



Management Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

Knowledge Accumulation, Privacy, and Growth in a Data Economy

Lin William Cong , Danxia Xie , Longtian Zhang

To cite this article:

Lin William Cong , Danxia Xie , Longtian Zhang (2021) Knowledge Accumulation, Privacy, and Growth in a Data Economy. Management Science 67(10):6480-6492. <https://doi.org/10.1287/mnsc.2021.3986>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2021, INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Knowledge Accumulation, Privacy, and Growth in a Data Economy

Lin William Cong,^a Danxia Xie,^b Longtian Zhang^c

^aSC Johnson College of Business, Cornell University, Ithaca, New York 14853; ^bInstitute of Economics, Tsinghua University, Beijing 100084, China; ^cSchool of International Trade and Economics, Central University of Finance and Economics, Beijing 102206, China

Contact: will.cong@cornell.edu,  <https://orcid.org/0000-0002-2617-2367> (LWC); xiedanxia@tsinghua.edu.cn,

 <https://orcid.org/0000-0003-4346-7084> (DX); zlt18@mails.tsinghua.edu.cn,  <https://orcid.org/0000-0002-8644-415X> (LZ)

Received: April 17, 2020

Revised: October 28, 2020

Accepted: December 15, 2020

Published Online in Articles in Advance:
August 24, 2021

<https://doi.org/10.1287/mnsc.2021.3986>

Copyright: © 2021 INFORMS

Abstract. We build an endogenous growth model with consumer-generated data as a new key factor for knowledge accumulation. Consumers balance between providing data for profit and potential privacy infringement. Intermediate good producers use data to innovate and contribute to the final good production, which fuels economic growth. Data are dynamically nonrival with flexible ownership while their production is endogenous and policy-dependent. Although a decentralized economy can grow at the same rate (but are at different levels) as the social optimum on the Balanced Growth Path, the R&D sector underemploys labor and overuses data—an inefficiency mitigated by subsidizing innovators instead of direct data regulation. As a data economy emerges and matures, consumers' data provision endogenously declines after a transitional acceleration, allaying long-run privacy concerns but portending initial growth traps that call for interventions.

History: Accepted by Kay Giesecke, finance.

Funding: This work was supported by the National Natural Science Foundation of China [Grant 71973076] and the Ewing Marion Kauffman Foundation [Grant RG-201805-4237].

Supplemental Material: Data files and the online appendix are available at <https://doi.org/10.1287/mnsc.2021.3986>.

Keywords: big data • data ownership • endogenous growth • innovation • nonrivalry • privacy regulation

1. Introduction

Data not only help produce new products and services, but also are used in research and development and knowledge creation, which in turn improves production efficiency.¹ Endogenous generation, dynamic nonrivalry, and flexible ownership of data distinguish them from labor and capital, with implications for labor market allocations and policies. Meanwhile, the proliferation of big data applications often comes at the expense of consumer privacy and is associated with discrimination and misuse.² Despite the introduction of data privacy laws ranging from the General Data Protection Regulation (GDPR) to the California Consumer Privacy Act (CCPA) to Japan's Act on the Protection of Personal Information, we know little about how data (mis)usage, digital infrastructure, and privacy regulation affect knowledge accumulation and the growth of a dynamic economy where data are the new key driver.

To fill in this gap, we build on Romer (1990) to develop an endogenous growth model of the data economy. Our key innovation lies in that consumer data add to R&D and knowledge accumulation. At the same time, data are by-products of economic activities with potential privacy issues, which differ from other

input factors such as labor or capital because growth can endogenously feed back to data generation. Consumers in our baseline model choose the quantity of data to sell to intermediate firms, cognizant of potential information leakage and misuses. Innovative intermediate firms utilize the raw data for research that contributes to the final good production. Specifically, data are transformed into intermediate goods (information products included)—a feature absent in other models. Data can generate spillovers through knowledge accumulation, which is further enhanced when they are traded over time and used by multiple parties with little reproduction costs (both static and dynamic nonrivalry). These effects are moderated in the model by the disutility from potential data privacy violation.

We show that a decentralized economy grows at the same rate as the social optimum on the Balanced Growth Path (BGP). But social welfare and consumer surplus are strictly lower due to underemployment and data overuse in the R&D sector. Monopolistic markups in the production of intermediate goods lead to a crowding out of labor in R&D (Jones 1995); producers then compensate the underemployment of labor in R&D by more aggressive utilization of data. The crowding-out of labor crowds in data usage in

R&D to a socially excessive level. R&D labor employment and data usage can thus deviate significantly from those in the social planner’s solution, especially in the initial phase of BGP (albeit less severe in the long run).

Different from recent studies such as Jones and Tonetti (2020), data are overused even when they are nonrival and owned by consumers who directly factor in the disutility from data leakage or abuse. Direct regulation on data usage comes at the cost of economic growth and constitutes a wealth transfer from future generations to current generations. Rather than taxing data overuse, our model reveals that subsidizing R&D wage or intermediate producers are more effective at mitigating the social inefficiency. Moreover, because data expand the innovation possibility frontier that exhibits diminishing returns to scale, historical data usage reduces the benefit of future data usage, potentially leading to a declining data provision per capita in the long run.

As the economy transitions into a steady growth, data provision by consumers can undergo accelerated growth before declining. Importantly, a low initial growth may limit data generation even in the planner’s solution, which further delays the transition to high growth stages in the long run—a form of growth trap. Interventions such as foreign aid for digital infrastructure development can help escape the trap, but only to the extent that the data generating constraint is still binding, that is, when economic activities translate into limited volumes of data.

Our paper primarily contributes to the emerging literature on information and the data economy. Relative to earlier studies on the social value, sales, and property rights of information (e.g., Hirshleifer 1971, Admati and Pfleiderer 1990, Murphy 1996), recent studies focus on connecting digital information with privacy issues (e.g., Akçura and Srinivasan 2005, Casadesus-Masanell and Hervas-Drane 2015) or the nonrivalry of data and competition (e.g., Easley et al. 2019, Ichihashi 2020a). We differ by being the first to connect data usage to knowledge accumulation and endogenous growth with privacy considerations. Furthermore, our model complements studies microfounding data privacy concerns (e.g., Ichihashi 2020b, Liu et al. 2020), and is broadly consistent with empirical patterns on the correlation of data economy size with regulation, privacy issues, and declining labor income share in the United States and around the world (Karabarbounis and Neiman 2014, Tang 2019, Barkai 2020, Liao et al. 2020, Abis and Veldkamp 2021).

As such, our theory adds to the large literature on economic growth, especially recent studies embedding data into growth models by allowing them to directly enter production. For example, Jones and Tonetti (2020) highlights the underutilization of data due to its

nonrivalry and the importance of giving data property rights to consumers; Farboodi and Veldkamp (2020) emphasize that data have bounded returns in long-run growth. We complement Jones and Tonetti (2020) by allowing data to facilitate knowledge accumulation in a semiendogenous growth model (Jones 1995, 2016) and accounting for the consumers’ data privacy concerns in the spirit of Stokey (1998) and Acemoglu et al. (2012).³ Through a mechanism different from Farboodi and Veldkamp (2020), we also find that data play a limited role in the long run.

2. The Model

We incorporate data production, data-based innovation, knowledge accumulation, and data privacy concerns into a macroeconomic model to characterize economic growth. Our dynamic “data economy” consists of representative agents who are both consumers and workers, innovative intermediate producers, and a final good producer. Time is continuous and infinite.

2.1. Representative Consumers

A population of homogeneous representative consumers (or households) grow at a constant rate n and is $L(t)$ at time t . Besides consumption, they each choose in each period to supply one unit of labor inelastically in either the R&D sector (intermediate good production) or final good production.

Each consumer produces data as by-products of consumption (e.g., Veldkamp 2005) and can sell the data to intermediate good producers.⁴ However, the data involve personal information and risk leakage and misuse, leading to a disutility that they consider when selling data. We follow the literature (e.g., Jones and Tonetti 2020) to allow data to fully depreciate in every period in the baseline model, but relax this assumption in Online Appendix 2.2 and further discuss it in Section 3.5.

