model, is

$$\text{data depreciation rate} = 1 - \frac{1}{\rho^2 + \sigma_\epsilon^2 \Omega_t}$$

A larger fraction of the stock of knowledge is lost to depreciation when the state changes lots from one period to the next (high σ_ϵ^2), when there is lots of knowledge to begin with (high Ω_t), and when high persistence makes the state a more variable process (high ρ).[4]

Depreciation rates are typically linear operators on the stock being depreciated. Appendix A.3 describes three types of economies where the data depreciation rate will be well-approximated by

[4]One might wonder why this depreciation rate can be negative for small values of $\rho^2 + \sigma_\epsilon^2 \Omega_t$. These are cases where the firms is so uncertain that its conditional variance is higher than the unconditional variance of next period's outcomes. This is not a scenario that ever arises in our model. If an agent were so uncertain, then simple mean-reversion should reduce their uncertainty. This natural reduction in uncertainty, without any additional data, is what would show up as a negative rate of depreciation.

a standard-looking multiplicative constant term.

Accounting rules depreciate all data like software, by amortizing it over three years. That is a depreciation rate of 30% per year. Our results suggest that the depreciation rate of data may vary widely, depending on whether the data is used to forecast something more static, like consumer location or tastes, or something more ephemeral like equity order flow.

2.2 A Law of Motion for Data

To get from this depreciation rate to the law of motion for the stock of knowledge requires adding new data from three sources: 1) data that was a by-product of production, 2) data that was bought or sold and 3) data that was inferred from a firm seeing its own quality at the end of the period. These pieces of information are incorporated into beliefs using Bayes' law.

The number of new data points generated by firm i's production, $n_{i,t}$ is assumed to be data mining ability times end of period physical output: $z_i k_{i,t}^{\alpha}$. Bayes law tells us that the posterior precision of a normal variable is the sum of the prior precisions and signal precisions. This means that the sum of the precisions of all the data points, $n_{i,t}\sigma_{\epsilon}^{-2}$, should be added to the stock of knowledge.

At the firm level, data inflows need to be adjusted for data trade. If a firm buys data ($\delta_{i,t} > 0$), we add all the newly-acquired data precision $\delta_{i,t}\sigma_{\epsilon}^{-2}$ to the stock of knowledge. If a firm sells data ($\delta_{i,t} < 0$), we subtract a fraction ι of that signal precision from their stock of knowledge. Since $\delta_{i,t}$ is negative, we add the negative number $\delta_{i,t}\sigma_{\epsilon}^{-2}$ to subtract off the lost knowledge.

Finally, the additional information learned from seeing one's own realization of quality $A_{i,t}$, at the end of period t has precision σ_a^{-2}.[5] This information is different from the purchased or produced data because the quality realization is a signal about θ_t, not about θ_{t+1}. Therefore, σ_a^{-2} gets added to the time-t stock of knowledge and depreciates, just like other time-t knowledge, that the firm takes with it to time $t+1$.

Lemma 1 puts the data depreciation and data inflows together. It tells us how the stock of knowledge evolves from one period to the next.

[5] We are approximating a mixture of normals here with a normal variable, whose mean and variance are equated to the mean and variance of the mixture variable. The mixture arises because the quadratic term inside the quality function g can be positive or negative. In many cases, the prior on θ_t will make one quadratic root much more likely than the other. In such cases, the mixture is almost a pure normal with $\sigma_a = \sigma_A$. See Appendix A for more details.

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. The proof is an application of Bayes' law, or equivalently, the Ricatti equation of a modified Kalman filter. Because the information structure is similar to that of a Kalman filter, the sequence of conditional variances, or their inverse, the sequence of precisions, is deterministic.

Non-rivalry as a negative bid-ask spread. Data non-rivalry adds a kink to the effective price of data, which shows up as indicator functions in (9). This 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.

Appendix A shows how to reformulate the problem where the choice variable is the amount of data added to a firm's stock of knowledge: $\omega_{i,t} = n_{i,t} + \delta_{i,t}$ for data purchases ($\delta_{i,t} > 0$) and $\omega_{i,t} = n_{i,t} + \iota\delta_{i,t}$ for data sales ($\delta_{i,t} < 0$). Then we can define an adjusted price of data, $\tilde{\pi}_{i,t} \equiv \pi/(\mathbf{1}_{\omega_{i,t}>n_{i,t}} + \iota\mathbf{1}_{\omega_{i,t}<n_{i,t}})$. When a firm sells data, $\tilde{\pi}$ is divided by $\iota < 1$, which raises the adjusted 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 forfeited. This insight allows us to import modeling and measurement tools from finance for a market equilibrium, with a bid-ask spread.

Information from aggregate prices. One might wonder why firms do not also learn from seeing aggregate price and the aggregate output. They reflect aggregate quality, which depends on 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 $a_{i,t}$. Since firms' actions are not observable, aggregate prices or quantities reveal

what other firms predicted well. But they convey no useful information about whether θ_t is high or low.

2.3 Valuing Data: A Recursive Representation

One of the most important valuation questions for modern economists, investors and accountants is how to value data. While some data is transacted and might be valued at its price, lots of data is retained by a firm, for its own use. A value function approach assigns a value to a firm with a given amount of data. While that is not a cookbook recipe for assigning a dollar value to data, it offers a first step, a clear way to think about data value and what its components are. Our value function can guide data valuation, in the same way that capital value functions have guided economists' measurement of capital values, for decades.

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:*

$$V(\Omega_{i,t}) = \max_{k_{i,t},\delta_{i,t}} P_t \mathbb{E}[A_{i,t}|\mathcal{I}_{i,t}]k_{i,t}^{\alpha} - \Psi(\Delta\Omega_{i,t+1}) - \pi_t\delta_{i,t} - rk_{i,t} + \left(\frac{1}{1+r}\right)V(\Omega_{i,t+1}) \tag{10}$$

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, $\Omega_{i,t}$, the stock of knowledge. 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.

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.

3 Short Run Features of a Data Economy

A key source of difference between a capital-based and a data economy is the 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. While that sounds positive, it also creates the possibility of a firm growth trap, with very slow growth and financial losses, early on the in the lifecycle of a new firm. As a result, the life-cycle path of book-to-market or Tobin's Q of data firms looks very different from capital-intensive firms. Finally, the fact that transactions generate data as a by-product explains why every exchange includes an element of barter, where goods are exchanged for data, frequently at a positive monetary price. But sometimes, the exchange of goods for data happens at a zero monetary price, in which case pure barter arises.

While these results may not be a surprising distance from our assumptions, they all demonstrate the ability of the framework to make sense of and re-interpret new data economy phenomena. Tools to model data phenomena can, in turn, be used to inform ongoing policy debates. Establishing that this is an economically-relevant collection of assumptions is important before using it for measurement or welfare analysis.

