Data-driven innovation and growth

Li, Hao and Wang, Gaowang and Yang, Liyang

13 October 2024

Data-Driven Innovation and Growth*

Hao Li† Shandong University

Gaowang Wang‡ Shandong University

Liyan Yang§ University of Toronto

October 13, 2024

Abstract

We develop an endogenous growth model where data drives innovation. In this model, big data fosters quality improvements by influencing the likelihood and magnitude of successful quality-enhancing innovations. It also promotes variety innovation through the efficient allocation of labor as a fixed cost, ultimately driving long-run economic growth. The social planner reduces the welfare costs associated with monopoly production and internalizes the externalities present in decentralized economies. As a result, the optimal growth rate exceeds the equilibrium growth rates under two data property rights regimes. Data property rights play a crucial role in determining long-run growth and steady-state welfare, which depend largely on two key model parameters: the weight for privacy and the frequency of creative destruction. This model also explores the interactions between quality innovation and variety innovation.

Keywords: data as innovation; endogenous growth; data property rights; interactions between quality innovation and variety innovation

JEL Classifications: E1, O3, O4

*We thank seminar participants of Chinese Academy of Sciences, Central University of Finance and Economics, Northeast Normal University, Shandong University, and University of International Business and Economics for helpful comments and suggestions. Gaowang Wang thanks the National Natural Science Foundation of China (72473082) and the Qilu Young Scholar Program of Shandong University for its financial support. All remaining errors are our responsibility.

†Center for Economic Research, Shandong University, Jinan, China. E-mail: hao.li.eco@gmail.com.

‡Center for Economic Research, Shandong University, Jinan, China. E-mail: gaowang.wang@sdu.edu.cn.

§Department of Finance, Joseph L. Rotman School of Management, University of Toronto, Toronto, Ontario M5S 3E6, Canada. E-mail: liyan.yang@rotman.utoronto.ca.

1 Introduction

The importance of data in the economy has become increasingly apparent in recent years. Big data refers to large volumes of data and the accompanied technological innovation used to gather, store, and process them (Farboodi and Veldkamp, 2023). Data are not only used for prediction and production but also for innovations, including both variety and quality innovations. As Farboodi and Veldkamp (2023) argue, big data fosters innovation: "By analyzing the data related to user behavior, companies can discover patterns that can identify the need for a new product or an upgrade of an existing one. Thus, big data fosters innovation as well." Many researchers explore how big data affect technological innovations and scientific discoveries from different perspectives.[1] In this paper, we develop an endogenous growth model in which big data affect quality innovation (by influencing the probability and magnitude of successful innovations) and variety innovation (through effective labor input), thereby driving long-run economic growth. We also examine the distortions driven by monopoly and externalities, the growth effects of data property rights, and the interactions between quality and variety innovations.

This article delivers three key messages. First, we introduce a new perspective on big data: data as innovation. Current research models data as prediction (Farboodi and Veldkamp, 2019), as production (Jones and Tonetti, 2020; Cong et al. 2022), or as variety innovation (Cong et al. 2021; Cong et al. 2022). We model data as both quality and variety innovation. In our model, individual quality innovations follow homogenous Poisson processes, where data increase both the probability and magnitude of successful quality innovations. Meanwhile, the creation of new varieties requires a certain amount of effective labor equipped with big data. This model also generalizes Dinopoulos and Thompson's (1998) Schumpeterian growth model without scale effects by incorporating big data. Unlike the exiting literature on data and growth, we provide an endogenous growth model with a microfoundation that explains how big data influence innovation and growth in a data-driven economy. Table 1 summarizes the perspectives of current research on big data and growth.

Table 1: Perspective and literature

[1] Nielsen (2012) illustrates the myriad ways in which big data and associated technologies are changing the mechanisms of discovery in science. Cockburn et. al. (2018) examines the implications of artificial intelligence and machine learning as general-purpose technologies (GPTs) for invention. Agrawal et. al (2018) develop a model to explore the importance of artificial intelligence in finding useful combinations in complex discovery spaces. Martin (2020) shows that most benefits of big data and related technologies involve improvements in the quality of products, services and customer access.

Perspective	Literature
Data as prediction	Farboodi & Veldkamp (2019)
Data as production	Jones & Tonetti (2020), Cong et al. 2022
Data as variety innovation	Cong et al (2021), Cong et al. (2022)
Data as innovation (variety and quality)	Our model

Second, optimal growth exceeds equilibrium growth, and data property rights significantly impact both growth and welfare. In the decentralized economy of the model, there are three types of distortions: (i) monopoly power over all varieties, (ii) a positive externality of the average quality level on all individual quality-enhancing innovations, and (iii) a negative externality on consumers' privacy conerns from firms' data sales when firms own the data, or a negative externality on firms' profits from consumers' data sharing due to creative destruciton when consumers own the data. When making allocation decisions, the social planner reduces the welfare cost of monopoly and internalizes all externatlities, thereby improving the equilibrium growth rate, regardless of whether data are owned by firms or consumers. This result differs from existing growth models involving data. For example, Jones and Tonetti (2020) shows that the optimal growth rate equals both equilibrium growth rates, as the long-run effects of market distortions in decentralized economies may cancel each other out under both data ownership regimes. Cong et al. (2021) find that the optimal growth rate equals the equilibrium growth rate when consumers own the data but is lower than the equilibrium growth rate when firms own the data. Cong et al. (2022) demonstrate that the optimal growth rate is higher than the equilibrium growth rate when consumers own the data.

We also show that data property rights matter for long-run growth, and the ranks/discrepancies between these two equilibrium growth rates largely depend on three key parameters: the weight of privacy concerns κ, the frequency of creative destruction δ_0, and the elasticity of substitution (EIS) for goods θ. Specifically, (1) If the weight for privacy κ is relatively large (small)-for instance, greater (less) than a critical value-then quality growth with consumers owning data is higher (lower) than with firms owning data. If the privacy weight κ equals the critical value, then the two equilibrium growth rates under different data property rights regimes are close to each other.

Intuitively, when firms own data, privacy concerns κ do not affect the optimizing behaviors of agents (i.e., consumers, firms, and data intermediaries) and therefore have no significant

impact on the economy. However, when consumers own data, changes in privacy concerns have two opposing effects on long-run quality growth. If privacy concerns increase, consumers sell less data. Fewer data sales and less data sharing reduce the frequency of creative destruction, thereby mitigating profit losses across varieties. For each variety, more resources are allocated to quality-enhancing activities, boosting quality growth. On the other hand, reduced data sales lowers consumers' revenues, diminishing per capita consumption expenditure and reducing market demand for each variety. As a result, each variety earns less profit, leading to fewer resources being devoted to innovation, which dampens quality growth.

Thus, if consumers value their privacy more than the critical threshold, the positive effect dominates, and the equilibirum growth rate is higher than when firms own data. Conversely, if consumers care less about privacy, the negative effect prevails, and the growth rate is lower than when firms own data. Additionally, if privacy concerns approach the critical value, the opposing effects nearly cancell each other out, and the two equilibrium growth rates converge.

(2) If the frequency of creative destruction δ_0 is relatively large (or small)-for example, greater (or less) than a critical value-then quality growth with consumers owning data is lower (or higher) than with firms owning data. If the frequecney of creative destruction equals the critical value, then the two equilibrium growth rates are close to each other.

Intuitively, an increase in the frequency of creative destruction has two opposing effects on quality growth. Given data sales, an increase in δ_0 directly raises the likelihood of creative destruction, which reduces each variety's profits or resources allocated to quality innovations, thus negatively affecting quality growth. On the other hand, an increase in δ_0 may decrease data sales and indirectly lowers the probability of creative destruction, which reduces profit losses from creative destruction, thereby positively affecting quality growth.

When firms own data, these two opposing effects cancel each other out, so δ_0 has no net effect on quality growth. However, when consumers own data and the frequency of creative destruction is relatively large, the negative effect dominates, leading to a lower equilibrium growth rate compared to when firms own data. Conversely, if δ_0 is less than its critical value, the positive effect dominates, and quality growth is higher than when firms own data. When the frequency of creative destruction equals the critical value, the two effects roughly cancell out and the two equilibrium growth rates are approximately equal.

(3) When the EIS for goods θ is relatively high, the quality growth associated with consumer-owned data exceeds that of firm-owned data. Conversely, when θ is relatively low, quality growth

with consumer-owned data is lower than with firm-owned data.

Intuitively, when firms own data, changes in the EIS for goods have two opposing effects on quality growth. A larger EIS reduces monopolistic prices, which leads to a decrease in equilibrium profits for each variety. This reduction in profits means fewer resources are allocated to quality innovations, negatively impacting quality growth. On the other hand, as product prices drop, each variety undervalues its original data and sells less of it. Reduced data sharing lowers the frequency of creative destruction, minimizing profit losses and positively influencing quality growth. These two opposing forces effectively cancel each other out, meaning that quality growth remains unaffected by the EIS for goods when firms own data.

When consumers own data, the same opposing effects apply. However, if the EIS for goods is relatively large, production prices are very low. Concerned about their privacy, consumers sell even less data, further reducing profit losses from creative destruction. In this case, the positive effect dominates, resulting in higher quality growth compared to when firms own data. On the other hand, if the EIS for goods is relatively small, consumers face higher prices for goods and sell more data. Here, the negative effect dominates, and quality growth is lower than firm-owned data.

Data property rights also impact welfare. In our model, the *Consumers Own Data (COD)* allocation is superior in relatively large parameter ranges, though by a relatively small margin, while the *Firms Own Data (FOD)* allocation yields higher welfare in relatively small parameter ranges, but by a relatively large margin. These results differ significantly from those of Jones and Tonetti (2020), who find that the COD allocation is generally superior, generating substantially higher welfare, while the FOD allocation only produces higher welfare in rare instances and by only a small amount.

Third, we propose an endogenous growth model that incorporates data and population growth, where quality and variety innovations interact. In this model, population growth has an expanding effect on the data market (a larger population implies larger market sizes and more data), which reinforces quality-enhancing innovations for all varieties and accelerates aggregate quality growth. By introducing big data into the framework of Dinopoulos and Thompson (1998), our model combines both quality and variety growth in an economy driven by data. Population growth directly influences variety growth and indirectly affects it through quliaty growth. Intuitively, higher population growth provides cheaper labor to the economy, directly facilitating variety innovations. Indirectly, a larger population creates greater market demand

and more data, promoting quality innovations across all varieties and generating additional resources (via creative destruction) to support further variety innovations.