Each consumer’s utility maximization problem is then as follows:

$$\max_{c(t), \varphi(t)} \int_0^{\infty} e^{-(\rho-n)t} \left[\frac{c(t)^{1-\gamma} - 1}{1-\gamma} - \varphi(t)^\sigma \right] dt, \quad (1)$$

subject to

$$\dot{a}(t) = (r(t) - n)a(t) + w(t) + p_\varphi(t)\varphi(t) - c(t), \quad \forall t \in [0, \infty), \quad (2)$$

and

$$\frac{\dot{\varphi}(t)}{\varphi(t)} \leq \frac{\dot{c}(t)}{c(t)} + s. \quad (3)$$

Here, $c(t)$ is the per capita consumption level at time t , and $\varphi(t)$ is the quantity of data a consumer provides to potential intermediate producers for R&D. ρ is the consumers’ discount rate, the reciprocal of γ (also

indistinguishable from the coefficient of risk aversion) is the elasticity of intertemporal substitution (EIS) of consumption, and σ parameterizes the disutility of data misuse or privacy violation, $\varphi(t)^\sigma$, which also depends on the quantity of data provided. s represents the tightness of this constraint and depends on the digital infrastructure, legal development, and privacy regulation policy of the country. We normalize s to zero in the baseline and discuss comparative statics with respect to s later to understand the impact of policy intervention.

In the budget constraint (2), $a(t)$ is the asset held by a consumer at time t and $r(t)$ is its interest rate. $w(t)$ and $p_\varphi(t)$ are the time- t wage for labor and price of data, respectively. The relevant variables for transitional dynamics are the growth rates of data contribution and consumption. To see this, we can derive the system's evolution in the form of Euler equations from the Hamilton:

$$\frac{\dot{c}(t)}{c(t)} = \frac{r(t) - \rho}{\gamma} \quad (4)$$

and

$$\frac{\dot{p}_\varphi(t)}{p_\varphi(t)} - (\sigma - 1) \frac{\dot{\varphi}(t)}{\varphi(t)} = r(t) - \rho. \quad (5)$$

Constraint (3) therefore requires the growth rate of data provision to be bounded by the corresponding growth rate of consumption, which is natural and ensures analytical tractability. Moreover, (3) directly implies that $\varphi(t) \leq \chi c(t)$ for some constant $\chi > 0$, which other recent studies feature. In other words, data are by-products of economic activities and cannot exceed a fixed proportion of consumption activities.

2.2. The Final Good Producer

A representative final good producer operates in a competitive environment with a production function,

$$Y(t) = L_E(t)^\beta \int_0^{N(t)} x(v,t)^{1-\beta} dv, \quad (6)$$

where $L_E(t)$ is the labor employed and $N(t)$ is the number of varieties of intermediate goods used in the final good production at time t . $x(v,t)$ is the total amount of intermediate good of variety v , which can only be used in the final good production for one period. We can view price $p_x(v,t)$ as the rental fee of patents. Finally, β is the elasticity coefficient of labor in the final good production. The final good producer's profit maximization over labor employed and the amount of each intermediate good yields the first-order conditions:

$$x(v,t) = \left[\frac{1-\beta}{p_x(v,t)} \right]^{\frac{1}{\beta}} L_E(t), \quad (7)$$

and

$$w(t) = \beta L_E(t)^{\beta-1} \int_0^{N(t)} x(v,t)^{1-\beta} dv. \quad (8)$$

2.3. Intermediate Producers

An unlimited pool of potential intermediate producers decide whether and how much to conduct research, the success of which gives them monopoly over the intermediate product developed. An intermediate producer therefore enters the market by conducting R&D and evaluating the prospective profit from success less the costs of labor and data as inputs for R&D. We solve the intermediate producers' problem through backward induction.

2.3.1. Production Phase. Upon R&D success, each entrant produces a distinct variety of intermediate goods in a monopolistic market. The present value of an intermediate good of variety v is

$$V(v,t) = \int_t^\infty \exp\left(-\int_t^\tau r(\tau) d\tau\right) \pi(v,s) ds, \quad (9)$$

where the profit from the intermediate good of variety v in a single period of time t is

$$\pi(v,t) = p_x(v,t)x(v,t) - \psi x(v,t). \quad (10)$$

Here, ψ is the marginal cost of this production process, which is set as constant in this economy.

Substituting (7) into (10) and taking derivative with respect to $p_x(v,t)$ yields the optimal price of each variety of intermediate good:

$$p_x(v,t) = \frac{\psi}{1-\beta}, \quad (11)$$

which is the same among different varieties and different periods. Next, substitute (11) into (7):

$$x(v,t) = \left[\frac{(1-\beta)^2}{\psi} \right]^{\frac{1}{\beta}} L_E(t) \equiv x(t), \quad (12)$$

which implies that production quantity is independent of the variety.

Substituting (11) and (12) into (6) and (10), respectively, we get

$$\pi(v,t) = \frac{\psi^{1-\frac{1}{\beta}} \beta}{(1-\beta)^{1-\frac{1}{\beta}}} L_E(t) \equiv \pi(t) \quad (13)$$

and

$$Y(t) = \left[\frac{(1-\beta)^2}{\psi} \right]^{\frac{1}{\beta}-1} N(t) L_E(t). \quad (14)$$

Also, we derive the wage rate as:

$$w(t) = \beta \left[\frac{(1-\beta)^2}{\psi} \right]^{\frac{1}{\beta}-1} N(t). \quad (15)$$

2.3.2. Entry and R&D Phase. Potential intermediate producers enter by conducting R&D using both labor (researchers, including those working on computing, and AI to allow us to use data more efficiently), $L_R(t)$, and data purchased from consumers, $\varphi(t)L(t)$.⁵ We assume that intermediate producers pay a “data processing cost” before they use data for R&D:

$$\theta\varphi(t)^\theta, \text{ where } \theta \geq 0 \text{ and } \phi > 1. \quad (16)$$

For simplicity, we set $\theta = 0$ in the baseline model, and solve in Section 3.6 the general case in which $\theta > 0$ and $\phi > 1$ ensure that the cost function is increasing and convex. We need $\theta > 0$ when we allow firm ownership of data because otherwise the solution is trivial: Firms certainly use up all the data if there is not any data processing cost.

We also specify the evolution of the aggregate innovation possibility frontier (number of varieties) as:

$$\begin{aligned} \dot{N}(t) &= \eta N(t)^\zeta [\varphi(t)L(t)]^\xi L_R(t)^{1-\xi} \\ &= \eta N(t)^\zeta \varphi(t)^\xi l_R(t)^{1-\xi} L(t), \end{aligned} \quad (17)$$

where $\eta > 0$ is an efficiency term of innovation, $\xi \in (0, 1)$ represents the relative contribution of data and labor in innovating process, $0 < \zeta < 1$ captures the spillover effect of knowledge, $\varphi(t)L(t)$ corresponds to data provided by all the consumers in the period, $L_R(t)$ is labor employed in R&D sector, and $l_R(t) = L_R(t)/L(t)$ denotes the fraction of labor employed in the R&D sector.⁶ Labor market clearing condition requires $L_E(t) + L_R(t) \leq L(t)$.

The fact that data enter the R&D of intermediate goods distinguishes our paper from studies such as Jones and Tonetti (2020) that have data only entering directly into final good productions (similar to the $x(v, t)$ in our model). While it holds for data-intensive industries (e.g., self-driving cars) that use data directly as inputs, other industries instead use nondata intermediate goods as inputs. Data, in addition to R&D labor, can be useful for creating those intermediate goods, which our model captures. As Romer (1990) aptly puts, “an intermediate-goods sector uses the designs from the research sector together with forgone output to produce the large number of producer durables that are available for use in final-goods production at any time.” Our model speaks to non-data-intensive industries as well.