3.1 Increasing Returns in the Short Run

While our model is a general equilibrium one, it is instructive to explore one firm growing. While all others are in steady state, we drop in one, atomless, low-data (low $\Omega_{i,t}$) firm and observe its growth and transition to a high-data firm. From this exercise, we learn about the forces are work in the model. This exercise also teaches us about barriers to new firm entrants. Of course, equilibrium effects are crucially important. We study those in the next section. But the nature of the one-firm problem is so different from a typical macro model, that it is worth taking the time to understand it in isolation.

For this section, we adopt a linear quality function, for simplicity: $g(x) = \bar{A} - x$. We relax this asssumption later on, when we discuss the long run.

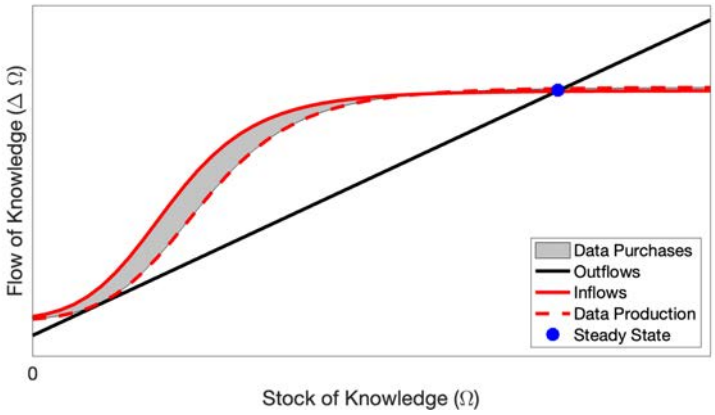

Figure 1: A single new firm grows slowly: Inflows and outflows of one firm's data.
Line labeled inflows plots the individual firm i version of the data inflows in Equation (13). Line labeled outflows plots the individual firm i version of the quantity in (14). Firm i is in an economy where all other firms are in steady state. Details in Appendix A.15.

Proposition 1 ***S-Shaped Accumulation of Knowledge*** *When all firms are in steady state, except for one firm i, then the firm's net data flow $\Omega_{i,t+1} - \Omega_{i,t}$*

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 $\hat{\hat{\Omega}}$.*

Figure 1 illustrates the inflows, outflows and dynamics of a single firm. This figure illustrates one possible economy. Data production may lie above or below the data outflow line. The difference between data inflows (solid line) and data production (dashed line) is data purchases. These purchases push the inflows line 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 1 traces out the S-shaped output path of a new entrant firm in this environment.

Firm size distribution One reason the S-shaped accumulation of data is interesting is that it also generates S-shaped path of firm size over time. When firms grow slowly in their early phase and then, only later grow quickly, this lends itself to a bifurcated firm size distribution. There are many new firms that are stuck small and data-poor. Then, there are firms that have reached the

explosive growth phase in the middle of the S-curve and grew large. In a world with increasing and then decreasing returns, firms do not remain mid-sized for long.

Single firms can have decreasing returns For some parameter values, the diminishing returns to data is always stronger than the data feedback loop. Proposition 8 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 4. But the appendix describes the set of parameters that make the data feedback loop sufficiently strong, to make data inflows convex at low levels of knowledge.

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:

$$\text{Profit}_t = P_t A_{i,t} k_{i,t}^{\alpha} - \Psi(\Delta\Omega_{i,t+1}) - \pi_t \delta_{i,t} - r k_{i,t}. \tag{11}$$

Proposition 2 ***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

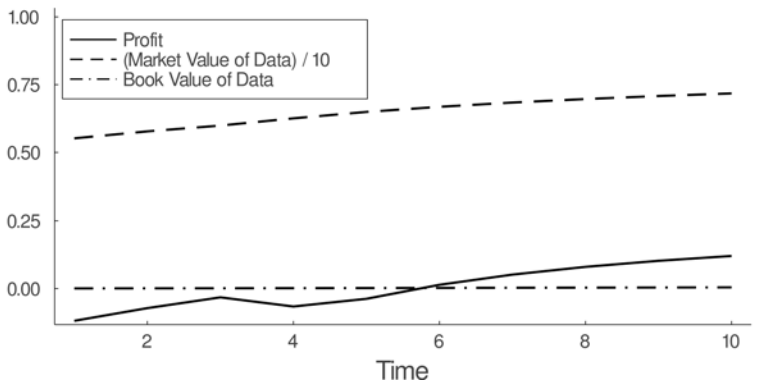

Figure 2: S-shaped growth can create initial profit losses and high market-to-book value of data. Market value is the value function of the firm, $V(\Omega_{i,t})$, divided by 10 to facilitate visualization. Knowledge stock defined in Lemma 1. Book value defined in (12). Parameters are discussed in Appendix A.15: $\rho = 1, r = 0.2, \beta = 0.97, \alpha = 0.3, \psi = 4, \overline{A} = 0.5, \sigma_a^2 = 0.05, \sigma_\theta^2 = 0.5, \sigma_\epsilon^2 = 0.1, z = 0.01, \pi = 0.002, P = 1, \iota = 1$

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 1. The two mechanisms of building stock of knowledge 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

$$\text{Book Value}_t = \sum_{\tau=0}^{t} \beta^{t-\tau} \pi_\tau \delta_\tau \mathbf{1}_{\delta_{i,\tau}>0}. \tag{12}$$

The market value of the data is the Bellman equation value function $V(\Omega)$ in (10).

Figure 2 plots the market value, 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 2, 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.

Figure 2 also 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.

3.3 Data Barter and Missing GDP

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.

The analysis also reveals that not only are zero-price transactions, like free apps, being missed, every transaction, in principle, has a data barter element to it. Every firm should charge slightly less for every product, because of the value of the data that accompanies its sale. 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.

Proposition 3 ***Bartering Goods for Data*** *It is possible that a firm will optimally choose positive production $k^{\alpha}_{i,t} > 0$, even if its price per unit is zero: $P_t = 0$.*

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. Our framework allows us to assign a value to such barter trades and partial-barter trades, despite their zero monetary price.

These results could enable better measurement of GDP. Investment in a stock of valuable knowledge is missing from aggregate measures of economic activity. Even if we cannot observe the data-adjusted true price of a transaction, if we can measure the value of the asset being generated, we can fill in this missing value. The value of the knowledge asset generated by all this barter trade is $V(\Omega_{i,t}) - V(\Omega_{i,t-1})$, for each firm i. Typical numerical approaches to approximating a value function could be applied to $V(\Omega_{i,t})$. Alternatively, one might use revenue data, use hiring and wages of workers who maintain data stocks and work with data, or look for the covariance of a firms' choices with the random variables it needs to forecast. A detailed discussion of the myriad of approaches to measure this value function is beyond the scope of this paper. However, frameworks like this are important inputs into digital economy measurement because they guide our thinking about what is missing and how to infer this missing aggregate economic activity.