The remainder of the paper is organized as follows. Section 2 presents the economic environment of the model. Section 3 examines the optimal allocation determined by the social planner. Section 4 analyzes the equilibrium allocation in a decentralized economy when firms own data. Section 5 explores the equilibrium allcoation when consumers own data. Section 6 develops the key insights of the model. Section 7 concludes. Section 8 contains the online appendix.

2 Economic environment

The economic environment discussed in this paper is summarized in Table 2. The model assumes a representative consumer with logarithmic utility over per capita consumption, $c(t)$. There are $m(t)$ varieties of consumption goods, which contribute to utility through a quality-adjusted Dixit-Stiglitz aggregator with a constant elasticity of substitution (EIS) θ (where $\theta > 1$), namely,

$$c(t) \equiv \left[\int_{i=0}^{m(t)} q(i,t)^{\frac{1}{\theta}} c(i,t)^{\frac{\theta-1}{\theta}} di \right]^{\frac{\theta}{\theta-1}}. \tag{1}$$

Here, $c(i,t)$ represents the instantaneous per capita consumption of good $i \in [0, m(t)]$, and $q(i,t)$ indicates the nominal quality of variety i, which follows independent homogeneous Poisson processes. There are $l(t)$ people in the economy, and the population grows exogenously at a rate of n. The time discount rate of the representative consumer is denoted by $\rho \in (0, \infty)$.

Table 2: The Economic Environment

$E_0\left[\int_{t=0}^{+\infty} e^{-(\rho-n)t} u\left(c\left(t\right), d\left(t\right)\right) dt\right]$	Utility
$u\left(c\left(t\right), d\left(t\right)\right) = \log c\left(t\right) - \frac{\kappa}{2} d\left(t\right)^2$	Flow utility
$c\left(t\right) \equiv \left[\int_{i=0}^{m(t)} q\left(i,t\right)^{\frac{1}{\theta}} c\left(i,t\right)^{\frac{\theta-1}{\theta}} di\right]^{\frac{\theta}{\theta-1}}$ with $\theta > 1$	Consumption per person
$J\left(i,t\right) = c\left(i,t\right) l\left(t\right) = y\left(i,t\right)$	Data creation
$d\left(i,t\right) = s\left(i,t\right) c\left(i,t\right)$	Variety i data shared
$d\left(t\right) = \int_{i=0}^{m(t)} d\left(i,t\right) di$	Per capita data shared
$y\left(i,t\right) = l_p\left(i,t\right) = c\left(i,t\right) l\left(t\right)$	Firm production
$dq\left(i,t\right) = \begin{cases} \gamma Q\left(t\right) d_v\left(i,t\right)^{\phi_2}, l_v\left(i,t\right)^b d_v\left(i,t\right)^{\phi_1} dt \\ 0, 1 - l_v\left(i,t\right)^b d_v\left(i,t\right)^{\phi_1} dt \end{cases}$	Quality-enhancing innovations
$\dot{m}\left(t\right) = l_m\left(t\right) l_h\left(t\right)^{-1}$	Variety-expanding innovations
$\mu = d_h\left(t\right)^{\psi} l_h\left(t\right)^{1-\psi}$	Production of effective labor
$D\left(t\right) \leq \chi\left[m\left(t\right)^{-1/\eta} \int_{i=0}^{m(t)} q\left(i,t\right)^{1/\eta} \left(l\left(t\right) d\left(i,t\right)\right)^{(\eta-1)/\eta} di\right]^{\eta/(\eta-1)}$	Data bundle
$\int_{i=0}^{m(t)} l_v\left(i,t\right) di + \int_{i=0}^{m(t)} c\left(i,t\right) l\left(t\right) di + l_m\left(t\right) = l\left(t\right)$	Labor resource allocation
$l\left(t\right) = l\left(0\right) e^{nt}$	Population growth
$\delta\left(s\left(i,t\right)\right) = 2^{-1} \delta_0 s\left(i,t\right)^2$	Creative destruction

Privacy considerations on personal data also influence flow utility. Data is a byproduct of economic activity, as discussed in the literature (Farboodi and Veldkamp, 2019; Jones and Tonetti, 2020; Cong, et al. 2021; Cong et al. 2022). Each unit of consumption generates one unit of data as a byproduct. Thus, $c\left(i,t\right)$ represents both the per capita consumption of variety i and the data produced from the consumption of variety i by each person. Let $s\left(i,t\right)$ denote the proportion of variety i data that is shared. Then $d\left(i,t\right) = s\left(i,t\right) c\left(i,t\right)$ represents the amount of variety i data shared by each agent. For simplicity, we assume that data from all varieties symmetrically affect utility. Thus, $d\left(t\right) = \int_{i=0}^{m(t)} d\left(i,t\right) di$ represents the total data shared for all varieties by each individual. Furthermore, we assume privacy costs follow a quadratic loss function. Overall, consumers gain utility from the consumption of each good but incur utility costs from data sharing, with $\kappa\left(>0\right)$ representing the weight placed on privacy relative to consumption. Additionally, let $J\left(i,t\right) = c\left(i,t\right) l\left(t\right) = y\left(i,t\right)$, where $J\left(i,t\right)$ represents the data created about variety i.

Firm i produces variety i using labor according to a linear production function: $y\left(i,t\right) = l_p\left(i,t\right)$, where $l_p\left(i,t\right)$ represents labor. Goods are produced using labor alone but may differ in two ways. First, there are quality differences, meaning identical quantities of different goods

can provide varying levels of utility. Second, labor productivity may vary regardless of quality, meaning identical quantities of different goods may require different amount of labor.

Firms invest in quality-enhancing R&D by hiring labor $l_v(i,t)$ and purchasing big data $d_v(i,t)$) to improve the quality of their products. Innovations in quality occur at random intervals, with a mean intensity that varies directly with R&D efforts. Specifically, the nominal quality levels $\{q(i,t)\}$ follow independent, homogeneous Poisson processes,[2] with an intensity of $l_v(i,t)^b d_v(i,t)^{\phi_1}$ and a magnitude of $\gamma Q(t) d_v(i,t)^{\phi_2}$, where $Q(t) \equiv m(t)^{-1} \int_{i=0}^{m(t)} q(i,t)\, di$ represents the mean quality level for all varieties at time t, and b, ϕ_1 and ϕ_2 are elasticity parameters for labor and data, satisfying $b, \phi_1, \phi_2 \in (0,1)$ and $b + \phi_1 + \phi_2 < 1$. Quality increments in this model are proportional to the current average quality level. In other words, the average quality level generates spillovers or positive externalities on individual quality-enhancing activities.[3] The proportionality feature is familiar from Romer (1990) and others. This type of externality is necessary to generate a stable distribution of relative quality in the balanced-growth equilibrium.

In contrast to the technology governing quality improvements, there are no spillovers in the creation of new varieties: the cost of developing a new variety remains constant, regardless of how many already exist. As argued by Romer (1990), developing a new variety is equivalent to incurring a fixed cost, which can be seen as a defining characteristic of the technology. We assume that a new variety is created upon the payment of a fixed effective labor cost, $\mu\,(>1)$, which aligns with the spirit of Romer (1990), Dinopoulos and Thompson (1998), and Jones and Tonetti (2020). However, unlike their models, the effective labor here is generated by combining original labor and big data through a production technology:

$$\mu = d_h(t)^{\psi} l_h(t)^{1-\psi}, \psi \in (0,1). \tag{2}$$

Furthermore, after a new variety is created, an initial relative quality is observed, drawn at random from the distribution of relative productivities for existing varieties, $F(\alpha)$.[4] By definition,

[2] The intensity of a Poisson process is the first-order approximation to the probability of a jump in the state variable within the next infinitesimally small interval of time. The mean time between consecutive jumps is the inverse of the intensity. The magnitude of a Poisson process refer to the size of the jump in the state variable, conditional on a jump occurring.

[3] If the magnitude of quality increments for variety i were proportional to its own quality $q(i,t)$, rather than $Q(t)$, firms that fall behind would have less incentives to engage in R&D, causing them to fall further behind. Such dynamics can lead to monopolistic outcomes.

[4] Thompson (1999) derived the characteristic function for F and showed that F is stationary and nondegenerate along the balanced growth path.

we know that $E(\alpha) = 1$.

The data used by all innovators is a bundle of data. How is this bundle created? Note that $l(t)\, d(i,t)$ represents the data about variety i shared by all raw data providers. This shared data is bundled together through a quality-adjusted CES production function, with an elasticity of substitution (EIS) denoted by η:

$$D(t) \leq \chi \left[m(t)^{-1/\eta} \int_{i=0}^{m(t)} q(i,t)^{1/\eta} \left(l(t)\, d(i,t) \right)^{(\eta-1)/\eta} di \right]^{\eta/(\eta-1)}, \eta > 1. \tag{3}$$

We assume the existence of a data intermediary sector with a data-aggregating-and-processing technology.[5] For simplicity, we setup the model so that data produced today is used for innovation today. We also assume that data fully depreciates at the end of each period. These two assumptions imply that data is not a state variable, which significantly simplifies the analysis.

We further assume that the integrated data $D(t)$ can be used repeatedly as a whole and cannot be divided into different parts, namely,

$$d_v(i,t) = d_h(t) = D(t). \tag{4}$$

This simplifying assumption has two merits. On one hand, it implies that data is nonrivalrous and can be sold repeatedly to different innovators. On the other hand, it ensures that the stationary solutions of the model economy are symmetric.

Labor is allocated among three uses: manufacturing, quality-enhancing innovations for existing varieties, and developing new varieties:

$$l(t) = \int_{i=0}^{m(t)} c(i,t)\, l(t)\, di + \int_{i=0}^{m(t)} l_v(i,t)\, di + \dot{m}(t) l_h(t). \tag{5}$$

The first term on the right-hand side of (5) represents the labor allocated to manufacturing. The second term, $l_v(t)\, m(t)$, denotes the labor devoted to quality-enhancing innovations. The third term, $\dot{m}(t) l_h(t)$, reflects the labor allocated to creating new varieties. For analytical convenience, we define $L_p \equiv l(t)^{-1} \int_{i=0}^{m(t)} c(i,t)\, l(t)\, di$, $L_v \equiv l(t)^{-1} \int_{i=0}^{m(t)} l_v(i,t)\, di$, and $L_m \equiv l(t)^{-1} \dot{m}(t) l_h(t)$ as the labor shares employed in these three sectors, respectively.

