A Growth Model of the Data Economy

Maryam Farboodi* and Laura Veldkamp†

October 29, 2019

Abstract

The rise of information technology and big data analytics has given rise to "the new economy." But are its economics new? This article constructs a classic growth model with data accumulation. Data has three key features: 1) Data is a by-product of economic activity; 2) data enhances firm productivity; and 3) data is information used for resolving uncertainty. The model can explain why data-intensive goods or services, like apps, are given away for free, why firm size is diverging, and why many big data firms are unprofitable for a long time. While these transition dynamics differ from those of traditional growth models, the long run features diminishing returns. Just like capital accumulation, data accumulation alone cannot sustain growth. Without improvements in non-data-productivity, data-driven growth will grind to a halt.

Does the new information economy have new economics, in the long run? When the economy shifted from agrarian to industrial, economists focused on capital accumulation and removed land from production functions. As we shift from an industrial to a knowledge economy, the nature of inputs is changing again. In the information age, production increasingly revolves around information and, specifically, data. Many firms are valued primarily for the data they have accumulated. As of 2015, global production of information and communications technology (ICT) goods and services was responsible for 6.5% of global GDP, and 100 million jobs (United Nations, 2017). Collection and use of data is as old as book-keeping. But recent innovations in computing and artificial intelligence (AI) allow us to use more data more efficiently. How will this new data economy

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

†Columbia Graduate School of Business, NBER, and CEPR, 3022 Broadway, New York, NY 10027; lv2405@columbia.edu.
PAPER IS PRELIMINARY AND INCOMPLETE.

evolve? Because data is non-rival, increases productivity and is freely replicable (has returns to scale), current thinking equates data growth with idea or technological growth. This article uses a simple framework to argue the contrary, that data accumulation is more like capital accumulation, which, by itself, cannot propel growth in the long run.

Data is information that can be encoded as a binary sequence of zeroes and ones. That broad definition includes literature, visual art and technological breakthroughs. We are focusing more narrowly on big data because that is where the technological breakthroughs have taken place that have spawned talk of a new information age or economy. Machine learning or artificial intelligence are prediction algorithms. They predict the probability of a customer buying, a picture being a cat, or anything else. Much of the big data firms use for these predictions are transactions data. It is personal information about online buyers, satellite images of traffic patterns near stores, textual analysis of user reviews, click through data, and other evidence of economic activity. Such data is used to forecast sales, earnings and the future value of firms and their product lines. Data is also used to advertise, which may create social value or might simply steal business from other firms. We will consider both possibilities. But the essential features of the data we consider are that it is user-generated and that it is used to predict uncertain outcomes.

Data and technology are not the same. They entail different benefits and different costs. Transactions data is information used to forecast unknown outcomes. The benefits of forecasting and inventing technologies differ: Because perfect foresight models never feature infinite profits, perfect forecasts must yield finite gains. If the gain to a perfect forecast is finite, the return to better and better forecasts must be diminishing. Production costs also differ. Producing new technology requires resources: skilled labor, a laboratory and prototypes. In contrast, data is a by-product of economic activity. Producing and selling generates data about the volume of sales, the means of payment, and characteristics of buyers. Sometimes collecting and processing the data to extract knowledge is costly. But data itself is not produced in a lab. More data comes from more economic activity. This difference in production matters. One of the fundamental insights of Romer (1990) is that monopolies are necessary to incentivize idea production. This is not true of data production. Because data is a by-product of economic transactions and data is less prone to leakage, no extra incentives are needed for its production.

The key features of the model reflect these differences. In the model, data is information, generated by past economic activity, used to forecast future unknown states. Like all information, it has returns to scale. Data can be used to make one unit of capital more valuable, or one thousand units more productive. The more productive capacity the data is matched with, the greater are the gains in output. In that sense, data is similar to technology. At the same time , the key insight of the model, that causes the data economy to resemble the manufacturing economy, is that data has diminishing returns. The diminishing returns arise from the fact that forecast errors give rise to profit losses. Data can mitigate those losses. But the best possible outcome is that data reduces forecast errors to zero. With perfect forecasting, zero operational mistakes, profits are large, but not infinite. In fact, many macro models have no uncertainty. Such environments are infinite-data limit economies.

Put differently, data cannot sustain long-run growth because data, like all information, is a means of reducing uncertainty. Uncertainty is bounded below by zero. Unless a perfect forecast gives a firm access to a pure, real, limitless arbitrage, the perfect forecast generates finite payoff. An arbitrage of real goods, at the aggregate level, is as real as alchemy. It is not possible. Thus, if the payoff to data is bounded above, the returns, at some point, must diminish, so as not to exceed that upper bound.

The frictionless data model looks much like a simple Solow (1956) model. There are inflows of data from new economic activity and outflows, as data depreciates. The depreciation comes from the fact that the state is constantly evolving. Firms are forecasting a moving target. Economic activity many periods ago was quite informative about the state at the time. However, since the state has random drift, such old data is less informative about what the state is today. When data is scarce, little is lost due to depreciation. As data stocks grow large, depreciation losses are substantial. The point at which data depreciation equals the inflow of new data is a data steady state. Firms with less data than their steady state grow in data, and therefore in productivity and investment. If a firm ever had more data than its steady state level, it should shrink. But without any other source of growth in the model, data-driven growth, like capital-driven growth eventually grinds to a halt.

Our result should not be interpreted to mean that data does not contribute to growth. It absolutely does, in the same way that capital investment does. If non-data-technology (referred to hereafter as "productivity" or "technology") continues to improve, data helps us find the most efficient uses of these new technologies. The accumulation of data may even reduce the costs of technological innovation by reducing its uncertainty, or increase the incentives for innovation by increasing the payoffs. The point is simply that, even if data is non-rival, freely replicable and productive, if it is used for forecasting, as most big data is, it cannot sustain infinite growth. We still need innovation for that.