4 Long-Run Features of a Data Economy

While the previous section emphasized the contrasts, this section highlights ways in which the long-run in this data economy is surprisingly similar to a capital-based production economy. Within the model, there is no long run growth because data has diminishing returns, a property documented empirically by (Bajari et al., 2018). Since data is a non-rival asset that contributes to productivity, one might question why it is fundamentally different from new ideas. To explore this, we describe a general class of models in which the accumulation of data used for forecasting does create long-run growth. This class of models has some unusual properties. To sustain long-run growth without innovation, perfect foresight must generate infinite output and all relevant future states must be deterministic. On the other hand, if we allow innovation, data that is an input into idea creation can easily generate growth. The take-away is that, just like with capital, we should distinguish between data that is used for research and development and data that is not.

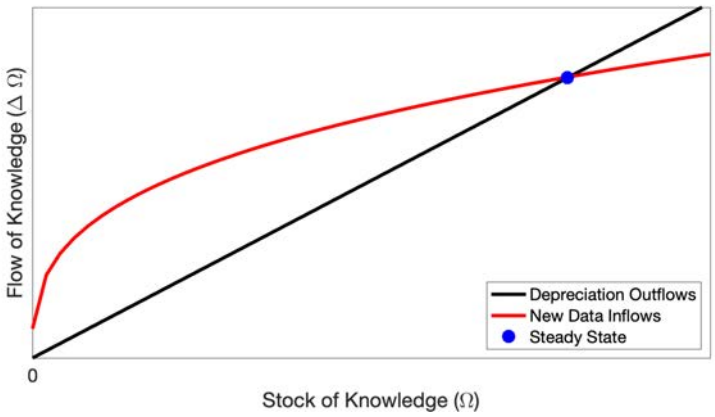

Figure 3: Economy converges to a data steady state: Aggregate inflows and outflows of data.
Line labeled inflows plots the quantity in (13) for the aggregate economy, for different levels of initial data stock and optimal capital choices. Line labeled outflows plots the quantity in (14). In all examples, 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: $\Delta\Omega_{i,t+1} = (\Omega_{i,t+1} - \Omega_{i,t})/\Omega_{i,t}$.

4.1 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. Consider an economy with firms that are symmetric in size, in data and in choice variables. Fora representative firm in this economy, we define the additions to the data stock that are generated by time-t economic activity to be inflows and define the total losses due to depreciation as outflows.

$$\text{Inflows:} \qquad \Omega_t^+ = \sigma_\epsilon^{-2} \int_i z_i k_{i,t}^\alpha di + \sigma_a^{-2} \tag{13}$$

$$\text{Outflows:} \qquad \Omega_t^- = \Omega_t + \sigma_a^{-2} - \int_i \left[(\rho^2(\Omega_{i,t} + \sigma_a^{-2}))^{-1} + \sigma_\theta^2\right]^{-1} di. \tag{14}$$

This is equivalent to the outflow and inflow for a representative firm i who operates in an economy populated with identical firms with no data trade.[6]

Figure 3 illustrates the inflows and outflows (eq.s 13 and 14), 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

[6]In an equilibrium where data is traded and $\iota < 1$, this would create an additional term in inflows that represents the net additions to the data stock from non-rivalry of traded data. Adding such data creation complicates the expressions and does not change any of the results in this section.

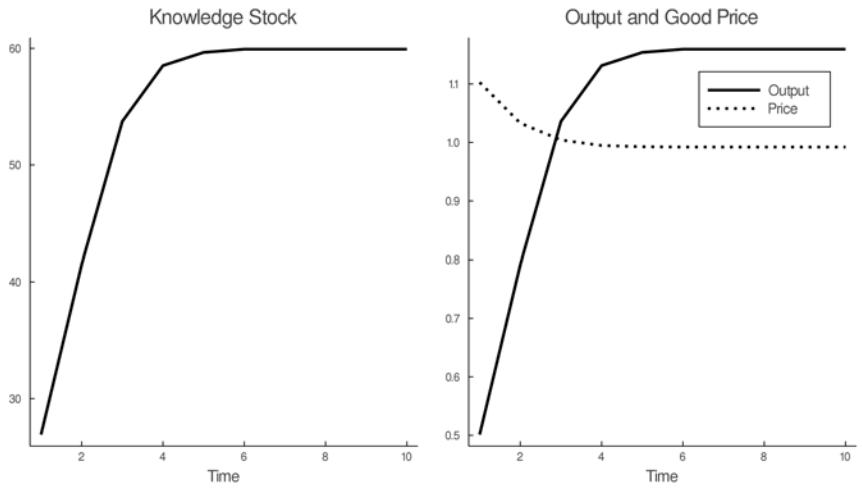

Figure 4: Aggregate growth dynamics: Data accumulation grows knowledge and output over time, with diminishing returns. Parameters: $\rho = 1, r = 0.2, \beta = 0.97, \alpha = 0.3, \psi = 0.4, \gamma = 0.1, \overline{A} = 1, \overline{P} = 1, \sigma_a^2 = 0.05, \sigma_\theta^2 = 0.5, \sigma_\epsilon^2 = 0.1, z = 5, \iota = 1$. See Appendix A.15 for parameter selection and numerical solution details.

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.

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 4. Without idea creation, data accumulation alone would generate slower and slower growth. Figure 1 differs from Figure 3 because it represents a single firm's transition, not the transition of a whole economy of symmetric firms, growing together. The difference between the two is the equilibrium price effect of goods prices and data prices.

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 and typically cannot even reach zero.

Long Run Growth Impossibility Results Is there some other model, without innovation, where forecast data accumulation can sustain growth? 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.[7] Both conditions are at odds with most theories.

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 4 ***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 $g(v) \to \infty$.*

Mathematically, this result is simple. If expected quality (g) does not approach infinity in the high-data limit, then output cannot become infinite. If output cannot be infinite, then it cannot grow at any rate $\underline{g} > 0$ forever.

However, 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.

For the next result, we first 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.

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

[7] 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.

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 5 ***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 4 does not. It says, even if you believe perfect one-period-ahead forecasts can produce infinite output, you still get diminishing returns because of the existence of fundamental, unlearnable randomness.

Thus, if one believes some events tomorrow are fundamentally random, data must 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.

4.2 General Equilibrium Effects

The difference between one firm entering when all other firms are in steady state (Proposition 1), and all firms growing together (Propositions 4 and 5), is prices. When all firms are data-poor, all goods are low quality. Since quality units are scarce, prices are high. The high price of 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 in the one-firm problem is that data may also be abundant, keeping the price of data low.

4.3 Endogenous 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 transforms our data economy version of Solow (1956), into a data-driven endogenous growth model. Of course, for this formulation to make sense, one needs to believe that information about who buys what can be used to discover growth-sustaining technologies. If that is true, then this augmented version of the model would be more useful for long-term economic forecasts.