[5] Kolanovic and Krishnamachari (2017) identify three types of intermediaries in the financial data market, based on the services they provide: data intermediaries (who collect data from providers and channel it to investors), technology intermediaries (who offer technology solutions to clients), and consultants (who advise firms on integrating big data and managing related legal issues). We assume an aggregate data intermediary sector that encompasses all three roles.

Aside from the privacy costs to individuals, data sharing has another downside for the economy: it increases the rate of creative destruction. The more potential competitors know about an incumbent firm, the greater the likelihood that the incumbent will be displaced by entrants. Similar to Jones and Tonetti (2020), we assume that ownership of variety i changes according to a Poisson process with an arrival rate $\delta(s(i,t))$. Since changes in ownership are not part of the technology faced by the planner, the social planner does not consider this aspect.

3 The optimal allocation

In this section, we investigate the optimal allocation in our environment. Based on the economic environment discussed above, we summarize the stochastic optimal control problem that the social planner solves as follows:

$$\max_{\{c(i,t),s(i,t)l_v(i,t)\}} E_0\left[\int_{t=0}^{+\infty} e^{-(\rho-n)t}\left(\log c(t)-\frac{\kappa}{2}d(t)^2\right)dt\right]$$

subject to the quality-enhancing innovation activities:

$$dq(i,t)=\begin{cases}\gamma Q(t)D(t)^{\phi_2}, l_v(i,t)^b D(t)^{\phi_1}dt\\ 0, 1-l_v(i,t)^b D(t)^{\phi_1}dt\end{cases},$$

and the variety-creating innovation activities:

$$dm(t)=\left(l(t)-\int_{i=0}^{m(t)} l_v(i,t)\,di+\int_{i=0}^{m(t)} c(i,t)\,l(t)\,di\right)\mu^{-\frac{1}{1-\psi}}D(t)^{\frac{\psi}{1-\psi}}dt.$$

Similar to Dinopoulos and Thompson (1998), we demonstrate in online Appendix 8.1 that there is no aggregate uncertainty in the model. Consequently, the above stochastic control problem of the social planner can be reformulated as the following equivalent deterministic programming problem:

$$\max_{\{l_v,l_p,s,m,Q\}}\int_{t=0}^{\infty} e^{-(\rho-n)t}\left[\ln\left(l_p(t)\,l(t)^{-1}Q(t)^{\frac{1}{\theta-1}}m(t)^{\frac{1}{\theta-1}}\right)-\frac{\kappa}{2}l_p(t)^2\,l(t)^{-2}\,s(t)^2\right]dt,$$

subject to the technology constraints:

$$\begin{aligned}
\dot{Q}(t) &= \gamma Q(t) l_v(t)^b D(t)^{\phi_1+\phi_2}, \\
\dot{m}(t) &= l_m(t) \mu^{-1/(1-\psi)} D(t)^{\psi/(1-\psi)}, \\
D(t) &= \chi s(t) l_p(t) Q(t)^{\frac{1}{\eta}},
\end{aligned}$$

and the resource constraint:

$$l(t) = l_v(t) m(t) + l_m(t) + l_p(t).$$

Solving this problem yields us the following

Proposition 1 (The Optimal Allocation) *If the knife-edge condition holds, i.e.,*

$$\frac{\phi_1+\phi_2}{b} = \frac{\psi}{1-\psi}, \tag{6}$$

then the optimal growth rate of the average quality level, g_Q, solves the following algebraic equation:

$$\frac{(\rho-n) g_Q^{\frac{1}{b}-1} - \left(b + \frac{\phi_1+\phi_2}{\eta-1}\right) g_Q^{\frac{1}{b}}}{(\rho-n) b + \left(b + \frac{\phi_1+\phi_2}{\eta-1}\right)(z_1 n + z_2 g_Q)} = \gamma^{\frac{1}{b}} \mu^{\frac{1}{1-\psi}}. \tag{7}$$

The optimal growth rates for other variables can be expressed as functions of g_Q as follows:

$$\begin{aligned}
g_m &= z_1 n + z_2 g_Q, && (8)\\
g_D &= \frac{b}{\phi_1+\phi_2}\left[(z_1-1) n + z_2 g_Q\right], && (9)\\
g_c &= \frac{1}{\theta-1}\left[z_1 n + (z_2+1) g_Q\right]. && (10)
\end{aligned}$$

The optimal steady-state values of $s(t)$, $d(t)$, and $e(t)$ are:

$$\begin{aligned}
s &= \sqrt{\frac{(\phi_1+\phi_2)}{b}\frac{1-L_p}{L_p}\kappa^{-1} L_p^{-2}}, && (11)\\
d &= \sqrt{\frac{(\phi_1+\phi_2)}{b\kappa}\frac{1-L_p}{L_p}}, && (12)\\
e &= \frac{\theta}{\theta-1} L_p. && (13)
\end{aligned}$$

The optimal labor shares employed in the production sector, quality-enhancing innovation

activities, and variety-expanding innovation activities are:

$$L_p = \frac{x}{x+y+1}, L_v = \frac{y}{x+y+1}, L_m = \frac{1}{x+y+1},$$

where

$$x = (\theta - 1)\left[\left(\rho - n + \gamma^{-\frac{1}{b}}\mu^{-\frac{1}{1-\psi}} g_Q^{\frac{1}{b}}\right)(z_1 n + z_2 g_Q)^{-1} + 1\right],$$

$$y = \gamma^{-\frac{1}{b}}\mu^{-\frac{1}{1-\psi}} g_Q^{\frac{1}{b}}(z_1 n + z_2 g_Q)^{-1},$$

$$z_1 = 1 + \frac{\phi_1 + \phi_2}{b}, z_2 = \frac{\phi_1 + \phi_2}{b}\frac{1}{\eta - 1}.$$

Proof The proof is provided in online Appendix 8.1. □

We solve the balanced growth path (BGP) of the model using the knife-edge condition (6). This condition has an intuitive explanation: the relative importance of the two inputs (i.e., labor and data) in both quality and variety innovations is equal. It also implies that two data elasticities in both innovation activities (i.e., ϕ and ψ) change in the same direction. To see this, define $\phi \equiv \phi_1 + \phi_2$ and $f(\psi) \equiv \psi/(1-\psi)$. Equation (6) indicates that both ϕ and ψ change in the same direction for any given value of b.

Equation (7) shows that the optimal growth rate is a nonliear function of the population growth rate, while equation (8) establishes that quality growth and variety growth interact with each other in our model. These results differ significantly from existing literature and will be further discussed in the following text.

4 Firms own data

We now explore one possible way to use market mechanisms to allocate resources. In this equilibrium, we assume that firms own the data and decide how much to sell. Data is bought and sold via a data intermediary that bundles data from all varieties and resells it to all innovators. Throughout the paper, data sellers always set prices when they have market power, while data buyers are always price takers.

4.1 Decision problems

Household problem. Each household supplies one unit of labor inelastically, receiving a wage $w(t)$ in return. We assume that labor is the numeraire, meaning all prices are expressed in units

of labor. The household holds assets that yield a return $r(t)$, where these assets represents claims on the value of the monopolistically competitive firms. The representative household then solves

$$\max_{\{c(i,t),a(t)\}} E_0 \left[\int_{t=0}^{+\infty} e^{-(\rho-n)t} \left(\ln c(t) - \frac{\kappa}{2} d(t)^2 \right) dt \right], \tag{14}$$

subject to

$$c(t) \equiv \left[\int_{i=0}^{m(t)} q(i,t)^{\frac{1}{\theta}} c(i,t)^{\frac{\theta-1}{\theta}} di \right]^{\frac{\theta}{\theta-1}}, \tag{15}$$

$$da(t) = [(r(t) - n) a(t) + w(t) - e(t)] dt, \tag{16}$$

where $e(t) \equiv \int_{i=0}^{m(t)} p(i,t) c(i,t) di$ denotes per capita consumption expenditure, and $p(i,t)$ is the price of good i. Note that households do not choose how much data is sold, as firms own the data in this allocation.

The optimal demand for good i is given by:

$$c(i,t) = \frac{q(i,t) p(i,t)^{-\theta} e(t)}{\int_{i=0}^{m(t)} q(i,t) p(i,t)^{1-\theta} di}, \tag{17}$$

which shows that the optimal demand for good i, $c(i,t)$, increases with its quality, $q(i,t)$, and per capita consumption expenditure, $e(t)$, while decreasing with its price, $p(i,t)$. Consumers favor high-quality goods, and high-quality goods generate more data.

The evolution of per capita expenditure follows the Euler equation:

$$g_e(t) \equiv \frac{\dot{e}(t)}{e(t)} = r(t) - \rho. \tag{18}$$

Firm problem. Current output has no impact on future profits, so pricing and research decisions can be analyzed independently. At each point in time, each firm produces one type of final product to maximize its profits, given its relative quality level. There is one input-labor- and one unit of labor is required to produce one unit of output of any kind, regardless of quality. Since labor is the numeraire and wages are set to unity, firm i's profit-maximizing problem is:

$$\max_{\{p(i,t),c(i,t)\}} [p(i,t) - 1] c(i,t) l(t),$$

subject to the market demand for good i, $c(i,t) l(t)$, where $c(i,t)$ satisfies the demand function

(17).

The profit-maximizing monopolistic price is a constant markup over marginal wage costs (i.e., $w(t) = 1$), regardless of quality:

$$p(i,t) = \frac{\theta}{\theta - 1} = p(t) > 1, \tag{19}$$

which yields instantaneous profits: $\pi(i,t) = \alpha(i,t)\lambda(t)/\theta$, where $\lambda(t) \equiv e(t)l(t)/m(t)$ is the average level of expenditure per variety, $\alpha(i,t) \equiv q(i,t)/Q(t)$ is the quality of variety i relative to the mean quality level, and $Q(t) \equiv \int_{i=0}^{m(t)} q(i,t)\,di/m(t)$ is the mean quality level for all varieties at time t.