When data has adjustment costs and purchased data is not as relevant as a firm's own data, a firm can remain data poor for a long time. This is similar to a growth poverty trap, but at the firm-level. Industries, or countries concentrated in a few industries may also get stuck in this trap. Thus, data could explain some of the increase in inequality, across firms, industries, or even countries that adopt data technologies at different rates.

To understand these trends, a theoretical framework is essential. Econometrics alone predicts the future based on past trends or correlations. In the midst of a structural transformation, when covariances are changing, such extrapolations are likely to be flawed. Therefore, we use a model to guide our thinking about which changes are logical outcomes, and which are not. The model also offers guidance for measurement. Measuring and valuing data is complicated by the fact that frequently, data is given away, in exchange for a free digital service. Our model makes sense of this pricing behavior and assigns a value to goods and data that have a zero transactions price. In so doing, it moves beyond price-based valuation, which often delivers misleading answers when valuing digital assets.

Related literature. Work on information frictions in business cycles, (Veldkamp (2005), Ordonez (2013) and Fajgelbaum et al. (2017)) have early versions of a data-feedback loop whereby more data enables more production, which in turn, produces more data. In each of these models, information is a by-product of economic activity; firms use this information to reduce uncertainty and guide their decision-making. These authors did not call the information data. But it has all the hallmarks of modern transactions data. The data produced was used by firms to forecast the state of the business cycle. Better forecasting enabled the firms to invest more wisely and be more profitable. These models restricted attention to data about aggregate productivity. In the data economy, that is not primarily what firms are using data for. But such modeling structures can be adapted so that production can also generate firm- or industry-specific information. As such, they provide useful equilibrium frameworks on which to build a data economy.

In the growth literature, our model builds on Jones and Tonetti (2018). They explore how different data ownership models affect the rate of growth of the economy. In both models, data is a by-product of economic activity and therefore grows endogenously over time. What is different here is that data is information, used to forecast a random variable. In Jones and Tonetti (2018), data contributes directly to productivity. It is not information. A fundamental characteristic of information is that it reduces uncertainty about something. When we model data as information, not technology, the long-run predictions reverse. Instead of long-run growth, there is long-run stagnation.

Other authors consider the interaction of artificial intelligence (AI) and innovation. Agrawal et al. (2018) develops a combinatorial-based knowledge production function and embeds it in the classic Jones (1995) growth model to explore how breakthroughs in AI could enhance discovery rates and economic growth. Lu (2019) embeds self-accumulating AI in a Lucas (1988) growth model and examines growth transition paths from an economy without AI to an economy with AI and how employment and welfare evolves. Aghion et al. (2017) explore the role of AI for the growth process and its reallocative effects. The authors argue that Baumol (1967)'s cost disease leads to the declining share of traditional industries' GDP, as they become automated. This decline is offset by the growing fraction of automated industry. In such an environment, AI may discourage future innovation for fear of imitation, undermining incentives to innovate in the first place.

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

In the finance literature, Begenau et al. (2018) grow the data processing capacity of financial investors, instead of modeling firms' use of their own data. Such studies complement this work by illustrating other ways in which abundant data is re-shaping the economy.

Finally, the model builds on the five-equation toy model in Farboodi et al. (2019), which was designed to explore the size distribution of heterogeneous firms. In this paper, we add features such as adjustment costs and tradeable data. These features make the model more suitable for answering questions about time-series dynamics and the long run.

1 A Data Economy Growth Model

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

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

The quality of a good depends on a firm's choice of a production technique $a_{i,t}$. Each period firm i has one optimal technique, with a persistent and a transitory components: $\theta_{i,t} + \epsilon_{a,i,t}$. Neither component is separately observed. The persistent component $\theta_{i,t}$ follows an AR(1) process: $\theta_{i,t} = \bar{\theta} + \rho(\theta_{i,t-1} - \bar{\theta}) + \eta_{i,t}$. The AR(1) innovation $\eta_{i,t}$ is $i.i.d.$ across time. We explore two possible correlations of $\eta_{i,t}$ across firms. First, we consider independent θ processes ($corr(\eta_{i,t}, \eta_{j,t}) = 0,\ \forall i \neq j$). Then we consider an aggregate θ process ($corr(\eta_{i,t}, \eta_{j,t}) = 1,\ \forall i, j$). The transitory shock $\epsilon_{a,i,t}$ is $i.i.d.$ across time and firms and is unlearnable.

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

$$A_{i,t} = \bar{A}_i \left[\hat{A} - (a_{i,t} - \theta_{i,t} - \epsilon_{a,i,t})^2 \right]. \tag{2}$$

The role of data is that it helps firms to choose better production techniques. One interpretation is that data can inform a firm whether blue or green cars or white or brown kitchens will be more valued by their consumers, and produce accordingly. Transactions help to reveal these preferences but they are constantly changing and firms must continually learn to catch up. Another interpretation is that the technique is inventory management, or other cost-saving activities. Observing many establishments with a range of practices and many customers provides useful information for optimizing business practices.

Specifically, data is informative about $\theta_{i,t}$. The role of the temporary shock ϵ_a is that it prevents firms, whose payoffs reveal their productivity $A_{i,t}$, from inferring $\theta_{i,t}$ 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 $\theta_{i,t-1}$ 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 $\theta_{i,t-1}$, but not on any data from before t. All past data is just a noisy signal about $\theta_{i,t-1}$, which the firm now knows. Thus preventing the revelation of $\theta_{i,t-1}$ keeps old data relevant and valuable.

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

$$n_{i,t} = z_i k_{i,t-1}^{\alpha}, \tag{3}$$

where z_i is the parameter that governs how much data a firm can mine from its cutomers. A data mining firm is one that harvests lots of data per unit of output.