Instead of Equation (2), assume the evolution of quality follows

$$A_{i,t} = A_{i,t-1} + \max\{0, \Delta A_{i,t}\} \qquad \text{with} \qquad \Delta A_{i,t} = \bar{A} - (a_{i,t} - \theta_t - \epsilon_{a,i,t})^2. \tag{15}$$

The solution inherits the same structure as before: the expected change in quality of firm i's good at time t is $\mathbb{E}[\Delta A_{i,t}|\mathcal{I}_{i,t}] = \bar{A} - \Omega_{i,t}^{-1} - \sigma_a^2$. The interpretation is that more data allows for more precisely targeted innovations, which increase the size of the technology advance. Depending on $\bar{A}$, data might make innovation viable ($\Delta A_{i,t}$ positive).

This extension teaches us that data used for research should be measured separately from data used for other purposes, just like economists typically do for capital expenditures.

5 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 then adds a business stealing externality, to examine the welfare properties of a data economy.

5.1 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

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

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, $\Phi_{it} := P_t A_{i,t} k_{i,t}^{\alpha} - \Psi(\Delta\Omega_{i,t+1}) - \pi_t \delta_{i,t} - r k_{i,t}$. 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:

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

Equilibrium An equilibrium is household choices of c_t and m_t that maximize (16), firm choices of capital $k_{i,t}$ and data $\delta_{i,t}$ that maximize (17) and prices P_t and π_t that clear markets. In addition, we restrict our attention to economies with λ, z_H and z_L such that there exists a symmetric, pure-strategy equilibrium, where all firms of the same type make the same choices; if $z_i = z_j$, then $\delta_{i,t} = \delta_{j,t}$ and $k_{i,t} = k_{j,t}$ $\forall t$.

These assumptions deliver the same inverse demand as in (7). But these foundations allow us to compare the decentralized equilibrium and optimal social planner outcomes, to see if inefficiencies arise.

Proposition 6 ***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 marginal social value. This aligns the private and social incentives for production.

5.2 Data for Business Stealing

Policy analysis needs to consider potential data externalities. For example, when data can be used for marketing or other forms of business stealing, firms' use of data harms others. Privacy costs

[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.

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 be made more relevant for welfare estimation and policy analysis.

Using data for business stealing can be represented through a quality externality:

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

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 capital choices, at a given set of prices. However, it does influence the aggregate quality of goods.

The baseline model is represented by $b = 0$. In this case, equations (18) and (2) are identical and there is no externality.

Alternatively, if $b > 0$, a firm's quality depends on the difference between the precision of its own prediction and the average precision of the predictions of all other firms. In other words, each firm achieves maximum quality when other firms have minimum quality. This captures the idea that when one firm uses data to market effectively, it reduces the ability of all other firms to generate value by reaching their preferred customers.

The extreme case where data does not have any social value is $b = 1$. The aggregate 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 7 ***Welfare with Business Stealing*** *If $b > 0$, there is over-investment in the steady state level of capital and excessive trade in the data market in equilibrium.*

Proposition 7 incorporates two distinct inefficiencies: excessive production and excessive data trade. While firms internalize the benefits of additional data from high production and additional revenue from data sales, they neglect the external effect: Higher data production and sales reduces

[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).

the quality of other firms' goods. Thus, in equilibrium, too much output is produced and too much data is traded.

The idea that firms sell too much data might appear counter-factual, since social networks and search engines do not primarily sell data directly. Instead, they use their data primarily to sell data services to their business customers. For example, Facebook revenue comes primarily from advertising, which is a data service. However, sales of data services is a type of data sales. A formal analysis of the equivalence between data services and data sales is in Admati and Pfleiderer (1990).

6 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 bifurcation in firm size? One possible source is the accumulation of data. The S-shaped dynamic of firm growth implies that firms stay small for a while and then grow large rapidly. During the convex, increasing returns portion of the growth trajectory, small initial differences in the initial data stock of firms get amplified into large differences in firm size.

Firm competition. Instead of assuming price taking behavior, one could model a finite number of firms that consider the price impact of their production decisions. In a static setting, the degree of imperfect competition modulates firms' use of data and firms' data affect the extent to which they compete (Eeckhout and Veldkamp, 2022). Working this out in a dynamic, recursive setting like this one, could give us insights about how data changes firms' dynamic 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. In the unknown state is a vector, then 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.

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, 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.

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.

Angeletos, George-Marios, Christian Hellwig, and Alessandro Pavan, "Signaling in a Global Game: Coordination and Policy Traps," *Journal of Political Economy*, 2006, *114* (3), 452–484.

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.

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

Bergemann, Dirk and Juuso Välimäki, "Experimentation in Markets," *The Review of Economic Studies*, 04 2000, *67* (2), 213–234.

Broer, Tobias, Alexandre Kohlhas, Kurt Mitman, and Kathrin Schlafmann, "Information and Wealth Heterogeneity in the Macroeconomy," *Working paper*, 2021.

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.

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

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

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

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

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.

Maćkowiak, Bartosz and Mirko Wiederholt, "Optimal sticky prices under rational inattention," *American Economic Review*, 2009, *99 (3)*, 769–803.

Matějka, Filip and Alisdair McKay, "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model," *American Economic Review*, January 2015, *105* (1), 272–98.

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.

Reis, Ricardo, "A Sticky-Information General Equilibrium Model for Policy Analysis," Working Papers Central Bank of Chile 495, Central Bank of Chile October 2008.

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

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, the firm takes as given its last-period beliefs, $\widehat{\mu}_{t-1} = \mathbb{E}\big[\theta_t \mid \mathcal{I}_{i,t-1}\big]$ and $\Omega_{t-1} = Var\big[\theta_t \mid \mathcal{I}_{i,t-1}\big]^{-1}$.

At the end of date t, each firm observes a signal derived from *observing their own output*. Date t output reveals good quality $A_{i,t} = y_{i,t}/k_{i,t}^{\alpha}$. Since the g function is known, it can be inverted. Thus, quality $A_{i,t}$ reveals $g^{-1}(A_{i,t})$, which is $(a_{i,t} - \theta_t - \epsilon_{a,i,t})^2$. The action $a_{i,t}$ is known to agent i. For the purposes of information content, this is like a constant that can be removed. That leaves a square of the sum of two unknown variables θ_t and $\epsilon_{a,i,t}$. The square has two roots, a positive and a negative one. Each root corresponds to a normal signal. That makes the signal inferred from quality a mixture of normals. None of the results that follow depend on the nature of the information revealed by output, as long as seeing one's output does not fully reveal θ_t. However, updating with mixtures of normals is not analytically tractable. Therefore, we use an approximation to solve the model. We replace the mixture normal signal with a normal signal that has its mean and variance adjusted to be equal to the mean and variance of the mixture variable. The variance of the mixture variable is what we call σ_a^2. It is the variance of the two underlying normal σ_A^2, plus a non-negative term that captures the dispersion in the means. In many cases, prior beliefs will put a large probability weight on either the positive or the negative root. In such cases, then this adjustment term will be small and $\sigma_a^2 \approx \sigma_A^2$.