An intermediate producer decides how much labor $L_R(t) = l_R(t)L(t)$ and data $\varphi(t)L(t)$ to employ and purchase to maximize the expected net profit:

$$\begin{aligned} \max_{L_R(t), \varphi(t)} \quad & \eta N(t)^\zeta \varphi(t)^\xi l_R(t)^{1-\xi} L(t) V(t) \\ & - w(t)l_R(t)L(t) - p_\varphi(t)\varphi(t)L(t). \end{aligned}$$

The first order conditions yield two free-entry conditions:

$$\eta \xi N(t)^\zeta \varphi(t)^{\xi-1} l_R(t)^{1-\xi} V(t) = p_\varphi(t), \quad (18)$$

and

$$\eta(1 - \xi)N(t)^\zeta \varphi(t)^\xi l_R(t)^{-\xi} V(t) = w(t). \quad (19)$$

Intermediate producers enter until the marginal benefits of adding data or labor equal to the marginal costs.

2.4. Equilibrium Definition

An equilibrium in our model is an allocation in which all intermediate producers choose $\{p_x(v, t), \varphi(t), L_R(t)\}_{v \in [0, N(t)], t=0}^\infty$ to maximize the discounted value of profits, the evolution of $\{N(t)\}_{t=0}^\infty$ is determined by free entry, the evolution of $\{r(t), w(t), p_\varphi(t)\}_{t=0}^\infty$, is consistent with market clearing, the evolution of $\{L_E(t), x(v, t)\}_{v \in [0, N(t)], t=0}^\infty$ is consistent with the final good producer’s profit maximization, and the evolution of $\{c(t), \varphi(t), L_E(t), L_R(t)\}_{t=0}^\infty$ is consistent with consumers’ utility maximization.

2.5. Distinguishing Features of Data

Although the functional forms of how data enter the production function of intermediate goods resemble that of R&D-specific labor, the role of data in economic growth is fundamentally different from labor or capital. It is worth clarifying the distinguishing features of data before we proceed to solve the model.

First, whereas population dynamics are exogenous in many growth models (such as the possible supply of labor), data are endogenized by consumption, which is itself endogenous and depends on data usage in knowledge accumulation. Privacy regulations such as the GDPR and CCPA can affect the endogenous production and usage of data while population growth is more organic and hard to regulate with immediate effects. Also, note that neither capital nor labor usage causes disutility from privacy concerns.

Another key distinguishing feature of data our paper highlights is dynamic nonrivalry. In Jones and Tonetti (2020), all firms need data for production and the nonrivalry is static and cross-sectional. In our model, only potential intermediate producer needs to use raw data in every period, although the entrants and incumbents are benefiting from the same data without incurring high reproduction costs. Because data are traded between intermediate producers entering in different periods, the focus is more on data nonrivalry over time.

This is implicit in our setup: Even though data are fully depreciated every period, they contribute to knowledge accumulation over time in terms of varieties of intermediate goods, which differs from the literature that only allows labor to enter the evolution.

Intuitively, a firm may innovate incrementally by observing other firms' previous data-based innovations. As such, data create knowledge spillovers to future periods by creating new varieties—a form of dynamic nonrivalry.

One more distinguishing feature of data from R&D-specific labor is that data can be owned by firms but labor cannot. Traditional firms use long-term labor contracts while recent on-demand labor or freelance services as seen in Uber, TaskRabbit, Scripted, and Amazon's Mechanical Turks require spot compensations (The Economist 2018). Neither entails firm ownership.

We further elaborate on data's endogenous production and usage as by-products of economic activities, nonrivalry under knowledge accumulation and creative destruction, and flexible ownership in Sections 3.4, 3.5, and 3.6, respectively, as we characterize the equilibrium.

3. Data Economy on the Balanced Growth Path

We first solve the model along the Balanced Growth Path (BGP), which requires that all variables are growing at the same constant rate—a steady state of transformed variables that the literature focuses on. Constant growth then implies $r(t) = r^*$ (Equation (4)). We then identify inefficient data overuse and underprovision of R&D labor, explore policy remedies, and discuss the implications of data nonrivalry and ownership.

3.1. Growth Rate and Labor Share in a Decentralized Economy

Proposition 1. *The economic growth of the decentralized economy on the balanced growth path does not exhibit scale effect, and the BGP growth rates can be expressed as follows*

$$g_c^* = g_y^* = g_N^* = g^* = \left[\frac{\sigma}{(1-\zeta)\sigma - \xi(1-\gamma)} \right] n. \quad (20)$$

The constraint on data provision per capita never binds under BGP, and its growth rate is:

$$g_\varphi^* = \frac{1-\zeta}{\xi} g^* - \frac{1}{\xi} n = \left[\frac{1-\gamma}{(1-\zeta)\sigma - \xi(1-\gamma)} \right] n < 0. \quad (21)$$

Online Appendix 1.2 contains the proof. Here, g_φ^* is negative: As the economy grows on a BGP, the data each person contributes steadily decrease in the long run, although the aggregate data use can still grow.

Because we recognize data as an input into the innovation possibility frontier, the BGP growth rates are related to not only the population growth (n), but also consumers' EIS ($1/\gamma$), relative contribution of data to innovation (ξ), and knowledge accumulation in the form of varieties of intermediate goods.

Specifically, note that all growth rates on BGP are ultimately driven by the exogenous population growth rate n ; growth rates become zeros when n becomes zero.⁷ In Online Appendix 2.1, we derive the parameter ranges for a BGP equilibrium to exist and to be unique. We restrict our discussions within these parameter ranges throughout the paper.

Next, when γ converges to 1, consumers' utility function converges to a logarithmic form as in Jones (1995), the BGP growth rate becomes, according to (20)

$$g_{\gamma \rightarrow 1}^* = \frac{n}{1-\zeta}. \quad (22)$$

Here, only knowledge accumulation and population growth rate influence the BGP growth rate. Yet, in contrast with $g_{Jones}^* = n(1-\xi)/(1-\zeta)$ in Jones (1995), our BGP growth rate is higher under the same set of parameters because data add positively to the innovation possibility frontier.

In general, when $\gamma > 1$ (as empirical studies in the macro and behavioral literature consistently estimate, e.g., Coen 1969, Lucas 1969, Vissing-Jørgensen 2002), we find that BGP growth rate increases with ξ , the extent that data influence innovation possibility frontier. Counter-intuitively, BGP growth rate also increases in σ , the severity of privacy concerns, and converges to (22). This general equilibrium effect stems from the fact that consumers in equilibrium require a higher growth rate to compensate for the disutility from greater privacy concerns or information leakage.⁸ Finally, the relationship between γ and growth rate is negative: Consumers prefer lower growth rate and thus less production when they are less willing to substitute current consumption with future consumption.

Besides growth rates, we derive the following result regarding labor shares in Online Appendix 1.3:

Proposition 2. *In this decentralized economy, the share of labor employed by the R&D sector s_D is constant in BGP, which is determined by*

$$s_D = \frac{1}{1+\Theta_D}, \quad \text{where } \Theta_D = \frac{g^*\gamma + \rho - n}{g^*(1-\xi)(1-\beta)}.$$

Here, the subscript "D" stands for decentralized. A larger growth rate g^* encourages firms to employ more labor in R&D. When the population stops growing ($g^* = 0$), labor in the R&D sector becomes zero because without growth, using labor in R&D only leads to disutility of consumers.

3.2. Growth Rate and Labor Share Under the Planner's Solution

We now derive the BGP growth rates and the shares of labor allocated in the two sectors under socially optimal allocations, which constitute a benchmark for comparison with those in the decentralized economy.

The equilibrium in a decentralized economy is not socially optimal because of monopolistic competition and knowledge spillover (the First Welfare Theorem fails here).

A social planner maximizes the utility of representative consumer/household, (1), subject to the resource constraint. The resource constraint requires that the aggregate consumption $C(t) = c(t)L(t)$ equals the aggregate net output, which we denote by $\tilde{Y}(t)$. In other words,

$$C(t) = \tilde{Y}(t) \equiv L_E(t)^\beta \int_0^{N(t)} x(v,t)^{1-\beta} dv - \int_0^{N(t)} \psi x(v,t) dv.$$

We first characterize the static allocation given $N(t)$ in every period. The social planner chooses the optimal level of intermediate goods input $[x(v,t)]_{v \in [0, N(t)]}$ at each time t given the time paths of $C(t)$, $\varphi(t)$, and $N(t)$, which is equivalent to maximizing $\tilde{Y}(t)$ with respect to $x(v,t)$. Thus, the optimal net output is:

$$\tilde{Y}_S(t) = \left(\frac{\psi}{1-\beta} \right)^{\frac{1}{\beta}} \beta N(t) L_E(t), \quad (23)$$

where the subscript “S” indicates the social planner’s problem. Relative to the decentralized economy (14), \tilde{Y}_S is larger with a markup of $(1-\beta)^{1-\beta/(2-\beta)}$ given the same level of labor and technology. This difference comes from the monopoly power in a decentralized economy, which lowers the provision of intermediate goods and thus, the net output of the final good.