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

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

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

In order to model data as a tradeable asset, we need to allow for the possibility that the amount of data a firm produces is not the same as the amount of data they use. The difference between the two is the amount of data purchased from or sold to other firms.[1] Let ω_{it} be the amount of

[1]Data here is used exclusively by one firm. This is not true to the idea that data is a non-rival good that can be used by many firms at once. However, in a purely competitive environment, any firm allowed to use and sell a piece of data would sell all its data to as many buyers as possible. Such a strategy would produce revenue from data sales and would, because of the perfect competition assumption, not affect any of the prices faced by the firm. In order to have a meaningful choice with non-exclusive use of data requires moving to a imperfect competition model. See Jones and Tonetti (2018) for an example. Such a model involves far more complexity than is required to make the

data used by firm i a time t. If $\omega_{it} < n_{it}$, the firm is selling data. If $\omega_{it} > n_{it}$, the firm purchased data. Let the price of one piece of data be π.

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

Each firm's flow of new data $n_{i,t}$ allows it to build up a stock of knowledge $\Omega_{i,t}$ that it uses to forecast future economic outcomes. Adjusting the level of data usage incurs a data adjustment cost. If a firm's data stock was $\Omega_{i,t}$ and becomes $\Omega_{i,t+1}$, the firm's period-t output is diminished by $\Psi(\Delta\Omega_{i,t+1}) = \psi(\Delta\Omega_{i,t+1})^2$, where Δ represents the percentage change: $\Delta\Omega_{i,t+1} = (\Omega_{i,t+1} - \Omega_{i,t})/\Omega_{i,t}$. The percentage change formulation is helpful because it makes doubling one's stock of knowledge equally costly, no matter what units data is measured in.

Firm Problem. A firm chooses a sequence of production, quality and data-use decisions $k_{i,t}, a_{i,t}, \omega_{i,t}$ to maximize

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left(P_t A_{i,t} k_{i,t}^{\alpha} - \Psi(\Delta\Omega_{i,t+1}) + \pi(n_{i,t+1} - \omega_{it}) - r k_{i,t} \right) \tag{5}$$

Firms update beliefs about $\theta_{i,t}$ 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 $a_{i,t}$ is $\mathcal{I}_{i,t} = [\{A_{i,\tau}\}_{\tau=0}^{t-1}; \{\{s_{i,\tau,m}\}_{m=1}^{n_{i,\tau}}\}_{\tau=0}^{t}]$.

Here, we take the rental rate of capital as given, just to show the data-relevant mechanisms as clearly as possible. It could be that this is a small open economy facing a world rate of interest r. But, of course, one should embed this problem in an equilibrium context where capital markets clear. Similarly, one can add labor markets and endogenize the demand for goods. The model here is only a sketch of an idea that should be explored in a fuller economic context.

simple points made here.

Equilibrium

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

$$P_t = \bar{P} Y_t^{-\gamma}, \tag{6}$$

$$Y_t = \int_i A_{i,t} k_{i,t}^{\alpha} di. \tag{7}$$

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

Solution The state variables of the recursive problem are the prior mean and variance of beliefs about $\theta_{i,t-1}$, last period's revenues, and the new data points. Taking a first order condition with respect to the technique choice, we find that the optimal technique is $a_{i,t}^* = \mathbb{E}_i[\theta_{i,t}|\mathcal{I}_{i,t}]$. Let the posterior precision of beliefs be $\Omega_{i,t} := \mathbb{E}_i[(\mathbb{E}_i[\theta_{i,t}|\mathcal{I}_{i,t}] - \theta_{i,t})^2]^{-1}$. Thus, expected quality is $\mathbb{E}_i[A_{i,t}] = \bar{A} - \Omega_{i,t}^{-1} - \sigma_a^2$. We can thus express expected firm value recursively.

Lemma 1 *The optimal sequence of capital investment choices $\{k_{i,t}\}$ and data use choices $\{\omega_{i,t}\}$ solves the following recursive problem:*

$$V(\Omega_{i,t}) = \max_{k_{i,t},\omega_{i,t}} \; P_t \left(\left(\bar{A} - \Omega_{i,t}^{-1} - \sigma_a^2 \right) k_{i,t}^{\alpha} - \Psi(\Delta\Omega_{i,t+1}) \right) + \pi(n_{i,t+1} - \omega_{it}) - rk_{i,t} + \beta V(\Omega_{i,t+1}) \tag{8}$$

where $n_{i,t} = z_i k_{i,t}^{\alpha}$ *and*

$$\Omega_{i,t+1} = \left[\rho^2 (\Omega_{i,t} + \sigma_a^{-2})^{-1} + \sigma_\theta^2 \right]^{-1} + \omega_{i,t} \sigma_\epsilon^{-2} \tag{9}$$

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

Valuing Data In this formulation of the problem, $\Omega_{i,t}$ can be interpreted as the amount of data a firm has. Technically, it is the precision of the firm's posterior belief. But according to Bayes' rule for normal variables, posterior precision is the discounted precision of prior beliefs plus the precision of each signal observed. In other words, the precision of beliefs, Ω is a linear transformation of the number of all past used data points, $\{\omega_{is}\}_{s=0}^{t}$. Ω_{it} captures the value of past observed data through the term for the discounted prior precision, $\Omega_{i,t-1}$.

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

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

The second means of measuring data is to look at what actions it allows firms to choose. A firm with more data can respond more quickly to market conditions than a firm with little data to guide them. To use this measurement approach, one needs to take a stand on what actions firms are using data to inform, what variable firms are using the data to forecast, and to measure both the variable and the action. One example is portfolio choice in financial markets Farboodi et al. (2018). Another example is firms' real investment David et al. (2016). Both measure the covariance between investment choices and future returns. That covariance between choices and unknown states reveals how much data investors have about the future unknown state. A similar approach could be to use the correlation between consumer demand and firm production, across a portfolio of goods, to infer firms' data about demand.

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

2 Results: Long Run Growth in a Data Economy

When data results from economic activity, reduces forecast errors and improves the value of firms' goods, there is both a force for increasing and a force for decreasing returns. This section explores how these two forces interact and what long-run economic trajectories are possible.