In short, at the end of date t, every firm receives a noisy signal about θ_t, derived from observed output. That signal is approximated by

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

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

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

$$\mathbb{E}\big[\theta_t \mid \mathcal{I}_{t-1}, s_{i,t-1}^{a}\big] = \frac{\Omega_{t-1}\widehat{\mu}_{t-1} + \sigma_a^{-2} s_{i,t-1}^{a}}{\Omega_{t-1} + \sigma_a^{-2}}$$
$$V\big[\theta_t \mid \mathcal{I}_{t-1}, s_{i,t-1}^{a}\big] = \big[\Omega_{t-1} + \sigma_a^{-2}\big]^{-1}$$

Next, use the law of motion $\theta_{t+1} = \bar{\theta} + \rho(\theta_t - \bar{\theta}) + \eta_{t+1}$ and take the expectation on both sides of the equation to get: $\mathbb{E}\big[\theta_t \mid \mathcal{I}_{t-1}, s_{i,t-1}^{a}\big] = \bar{\theta} + \rho \cdot \big(\mathbb{E}\big[\theta_{t-1} \mid \mathcal{I}_{t-1}, s_{i,t-1}^{a}\big] - \bar{\theta}\big)$. If we take the variance of both sides of the equation, we get $V\big[\theta_t \mid \mathcal{I}_{t-1}, s_{i,t-1}^{a}\big] = \rho^2\big[\Omega_{t-1} + \sigma_a^{-2}\big]^{-1} + \sigma_\theta^2$.

The other type of signal the firm observes is *data points*. These are different from the s^a signal because they are about next period's state θ_{t+1}. Here, we introduce a new piece of notation: 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 $\omega_{i,t} = n_{i,t} = z k_{i,t}^{\alpha}$. 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 $\bar{s}_t = \theta_{t+1} + \epsilon_{s,t}$, where $\epsilon_{s,t} \sim \mathcal{N}(0, \sigma_\epsilon^2/\omega_{it})$.

Then, the final step is to use the mean and variance above as prior beliefs and use Bayes law to update them with the average signal $\bar{s}_t$:

$$\widehat{\mu}_t = \mathbb{E}\big[\theta_t \mid \mathcal{I}_t\big] = \frac{\Big[\rho^2\big[\Omega_{t-1} + \sigma_a^{-2}\big]^{-1} + \sigma_\theta^2\Big]^{-1} \cdot \mathbb{E}\big[\theta_t \mid \mathcal{I}_{t-1}, s_{i,t-1}^{a}\big] + \omega_t \sigma_\epsilon^{-2} \bar{s}_t}{\Big[\rho^2\big[\Omega_{t-1} + \sigma_a^{-2}\big]^{-1} + \sigma_\theta^2\Big]^{-1} + \omega_t \sigma_\epsilon^{-2}} \tag{20}$$

$$\Omega_t^{-1} = Var\big[\theta_t \mid \mathcal{I}_t\big] = \Big\{\Big[\rho^2\big[\Omega_{t-1} + \sigma_a^{-2}\big]^{-1} + \sigma_\theta^2\Big]^{-1} + \omega_t \sigma_\epsilon^{-2}\Big\}^{-1}. \tag{21}$$

Equations (20) and (21) 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 (21), 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:

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

Next, we 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 problem becomes

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

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

Since $\frac{\partial \Omega_{i,t,t+1}}{\partial \omega_{i,t}} = \sigma_\epsilon^{-2}$, the first order condition for the optimal $\omega_{i,t}$ is

$$FOC[\omega_{i,t}]: \quad -\Psi'(\cdot)\sigma_\epsilon^{-2} - \tilde{\pi} + \left(\frac{1}{1+r}\right) V'(\Omega_{i,t+1})\sigma_\epsilon^{-2} = 0 \tag{24}$$

where $\tilde{\pi} \equiv \pi/(\mathbf{1}_{\omega_{i,t}>n_{i,t}} + \iota \mathbf{1}_{\omega_{i,t}<n_{i,t}})$ is the price of data, adjusted for non-rivalry. It is lower for data sales since less data is lost per unit of data sold.

A.3 Lemma 3, 4, 5: Linearity of Data Depreciation

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

Lemma 3 ***Linear Data Outflow with Low Knowledge*** *$\exists \epsilon > 0$ such that $\forall \Omega_{i,t} \in B_\epsilon(0)$, data outflow is approximately linear and the approximation error is bounded from above by $\frac{\rho^4 \sigma_\theta^2}{1+\rho^2\sigma_\theta^2\sigma_a^{-2}} \frac{\epsilon^2}{1+\rho^2\sigma_\theta^2(\epsilon+\sigma_a^{-2})}$. The approximation error is small when ρ or σ_θ is small, or when $\Omega_{i,t}$ is very close to 0.*

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

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

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

Lemma 5 ***Linear Data Outflow with Abundant Knowledge*** *When $\Omega_{i,t} \gg \sigma_\theta^{-2}$, discounted data stock is very small relative to $\Omega_{i,t}$, so that data outflows are approximately linear. The approximation error is small when ρ is small or when σ_θ is sufficiently large.*

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

A.4 Deterministic Aggregate Output.

Why is there no expectation operator around aggregate output, profits or prices? 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.

A.5 Proof of Proposition 1: S-shaped Accumulation of Knowledge

We proceed in two parts: convexity and then concavity.

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

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

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

Since we assumed that Ψ is small, consider the case where $\psi = 0$. In this case, the data adjustment term drops out and the capital first-order condition reduces to

$$k^{1-\alpha}_{i,t} = \frac{\alpha}{r}\left(P_t A_{i,t} + z_i\sigma_{\epsilon}^{-2}\frac{1}{1+r}V'(\Omega_{i,t+1})\right). \tag{25}$$

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

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

Differentiating (25) with respect to $\Omega_{i,t}$ and using the relationships $\frac{\partial A_{i,t}}{\partial \Omega_{i,t}} = {\Omega_{i,t}}^{-2}$ and $\frac{\partial \Omega_{i,t+1}}{\partial \Omega_{i,t}} = \rho^2[\rho^2 + \sigma^2_{\theta}(\Omega_{i,t} + \sigma_a^{-2})]^{-2}$, yields