Quality-enhancing innovators. At each point in time, owners of existing varieties engage in quality-enhancing innovations, to maximize the expected present value of their ownership. Each incumbent firm decides how much data to buy and sell, and how much labor to hire. Firm i hires labor $l_v(i,t)$ and purchases integrated data $d_v(i,t)$ from the data intermediary at the price $p_D^v(t)$, which it takes as given. However, firm i sells original data $d(i,t)$ to the data intermediary at a price $p_d(i,t)$, that it sets through monopolistic competition. Firms balance the threat of creative destruction with the revenues from selling data, without considering potential infringements on consumers' privacy. As a result, firms' data sharing imposes negative externalities on consumers' utility.

The nominal quality levels of any variety i, $q(i,t)$, follow independent homogeneous Poisson processes with intensity $l_v(i,t)^b d_v(i,t)^{\phi_1}$ and magnitude $\gamma Q(t) d_v(i,t)^{\phi_2}$, where $b \in (0,1)$ measures the effect of labor on the probability of successful quality-enhancing innovations, and $\phi_1,\ \phi_2 \in (0,1)$ measure the effects of big data on the probability and magnitude of successful quality-enhancing innovations, respectively. We assume that quality-enhancing innovations exhibit decreasing returns to scale with respect to both inputs (i.e., labor and data), specifically, $b + \phi_1 + \phi_2 < 1$. Under these assumptions, the evolution of firm i's relative quality level, $\alpha(i,t)\,(\equiv q(i,t)/Q(t))$, follows a stationary shot-noise process[6] The process is described by the following stochastic differential equation:

$$d\alpha(i,t) = -\alpha(i,t)\, g_Q(t)\, dt + \frac{1}{Q(t)} dq(i,t), \tag{20}$$

[6] A shot-noise process is characterized by discrete increments to a variable occurring randomly at irregular intervals, with each increment decaying deterministically at an exponential rate. In this model, decay is driven by innovations in other product lines.

where $g_Q(t) \equiv \dot{Q}(t)/Q(t)$ is the growth rate of the average quality level.

To summarize, the owner of firm i solves the following problem:

$$J(\alpha(i,t),t) \equiv \max_{\{l_v(i),d_v(i),d(i)\}} E_t\left[\int_{u=t}^{+\infty} e^{-\int_{\tau=t}^{u} r(\tau)d\tau} \begin{pmatrix} \pi(i,u) - l_v(i,u) - p_D^v(u)\, d_v(i,u) \\ +p_d(i,u)\, d(i,u) - \delta(s(i,u))\, \pi(i,u) \end{pmatrix} du\right], \tag{21}$$

subject to the evolutionary process of its relative quality level $\alpha(i,t)$, i.e., equation (20), and the downward-sloping demand curve for its original data from the data intermediary, which will be described below:

$$p_d(i,t) = \left(p_D^v(t)\, m(t) + p_D^h(t)\, \dot{m}(t)\right) \chi\, [\Xi]^{\frac{1}{\eta-1}}\, m(t)^{-\frac{1}{\eta}}\, q(i,t)^{\frac{1}{\eta}}\, (d(i,t)\, l(t))^{-\frac{1}{\eta}}, \tag{22}$$

where

$$\Xi \equiv m(t)^{-\frac{1}{\eta}} \int_{i=0}^{m(t)} q(i,t)^{\frac{1}{\eta}}\, (l(t)\, d(i,t))^{\frac{\eta-1}{\eta}}\, di.$$

Each firm seeks to sell data on its variety to the data intermediary, but this desire is constrained by the threat of creative destruction. When more information about a firm's variety becomes available to other firms, the firm is vulnerable to innovation activities by its competitors. The term $\delta(s(i,t))\,\pi(i,t)$ in equation (21) reflects this downside of data sharing. The rate of creative destruction follows a Poisson process with an arrival rate $\delta(s(i,t)) = 0.5\delta_0 s(i,t)^2$, indicating that the more firm i sells its data, the greater the probability of successful innovations by other firms. When successful innovations occur, existing varieties experience ownership changes or profit losses due to their lower relative quality. Specifically, we assume that variety i loses a portion of its production profits, represented by $\delta(s(i,t))\,\pi(i,t)$. Firms may also wish to purchase bundles of data from other firms, weighing the cost of such purchases against the potential gains in relative quality and increased sales.

Following Merton (1971) and Dixit and Pindyck (1994), we write the Bellman equation as:

$$0 = \max_{\{l_v,d_v,p_d,d\}} \left\{ \begin{array}{c} e^{-\int_{u=0}^{t} r(u)du} \begin{pmatrix} (1-\delta(s(i,t)))\,\pi(i,t) - l_v(i,t) - \\ p_D^v(t)\, d_v(i,t) + p_d(i,t)\, d(i,t)\, l(t) \end{pmatrix} \\ -J_\alpha(\alpha(i,t),t)\,\alpha(i,t)\, g_Q(t) + J_t(\alpha(i,t),t) + \\ l_v(i,t)^b\, d_v(i,t)^{\phi_1} \left[J\left(\alpha(i,t) + \gamma d_v(i,t)^{\phi_2}, t\right) - J(\alpha(i,t),t)\right] \end{array} \right\}, \tag{23}$$

where $J(\alpha(i,t),t)$ is the expected discounted present value of firm i, and $J_t(\alpha,t)$ and $J_\alpha(\alpha,t)$

denote the partial derivatives of J with respect to time t and relative quality α, respectively. Following Dinopoulos and Thompson (1998), we focus on balanced growth solutions, from which we know that $J(\alpha(i),t) = e^{-rt}V(\alpha(i),\lambda(t),p_D^v(t))$ on the balanced growth path (BGP). We conjecture the following form for the value funtion:

$$V(\alpha(i),\lambda(t),p_D^v(t)) = y_1\alpha(i)\lambda(t) + y_2 p_D^v(t)^{x_1}\lambda(t)^{x_2}, \tag{24}$$

where x_1, x_2, y_1, y_2 are four undetermined coefficients. From this, we derive the following optimality conditions:

$$l_v(i,t) = \left[\gamma b y_1 \lambda(t) d_v(t)^{\phi_1+\phi_2}\right]^{\frac{1}{1-b}} = l_v(t), \tag{25}$$

$$d_v(i,t) = \left[\frac{\phi_1+\phi_2}{b}\left(b\gamma y_1\lambda(t)\right)^{\frac{1}{1-b}} p_D^v(t)^{-1}\right]^{\frac{1-b}{1-b-(\phi_1+\phi_2)}} = d_v(t), \tag{26}$$

$$p_d(i,t) = \left(p_D^v(t)m(t) + p_D^h(t)\dot{m}(t)\right)\chi Q(t)^{\frac{1}{\eta-1}} = p_d(t), \tag{27}$$

$$s(i,t) = \delta_0^{-1}\left(1-\frac{1}{\eta}\right)(\theta-1)p_d(t). \tag{28}$$

Equations (25)-(28) establish that all product lines use the same levels of labor and data for quality-enhancing innovations, sell the same amount of their original data, and set a uniform price level for their data.

Variety-expanding innovators. To produce any variety i, the variety-expanding sector purchases labor and data, solving the following cost-minimization problem:

$$\min_{\{l_h(t),d_h(t)\}} l_h(t) + d_h(t)p_D^h(t), \text{ s.t., } \mu = d_h(t)^{\psi} l_h(t)^{1-\psi}. \tag{29}$$

Combining the first-order conditions with respect to $l_h(t)$ and $d_h(t)$ leads to:

$$p_D^h(t) = \frac{\psi}{1-\psi}\frac{l_h(t)}{d_h(t)}, \tag{30}$$

which shows that the marginal productivities of the two inputs (labor and data) in generating effective labor equal their price ratios. Combining (30) and the production technology of effective labor (i.e., $\mu = d_h(t)^{\psi} l_h(t)^{1-\psi}$) yields optimal amounts of the two inputs:

$$l_h(t) = \mu\left(\frac{1-\psi}{\psi}\right)^{\psi} p_D^h(t)^{\psi}, d_h(t) = \mu\left(\frac{\psi}{1-\psi}\right)^{1-\psi} p_D^h(t)^{\psi-1}. \tag{31}$$

After a new variety is created, its initial relative quality is drawn at random from the distribution of relative productivities for existing varieties, $F(\alpha)$.[7] Thus, we know that $E(\alpha) = 1$. Additionally, new entrants benefit from business stealing: they capture profit flows from existing varieties affected by creative destruction. Since V is linear in α, a zero-profit condition for variety-expanding R&D implies:

$$l_h(t) + d_h(t) p_D^h(t) = V(1, \lambda(t), p_D^v(t)) + \left(\dot{m}(t)\right)^{-1} \int_{i=0}^{m(t)} \delta(s(i,t)) \pi(i,t) \, di, \tag{32}$$

where $V(1, \lambda(t), p_D^v(t))$ is defined in (24).

Combining (30) and (32) gives rise to

$$\begin{aligned} V(1, \lambda(t), p_D^v(t)) &= \frac{1}{1-\psi} \mu^{\frac{1}{1-\psi}} d_h(t)^{-\frac{\psi}{1-\psi}} - \left(\dot{m}(t)\right)^{-1} \int_{i=0}^{m(t)} \delta(s(i,t)) \pi(i,t) \, di \\ &= \frac{1}{1-\psi} l_h(t) - \left(\dot{m}(t)\right)^{-1} \int_{i=0}^{m(t)} \delta(s(i,t)) \pi(i,t) \, di \\ &= \frac{1}{\psi} d_h(t) p_D^h(t) - \left(\dot{m}(t)\right)^{-1} \int_{i=0}^{m(t)} \delta(s(i,t)) \pi(i,t) \, di. \end{aligned} \tag{33}$$

Data intermediary. The non-rivalrous nature of data makes a perfectly competitive data intermediary sector impossible. Following Jones and Tonetti (2020), we assume that the data intermediary operates as a monopolist, subject to free entry at a vanishingly small cost. Actual and potential data intermediaries take the price at which they buy data from firms, $p_d(i,t)$, as given. This setup results in a limit pricing condition under which the data intermediary earns zero profits, despite the non-rival nature of data.