Our goal is not to make precise quantitative foreacasts of future growth. The model is far too simple for that. Rather, the goal is to illustrate the forces at work and to compare and contrast them with more traditional growth models. Illustrating with figures is helpful. Figures require putting numbers on parameters. Table 1 lists the parameters, which were chosen to be simple, to conform with standing economic intuition, or because they delivered the trade-off we sought to illustrate. The data mining parameter z_i is varied across exercises.

Table 1: Parameters

ρ	r	β	α	ψ	γ	$\bar{A}$	$\bar{P}$	π	σ_a^2	σ_θ^2	σ_ϵ^2
0.8	0.2	0.98	0.5	2.5	2	1	0.025	0.01	0.25	0.025	1

Aggregate data inflows and outflows. 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. The inflows of data are new pieces of data that are generated by economic activity. The number of new data points $n_{i,t}$ was assumed to be data mining ability times last-period's physical output: $z_i k_{i,t-1}^\alpha$. By Bayes' law for normal variables, the total precision of that information is the sum of the precisions of all the data points: $n_{i,t}\sigma_\epsilon^{-2}$. That is the quantity represented by the line inflows in Figure 1. How can data flow out? It is not really leaving. It is just depreciating. Data depreciates because data generated at time t is about next period's optimal technique θ_{t+1}. But that means that data generated s periods ago is about θ_{t-s+1}. Since θ is an AR(1) process, it is constantly evolving. Data from many periods ago, about a θ realized many periods ago is not as relevant as more recent data. So, just like capital, data depreciates.

Note that since we are examining the aggregate amount of data in the economy, we ignore data purchases and sales, because they just move data around, rather than changing the aggregate amount. If we allowed data to have multiple simultaneous users, this would need to change.

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

$$\text{Outflows} = \Omega_{i,t} + \sigma_a^{-2} - \left[(\rho^2(\Omega_{i,t} + \sigma_a^{-2}))^{-1} + \sigma_\theta^2\right]^{-1} \tag{10}$$

Figure 1 illustrates the inflows and outflows, in a form that looks just like the traditional Solow model with capital accumulation. Depreciation is not linear, but is very close to linear. For other parameter values, sometimes it has a noticeable concavity at low levels of knowledge Ω. The inflows have diminishing returns. There are two reasons for this. The first is that more data raises efficiency, which incentivizes more capital investment. But capital has diminishing returns because the exponent in the production function is $\alpha < 1$. But that is not the only reason. Even if capital did not have diminishing marginal returns, inflows would still exhibit concavity. The reason is that the returns to data are bounded. With infinite data, all learnable uncertainty about θ can be resolved. With a perfect forecast of θ, the expected good quality is $\left(\bar{A} - \sigma_a^2\right)$, which is finite. Thus,

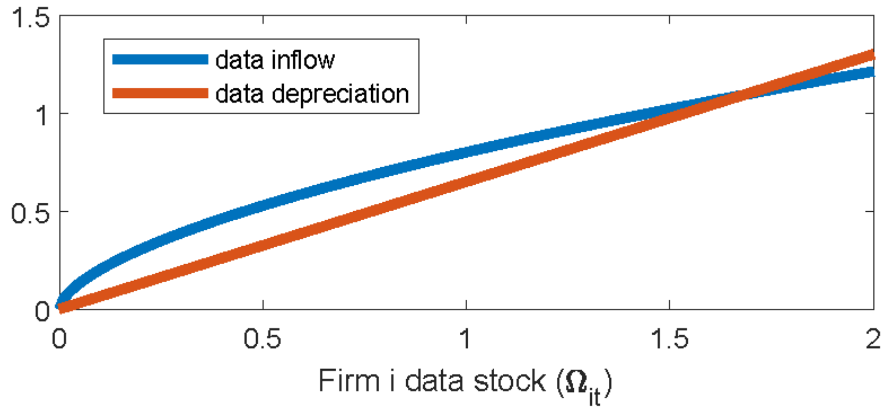

Figure 1: Economy converges to a data steady state: Aggregate inflows and outflows of data. Line labeled inflows plots $z_i k_{i,t}^{\alpha} \sigma_{\epsilon}^{-2}$ for a firm i, that makes an optimal capital decision $k_{i,t}^{*}$, with different levels of initial data stock. Line labeled outflows plots the quantity in (10).

the optimal capital investment is finite. Since a function that is continuous and not concave will always cross any finite upper bound, productivity, investment and data inflows must all be concave in the stock of knowledge Ω.

Diminishing returns arise because we model data as information, not directly as an addition to productivity. Information is used to forecast random variables. With infinite information, the best forecast has perfect foresight. But perfect foresight does not typically mean infinite investment or profits. Information has diminishing returns because its ability to reduce variance gets smaller and smaller as beliefs become more precise.

Growth without data sales. We begin by looking at a simple version of the model where firms' optimal techniques are independent ($corr(\eta_i, \eta_j) = 0 \; \forall i \neq j$). This implies that firms can only learn from data that they themselves produce: $\omega_{it} = n_{it}, \; \forall i, t$. They are allowed to buy data. But since such data is not informative about their optimal action, it has no value to them.