Given (23), the social planner solves (dropping the subscript “S” to simplify notation):

$$\max_{c(t), \varphi(t)} \int_0^\infty e^{-(\rho-n)t} \left[\frac{c(t)^{1-\gamma} - 1}{1-\gamma} - \varphi(t)^\sigma \right] dt, \quad (24)$$

subject to

$$\dot{N}(t) = \eta N(t)^\zeta \varphi(t)^\xi l_R(t)^{1-\xi} L(t), \quad (25)$$

$$c(t) = \left(\frac{\psi}{1-\beta} \right)^{\frac{1}{\beta}} \beta N(t) l_E(t), \quad (26)$$

$$l_R(t) + l_E(t) = 1. \quad (27)$$

Here, (25) is the innovation possibility frontier, (26) is the simplified resource constraint, and (27) requires labor market clearing. In Online Appendix 1.4, we derive the following proposition:

Proposition 3. *With $\gamma > 1$, BGP growth rates in the social planner’s problem are the same as those in the decentralized economy.*

Despite the lower net aggregate output relative to the planner’s solution, the decentralized data economy grows at the same rate. The growth rates are

determined by the final goods production and the usage of data. The former is the same under the decentralized and the planner’s solutions. The latter exhibits a gap due to the markup created by the data price mechanism in the decentralized economy (as opposed to the planner’s directly setting data usage), which causes the decentralized equilibrium to be less socially efficient. But, the gap is a constant and does not manifest in growth rates.

Reminiscent of Jones (1995), growth rates alone cannot fully characterize the performance of an economy. We thus also examine the labor share in R&D and derive in Online Appendix 1.5:

Proposition 4. *In the social planner’s problem, the share of labor allocated in the R&D sector s_S is constant in BGP, which is determined by*

$$s_S = \frac{1}{1 + \Theta_S}, \quad \text{where} \\ \Theta_S = \frac{(\sigma - \xi)n + \xi\rho}{\xi(1 - \xi)g_S^*} - \frac{(\sigma - \xi)(1 - \zeta)}{\xi(1 - \xi)}. \quad (28)$$

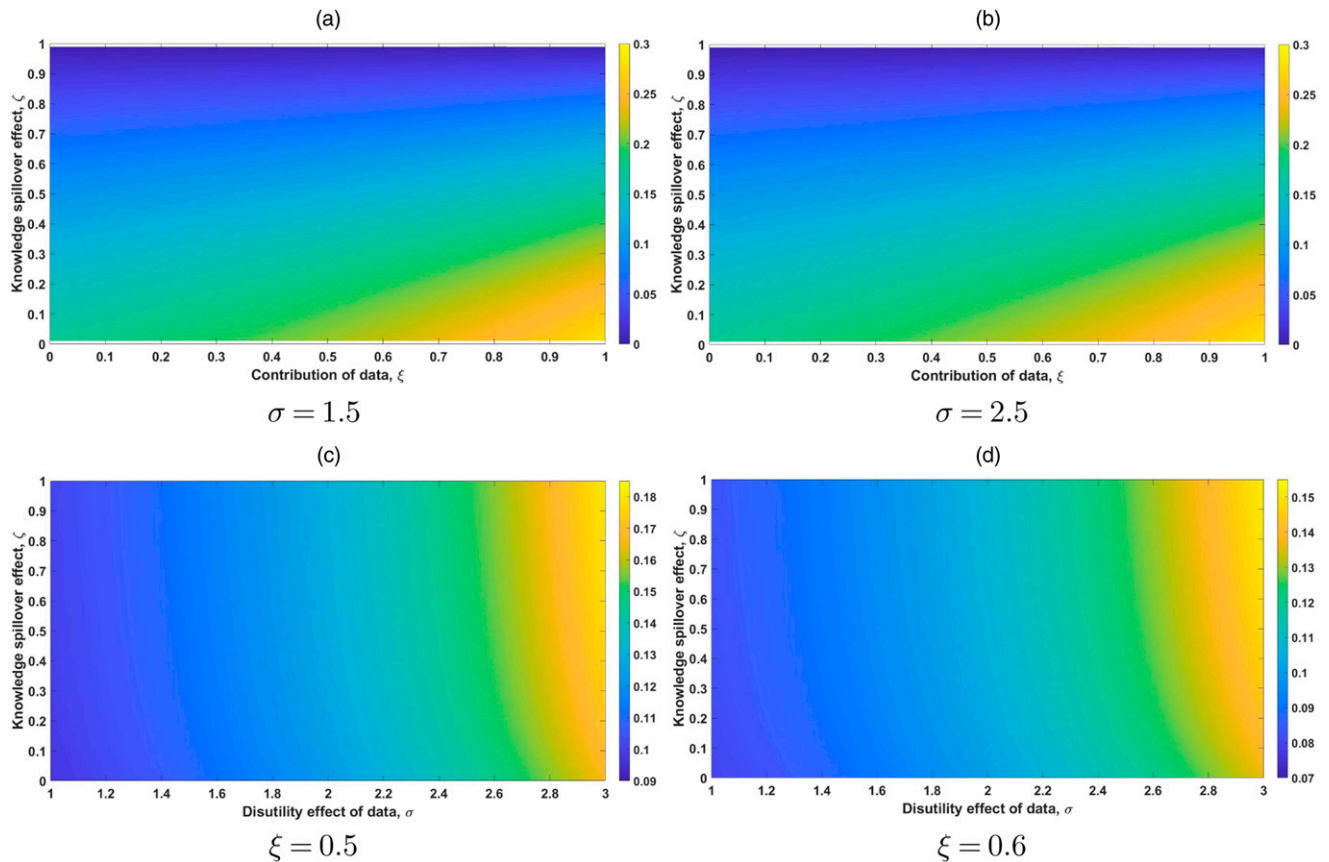
Without monopolistic production, the result differs from that in the decentralized economy. To ensure $s_S \in [0, 1]$, that is, $\Theta_S \geq 0$, we also note that the BGP growth rate under the planner’s solution cannot be too high:

$$0 < g_S^* \leq \frac{1}{1 - \zeta} \left(n + \frac{\xi}{\sigma - \xi} \rho \right). \quad (29)$$

This upper limit consists of three components: knowledge accumulation, population growth, and a data-related term. Intuitively, a social planner does not want the growth of the economy to be too fast since higher growth rates require more data usage, which can create excessive data leakage or privacy violations in expectation. When $\gamma > 1$, g_S^* is given by (20) and (29) always holds. But when $\gamma < 1$, it is possible that the upper limit binds and the planner’s solution features slower growth.

3.3. Misallocation and Data Overuse in a Decentralized Economy

Endogenous labor allocations between production and R&D sectors influence other variables in equilibrium. In particular, with n, β, γ and ρ taking on standard values from the existing literature, the labor allocation in the R&D sector in the social planner’s problem is always larger than that in the decentralized economy. Figure 1 plots this difference as we vary the influence of data in innovation possibility frontier (ξ), knowledge accumulation through innovation possibility frontier (ζ), and the severity of privacy concerns in the consumers’ utility function (σ). The larger σ is, the more the failure to internalize privacy concerns creates

Figure 1. (Color online) Difference in Labor Employed in the R&D Sector Between the Two Cases

Notes. The figure shows the difference of labor allocated in R&D sector between decentralized economy and the social planner's problem. Light color represents larger differences (greater misallocation) and dark represents smaller ones. Other parameters are set as $n = 0.02$, $\beta = 2/3$, $\gamma = 2.5$, $\rho = 0.03$, which are standard values used in existing literature. $\xi \in [0, 1]$, $\sigma > 1$ ensures the convexity of disutility term in consumer's utility function and $\zeta \in [0, 1]$ ensures the existence of BGP.

misallocation; the overuse of data and underallocation of labor in R&D are especially severe when data are important for innovation (large ξ) and knowledge accumulation is slow (small ζ).