The data intermediary accepts its purchase price of data, $p_d(i,t)$, as given and maximizes profits by choosing the quantity of data to purchase from each variety ($d(i,t) l(t)$), the quantity of the integated data to sell to the two kinds of innovators ($d_v(i,t), d_h(t)$), and the price at which it sells bundles of data to innovators ($p_D^v(t), p_D^h(t)$):

$$\max_{\{d(i), d_v(i), d_h(t), p_D(t)\}} \int_{i=0}^{m(t)} p_D^v(t) d_v(i,t) \, di + \dot{m}(t) d_h(t) p_D^h(t) - \int_{i=0}^{m(t)} p_d(i,t) d(i,t) l(t) \, di,$$

[7] Thompson (1999) derived the characteristic function for F, showing that F is stationary and non-degenerate along the balanced growth path.

subject to

$$D(t) \leq \chi \left[m(t)^{-1/\eta} \int_{i=0}^{m(t)} q(i,t)^{1/\eta} \left(l(t) d(i,t) \right)^{(\eta-1)/\eta} di \right]^{\eta/(\eta-1)}, \eta > 1, \tag{34}$$

$$d_v(i,t) = d_h(t) = D(t), \tag{35}$$

$$p_D^v(t) \leq p_D^*(t), p_D^h(t) \leq p_D^*(t), \tag{36}$$

where $p_D^*(t)$ is the limit price associated with the zero-profit condition arising from free entry. The first two terms in the profit expression incorporate the fact that the data intermediary can buy data once and sell it multiple times, reflecting the non-rivalrous nature of data. For example, location data from consumers can, technologically, be sold to each firm in the economy, not just to the store where consumers happen to be shopping at the moment. Equation (34) displays the data processing and intergrating technology, which also serves as the resource constraint on data. Furthermore, we assume that the intergrated data $D(t)$ can be sold as a whole and cannot be divided into different parts, as implied in equation (35).

Substituting (35) into the objective function, we can rewrite the data intermediary's problem as follows:

$$\max_{\{p_D(t), d(i,t)\}} \left(p_D^v(t) m(t) + p_D^h(t) \dot{m}(t) \right) D(t) - \int_{i=0}^{m(t)} p_d(i,t) d(i,t) l(t) di,$$

subject to (34) and (36). Solving the first-order condition with respect to $d(i,t)$ leads to (22):

$$p_d(i,t) = \left(p_D^v(t) m(t) + p_D^h(t) \dot{m}(t) \right) \chi \left[\Xi \right]^{\frac{1}{\eta-1}} m(t)^{-\frac{1}{\eta}} q(i,t)^{\frac{1}{\eta}} \left(d(i,t) l(t) \right)^{-\frac{1}{\eta}}.$$

Free entry results in zero profit in the data intermediary sector, namely:

$$\left(p_D^v(t) m(t) + p_D^h(t) \dot{m}(t) \right) D(t) = p_d(t) l(t) s(t) e(t) \frac{\theta - 1}{\theta}. \tag{37}$$

4.2 Equilibrium and the balanced growth path (BGP)

In equilibrium, labor is allocated among three uses: manufacturing, quality-enhancing innovations for existing varieties, and developing new varieties:

$$l(t) = \int_{i=0}^{m(t)} c(i,t) l(t) di + \int_{i=0}^{m(t)} l_v(i,t) di + \dot{m}(t) l_h(t).$$

The market-clearing condition for the asset market is

$$a(t)\,l(t) = \int_{i=0}^{m(t)} V\left(\alpha(i), \lambda(t), p_D^v(t)\right) di \left(= m(t)\, V\left(1, \lambda(t), p_D^v(t)\right)\right).^{8} \tag{38}$$

This condition establishes that the stock of total assets equals the present value of total profits earned by quality-enhancing innovators.

We then solve the balanced growth path (BGP) of the market economy, where some variables $(r(t), e(t), d(t), s(t), p_d(t))$ have constant values (r, e, d, s, p_d), and other variables $(Q(t), m(t), p_D(t), D(t),$ exhibit constant growth rates (g_Q, g_m, g_{p_D}, g_D, g_c, g_λ).

As argued in Dinopoulos and Thompson (1998), the change in $Q(t)$ over the interval $[t, t+dt]$ is given by the product of the intensity of the Poisson process, $l_v(t)^b D(t)^{\phi_1} dt$, and its magnitude, $\gamma Q(t) D(t)^{\phi_2}$. Thus, we have $dQ(t) = l_v(t)^b D(t)^{\phi_1} dt * \gamma D(t)^{\phi_2}$. Rearranging gives us

$$g_Q(t) \equiv \frac{\dot{Q}(t)}{Q(t)} = \gamma l_v(t)^b D(t)^{\phi_1+\phi_2}. \tag{39}$$

Then, we have the following

Proposition 2 (Firms own data) [9] *If the knife-edge condition holds, i.e., (6), the BGP growth rate of average quality level, g_Q, is determined by the following algebraic equation:*

$$\left(\frac{1}{bg_Q} + \frac{(1-b-(\phi_1+\phi_2))}{b\left[\rho + (z_1-1)n + z_2 g_Q\right]}\right)\gamma^{-\frac{1}{b}} g_Q^{\frac{1}{b}} = \frac{\mu^{\frac{1}{1-\psi}}}{1-\psi} - \frac{(\eta-1)\psi}{2\eta(1-\psi)}\left(\frac{\gamma^{-\frac{1}{b}} g_Q^{\frac{1}{b}}}{(z_1 n + z_2 g_Q)} + \mu^{\frac{1}{1-\psi}}\right). \tag{40}$$

The BGP growth rates of all other variables can be expressed in terms of g_Q as follows:

$$g_m = z_1 n + z_2 g_Q, \tag{41}$$

$$g_D = n + \frac{1}{\eta-1} g_Q, \tag{42}$$

$$g_c = \frac{1}{\theta-1}\left[z_1 n + (1+z_2) g_Q\right], \tag{43}$$

$$g_{p_D} = -z_1\left(n + \frac{1}{\eta-1} g_Q\right), \tag{44}$$

$$g_\lambda = -\frac{\phi_1+\phi_2}{b}\left(n + \frac{1}{\eta-1} g_Q\right), \tag{45}$$

[8] The second equality follows from the linearity of $V(\cdot)$ with respect to $\alpha(i)$.

[9] If there is no big data (i.e., $\phi_1 = \phi_2 = \psi = 0$), our model degenerates to Dinopoulos and Thompson (1998).

where

$$z_1 \equiv 1 + \frac{\phi_1 + \phi_2}{b}, z_2 \equiv \frac{\phi_1 + \phi_2}{b} \frac{1}{\eta - 1}.$$

The steady-state value of the equilibrium interest rate is:

$$r = \rho.$$

The steady-state value of e is determined by the following equation:

$$beg_Q \frac{1 + (\theta - 1)\left[1 - \frac{1}{2}\left(1 - \frac{1}{\eta}\right)\right] \frac{\phi_1 + \phi_2}{b} \frac{\left(1 - \frac{\theta - 1}{\theta} e\right)}{\frac{\theta - 1}{\theta} e}}{\theta \left[\rho + (z_1 - 1) n + (z_1 + 1) g_Q\right]} = \frac{1 - \frac{\theta - 1}{\theta} e}{1 + (z_1 n + z_2 g_Q) g_Q^{-1/b} \gamma^{1/b} \mu^{1/(1-\psi)}}. \tag{46}$$

The steady-state values of p_d, s, and d are:

$$s = \sqrt{\frac{\theta - 1}{\delta_0} \frac{\eta - 1}{\eta} \frac{\phi_1 + \phi_2}{b} \left(e^{-1} \frac{\theta}{\theta - 1} - 1\right)}, \tag{47}$$

$$p_d = \frac{\delta_0}{\theta - 1} \frac{\eta}{\eta - 1} s, \tag{48}$$

$$d = s \frac{\theta - 1}{\theta} e. \tag{49}$$

The steady-state labor shares employed in the production sector, the quality-enhancing innovation activities, and the variety-expanding innovation activities are:

$$L_p = \frac{\theta - 1}{\theta} e, L_v = \frac{1 - \frac{\theta - 1}{\theta} e}{1 + (z_1 n + z_2 g_Q) g_Q^{-1/b} \gamma^{1/b} \mu^{1/(1-\psi)}}, L_m = 1 - L_p - L_v. \tag{50}$$

Proof The proof is provided in online Appendix 8.2. □

In Section 6, we compare this equilibrium allocation with the optimal allocation and alternative equilibrium allocations. Before doing so, we define another equilibrium allocation in which consumers own data, enabling us to efficiently make all comparisons at once. Therefore, we now turn to the equilibrium where consumers own data.

5 Consumers own data

We now consider an allocation in which consumers own data associated with their purchases. They can sell data to a data intermediary and choose how much data to sell to balance the gain

in income versus the cost to privacy. Firms own zero data as it is created but can purchase data from the data intermediary.

5.1 Decision problems

Changing the data property rights alters the optimization problems of households and quality-enhancing innovators, while leaving unchanged the problems of producers, variety-expanding innovators, and data intermediaries. Therefore, we now present the details of these changes.

Household problem. The household problem is similar to the case where firms own the data, except that the household now decides how much data to sell. Consumers balance their privacy concerns against the economic benefits of selling data to intermediaries, regardless of the business-stealing effects this might have on firms. As a result, consumers' data sales impose negative externalities on firms' behavior. The representative household solves:

$$\max_{\{c(i),d(i),a\}} E_0 \left[\int_{t=0}^{+\infty} e^{-(\rho-n)t} \left(\ln c\left(t\right) - \frac{\kappa}{2} d\left(t\right)^2 \right) dt \right], \tag{51}$$

subject to

$$c\left(t\right) \equiv \left[\int_{i=0}^{m(t)} q\left(i,t\right)^{\frac{1}{\theta}} c\left(i,t\right)^{\frac{\theta-1}{\theta}} di \right]^{\frac{\theta}{\theta-1}}, \tag{52}$$

$$\begin{aligned} da\left(t\right) &= \left[\left(r\left(t\right)-n\right) a\left(t\right) + w\left(t\right) - e\left(t\right) + \int_{i=0}^{m(t)} p_d\left(i,t\right) d\left(i,t\right) \right] dt \\ &= \left[\left(r\left(t\right)-n\right) a\left(t\right) + w\left(t\right) - \int_{i=0}^{m(t)} \widetilde{p}\left(i,t\right) c\left(i,t\right) \right] dt, \end{aligned} \tag{53}$$

where $\widetilde{p}\left(i,t\right) \equiv p\left(i,t\right) - p_d\left(i,t\right) s\left(i,t\right)$ is the effective price of consumption, accounting for the fact that the fraction $s\left(i,t\right)$ of each good consumed generates income when the associated data is sold.