Consider an economy of symmetric firms, all producing and accumulating data at the same rate. Figure 2 shows how firms accumulate data rapidly. This accumulation helps them to produce higher quality goods. The raw units of non-quality-adjust goods production in this example grows proportionately to Y_t. Production net of data adjustment cost is more volatile. When goods are higher quality, firms produce more units of them. More production generates more transactions data. The accumulation of data makes each unit of production generate more value. This can

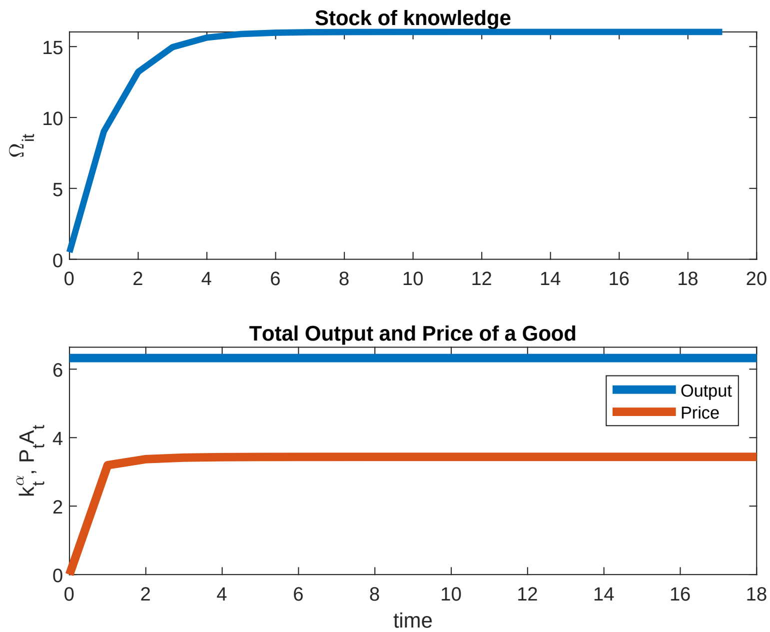

Figure 2: Data, Output and Price for firms that cannot sell their data.
Top panel plots Ω_{it} for symmetric firms i, over time t. Bottom panel plots the market price of an average unit of the good, which is price per quality unit, times the quality: $P_t A_t$ and aggregate output Y_t, which is measured in quality units. Parameters listed in Table 1, with the data mining ability set to $z = 10$.

be seen in the rise in price for the higher-quality goods. The price per quality unit, P_t is falling monotonically, as output rises. But the price for a unit of the good, $P_t A_{it}$, rises as the goods' quality, A_{it}, improves.

Growth with data sales. Next, consider an economy with perfectly correlated optimal techniques ($corr(\eta_i, \eta_j) = 1 \; \forall i, j$). In this economy, purchased data is just as relevant as the firms' own data. Now, firms will want to buy or sell data. The price of data π is fixed. This could be seen as a small open economy buying and selling into a large international data market. When firms sell data, one piece of their own data is perfectly substitutable for a piece of data generated by another firm. For two firms selling an identical good to the same market, this might be close to true. In many cases, a firms' own data is more relevant for forecasting the variables of interest to that firm: its demand, its future cost or its own future productivity. But certainly external data

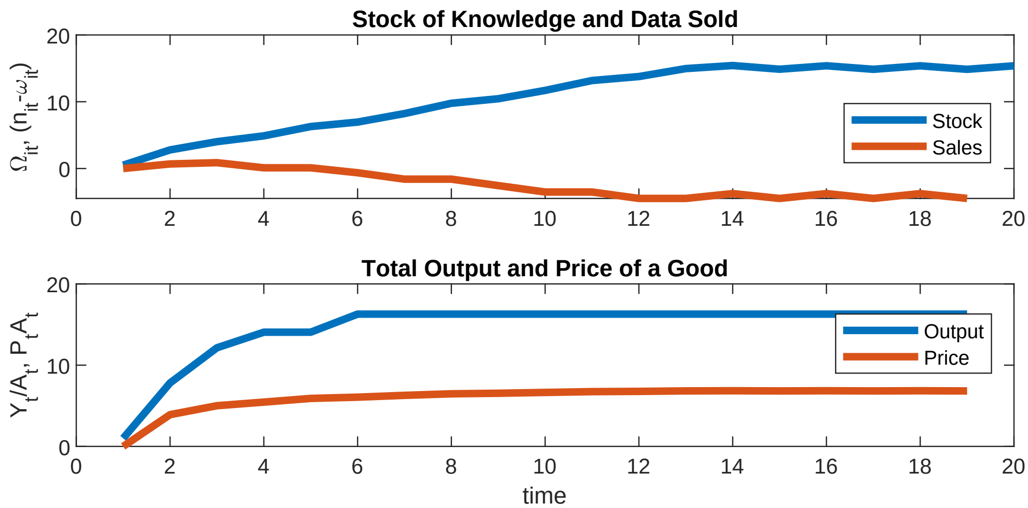

Figure 3: Data, Output and Price for poor data miners ($z_i = 1$).
Top panel plots Ω_{it} for symmetric firms i, over time t. Bottom panel plots the market price of an average unit of the good, which is price per quality unit, times the quality: $P_t A_t$ and aggregate output Y_t, which is measured in quality units, divided by A_t, to report units produced. Parameters listed in Table 1, with the data mining ability set to $z = 1$.

is relevant and is frequently purchased. Thus the truth likely lies in between this case and the previous, non-tradeable data one.

There are two interesting cases here. The first case is where the industry is not very good at data mining ($z = 1$). In this case, the industry will be a net purchaser of data. Figure 3 illustrates that firms in such an industry will accumulate data (rise in Ω_{it} in the top panel), but will do so by purchasing data $n_{it} - \omega_{it} < 0$. Firms that are poor data miners need to buy data from others to raise the quality of their production.

With a data market, an unproductive data producer is not condemned to produce low-quality goods. Instead, they can purchase data from others. The firms in the example continue to purchase data throughout their lifespan. Early on, data purchases increase the stock of knowledge. Once the stock is sufficiently large, the same data purchases are just enough to offset the depreciation of old data. As the net inflows of data slows, price and output level off as well.

The opposite case is a highly productive data miner. In Figure 4, the firms in the economy all have $z = 25$, meaning that every unit of production generates 25 signals. These firms do not use most of their data. Instead, they sell most of their data. In the top panel of Figure 4, firms' data sales are positive and far exceed their data stock. These firms do not produce higher quality goods. Instead, they produce a larger volume of low-quality goods. They do this because their main source of revenue is not from their goods sales, it is from their data sales. In period 1, these firms set a price of zero. They give goods away for free, like free apps for download, in order to mine data from their customers and sell that data. This is optimal because the price of data is sufficiently high. If all producers created lots of data, the equilibrium price of data would fall. But this illustrates how a data producer who is more data-productive than other firms can endogenously become a data provider.