Our model thus reveals a new source of inefficiency in the decentralized economy, even though the growth rates are the same as those in the planner's solution: undersupply of labor and overuse of data in the R&D sector. This finding contrasts and complements recent studies such as Jones and Tonetti (2020), which predicts an underutilization of data due to its nonrival nature. In our setting, data are overused even when owned by consumers who directly factor in the disutility from data leakage and/or abuse.

Similar to Jones (1995), the underallocation of labor in the R&D sector comes from monopolistic markups in the production of intermediate goods. The final good producer employs more labor to compensate for the lower production and usage of intermediate goods, which in turn crowds out labor employed in R&D. The crowding out of labor in R&D is exacerbated because intermediate producers are less reliant on

labor, once data enter the evolution of the innovation possibility frontier (unlike Jones 1995, in which labor is the only input for the innovation possibility frontier). To maintain the same growth rate as in the planner's solution, intermediate producers have to compensate the underemployment of labor in R&D by more aggressive utilization of data. In other words, this crowding-out effect of labor crowds in data usage in the R&D sector to a socially excessive level. Given the parameter set we have, data usage is four to five times higher in the decentralized economy than in the social planner's problem.

3.4. Data Generation and Regulatory Policies

Different from R&D-specific labor, the production of data is endogenized by consumption, which itself is also endogenous. More data supplied add to innovations on variety, which in turn spurs consumption (e.g., in BGP), which then further relaxes the data generating constraint. Such a feedback is absent when it comes to labor or capital in that consumption is typically

decoupled from the exogenous evolution of population size or the growth in potential capital to be allocated.

As such, privacy regulations reducing s in (3) may affect the production and thus the usage of data.⁹ However, reducing s would reduce growth rate and would not lead to welfare improvement without sacrificing growth, if it improves welfare at all. A full analytical characterization is not tractable, not to mention that such an intervention entails intergenerational transfers in practice, which can be controversial. We therefore restrict our discussion in this section to policy interventions that preserve growth rates in the planner's BGP solution.

We find that levying a tax on the usage of data alters the transitional dynamics but is ineffective in bringing the equilibrium allocations in decentralized economy closer to the social planner's solution because as discussed in Online Appendix 3.1.1, it does not solve the underemployment of labor in R&D but only slows down the economy before it eventually returns to the original BGP path. However, subsidizing labor wage rate in the R&D sector or subsidizing intermediate producers in terms of profit proves to be effective. Appendices 3.1.2 and 3.1.3 derive these optimal subsidies.

The intuition is that labor wage is pinned down by both the R&D and production sectors, and a subsidy directly affecting the wage level alters the labor share. Specifically, because labor allocations are derived from equalizing wages in the R&D sector and production sector, a government can apply taxes or subsidies to adjust the prices of factors to alleviate the overuse of data in intermediate producers' R&D. For example, a subsidy of rate $0 < \tau(t) < 1$ to the R&D sector for employing labor would modify (19) into

$$\eta(1 - \xi)N(t)^\zeta \varphi(t)^\xi l_R(t)^{-\xi} V(t) = \tau(t)w(t),$$

while (15) remains unchanged. Then, one can derive the desirable $\tau(t)$ by equating the labor share to that in the social planner's problem.¹⁰ To avoid overuses of data and potential privacy violations, intermediate producers should be incentivized to employ more labor for innovation with subsidies on wages in the R&D sector, which would lower data usage.

In contrast, data price is pinned down by intermediate producers and consumers. The consumers' joint decision on consumption and data provision implies that they care about the growth rate of data provision, not the level, as seen in (5). In a sense, a direct tax on data purchase is decoupled from the data provision and because of that, would not alter the equilibrium labor share.¹¹ Consequently, the underemployment in the R&D and data overuse persist.

Our findings have important policy implications because the current debate has centered on privacy

regulations, which decisively affect the data economy's growth. We demonstrate how R&D labor relates to the increasing use of data and how labor market policies can effectively reduce excessive data overuse to improve the social welfare.

3.5. Historical Data and Dynamic Data Nonrivalry

One unique and important property of data we highlight thus far is dynamic nonrivalry. Beyond its implicit manifestation through the evolution of intermediate good varieties in the baseline specification, we now extend our discussion of dynamic nonrivalry by explicitly modeling the trading of historical data and the associated creative destruction. Specifically, instead of one potential intermediate producer using data for research (and entry) with full depreciation, we now assume:

- Data generated at time t can also be used and traded in the following $M > 0$ periods. Historical data depreciate at a rate of δ .
- Potential intermediate producers can purchase not only data from current consumers, but also historical data from existing intermediate producers entering in the past M periods. The new and historical data bundles are perfect substitutes.
- The owners of historical data determine the proportion of data sold to newcomers, fully recognizing potential creative destruction.
- If historical data are reused by new intermediate producers, disutility of contributing consumers is also cumulative according to the usage.

In a sense, data become a stock variable, whereas labor is a flow variable in the intermediate good production function. Online Appendix 3.2 contains detailed derivation and discussion, which are summarized here:

Proposition 5. *Dynamic data nonrivalry under creative destruction as specified does not change the BGP growth rates or labor share in the decentralized economy and social planner's problem. However, incumbent intermediate producers are always willing to sell more than the social optimal level of historical data to entrants in the decentralized economy, as long as the negative effect of creative destruction does not come to an extreme level.*

The implicit dynamic nonrivalry in our baseline model still exists: Data are transformed into new varieties of intermediate goods, and the existing level of varieties can affect the R&D process in future. But, the proposition demonstrates that the key findings are robust to explicitly modeling data nonrivalry and the direct use of historical data. Therefore, the differences between our findings and Jones and Tonetti (2020)'s are driven by the dynamic nature of data usage we model, not by the simple consideration of nonrivalry. Moreover, the proposition has implications for regulating data resale among intermediate producers, an interesting topic for future studies.

3.6. Data Ownership

So far, we have allowed consumers to own data and the firms (intermediate producers) to acquire data by paying the consumers. We now consider cases with positive data processing cost $\theta > 0$ under both consumer ownership and firm ownership (detail derivation in Online Appendix 3.3). $\theta > 0$ would not affect much of the equilibrium outcomes when consumers own data as long as the knowledge spillover effect is moderate, but consumers' EIS are reasonably moderate and the data processing cost is sufficiently convex (which holds true due to, e.g., curse of dimensionality). The main new insight from the analysis is the positive BGP growth rate of data provision when firms own data. In other words, data usage under BGP could be increasing when firms own data, which is different from the decreasing trend when consumers own data.

Intuitively, because firms no longer need to pay consumers for data usage, they no longer bear the disutility of potential privacy violation. Although they still pay a mechanical "data processing cost", it does not push the BGP growth rate of data provision to negative as we have seen in the case in which consumers own data. That said, under standard parameter values, for example, $\xi = 0.5$, $\zeta = 0.85$, and $\phi > 1$, the BGP growth rate of other variables, g^* , is still larger than g_φ^* . This is consistent with our baseline model: Data usage becomes trivial in BGP compared with consumption as time passes.

Moreover, Sections 3.5 and 3.6 demonstrate that our finding on data overuse is robust to nonrivalry and ownership variation. This result differs from and complements Jones and Tonetti (2020), which emphasizes the underutilization of data due to its cross-sectional nonrivalry and the importance of consumer data ownership. Because data add to innovation over time and have persist benefits in our model (instead of only entering contemporary production as in Jones and Tonetti (2020)), firms are more willing to utilize and purchase data if consumers own data.

4. Transitional Dynamics: A Numerical Analysis

The transitional dynamics of the data economy onto a BGP also reveal interesting patterns, as we demonstrate numerically with the baseline model setup. Even without regulations restricting the use of data, a decline in data usage could occur as the economy

grows, which mitigates the concern of data privacy. At the same time, for data economies with minimal initial growth, the process of data generation dictates that they may be trapped in low-growth regimes for a long time without any intervention to improve digital infrastructure or relax privacy regulation.