The optimal demand function and the Euler equation remain the same as in the scenario where firms own data. Additionally, we derive a first-order necessary condition with respect to $d\left(i,t\right)$:

$$p_d\left(i,t\right) = \kappa d\left(t\right) e\left(t\right) = p_d\left(t\right), \tag{54}$$

which implies that the equilibrium prices for all original data are identical.

Quality-enhancing innovators. While the production of each variety remains unchanged,

firms own no data to sell when engaging in quality-enhancing innovation activities. Firm i hires labor, $l_v(i,t)$, and purchases integrated data, $d_v(i,t)$, to solve:

$$J(\alpha(i,t),t) \equiv \max_{\{l_v(i),d_v(i)\}} E_t \left[\int_{u=t}^{+\infty} e^{-\int_{\tau=t}^{u} r(\tau)d\tau} \left(\begin{array}{c} \pi(i,u) - l_v(i,u) - \\ p_D^v(u)\, d_v(i,u) - \delta(s(i,u))\, \pi(i,u) \end{array} \right) du \right], \tag{55}$$

subject to the evolutionary equation of its relative quality level $\alpha(i,t)$, i.e., (20). The Bellman equation is modified accordingly:

$$0 = \max_{\{l_v,d_v\}} \left\{ \begin{array}{c} [1-\delta(s(i))]\,\pi(i,t) - l_v(i,t) - p_D(t)\, d_v(i,t) \\ -(\rho+g_e)\, V(\alpha(i),\lambda(t),p_D^v(t)) + V_\lambda(\alpha(i),\lambda(t),p_D^v(t))\, g_\lambda \lambda(t) \\ +V_{p_D^v} g_{p_D} p_D(t) - \alpha(i)\, g_Q V_\alpha + \\ l_v(i,t)^b\, d_v(i,t)^{\phi_1} \left[V\left(\alpha(i) + \gamma d_v(i,t)^{\phi_2}\right) - V(\alpha(i)) \right] \end{array} \right\}. \tag{56}$$

Conjecture that the value function takes the following form:

$$V(\alpha(i),\lambda(t),p_D^v(t)) = y_1 \alpha(i)\,\lambda(t) + y_2 p_D^v(t)^{x_1}\,\lambda(t)^{x_2}, \tag{57}$$

where x_1, x_2, y_1, and y_2 are undetermined coefficients. The first order conditions are:

$$l_v(i,t) = \left[\gamma b y_1 \lambda(t)\, d_v(t)^{\phi_1+\phi_2}\right]^{\frac{1}{1-b}}, \tag{58}$$

$$p_D^v(t)\, d_v(i,t) = \frac{\phi_1+\phi_2}{b} \left[b\gamma y_1 \lambda(t)\, d_v(t)^{\phi_1+\phi_2}\right]^{\frac{1}{1-b}} = \frac{\phi_1+\phi_2}{b} l_v(i,t). \tag{59}$$

To proceed, substituting (57)-(59) into the Hamilton-Jacobi-Bellman equation (56) to solve for the desired expressions:

$$x_1 = -\frac{\phi_1+\phi_2}{1-b-(\phi_1+\phi_2)}, x_2 = \frac{1}{1-b-(\phi_1+\phi_2)}, \tag{60}$$

$$y_1 = \frac{1-\delta(s(i))}{\theta(\rho-n+g_m+g_Q)}, y_2 = \frac{\frac{1-b-(\phi_1+\phi_2)}{b}(\gamma b y_1)^{x_2}\left(\frac{\phi_1+\phi_2}{b}\right)^{-x_1}}{\rho+g_e-x_2 g_\lambda - x_1 g_{p_D^v}}. \tag{61}$$

5.2 The equilibrium when consumers own data

Proposition 3 (Consumers own data) [10] *If the knife-edge condition holds, i.e., (6), the steady-state levels of the growth rate of average quality level g_Q and per capita consumption expenditure e are determined by the following two algebraic equations:*

$$\left[\frac{1}{bg_Q}+\frac{1-b-(\phi_1+\phi_2)}{b\left(\rho+(z_1-1)n+z_2g_Q\right)}\right]\gamma^{-\frac{1}{b}}g_Q^{\frac{1}{b}}=\frac{\mu^{\frac{1}{1-\psi}}}{1-\psi}-\frac{\delta_0\psi\left(\frac{\theta}{\theta-1}\right)^2}{2\left(1-\psi\right)\kappa e^2}\left(\mu^{\frac{1}{1-\psi}}+\frac{\gamma^{-\frac{1}{b}}g_Q^{\frac{1}{b}}}{z_1n+z_2g_Q}\right), \tag{62}$$

$$bg_Qe\frac{1-\frac{\delta_0}{2}\left(\frac{\theta e^{-1}}{\theta-1}\right)^2\left(e^{-1}-\frac{\theta-1}{\theta}\right)\frac{\psi\kappa^{-1}}{1-\psi}}{\theta\left[\rho+(z_1-1)n+(z_2+1)g_Q\right]}=\frac{\left(1-\frac{\theta-1}{\theta}e\right)}{1+\gamma^{\frac{1}{b}}g_Q^{-\frac{1}{b}}\mu^{\frac{1}{1-\psi}}\left(z_1n+z_2g_Q\right)}. \tag{63}$$

The BGP growth rates of all other variables can be expressed as functions of g_Q as follows:

$$g_m = z_1n+z_2g_Q, \tag{64}$$

$$g_D = n+\frac{1}{\eta-1}g_Q, \tag{65}$$

$$g_c = \frac{1}{\theta-1}\left[z_1n+(1+z_2)g_Q\right], \tag{66}$$

$$g_{p_D} = -z_1\left(n+\frac{1}{\eta-1}g_Q\right), \tag{67}$$

$$g_\lambda = -\frac{\phi_1+\phi_2}{b}\left(n+\frac{1}{\eta-1}g_Q\right), \tag{68}$$

where

$$z_1=1+\frac{\phi_1+\phi_2}{b}, z_2=\frac{\phi_1+\phi_2}{b}\frac{1}{\eta-1}.$$

The steady-state value of the equilibrium interest rate is

$$r=\rho.$$

The steady-state values of d, s, and p_d are expressed as functions of e as follows:

$$d=\sqrt{\frac{1}{\kappa}\frac{\psi}{1-\psi}\left(\frac{1}{e}-\frac{\theta-1}{\theta}\right)}, s=\frac{\theta}{\theta-1}\frac{d}{e}, p_d=\kappa de.$$

[10] Similar to the case with firms owning data, if there is no big data (i.e., $\phi_1=\phi_2=\psi=0$), the model also degenerates to the Dinopoulos and Thompson (1998) model.

The steady-state labor shares employed in the production sector, the quality-enhancing innovation activities, and the variety-expanding innovation activities are given by:

$$L_p = \frac{\theta - 1}{\theta} e, L_v = beg_Q y_1, L_m = 1 - L_p - L_v,$$

where

$$y_1 = \frac{1 - \frac{\delta_0}{2} s^2}{\theta \left[\rho + (z_1 - 1) n + (z_2 + 1) g_Q\right]}.$$

Proof The proof is provided in online Appendix 8.3. □

6 Key insights and model implications

6.1 Economic growth

In this subsection, we present growth performances under three different allocations and compare them with the existing literature. By combining the theoretical results from Proposition 1, Propostion 2, and Proposition 3, we arrive at the following theorem.

Theorem 1 *In our model with data as innovation, the long-run growth rates differ under three allocations: the optimal allocation, the equilibrium allocation when firms own data, and the equilibrium allocation when consumers own data. Under benchmark parameter values, the optimal growth rate exceeds both equilibrium growth rates. The comparison of growth rates under the two data property right regimes depends critically on the parameter pair governing privacy concerns and creative destruction (i.e., (κ, δ_0)) as well as the elasticity of substitution between goods (i.e., θ).*

6.1.1 Optimal growth and equilibrium growth

As shown in equations (7), (40), and (62), the average quality growth rates under the three allcoations differ. Specifically, the optimal growth rate is distinct from the equilibrium growth rates, whether data are owned by firms or by consumers.[11] Under benchmark parameter values, Figure 1 shows that the optimal growth rate (dotted dashed red line) is higher than both the equilibrium growth rate when firms own data (dashed green line) and when consumers own

[11] In online Appendix 8.5, we examine the growth effects of subsidies and taxes when firms own data. It is demonstrated that quality growth depends not on subsidies for data inputs in both innovation activities, but on subsidies for labor inputs. We also explore the optimal interventions that enable equilibrium growth to reach optimal growth.

data (solid green line). Intuitively, this is due to three types of distortions in decentralized allocations:

(1) The monopoly power over varieties results in welfare losses.

(2) Positive externalities from the average quality level benefit quality-enhancing innovations for all varieties.

(3) There are negative externalities associated with data sales: firms' data sales negatively impact consumers' privacy when firms own data, and consumers' data sales reduce firms' profits when consumers own data.

The social planner can mitigate welfare losses due to monopoly power and effectively internalize the externalities related to the average quality level and data sales, thus enhancing the equilibrium growth rates. As a result, the optimal growth rate exceeds the two equilibrium growth rates.[12]

Table 3: Benchmark parameter values[13]

Description	Parameter	Value
Time preference rate	ρ	0.025
Population growth rate	n	0.02
Elasticity of substitution (goods)	θ	4
Weight of privacy	κ	0.2
Elasticity of labor in quality-enhancing R&D	b	0.7
Elasticity of data in quality-enhancing R&D	ϕ	0.1
Size of innovations	γ	0.1
Frequency of creative destruction	δ_0	0.4
Elasticity of substitution (data)	η	50
Cost of developing a new variety	μ	10

[Insert Figure 1 here.]

6.1.2 Impact of data ownership on quality growth

We further examine the growth performances under different data property rights regimes. Propositions 2 and 3 establish that the average growth rate crucially depends on the parameter

[12] In online Appendix 8.2, we further analyze how subsidies and taxes influence equilibrium growth and investigate the optimal interventions in the decentralized equilibrium when firms own data.