Valuing zero-price goods (data barter). In many of these examples, the optimal price of a firm's good was initially close to zero (e.g., the period 1 price in Figures 2, 3 and 4). The reason for the near-zero price is that the firm wants to sell many units, in order to accumulate data. Data will boost the productivity of future production and enable future profitable goods sales. It could also be that the firm wants to accumulate data, in order to sell it. Lowering the price is a costly investment in a productive asset that yields future value. This is what makes a zero optimal price possible.

Data poverty traps. In some cases, firms can experience slow growth for a while, before they get enough data to produce productively, and their growth takes off. This is a special case that arises for a subset of parameters. But it is an important case because this slow growth phase, in which the firm produces low-quality goods, can be a barrier to entry. We will examine the simplest version of our problem: One firm growing in an economy with no data sales. The previous results illustrated what happens when a whole industry of competitive firms accumulates data and produces higher quality goods. But sometimes one firm leaps ahead in their use of data, while others' business practices are stagnant. The next result explores the consequences of relative data mining prowess. The difference between this exercise and the previous one is that with only one firm changing, we hold the market price of goods, P_t fixed. Doing this also reveals what role the equilibrium price of

Figure 4: Data, Output and Price for productive data miners ($z_i = 25$).
Top panel plots Ω_{it} for symmetric firms i, over time t. Bottom panel plots the market price of an average unit of the good, which is price per quality unit, times the quality: $P_t A_t$ and aggregate output Y_t, which is measured in quality units. Parameters listed in Table 1, with the data mining ability set to $z = 25$.

goods plays in our results.

Figure 5 illustrates such a case. It reveals that firm growth dynamics take on an S-shape. The firm's early growth is constrained by its ability to produce data at a small size. Because the firm is small, it produces little data. Because the firm produces little data, its optimal size is small. The adjustment cost keeps the firm from incurring a one-time fixed cost and jumping to its optimal size. But the fact that this is just one firm also plays a role. When all firms are unproductive, quality units are scarce. The scarcity drives up the equilibrium price of goods. High prices induce firms to produce more, even though they are unproductive. But when only one firm starts accumulating data and other already have, that equilibrium price mechanism is shut off, slowing the data transition.

The firm in this example makes losses for the first four periods. It produces goods even though its cost is higher than the price, because doing so generates data. This negative value production is a costly investment in data, which enables future profitable production. In this example, that strategy works for the firm because it faces no financing constraint. In reality, many firms that make losses for years on end lose their financing and exit.

These results suggest that financially constrained young firms may never enter or make the investment in data systems. By not paying the data adjustment costs, they can remain financially viable. But by not collecting and using data, they never escape the trap of poor data and low value-added.

In a country where data science skills are scarce, the labor cost of hiring a data analyst may make this data adjustment cost very high. Thus scare data skilled labor might condemn an entire economy of firms to this data poverty trap.

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

Why is rational inattention a natural complement to this model? Following Sims (2003), rational inattention problems consider what types of information or data is most valuable to process, subject to a constraint on the mutual information of the processed information and the underlying uncertain economic variables. The idea of using mutual information as a constraint, or the basis

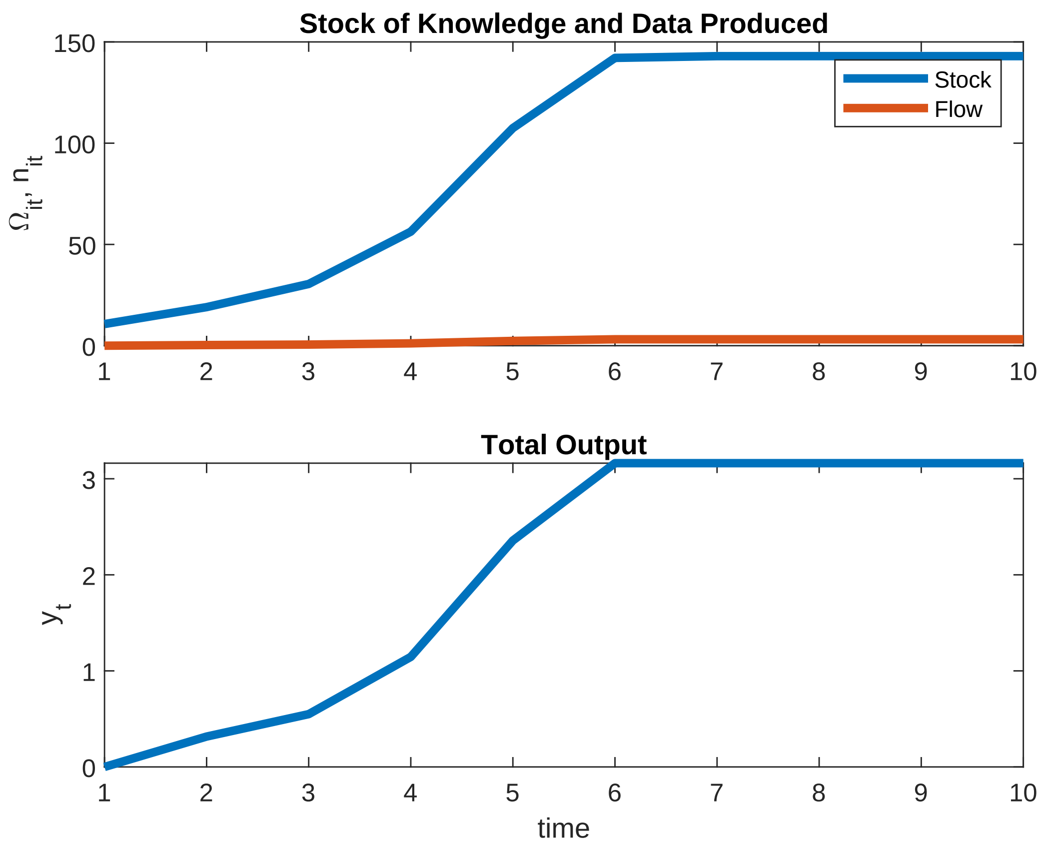

Figure 5: Data and Output with S-shape dynamics. Top panel plots Ω_{it} for a single firm i, over time t. Bottom panel plots the firm's raw output k_t^α. $\mathrm{Corr}(\eta_i, \eta_j) = 0$, so that there is no data trade. Parameters listed in Table 1, with $\sigma_a^2 = 0.25$, $\sigma_t^2 = 0.05$, $\sigma_e^2 = 0.025$, and the data mining ability set to $z = 1$.