4.1. Methodology and Calibration

Similar to Jones (2016), we focus on the planner's solution in this numerical exercise mainly for tractability. This would be the situation if a benevolent government implements policies, balancing privacy protection and growth, such as the subsidy scheme we mentioned earlier. Instead of a formal calibration designed to replicate any country's data, this analysis is best viewed as an illustration of the basic transitional dynamics that are possible in our theoretical framework.

We derive in Online Appendix 2.3 a system of differential equations that describe the dynamics of the planner's economy. They consist of a constraint on data generation (3) and three state-like variables whose interpretations and steady state values are shown in Table 1. Other variables can be derived from them. We solve the system of differential equations using "reverse shooting" (Judd 1998).¹² Table 2 provides a summary of parametrization.

4.2. Results and Discussions

We can simulate the paths of the three state-like variables before reaching equilibrium for different values of σ , which stands for the extent of disutility caused by privacy leakage. We can then derive the paths of other variables like $c(t)$ and $\varphi(t)$. Figure 2 provides an illustration of the transitional dynamics for $\sigma = 1.5$ with and without the data provision constraint here. Other cases exhibit similar patterns, which we discuss further in Online Appendix 2.4.

Several robust patterns emerge. No matter where an economy starts before BGP, the growth rate of consumption and variety both move to the steady state levels, and a transition can happen relatively quickly once the economy reaches nontrivial growth (sufficiently away from zero), as (a) and (c) show. Moreover, as shown in (b) and (d), the growth rate of data provision decreases from positive to negative in BGP. For an economy starting from low growth, the provision of data increases rapidly for the accumulation of varieties of intermediate goods, which contributes to

Table 1. State-like Variables for Studying Transitional Dynamics

Variable	Meaning	Steady state value
$g_N(t)$	Growth rate of variety of intermediate goods	(20)
$g_\mu(t)$	Growth rate of shadow price corresponding to technology change	$g_\mu^* = \frac{\sigma(1-\zeta)-\xi}{\xi} g_N^* - \frac{\sigma}{\xi} n$
$l_E(t)$	Ratio of labor employed in production sector	(28)

Table 2. Parameters for Studying Transitional Dynamics

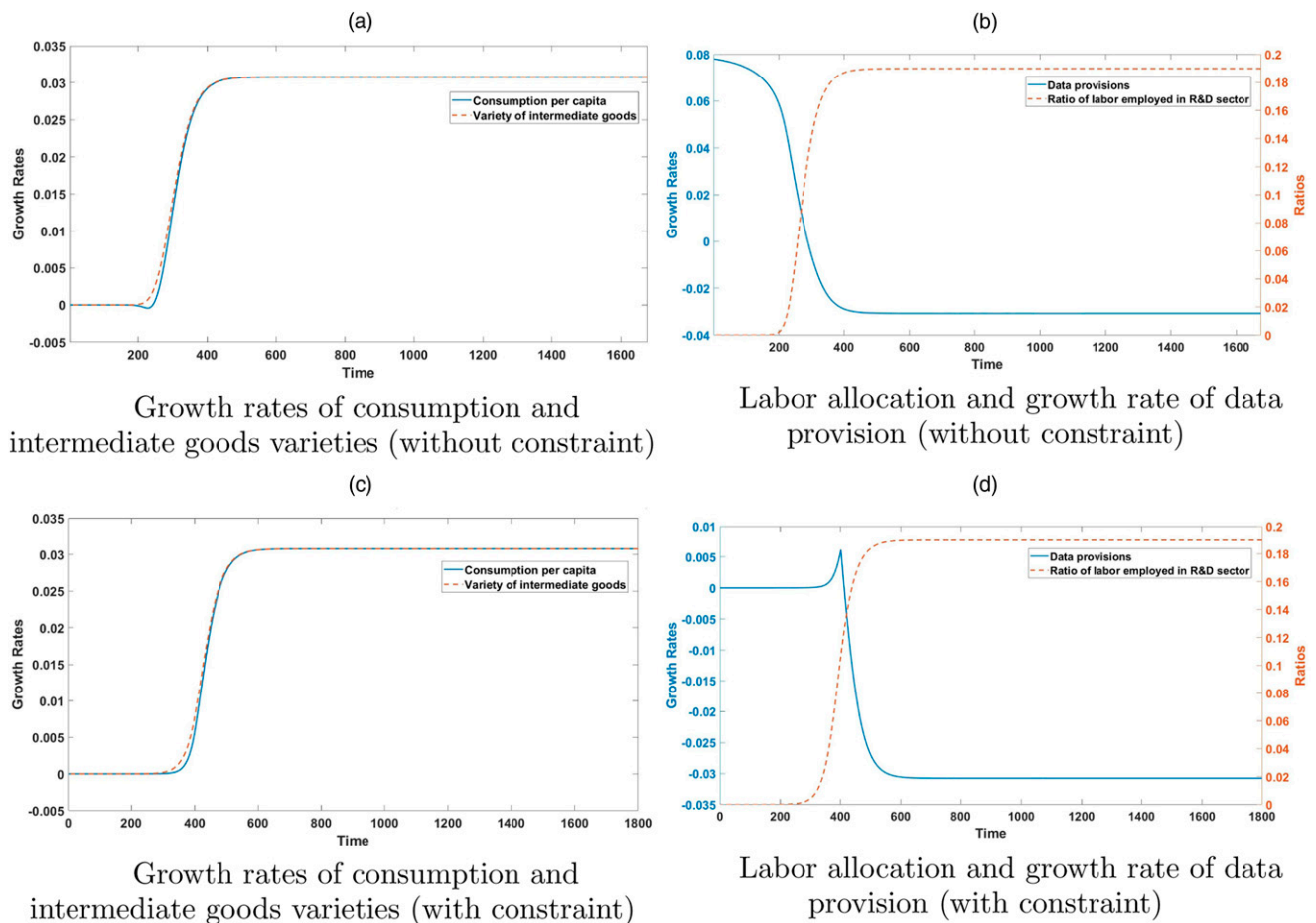
Variable	Meaning	Value	Source
β	Contribution of labor in final good production function	2/3	Standard
γ	Reciprocal of Elasticity of Intertemporal Substitution	2.5	Standard
ρ	Subjective discount factor	0.03	Standard
ξ	Contribution of data in innovation possibility frontier	0.5	Discretionary
ζ	Knowledge accumulation through innovation possibility frontier	0.85	Discretionary
σ	Severity of consumers' privacy concern	1.5	Discretionary
n	Population growth rate	0.02	Standard
η	Efficiency term in innovation possibility frontier	1	Standard

the final good production and growth in consumption. Along this transitional path, labor moves from production sector to R&D sector, which reflects how labor is used to compensate for the decreasing provision of data.

Notice that in Panel (a), different from the growth rate of variety, the growth rate of consumption undergoes some periods of negative growth before increasing to the positive growth rate without constraints on data provision. The temporary pain is common in the growth

literature (e.g., Brock and Taylor 2010) and indicates that at the beginning of the adjustment to high BGP growth, labor moves out from the production sector, which causes the decrease of output and consumption. But data are unique because economic activities constrain their supply, as is clear from (d) in which data provision is binding from time 0 to 400. The temporary pain is absent because moving labor away from production is not as costly as before, due to declines in data provision (which lead to slower growth rates).

Figure 2. (Color online) Key Variables Along Transitional Path When $\sigma = 1.5$



Notes. These figures show the transitional dynamics of the social planner's problem when $\sigma = 1.5$, with and without the constraint of data provision. The economies undergo long but relatively steady states before finally reaching BGP at the end points.

The transitional dynamics is also broadly consistent with the empirical patterns in Abis and Veldkamp (2021) that the labor income share in knowledge work is decreasing. Formally, labor income share is decreasing in $l_R(t)$ in our model:

$$\begin{aligned} \text{Labor Income Share} &= \frac{w(t)L(t)}{p_\varphi(t)\varphi(t)L(t)} \\ &= \frac{\eta(1-\xi)N(t)^\zeta\varphi(t)^\xi l_R(t)^{1-\xi}V(t)L(t)}{\eta\xi N(t)^\zeta\varphi(t)^\xi l_R(t)^{1-\xi}V(t)L(t)} \frac{1}{l_R(t)} \\ &= \frac{1-\xi}{t\xi} \frac{1}{l_R(t)}. \end{aligned}$$

With similar extent of labor allocation in intermediate and final goods in production sectors under the decentralized economy, our numerical simulations reveal this decreasing trend of labor income share as $l_R(t)$ increases in transitional states.