[13] This should not be viewed as a formal calibration that can be compared quantitatively to facts about the US economy.

pair (κ, δ_0) and the elasticity of intertemporal substitution (EIS) parameter θ among goods when consumers own data, whereas quality growth does not depend on either parameter when firms own data. As a result, average quality growth rates differ between the two data property rights regimes.

To analyze how model parameters-especially $(\kappa\ , \delta_0)$ and θ-affect growth under the two data property rights regimes, we simulate the model and present the results in Figures 2-6. These figures demonstrate that the ranking of quality growth between the two regimes largely hinges on the parameter pair $(\kappa\ , \delta_0)$ and the parameter θ, while being relatively insensitive to other model parameters.

Impact of weight for privacy (κ) on quality growth. Given the benchmark parameter values in Table 3, we identify a pair of critical values for the parameter pair (κ, δ_0): $(\kappa^*, \delta_0^*) = (0.1636, 0.4874)$. Given the benchmark value for $\delta_0 = 0.4$, the weight for privacy κ significantly (κ) influences quality growth, depending on its relationship to the critical value (κ^*). If the weight for privacy κ is relatively large (i.e., $\kappa = 0.2 > \kappa^*$), the quality growth associated with data owned by consumers exceeds that of data owned by firms $(g_Q^c > g_Q^f)$ for any given values of ϕ, b, γ, μ, ρ, and n. This relationship is illustrated in Figure 2. Conversely, if κ is relatively small (i.e., $\kappa = 0.1 < \kappa^*$), then the quality growth with consumer-owned data is lower than that with firm-owned data $(g_Q^c < g_Q^f)$, as depicted in Figure 3. When the weight for privacy κ equals its critical value (i.e., $\kappa = \kappa^*$), the quality growth under both data property rights regimes becomes nearly equal $(g_Q^c \approx g_Q^f)$, as shown in Figure 4. Overall, these observations underscore the critical role that the parameter (κ, δ_0) play in determining growth performances, particularly how ownership of data influences quality growth.

Intuitively, when firms own data, privacy concerns do not significantly influence the optimizing behaviors of all agents (i.e., consumers, firms and data intermediaries), resulting in negligible effects on the economy. However, when consumers own their data, an increase in privacy concerns produces two opposing effects on long-term quality growth. Increased privacy concerns lead consumers to sell less data. Reduced data sales and sharing diminish the frequency of creative destruction, thereby lowering profit losses across all varieties. Consequently, for each variety, more resources can be allocated to quality-enhancing activities, positively influencing quality growth. Conversely, reduced data sales also decrease consumers' revenues from data sales. This decline diminishes per capita consumption expenditure, leading to a reduced market demand for each variety. As a result, each variety experiences lower profits, resulting in fewer

resources being devoted to innovation activities, which negatively impacts quality growth.

When consumers place a higher value on privacy than the critical threshold, the positive effect on quality growth prevails, leading to an equilibirum growth rate greater than that of firm-owned data ($g_Q^c > g_Q^f$). Conversely, if consumers prioritize privacy less than this critical value, the negative effect dominates, resulting in a lower growth rate compared to firm ownership ($g_Q^c < g_Q^f$). Finally, as privacy concerns approach the critical value, the two opposing effects nearly cancell each other out, resulting in equilibrium growth rates that are close to one another ($g_Q^c \approx g_Q^f$).

[Insert Figure 2, Figure 3, and Figure 4 here.]

Impact of frequency of creative destruction (δ_0) on quality growth. Given the benchmark value for $\kappa = 0.2$, the frequency of creative destruction (δ_0) significantly influences quality growth, depending on its relationship to the critical value (δ_0^*). When δ_0 is relatively large (i.e., $\delta_0 = 0.6 > \delta_0^*$), quality growth with data owned by consumers is lower than that with firms owning data ($g_Q^c < g_Q^f$) for any given values of ϕ, b, γ, μ, ρ, and n. This relationship is illustrated in Figure 5. Conversley, if δ_0 is relatively small (i.e., $\delta_0 = 0.4 < \delta_0^*$), quality growth with consumer-owned data exceeds that of firm-owned data ($g_Q^c > g_Q^f$), as shown in Figure 2. If δ_0 equals its critical value (i.e., $\delta_0 = \delta_0^*$), the quality growth under both data property rights regimes is nearly equal ($g_Q^c \approx g_Q^f$), as depicted in Figure 6. Overall, these observations highlight the critical role of the frequency of creative destruction in determining the dynamics of quliaty growth based on data ownership.

Intuitively, when firms own data, an increase in the frequency of creative destruction (δ_0) has two opposing effects on quality growth: direct (negative) effect and indirect (positive) effect. Given constant data sales (s remains unchanged), an increase in δ_0 directly raises the likelihood of creative destruction ($\delta(s(i)) \uparrow$), which reduces profits for each variety and decreases the resources allocated to quality innovations. This negatively impacts quality growth. Conversely, in response to a higher frequency of creative destruction, each variety reduces its data sales ($s(i) \downarrow$), which in turn decreases the occurance of creative destruction ($\delta(s(i)) \downarrow$). This reduction in profit losses from creative destruction positively influences quality growth.

When firms own data, these two opposing effects exactly cancel each other out, leaving δ_0 with no net impact on quality growth, as shown in Proposition 2. When consumers own data, the same two opposing effects are present. However, the balance between these effects changes depending on the magnitude of δ_0. If δ_0 exceeds its critical value, the direct, negative

effect dominates, resulting in a lower equilibrium growth rate compared to firm-owned data ($g_Q^c < g_Q^f$). If δ_0 is below its critical value, the indirect, positive effect prevails, leading to a higher growth rate than with firm-owned data ($g_Q^c > g_Q^f$). When $\delta_0 = \delta_0^*$, the two effects roughly cancel each other out, making the equilibrium growth rates under both data ownership regimes approximately equal (($g_Q^c \approx g_Q^f$)).

[Insert Figure 5 and Figure 6 here.]

To further examine the impact of data property rights on quality growth, we plot Figure 7, which illustrates the differences in average quality growth rates under two data ownership regimes: consumer-owned data (g_Q^c) and firm-owned data (g_Q^f). The difference is measured as $g_Q^c - g_Q^f$, and its dependence on the parameter pair (κ, δ_0) is shown. When this difference is greater than zero, quality growth with consumer-owned data exceeds that with firm-owned data ($g_Q^c > g_Q^f$). In Figure 7, this occurs over relatively large parameter ranges, indicating that consumer data ownership is more likely to result in higher quality growth in such cases. Conversely, when the difference is less than zero, quality growth with firm-owned data surpasses that with consumer-owned data , which occurs in relatively small parameter ranges. This suggests that firm data ownership is less likely lead to higher growth, but when it does, the effect is more pronouced. Moreover, for different values of ϕ (i.e., 0.05, 0.1, and 0.2), the maximum differences observed are 0.0032, 0.0095, and 0.0489, respectively, while the minimum differences are -0.0263, -0.0357, and -0.1014, respectively. These results indicate that: Consumer-owned data tends to generate higher growth in larger parameter ranges, though the differences are relatively small; Firm-owned data generates higher growth in smaller parameter ranges, buth the differences are relatively large.

[Insert Figure 7 here.]

Notes: *Growth and data property right.* The plots depict the difference in average quality growth rates, $g_Q^c - g_Q^f$, across various combinations of κ and δ_0. The Consumers Own Data allocation produces higher growth in relatively large parameter ranges, where the difference $g_Q^c - g_Q^f$ is positive. In constrast, the Firms Own Data allocation leads to higher growth in relatively small parameter ranges, where the difference $g_Q^c - g_Q^f$ is negative. Additionally, the plots provide the largest, smallest, and average values of $g_Q^c - g_Q^f$ in each case.

ϕ	Maximum	Minimum	Mean
0.05	0.0032	-0.0263	4.35E-04
0.10	0.0095	-0.0357	0.0023
0.20	0.0489	-0.1014	0.0150

Impact of elasticity of intertemporal substitution (θ) on quality growth. Equation (62) demonstrates that quality growth depends on the EIS for goods when consumers own data, whereas equation (40) shows that quality growth is independent of the EIS when firms own data. This contrast in dependencies is visually supported by Figures 2-6. When the EIS for goods (θ) is relatively high, the quality growth associated with consumer-owned data exceeds that of firm-owned data ($g_Q^c > g_Q^f$). This suggests that consumer data ownership leads to stronger growth under conditions of greater substitutability between goods. Conversely, if θ is relatively low, the quality growth with consumer-owned data falls below that with firm-owned data ($g_Q^c < g_Q^f$), indicating that firm ownership of data results in better growth outcomes when the EIS is lower.

Intuitively, when firms own data, changes in the EIS for goods have two opposing effects on quality growth. A larger EIS reduces the monopolistic price ($p \downarrow$), leading to a decrease in the equilibrium profit for each variety ($\pi \downarrow$). This reduction in profits leads to fewer resources being allocated toward quality innovations, negatively impacting quality growth ($g_Q \downarrow$). On the other hand, as prices for each product drop, each variety undervalues its original data ($p_d \downarrow$) and consequently sells less of it ($s \downarrow$). Reduced data sharing lowers the frequency of creative destruction ($\delta(s) \downarrow$), which minimizes profit losses and positively influences quality growth ($g_Q \uparrow$). These two opposing forces cancel each other out, resulting in quality growth that does not depend on the EIS for goods when data is owned by firms, as demonstrated in equation (40).

When consumers own data, the same two opposing effects are present. However, if the EIS for goods is relatively large, production prices are very low. Concerned about their privacy, consumers sell even less data, further reducing profit losses from creative destruction. In this scenario, the positive effect dominates, and quality growth is higher than when firms own data. Conversely, if the EIS for goods is relatively small, consumers face higher prices for goods and, as a result, sell more data than firms. In this case, the negative effect dominates, and quality growth is lower than when firms own data.

6.1.3 Comparing with the literature

We now compare our findings with the existing literature. In Jones and Tonetti (2020)'s model, the optimal growth rate is equal to the equilibrium growth rate under both data property rights regimes (consumer-owned and firm-owned). They argue that, compared to optimal growth, the inefficiencies of equilibrium growth only affect labor allocation and data usage, without impacting long-run growth. Moreover, they posit that data property rights are neutral for long-run growth. That is, while data property rights influence labor allocation, data usage, and welfare in the long run, they do not affect the long-run growth rate itself. Cong et al. (2021) show that the optimal growth rate equals the equilibrium growth rate when data is owned by consumers, while the equilibrium growth rate is higher when data is owned by firms. This suggests that firm-owned data yields greater growth in equilibrium than consumer-owned data or the optimal growth rate. In a subsequenct study, Cong et al. (2022) demonstrate that the equilibrium growth rate with consumers-owned data is lower than the optimal growth rate, implying an inefficiency in this regime.