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

Therefore,

$$\begin{aligned}\frac{\partial d\Omega^+_{i,t}}{\partial \Omega_{i,t}} &= z_{i,t}k^{2\alpha-1}_{i,t}\sigma_{\epsilon}^{-2}\frac{\alpha^2}{(1-\alpha)r}\left(P_t{\Omega_{i,t}}^{-2} + z_i\sigma_{\epsilon}^{-2}\frac{1}{1+r}V''(\Omega_{i,t+1})\rho^2[\rho^2 + \sigma^2_{\theta}(\Omega_{i,t} + \sigma_a^{-2})]^{-2}\right) \\ &= z_{i,t}k^{2\alpha-1}_{i,t}\sigma_{\epsilon}^{-2}\frac{\alpha^2}{(1-\alpha)r}P_t{\Omega_{i,t}}^{-2} + z^2_{i,t}k^{2\alpha-1}_{i,t}\sigma_{\epsilon}^{-4}\frac{\alpha^2}{1-\alpha}\frac{1}{r(1+r)}V''(\Omega_{i,t+1})\rho^2[\rho^2 + \sigma^2_{\theta}(\Omega_{i,t} + \sigma_a^{-2})]^{-2}.\end{aligned} \tag{26}$$

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

$$\frac{\partial d\Omega^-_{i,t}}{\partial \Omega_{i,t}} = 1 - \frac{1}{\rho^2(\Omega_{i,t} + \sigma_a^{-2})^2(\sigma^2_{\theta} + \rho^{-2}(\Omega_{i,t} + \sigma_a^{-2})^{-1})^2}. \tag{27}$$

The derivatives of net data flow is then

$$\begin{aligned}\frac{\partial d\Omega^+_{i,t}}{\partial \Omega_{i,t}} - \frac{\partial d\Omega^-_{i,t}}{\partial \Omega_{i,t}} &= z_{i,t}k^{2\alpha-1}_{i,t}\sigma_{\epsilon}^{-2}\frac{\alpha^2}{(1-\alpha)r}P_t{\Omega_{i,t}}^{-2} + z^2_{i,t}k^{2\alpha-1}_{i,t}\sigma_{\epsilon}^{-4}\frac{\alpha^2}{1-\alpha}\frac{1}{r(1+r)}V''(\Omega_{i,t+1})\rho^2[\rho^2 + \sigma^2_{\theta}(\Omega_{i,t} + \sigma_a^{-2})]^{-2} \\ &\quad + \frac{1}{\rho^2(\Omega_{i,t} + \sigma_a^{-2})^2(\sigma^2_{\theta} + \rho^{-2}(\Omega_{i,t} + \sigma_a^{-2})^{-1})^2} - 1.\end{aligned} \tag{28}$$

For notational convenience, denote the first term in (28) 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

$$z_{it}k_{it}^{2\alpha-1}\frac{\alpha^2}{(1-\alpha)r}\sigma_\epsilon^{-2}\left[-2\bar{P}\Omega_{it}^{-3} - z_{it}\sigma_\epsilon^{-2}\frac{1}{1+r}\nu\rho^2\frac{2\sigma_\theta^2}{\left(\rho^2+\sigma_\theta^2\left(\Omega_{it}+\sigma_a^{-2}\right)\right)^3}\right] > 0 \tag{29}$$

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

$$f(\Omega_{i,t}) := -z_{it}\sigma_\epsilon^{-2}\tfrac{1}{1+r}\nu\rho^2\Omega_{it}^3\tfrac{\sigma_\theta^2}{\left(\rho^2+\sigma_\theta^2\left(\Omega_{it}+\sigma_a^{-2}\right)\right)^3} \tag{30}$$

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

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

$$h(\Omega_{i,t}) = 1 - M_3, \tag{31}$$

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 (26), 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.6 Proof of Proposition 2: New Firms Earn Negative Profits

Without any production or any data purchased, $\Omega_0 = \sigma_\theta^{-2}$, 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 (11), 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.7 Proof of Proposition 3: Firms Sell Goods at Zero Price (Data Barter)

Proof: 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 (10) 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.8 Data Accumulation Can be Purely Concave

Data accumulation is not always S-shaped, only for some parameter values. For others, it can be that data accumulation is purely concave. Instead, the net data flow (the slope) decreases with $\Omega_{i,t}$, right from the start.

Proposition 8 ***Concavity of Data Inflow*** *$\exists \epsilon > 0$ such that $\forall \Omega_{i,t} \in B_\epsilon(0)$, the net data flow decreases with $\Omega_{i,t}$ if $\sigma_\theta^2 > \sigma_a^2$.*

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

Step 1: Data outflows are approximately linear when $\Omega_{i,t}$ is small.
This is proven separately in Lemma 3.

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

Recall that the derivative of data inflows with respect to the current stock of knowledge Ω_t is $\frac{\partial d\Omega^+_{i,t}}{\partial \Omega_{i,t}} = \rho^2 \left[\rho^2 + \sigma_\theta^2(\Omega_{i,t} + \sigma_a^{-2})\right]^{-2} > 0$ (see the Proof of Proposition 1 for details). Thus

$$\frac{\partial d\Omega^+_{i,t}}{\partial \Omega_{i,t}} - \frac{\partial d\Omega^-_{i,t}}{\partial \Omega_{i,t}} \approx \rho^2 \left[\rho^2 + \sigma_\theta^2(\Omega_{i,t} + \sigma_a^{-2})\right]^{-2} - 1 + \rho^2(1 + \rho^2 \sigma_\theta^2 \sigma_a^{-2})^{-2}. \tag{32}$$

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.9 Proof of Proposition 4: 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 $t + s$, $\forall s > 0$:

$$Y_{t+s} = \int_i y_{i,t} di < M_y \int_i di < M_y N < \infty. \tag{33}$$

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 $t+s$, $\forall s > 0$, output growth is $\ln(Y_{t+s}) - \ln(Y_t) = [\ln(Y_{t+s}) - \ln(Y_{t+s-1})] + \cdots + [\ln(Y_{t+1}) - \ln(Y_t)] > \underline{g}s$. This implies that $Y_{t+s} > Y_t e^{\underline{g}s}$.. Thus for $\forall s > \underline{s} \equiv \lceil \frac{\ln(MN) - \ln(Y_t)}{\underline{g}} \rceil$,

$$Y_{t+s} > Y_t e^{\underline{g}s} > Y_t e^{\underline{g}\underline{s}} > Y_t e^{\underline{g}\frac{\ln(M_y N) - \ln(Y_t)}{\underline{g}}} = M_y N, \tag{34}$$

which contradicts (33).

A.10 Proof of Proposition 5: Data 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 4, 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.11 Competitive Equilibrium

In order to prove our welfare result, 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: $P_t = u'(c_t)$.