4.3. Implications for Growth Trap and Privacy Concerns

Comparing the cases with and without the constraint of data provision, we find that constraint (3) binds for an economy with low initial growth rate. The per capita data contribution is also much lower than the case without the constraint. Importantly, for an economy starting with a growth rate close to zero, it takes almost 200 additional time periods (years) for it to reach BGP. Even after reaching BGP growth, the economy's output could be significantly lower due to the delay of accelerated growth. The intuition is straightforward: Initial low growth limits data generation, which negatively feeds back to growth because data constitute an input factor for innovations in intermediate good variety.

An intervention to boost the initial growth rate is crucial for escaping such a growth trap. Meanwhile, we should notice that the social planner problem is at a national level. Thus, intervention to relax the data provision constraint (3) here is more at an international level, for example, actions taken by the World Bank or IMF or fellow countries in European Union to improve digital infrastructure and data collection/storage efficiency or share expertise, which relax (3). Also, if the central planner is originally too budget/cash constrained to improve data infrastructure, foreign subsidies might help relax the binding constraint temporarily.

Global waves of technological innovations such as the Internet and AI take place when countries are having different population growth and different environments for technology transfer and spillover. This can lead to vastly different transitional dynamics. The tightness of data constraint s does not affect the BGP growth rates, but affects transitional dynamics. For

example, the United States and Europe are in the same stage of development when data economy emerges, but the difference in privacy protection (different s) makes them tread different paths, reaching BGP at different times.

First, as the economy just starts to grow, data can significantly facilitate the transition to BGP, as illustrated in Figure 3: Economies 1 and 2 have the same parameters but start at different stages in their transitions to BGP, with Economy 1 trapped in near zero growth for a prolonged period of time. An intervention to enhance data facilities and digital infrastructure or to relax privacy regulation (increasing s) can help escape the growth trap earlier. However, even when such interventions fully relax (3), that is, $s \rightarrow \infty$, they compensate the lagging behind Economy 2 only partially.

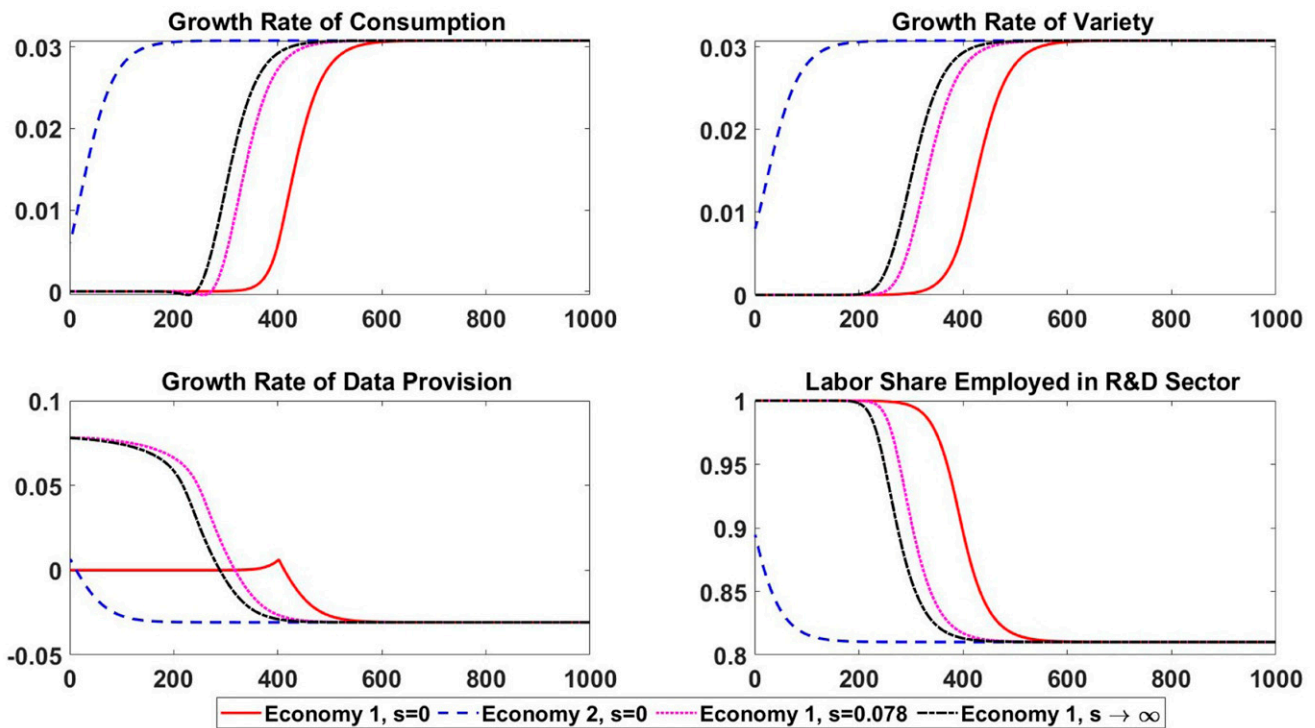
In the long run, because the knowledge accumulation parameter ζ is set to be smaller than one, it becomes less effective to provide more data for creating new varieties of intermediate goods as the existing varieties accumulate. Thus, the growth rate of data provision decreases after some periods of increase. As data become less productive in the innovation process, the economy substitutes the use of data with more labor to focus on better exploitation of data. Finally, as the economy matures (transition to BGP), the benefit of using data are diminishing but individual's privacy concerns remain, which reduce each consumer's endogenous data contribution. In that regard, privacy issues in the long run may not be as severe as current debates indicate. Instead, regulatory policies could focus on the overuse of data in the R&D sector, as discussed in Section 3.4.

5. Conclusion

We develop an endogenous growth model against the backdrop of the rise of big data and digital economy. Although a decentralized economy on the balanced growth path grows at the same rate as in the socially optimal allocations, data are inefficiently overused and R&D is understaffed. Consumers suffer because they are inadequately compensated for potential information leakage and privacy violation. When consumers own data, data privacy concerns become allayed in the long run because the use of data eventually declines. However, less developed economies with low growth at the dawn of the data economy may face a new form of poverty trap that potentially warrants interventions.

For the first time, we treat data as an input factor besides labor in the process of creating new varieties of intermediate goods, which subsequently fuel the production of final good and long-run growth. We highlight data's endogenous generation as by-products of economic activities, nonrivalry in a dynamic environment, and flexible ownership. For tractability and

Figure 3. (Color online) Different pre-BGP Paths Given Different Starting Points and Different Privacy Policies



Notes. These panels show the transitional dynamics of two economies with the same parameters but different starting points. The full line and the dash line show the paths of the two economies with $s = 0$, the dotted line shows the path of Economy 1 when its privacy policy becomes loose ($s = 0.078$), and the dash dotted line shows the path of Economy 1 when privacy regulation is completely relaxed ($s \rightarrow \infty$).

focus, we necessarily leave out certain aspects of the data economy such as final goods differentiation. Therefore, our findings should be taken as first-order benchmark results rather than foregone conclusions.

Acknowledgments

The authors thank Jasmine Cheng, Kay Giesecke, Shiyang Huang, Laura Veldkamp, Ji Yan, Buyuan Yang, Liyan Yang, and the anonymous associate editor and referees for their feedback. The authors also thank conference and seminar participants at the Midwest Macroeconomics Meeting, Australasian Finance and Banking Conference, Delhi School of Economics and the Econometric Society Winter School, Machine Lawyering’s Conference for helpful comments and discussions. The contents of this publication are solely the responsibility of the authors.

Endnotes

¹ Data have become a key input to many technology and financial firms (e.g., Arrieta-Ibarra et al. 2018, Cong et al. 2019, Veldkamp and Chung 2019); Manyika et al. (2011) estimate that big data research is believed to save over 100 billion Euros for Europe, and reduce medical care cost of the United States by 8% or 300 billion dollars annually.