of a cost function, comes from the computer science literature on information theory. The mutual information of a signal and a state is an approximation to the length of the bit string or binary code necessary to transmit that information (Cover and Thomas (1991)). While the interpretation of rational inattention in economics has been mostly as a cognitive limitation on processing information, the tool was originally designed to model computers' processing of data. Therefore, it would be well suited to explore the data processing choices of firms.

3 Data for Business Stealing

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

$$A_{i,t} = \hat{A} - \left(a_{i,t} - \theta_{i,t} - \epsilon_{a,i,t}\right)^2 + \int_{j=0}^{1} \left(a_{j,t} - \theta_{j,t} - \epsilon_{a,j,t}\right)^2 dj \tag{11}$$

This captures the idea that when one firm uses data to reduce the distance between their chosen technique a_{it} and the optimal technique $\theta + \epsilon$, that firm benefits, but all other firms lose a little bit. These gains and losses are such that, when added up to compute aggregate productivity, they cancel out: $\int A_{it} = \hat{A}$. This represents an extreme view that data processing contributes absolutely nothing to social welfare. While that is unlikely, examining the two extreme cases is illuminating.

What we find is that reformulating the problem this way makes very little difference for most of our conclusions. The externality does reduce the productivity of firms and does reduce welfare, relative to the case without the externality. But it does not change firms' choices. Therefore, it does not change data inflows, outflows or accumulation. It does not change firm dynamics. The reason there is so little change is that the externality does not enter in a firm's first order condition. It does not change its optimal choice of anything. Firm i's actions have an infinitesimal, negligible effect on the average productivity term $\int_{j=0}^{1} \left(a_{j,t} - \theta_{j,t} - \epsilon_{a,j,t}\right)^2 dj$. Because firm i is massless in a competitive industry, its actions do not affect that aggregate term. So the derivative of that term with respect to i's choice variables is zero. If the term is zero in the first order condition, it means it has no effect on choices of the firm.

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

4 Conclusions

The economics of transactions data bears some resemblance to technology and some to capital. It is not identical to either. Yet, when economies accumulate data alone, the aggregate growth economics are similar to an economy that accumulates capital alone. Diminishing returns set in and the gains are bounded. Yet, the transition paths differ. There can be regions of increasing returns that create possible poverty traps. Such traps arise with capital externalities as well. Data's production process, with its feedback loop from data to production and back to data, makes such increasing returns a natural outcome. When markets for data exist, some of the effects are mitigated, but the diminishing returns persist. Even if data does not increase output at all, but is only a form of business stealing, the dynamics are unchanged. Thus, while the accumulation and analysis of data may be the hallmark of the "new economy," this new economy has many economic forces at work that are old and familiar.

References

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

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

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

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

Cover, Thomas and Joy Thomas, *Elements of information theory*, first ed., John Wiley and Sons, New York, New York, 1991.

David, Joel M, Hugo A Hopenhayn, and Venky Venkateswaran, "Information, misallocation, and aggregate productivity," *The Quarterly Journal of Economics*, 2016, *131* (2), 943–1005.

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

Farboodi, Maryam, Adrien Matray, and Laura Veldkamp, "Where has all the big data gone?," 2018.

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

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

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

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

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

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

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

Romer, Paul, "Endogenous Technological Change," *Journal of Political Economy*, 1990, *98* (5), S71–102.

Sims, Christopher, "Implications of rational inattention," *Journal of Monetary Economics*, 2003, *50(3)*, 665–90.

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

United Nations, "Information Economy Report: Digitalization, Trade and Development," Technical Report, United Nations Conference on Trade and Development 2017.

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

5 Appendix

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

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

At time t, each firm observes two types of signals. First, date $t-1$ output provides a noisy signal about θ_{t-1}:

$$y_{t-1} = \theta_{t-1} + \epsilon_{a,t-1}, \tag{12}$$

where $\epsilon_{a,t} \sim \mathcal{N}(0, \sigma_a^2)$. We provide model detail on this step below. Second, the firm observes $n_t = zk_t^\alpha$ data points as a bi-product of its economic activity. The set of signals $\{s_{t,m}\}_{m\in[1:n_{i,t}]}$ are equivalent to an aggregate (average) signal $\bar{s}_t$ such that:

$$\bar{s}_t = \theta_t + \epsilon_{s,t}, \tag{13}$$

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

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

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

At time, t, the firm takes as given:

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

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

We update the state variable sequentially, using the two signals. First, combine the priors with y_{t-1}:

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

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

$$\widehat{\mu}_t = \mathbb{E}\big[\theta_t \mid \mathcal{I}_t\big] = \frac{\Big[\rho^2\big[\Sigma_{t-1}^{-1} + \sigma_a^{-2}\big]^{-1} + \sigma_\theta^2\Big]^{-1} \cdot \mathbb{E}\big[\theta_t \mid \mathcal{I}_{t-1}, y_{t-1}\big] + n_t\sigma_\epsilon^{-2}\bar{s}_t}{\Big[\rho^2\big[\Sigma_{t-1}^{-1} + \sigma_a^{-2}\big]^{-1} + \sigma_\theta^2\Big]^{-1} + n_t\sigma_\epsilon^{-2}} \tag{14}$$

$$\Sigma_t = Var\big[\theta \mid \mathcal{I}_t\big] = \Big\{\Big[\rho^2\big[\Sigma_{t-1}^{-1} + \sigma_a^{-2}\big]^{-1} + \sigma_\theta^2\Big]^{-1} + n_t\sigma_\epsilon^{-2}\Big\}^{-1} \tag{15}$$