Firm Problem Firms' sequential optimization problem is

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

Equivalently, in recursive form

$$V(\Omega_{i,t}) = \max_{k_{i,t},\delta_{i,t}} P_t\mathbb{E}[A_{i,t}|\mathcal{I}_{i,t}]k_{i,t}^{\alpha} - \Psi(\Delta\Omega_{i,t+1}) - \pi_t\delta_{i,t} - rk_{i,t} + \frac{V(\Omega_{i,t+1})}{1+r} \tag{35}$$

$$\text{s.t. } \Omega_{i,t+1} = \left(\rho^2(\Omega_{i,t}+\sigma_a^{-2})^{-1} + \sigma_\theta^2\right)^{-1} + \left(z_i k_{i,t}^{\alpha} + \left(\mathbf{1}_{\delta_{i,t}>0} + \iota\mathbf{1}_{\delta_{i,t}<0}\right)\delta_{i,t}\right)\sigma_\epsilon^{-2} \tag{36}$$

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

Market Clearing (Resource Constraint)

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

The adjustment cost Ψ is incorporated in the market clearing/resource constraint for the numeraire good so that it shows up in the planner's objective function.

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 $P^{eq} = u'(c^{eq})$. In steady state, the market clearing conditions simplify to

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

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 $k_{i,t}$:

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

Substitute

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

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:

$$\alpha Pk_i^{\alpha-1}\Big(\mathbb{E}[A_i] + \frac{z_i\sigma_\epsilon^{-2}}{1+r}\Omega_i^{-2}g\prime k_i^{\alpha}\Big) = r. \tag{37}$$

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 $\pi_t = \frac{1}{1+r}V'(\Omega_{i,t+1})\sigma_\epsilon^{-2}(\mathbb{1}_{\delta_{i,t}>0} + \iota\mathbb{1}_{\delta_{i,t}<0})$, which can be rearranged as

$$V'(\Omega_{i,t+1}) = \frac{(1+r)\pi_t}{\sigma_\epsilon^{-2}(\mathbb{1}_{\delta_{i,t}>0} + \iota\mathbb{1}_{\delta_{i,t}<0})} \tag{38}$$

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:

$$V'(\Omega_{i,t}) = P_t k_{i,t}^{\alpha} \Omega_{i,t}^{-2} + \frac{1}{1+r} V'(\Omega_{i,t+1}) \frac{\partial \Omega_{i,t+1}}{\Omega_{i,t}}, \tag{39}$$

Differentiating Equation (36) with respect to $\Omega_{i,t}$,

$$\frac{\partial \Omega_{i,t+1}}{\partial \Omega_{i,t}} = \frac{\rho^2}{\left(\rho^2 + \sigma_\theta^2(\Omega_{i,t} + \sigma_a^{-2})\right)^2}. \tag{40}$$

Substitute Equation (38) for $V'(\Omega_{i,t}) = V'(\Omega_{i,t+1})$ (in steady state) in (39):

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

Next substitute for $V'(\Omega_{i,t})$ in Equation (38), using the expression for $\frac{\partial \Omega_{i,t+1}}{\partial \Omega_{i,t}}$ from Equation (40). 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:

$$\left(1 + r - \frac{\rho^2}{(\rho^2 + \sigma_\theta^2(\Omega_i + \sigma_a^{-2}))^2}\right) \frac{\pi}{P\sigma_\epsilon^{-2}(\mathbb{1}_{\delta_i>0} + \iota \mathbb{1}_{\delta_i<0})} = k_i^{\alpha} \Omega_i^{-2} \qquad i = L, H \tag{41}$$

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 (44).

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 (42) and (43) represent two equations each.

$$\Omega_i^{eq} = \left[\rho^2 (\Omega_i^{eq} + \sigma_a^{-2})^{-1} + \sigma_\theta^2\right]^{-1} + \left(z_i (k_i^{eq})^{\alpha} + \delta_i^{eq} (\mathbf{1}_{\delta_i^{eq}>0} + \iota \mathbf{1}_{\delta_i^{eq}<0})\right) \sigma_\epsilon^{-2} \qquad i = L, H \tag{42}$$

$$r = \alpha \bar{P} (c^{eq})^{-\gamma} (k_i^{eq})^{\alpha-1} \left[\mathbb{E}[A_i^{eq}] + \frac{z_i \sigma_\epsilon^{-2}}{1+r} (k_i^{eq})^{\alpha} (\Omega_i^{eq})^{-2}\right] \qquad i = L, H \tag{43}$$

$$\frac{\frac{(k_L^{eq})^{\alpha}}{(\Omega_L^{eq})^2}}{\iota \frac{(k_H^{eq})^{\alpha}}{(\Omega_H^{eq})^2}} = \frac{1 + r - \frac{\rho^2}{(\rho^2 + \sigma_\theta^2 (\Omega_L^{eq} + \sigma_a^{-2}))^2}}{1 + r - \frac{\rho^2}{(\rho^2 + \sigma_\theta^2 (\Omega_H^{eq} + \sigma_a^{-2}))^2}} \tag{44}$$

$$\lambda \delta_L^{eq} + (1 - \lambda) \delta_H^{eq} = 0. \tag{45}$$

Equation (42) represents the two law of motions for stock of knowledge, one for each type of firm $i = L, H$. Equation (43) comes from (37) with $P' = u(c)$ and $u'(c) = \bar{P} c^{-\gamma}$ substituted in. It represents the two first order conditions for capital choice, one for each type of firm $i = L, H$. (44) 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 (45) is the resource constraint for the total traded data, which should be zero.

A.12 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

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

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

$$
r\lambda_i = \frac{\partial u(c_t)}{\partial k_{i,t}} + \frac{1}{1+r}\frac{\partial u(c_{t+1})}{\partial k_{i,t}} \qquad \text{for } \ i = L, H \tag{46}
$$

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, $c_t = \lambda\mathbb{E}[A_{L,t}]k_{L,t}^{\alpha} + (1-\lambda)\mathbb{E}[A_{H,t}]k_{H,t}^{\alpha}$.Thus

$$
\begin{aligned}
\frac{\partial c_t}{\partial k_{i,t}} &= \alpha\lambda_i\mathbb{E}[A_{i,t}]k_{i,t}^{\alpha-1} \\
\frac{\partial c_t}{\partial \Omega_{i,t}} &= \lambda_i\frac{\partial\mathbb{E}[A_{i,t}]}{\partial\Omega_{i,t}}k_{i,t}^{\alpha} = \lambda_i\Omega_{i,t}^{-2}g\prime k_{i,t}^{\alpha}
\end{aligned} \tag{47}
$$

In steady state, substitute in the expressions above into (46),

$$
r = \alpha\bar{P}(c^{opt})^{-\gamma}(k_i^{opt})^{\alpha-1}\left[\mathbb{E}[A_i^{opt}] + \frac{z_i\sigma_\epsilon^{-2}}{1+r}(k_i^{opt})^{\alpha}(\Omega_i^{opt})^{-2}g\prime\right] \qquad i = L, H \tag{48}
$$

This is the same as Equation (43). 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 $\Omega_{i,t}$, $i = L, H$. To solve for V_i^P in steady state, differentiate the value function and apply the envelope condition to get:

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

The data first order condition reveals that the Lagrange multiplier η_t on the data constraint is

$$
\lambda_i\eta_t = \frac{1}{1+r}V_i^P(\Omega_{L,t+1},\Omega_{H,t+1})\sigma_\epsilon^{-2}(\mathbb{1}_{\delta_{i,t}>0} + \iota\mathbb{1}_{\delta_{i,t}<0}). \tag{49}
$$

In steady state, $V_i^p(\Omega_{i,t},\Omega_{-i,t}) = V_i^P(\Omega_{i,t+1},\Omega_{-i,t+1})$. Use this equality and Equations (40) and (49) 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

$$
\left(1 + r - \frac{\rho^2}{(\rho^2 + \sigma_\theta^2(\Omega_{i,t}+\sigma_a^{-2}))^2}\right)\frac{\eta_t\lambda_i}{\sigma_\epsilon^{-2}(\mathbb{1}_{\delta_{i,t}>0} + \iota\mathbb{1}_{\delta_{i,t}<0})} = \frac{\partial u(c_t)}{\partial\Omega_{i,t}} \qquad i = L, H \tag{50}
$$

which in steady state can be written as

$$
\left(1 + r - \frac{\rho^2}{(\rho^2 + \sigma_\theta^2(\Omega_i^{opt}+\sigma_a^{-2}))^2}\right)\frac{\eta\lambda_i}{\bar{P}(c^{opt})^{-\gamma}\sigma_\epsilon^{-2}(\mathbb{1}_{\delta_i^{opt}>0} + \iota\mathbb{1}_{\delta_i^{opt}<0})} = \lambda_i(k_i^{opt})^{\alpha}(\Omega_i^{opt})^{-2}g\prime \qquad i = L, H \tag{51}
$$

In steady state, H firms sell data. For them, $(\mathbb{1}_{\delta_i^{opt}>0} + \iota\mathbb{1}_{\delta_i^{opt}<0}) = \iota$. In steady state, L firms buy data. For them, $(\mathbb{1}_{\delta_i^{opt}>0} + \iota\mathbb{1}_{\delta_i^{opt}<0}) = 1$. Next take the ratio of the H and L conditions from (51). $(c^{opt})^{-\gamma}$ and the Lagrange multiplier η both drop out of the resulting equation, thus we have

$$\frac{\frac{(k_L^{opt})^\alpha g'}{(\Omega_L^{opt})^2}}{\iota\frac{(k_H^{opt})^\alpha g'}{(\Omega_H^{opt})^2}} = \frac{1+r-\frac{\rho^2}{(\rho^2+\sigma_\theta^2(\Omega_L^{opt}+\sigma_a^{-2}))^2}}{1+r-\frac{\rho^2}{(\rho^2+\sigma_\theta^2(\Omega_H^{opt}+\sigma_a^{-2}))^2}}, \tag{52}$$

which is the same as Equation (44).

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.13 Proof of Proposition 6: Efficient Equilibrium

The decentralized equilibrium is characterized by Equations (42) for $i = L, H$, (44), (45), and (43) for $i = L, H$.

The social planner's optimum is characterized by Equations (42) for $i = L, H$ and (45) (all for optimum variables), Equation (48) for $i = L, H$, and Equation (52).

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 Proof of Proposition 7: Inefficiency with Business Stealing

With business stealing externality, i.e. when $b = 1$, the only difference is that A_i is determined by Equation (18). Thus in a symmetric allocation, with 2 types, where all firms of type i are the same, in equilibrium we have

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

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 (42) for $i = L, H$, (44), (45), and (43) for $i = L, H$.

For the optimum, Equations (42) for $i = L, H$ and (45) 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's Optimal 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:

$$\begin{aligned}\frac{\partial c_t}{\partial \Omega_{i,t}} &= \lambda_i k_{i,t}^\alpha \frac{\partial \mathbb{E}[A_{i,t}]}{\partial \Omega_{i,t}} + (1-\lambda_i)k_{-i,t}^\alpha \frac{\partial \mathbb{E}[A_{-i,t}]}{\partial \Omega_{i,t}} = \lambda_i(1-\lambda_i)k_{i,t}^\alpha \Omega_{i,t}^{-2} - (1-\lambda_i)^2 k_{-i,t}^\alpha \Omega_{i,t}^{-2} \\ &= (1-\lambda_i)\Omega_{i,t}^{-2}\left(\lambda_i k_{i,t}^\alpha - (1-\lambda_i)k_{-i,t}^\alpha\right).\end{aligned} \tag{53}$$

Comparing Equations (47) and (53) 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 (51) and (52) change to

$$\left(1+r-\frac{\rho^2}{(\rho^2+\sigma_\theta^2(\Omega_i^{opt}+\sigma_a^{-2}))^2}\right)\frac{\eta\lambda_i}{\bar{P}(c^{opt})^{-\gamma}\sigma_\epsilon^{-2}(\mathbb{1}_{\delta_i^{opt}>0}+\iota\mathbb{1}_{\delta_i^{opt}<0})} = (1-\lambda_i)(\Omega_i^{opt})^{-2}\left(\lambda_i(k_i^{opt})^\alpha - (1-\lambda_i)(k_{-i}^{opt})^\alpha\right) \quad \forall i \tag{54}$$

$$\left(\frac{1-\lambda}{\lambda}\right)^2 \frac{\frac{\lambda(k_L^{opt})^\alpha + (1-\lambda)(k_H^{opt})^\alpha}{(\Omega_L^{opt})^2}}{\iota\frac{(1-\lambda)(k_H^{opt})^\alpha + \lambda(k_L^{opt})^\alpha}{(\Omega_H^{opt})^2}} = \frac{1+r-\frac{\rho^2}{(\rho^2+\sigma_\theta^2(\Omega_L^{opt}+\sigma_a^{-2}))^2}}{1+r-\frac{\rho^2}{(\rho^2+\sigma_\theta^2(\Omega_H^{opt}+\sigma_a^{-2}))^2}}, \tag{55}$$

Equation (55) is different from equilibrium Equation (44).

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's Optimal Capital with Business Stealing The first order condition for the planner's capital choice becomes $r\lambda_i = \partial u(c_t)/\partial k_{i,t} + \frac{1}{1+r}\partial u(c_{t+1})/\partial k_{i,t}$ for $i = L, H$. Substituting in the same expressions for marginal utility as before yields

$$r = \alpha \bar{P}(k_i^{opt})^{\alpha-1}(c^{opt})^{-\gamma}\left[\mathbb{E}[A_i^{opt}] + \frac{z_i\sigma_\epsilon^{-2}(1-\lambda_i)}{1+r}\Big((k_i^{opt})^\alpha - \frac{1-\lambda_i}{\lambda_i}(k_{-i}^{opt})^\alpha\Big)(\Omega_i^{opt})^{-2}\right]. \qquad i = L, H \tag{56}$$

Equation (56) is different from Equation (43).

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.

A.15 Numerical Examples

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

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

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.

Computational Procedure Figure 4 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 2 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.