² Cambridge Analytica and Facebook’s data scandal presents a salient example (Confessore 2018). The United States has the highest ratio of data records stolen relative to population (over 6 billion stolen records, exceeding the population by 19 times, Sobers 2020). Data from *China Judgements Online* and *China Academy of Information*

and *Communications Technology* also reveal that civil and criminal cases of data leakage have been increasing in lock-steps with the growth of China’s digital economy.

³ Unlike the growth literature focusing on firms’ property rights, we allow consumers to own and sell data.

⁴ “Selling” can be broadly interpreted: For example, not having to pay for the use of Gmail can be viewed as compensation to consumers for allowing Google to use their data in ways delineated in the usual terms and conditions. In Section 3.6, we allow firms to own data as in Jones and Tonetti (2020) and derive additional insights in addition to showing robustness of the main results.

⁵ In some sense, data acquired by firms are their intangible capital, akin to customer capital discussed in recent studies (e.g., Dou et al. 2021).

⁶ We follow the convention in the literature to maintain constant return to scale of inputs; $\zeta > 0$ corresponds to the positive external returns due to the cumulative nature of innovation (e.g., Chang 1995, Cong and Howell 2021), whereas $\zeta < 1$ follows the innovation of Jones (1995) over Romer (1990) to reflect that it becomes harder to come up with new varieties of intermediate goods as the variety expands.

⁷ Like Jones (1995), our model does not suffer from the scale effect in, for example, Romer (1990), whose endogenous growth rates depend on population level.

⁸ Similar phenomenon can also be found in Stokey (1998): Higher growth rates are needed in equilibrium when there is greater pollution because people dislike the harm caused by pollution and require compensation.

⁹ Both the GDPR and the CCPA give individuals certain rights over the collection and usage of their personal information, but they

differ in multiple aspects (e.g. Hospelhorn 2020). For example, California's being a much larger economy implies more severe penalties than that of the GDPR, which maps to a tighter constraint (3).

¹⁰ In Online Appendix 3.1.2, we show that this subsidy rate should be a constant. Because only the changing rate of variables may influence growth rates (as seen in Online Appendix 1.2), applying a fixed subsidy rate does not introduce further distortions into the model.

¹¹ A nonconstant tax rate would not work either because it alters the BGP growth rate in the decentralized economy, which is originally the same as that in the social planner's solution.

¹² We start from the steady state and run the system backward according to the three differential equations. We first find the values of the three state-like variables that minimize the distance between their growth rates and zero, because they all converge to nonzero constants in the steady state. For other parameters, we choose the standard values in the existing literature or plausible values if they are not discussed in extant studies (ξ , σ , and ζ). We set $\xi = 0.5$ to indicate that data and labor contribute equally in creating new varieties; we discuss various choices of σ in Online Appendix 2.4; we set ζ to be a value less than one following Jones (1995).

References

- Abis S, Veldkamp L (2021) The changing economics of knowledge production. Working paper, Columbia Business School, New York.
- Acemoglu D, Aghion P, Bursztyn L, Hemous D (2012) The environment and directed technical change. *Amer. Econom. Rev.* 102(1):131–166.
- Admati AR, Pfleiderer P (1990) Direct and indirect sale of information. *Econometrica* 58(4):901–928.
- Akçura MT, Srinivasan K (2005) Research note: Customer intimacy and cross-selling strategy. *Management Sci.* 51(6):1007–1012.
- Arrieta-Ibarra I, Goff L, Jiménez-Hernández D, Lanier J, Weyl EG (2018) Should we treat data as labor? Moving beyond “free.” *AEA Pap. Proc.* 108:38–42.
- Barkai S (2020) Declining labor and capital shares. *J. Finance* 75(5):2421–2463.
- Brock WA, Taylor MS (2010) The Green Solow Model. *J. Econom. Growth* 15(2):127–153.
- Casadesus-Masanell R, Hervas-Drane A (2015) Competing with privacy. *Management Sci.* 61(1):229–246.
- Chang HF (1995) Patent scope, antitrust policy, and cumulative innovation. *RAND J. Econom.* 26(1):34–57.
- Coen RM (1969) Tax policy and investment behavior: A comment. *Amer. Econom. Rev.* 59(3):370–379.
- Confessore N (2018) Cambridge Analytica and Facebook: The Scandal and the Fallout So Far. *New York Times* (April 4).
- Cong LW, Howell ST (2021) Policy uncertainty and innovation: Evidence from IPO interventions in China. *Management Sci.* Forthcoming.
- Cong LW, Li B, Zhang T (2019) Alternative data for FinTech and business intelligence. Preprint, submitted February 11, <http://dx.doi.org/10.2139/ssrn.3521349>.
- Dou W, Ji Y, Reibstein D, Wu W (2021) Inalienable customer capital, corporate liquidity, and stock returns. *J. Finance* 76(1):211–265.
- Easley D, Huang S, Yang L, Zhong Z (2019) The economics of data. Working paper, Cornell University, Ithaca, NY.
- Economist, The* (2018) Worries about the rise of the gig economy are mostly overblown. *The Economist* (October 4), 64–66.
- Farboodi M, Veldkamp L (2020) Long-run growth of financial data technology. *Amer. Econom. Rev.* 110(8):2485–2523.
- Hirshleifer J (1971) The private and social value of information and the reward to inventive activity. *Amer. Econom. Rev.* 61(4):561–574.
- Hospelhorn S (2020) California Consumer Privacy Act (CCPA) vs. GDPR. *Inside Out Security Blog*, June 17, <https://www.varonis.com/blog/ccpa-vs-gdpr/#>.
- Ichihashi S (2020a) Competing data intermediaries. Working paper, Bank of Canada, Ottawa, Ontario.
- Ichihashi S (2020b) Online privacy and information disclosure by consumers. *Amer. Econom. Rev.* 110(2):569–595.
- Jones CI (1995) R&D-based models of economic growth. *J. Political Econom.* 103(4):759–784.
- Jones CI (2016) Life and growth. *J. Political Econom.* 124(2):539–578.
- Jones CI, Tonetti C (2020) Nonrivalry and the economics of data. *Amer. Econom. Rev.* 110(9):2819–2858.
- Judd KL (1998) *Numerical Method in Economics* (MIT Press, Cambridge).
- Karabarbounis L, Neiman B (2014) The global decline of the labor share. *Quart. J. Econom.* 129(1):61–103.
- Liao L, Wang Z, Yan H, Yang J, Zhou C (2020) Angry borrowers: Negative reciprocity in a financial market. Preprint, submitted December 5, <http://dx.doi.org/10.2139/ssrn.3415808>.
- Liu JZ, Sockin M, Xiong W (2020) Data privacy and temptation. NBER Working Paper No. 27653, National Bureau of Economic Research, Cambridge, MA.
- Lucas RE (1969) Labor-capital substitution in U.S. manufacturing. Harberger A, Bailey MJ, eds. *The Taxation of Income from Capital* (The Brookings Institution, Washington, DC), 223–274.
- Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, Byers AH (2011) Big data: The next frontier for innovation, competition and productivity. Technical report, McKinsey Global Institute (McKinsey & Company).
- Murphy RS (1996) Property rights in personal information: An economic defense of privacy. *Georgetown Law J.* 84(7):2381–2417.
- Romer PM (1990) Endogenous technological change. *J. Political Econom.* 98(5):S71–S102.
- Sobers R (2020) The world in data breaches. *Inside Out Security Blog* (March 29), <https://www.varonis.com/blog/the-world-in-data-breaches/>.
- Stokey NL (1998) Are there limits to growth? *Internat. Econom. Rev.* 39(1):1–31.
- Tang H (2019) The value of privacy: Evidence from online borrowers. Working paper, London School of Economics, London.
- Veldkamp L (2005) Slow boom, sudden crash. *J. Econom. Theory* 124(2):230–257.
- Veldkamp L, Chung C (2019) Data and the aggregate economy. Working paper, Columbia Business School, New York.
- Vissing-Jørgensen A (2002) Limited asset market participation and the elasticity of intertemporal substitution. *J. Political Econom.* 110(4):825–853.