Multiply and divide equation (14) by Σ_t as defined in equation (15) to get

$$\widehat{\mu}_t = (1 - n_t\sigma_\epsilon^{-2}\Sigma_t)\left[\bar{\theta}(1-\rho) + \rho\left((1-M_t)\mu_{t-1} + M_t\tilde{y}_{t-1}\right)\right] + n_t\sigma_\epsilon^{-2}\Sigma_t\bar{s}_t, \tag{16}$$

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

Equations (15) and (16) constitute the Kalman filter describing the firm dynamic information problem. Importantly, note that Σ_t is deterministic.

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

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

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

$$k_{it} = \left[\frac{\alpha}{r}\left(P_t A_{it} + z_i\sigma_\epsilon^{-2}(\beta V'(\cdot) - P_t\psi(\cdot))\right)\right]^{1/(1-\alpha)} \tag{17}$$

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

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

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

To solve this, we start with a guess of V' and then solve the non-linear equation above for k_{it}. Then, update our guess of V.

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

$$n_{ss}\sigma_{\epsilon}^{-2} + \left[\rho^2(\Omega_{ss} + \sigma_a^{-2})^{-1} + \sigma_\theta^2\right]^{-1} = \Omega_{ss} \tag{18}$$

This is equating the inflows of data $n_{it}\sigma_\epsilon^{-2}$ with the outflows of data $\left[\rho^2(\Omega_{i,t} + \sigma_a^{-2})^{-1} + \sigma_\theta^2\right]^{-1} - \Omega_{it}$. Given a number of new data points n_{ss}, this pins down the steady state stock of data. The number of data points depends on the steady state level of capital. The steady state level of capital is given by (17) for A_{ss} depending on Ω_{ss} and a steady state level of V'_{ss}. We solve for that steady state marginal value of data next.

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

$$V'_{ss} = \left[1 - \beta\frac{\partial\Omega_{t+1}}{\partial\Omega_t}|_{ss}\right]^{-1} P_t k_{ss}^{\alpha}\Omega_{ss}^{-2} \tag{19}$$

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

The equations (17), (18) and (19) form a system of 3 equations in 3 unknowns. The solution to this system delivers the steady state levels of data, its marginal value and the steady state level of capital.

Backwards induction solution Start by solving for the steady state where k, Ω and therefore V and V' are equal at dates t, $t+1$ and forever in the future. Then, work backwards. What is the $t-1$ level of k and Ω that would result in t level being the steady state level.

In the period before the economy reaches steady state ($ss-1$), the marginal value of data is

$$V'_{ss-1} = P_t k_{ss-1}^{\alpha}\Omega_{ss-1}^{-2} - \Psi'(\cdot)\left(\frac{\partial\Omega_{ss}}{\partial\Omega_{ss-1}} - 1\right) + \beta V'_{ss}\frac{\partial\Omega_{ss}}{\partial\Omega_{ss-1}}$$

Since V'_{ss} is known, if we can solve for k_{ss-1} and Ω_{ss-1}, we can retrieve the marginal value of data. We need two more equations to solve for the three unknowns jointly: V'_{ss-1}, k_{ss-1} and Ω_{ss-1}. The other two equations are the first order condition for capital (17) and the low of motion for Ω, (9). These three equations form a backwards recursion that allows us to solve for equilibrium in the date before steady state. Performing the same recursion allows us to solve backwards for a time path of data and capital, leading up to steady state.

Proof of lemma 1 *Lemma.* The sequence problem of the firm can be solved as a non-stochastic recursive problem with one state variable. Consider the firm sequential problem:

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

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:

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

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:

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

with $n_t = k_{t-1}^{\alpha}$. 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

$$\begin{aligned} V(\{s_{t,m}\}_{m\in[1:n_t]}, y_{t-1}, \hat{\mu}_{t-1}, \Sigma_{t-1}) = \max_{k_t} \mathbb{E}\Big[&\big(\bar{A} - (\mathbb{E}[\theta_t | I_t] - \theta_t - \epsilon_{a,t})^2\big) k_t^{\alpha} - r k_t \\ &+ \beta V(\{s_{t+1,m}\}_{m\in[1:n_{t+1}]}, y_t, \hat{\mu}_t, \Sigma_t) | \mathcal{I}_{t-1}\Big]. \end{aligned}$$

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

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

Notice that $\mathbb{E}[(\mathbb{E}[\theta_t|I_t] - \theta_t)^2|I_{t-1}]$ is the time-t conditional (posterior) variance of θ_t, and the posterior variance of beliefs is $\mathbb{E}[(\mathbb{E}[\theta_t|I_t] - \theta_t)^2] := \Sigma_t$. Thus, expected productivity is $\mathbb{E}[A_t] = \bar{A} - \Sigma_t - \sigma_a^2$, which determines the within period expected payoff. Additionally, using the Kalman system equation (15), this posterior variance is

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

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

$$\begin{aligned} V(\Sigma_t) &= \max_{k_t} \Big[(\bar{A} - \Sigma_t - \sigma_a^2)k_t^\alpha - rk_t + \beta V(\Sigma_{t+1})\Big] \\ \text{s.t.} \qquad \Sigma_{t+1} &= \Big[\big[\rho^2(\Sigma_t^{-1} + \sigma_a^2)^{-1} + \sigma_\theta^2\big]^{-1} + zk_t^\alpha\sigma_\epsilon^{-2}\Big]^{-1} \end{aligned}$$

Finally, note that when we add data usage choice or adjustment costs, neither of these creates a new state variable. Data use in incorporated in the stock of knowledge through (9), which still represents one state variable.