In contrast to Jones and Tonetti (2020), we find that inefficiencies in equilibrium growth manifest not only in labor reallocation and data usage but also in long-run growth. Additionally, unlike the neutral stance of Jones and Tonetti, our results indicate that data property rights do matter for long-run growth. The relative ranking of growth outcomes under different data property regimes depends on key model parameters such as the weight on privacy (κ), the frequency of creative destruction (δ_0), and the EIS for goods (θ). Furthermore, our model shows that the social planner reduces welfare losses from monopoly pricing and effectively internalizes the externalities arising from aggregate quality and data sharing. As a result, the optimal growth rate in our model is higher than the equilibrium growth rate under both data property right regimes. We summarize the key comparisons among the literature in the following Table 4.

Table 4: Comparisons with the literature

Farboodi & Veldkamp (2019)	$\Delta\Omega^f = 0$, i.e., no long-run growth
Jones & Tonetti (2020)	$g^{sp} \begin{cases} = g^c \\ = g^f \end{cases}$ $g^c = g^f$
Cong et al (2021)	$g^{sp} \begin{cases} = g^c \\ < g^f \end{cases}$ $g^c < g^f$
Cong et al. (2022)	$g^{sp} > g^c$
Our model	$g^{sp} \begin{cases} > g^c \\ > g^f \end{cases}$ $g^c \gtreqless g^f$ hinges on (δ_0, κ) and θ

6.2 Welfare and data property rights

In this section, we compare the steady-state welfare under the consumers-own-data and firms-own-data property right regimes. Consumers value both consumption and privacy, and the steady-state welfare accounts for both factors. Along a balanced growth path, steady-state welfare is given by

$$W_{ss}^{alloc} = \frac{1}{\widetilde{\rho}} \left(\log c\left(0\right)^{alloc} - \frac{\kappa}{2} d_{alloc}^2 + \frac{1}{\widetilde{\rho}} g_c^{alloc} \right), \tag{69}$$

where $\widetilde{\rho} \equiv \rho - n$, and $g_c^{alloc} = (\theta - 1)^{-1} \left(g_Q^{alloc} + g_m^{alloc} \right)$. The difference in steady-state welfare between the two data property rights regimes is given by:

$$W_{ss}^c - W_{ss}^f = \underbrace{\widetilde{\rho}^{-1} \left(\log c\left(0\right)^c - \log c\left(0\right)^f \right)}_{\text{Level term}} - \underbrace{\widetilde{\rho}^{-1} \frac{\kappa}{2} \left(d_c^2 - d_f^2 \right)}_{\text{Privacy term}} + \underbrace{\widetilde{\rho}^{-2} \left(g_c^c - g_c^f \right)}_{\text{Growth term}}. \tag{70}$$

This expression shows that the welfare discrepancies between the two data property rights regimes can be decomposed into three terms, reflecting differences in the level of consumption, the degree of privacy, and the consumption growth rate.

Figure 8 shows the discrepancies in steady-state welfares, $W_{ss}^c - W_{ss}^f$, for various combinations of ϕ (where $\phi = \phi_1 + \phi_2$), κ and δ_0. When this difference is less than zero, the steady-state welfare with data owned by firms is higher than that with data owned by consumers. In relatively large parameter value ranges, this difference is greater than zero, indicating that the Consumers Own Data (COD) allocation is more likely to be superior. Conversely, the Firms

Own Data (FOD) allocation generates higher welfare in relatively small parameter value ranges, where this difference is less than zero. Furthermore, when ϕ =0.05, 0.10, and 0.20, the largest differences are 0.4290, 0.7867, and 3.2701, while the smallest differences are -1.5576, -2.1559, and -6.6942. Taken together, in the relatively large ranges where the COD allocation generates higher welfare, it does so by a relatively small amount; whereas in the relatively small ranges where the FOD allocation generates higher welfare, it does so by a relatively large amount.

[Insert Figure 8 here.]

Note: The plots display the differences in steady-state welfares, $W_{ss}^c - W_{ss}^f$, across various combinations of κ and δ_0. The Consumers Own Data allocation is superior in relatively large parameter ranges, whereas the Firms Own Data allocation is superior in relatively small parameter ranges. The largest, smallest and average values of $W_{ss}^c - W_{ss}^f$ in each plot are as follows:

ϕ	Maximum	Minimum	Mean
0.05	0.4290	-1.5576	0.0418
0.10	0.7867	-2.1559	0.1593
0.20	3.2701	-6.6942	0.9773

These results differ significantly from those in Jones and Tonetti (2020). Jones and Tonetti (2020) show that the Consumers Own Data allocation is generally superior and typically generates substantially higher welfare, while the Firms Own Data allocation only generates higher welfare in relatively rare instances, and by a small amount. The key reason for this discrepancy between our model and Jones and Tonetti (2020) relates to the growth term in the welfare decomposition (i.e., equation (70)). In Jones and Tonetti (2020), the growth term disappears, and the welfare comparisons depend on the level and privacy terms, since the BGP growth rates under the three allocations are equal. However, in our model, the growth term persists and dominates the other two terms, as shown in Figures 7 and 8. We summarize the comparisons with Jones and Tonetti (2020) in Table 5.

Table 5: Comparison with Jones & Tonetti (2020)

Literature	Welfare and property rights
Jones&Tonetti (2020)	$U_{ss}^c > U_{ss}^f$ for the majority of the plot, substantially higher
	$U_{ss}^c < U_{ss}^f$ for relatively rare instances, only small amounts
Our model	$W_{ss}^c > W_{ss}^f$ in relatively large ranges, relatively small amounts
	$W_{ss}^c < W_{ss}^f$ in relatively small ranges, relatively large amounts

6.3 Endogenous growth with interactions of quality and variety R&Ds

6.3.1 Endogenous growth with population growth

In the endogenous growth literature, a major issue is whether long-run growth driven by R&D is endogenous or semi-endogenous. Semi-endogenous growth means that (i) technological change is endogenous, and (ii) long-run growth is determined by exogenous population growth. Based on this definition, Jones and Tonetti (2020) and Cong et al. (2021) present semi-endogenous growth models with data. In these models, the long-run growth rate is proportional to population growth. This implies that if the population growth rate is zero, the long run growth rate will also be zero. In constrast, Cong et al. (2022) propose an endogenous growth model, which incorporates data and assumes no population growth. Building on Dinopoulos and Thompson (1998), we develop an endogenous growth model where population growth positively and nonlinearly affects long-run economic growth by incorporating the role of big data.

Proposition 4 (Endogenous growth) *If the population growth rate is zero, then the average quality growth rate is positive. If the population growth rate is positive, then the average quality growth rate is a nonlinear function of the population growth rate.*

In Dinopoulos and Thompson (1998), long-run growth is independent of population growth. In our extended model with big data, quality growth (i.e., g_Q) is an increasing and nonlinear function of the population growth rate (i.e., n), as shown in Figure 9.[14] Unlike the existing literature (e.g., Jones, 1995; Kortum, 1997; Segerstrom, 1998; Jones and Tonetti, 2020; Cong et al., 2021), quality growth in our model is not proportional to population growth. Intuitively, population growth in our model induces an expansion in the data market: a larger population implies greater market demand/size and thus more data ($c(i,t)l(t)$ or $D(t)$), which drives quality-enhancing innovations for all varieties and accelerates aggregate quality growth. For

[14] When firms own data, we can prove that the average quality growth rate increases with the population growth rate.

ease of comparison, Table 6 summarizes the growth models with data and their nature of growth.

[Insert Figure 9 here.]

Table 6: Semi-endogenous or endogenous growth models with data

Model	Growth rate	Nature of growth
Jones & Tonetti(2020)	$g^{sp}=g^{c}=g^{f}=\left(\frac{1}{\sigma-1}+\frac{\eta}{1-\eta}\right)g_{L}$	Semi-endogenous
Cong et al.(2021)	$g^{sp}=g^{c}=\left(\frac{\sigma}{(1-\zeta)\sigma-\xi(1-\gamma)}\right)n, g^{f}=\left(\frac{\xi+\phi}{\phi(1-\zeta)-\xi}\right)n$	Semi-endogenous
Cong et al.(2022)	$g^{sp}=(\gamma-1)\rho\xi^{-1}\left(\kappa d_{sp}^{2}-\eta\right), g^{c}=\frac{\rho\left[\kappa d_{c}^{2}-\left(1-\gamma^{-1}\right)\eta\right]}{\xi\Gamma-\left[\kappa d_{c}^{2}-(1-\gamma^{-1})\eta\right]}$	Endogenous
Our model	g^{sp}, g^{f}, g^{c} are increasing and nonlinear in n	Endogenous

6.3.2 Interactions between quality and variety growth

In the Dinopoulos and Thompson (1998) model without data, quality growth is independent of population growth, while variety growth equals population growth. This implies that there is no connection between quality growth and variety growth. By introducing big data into both quality and variety innovations in the Dinopoulos and Thompson (1998) model, our approach integrates quality growth with variety growth in a data-driven growth economya. Equations (8), (41), and (64) illustrate this relationship:

$$g_{m}=\underbrace{z_{1}n}_{\text{direct effect}}+\underbrace{z_{2}g_{Q}}_{\text{indirect effect}}, \tag{71}$$

where $z_{1}=1+\left(\phi_{1}+\phi_{2}\right)/b$ and $z_{2}=\left(\phi_{1}+\phi_{2}\right)/b\left(\eta-1\right)$. Equation (71) establishes that variety growth positively depends on both population growth and quality growth. Using Propostion 4 and equation (71), we can decompose the effect of population growth on variety growth into two components: a direct, linear effect via population growth itself, and an indirect, nonlinear effect through quality growth. Intuitively, higher population growth provides cheaper labor, which directly facilitates variety innovations. Indirectly, a larger population implies greater market demand/size and more data, promoting quality innovations for existing varieties, thereby supplying more resources for variety innovations (via creative destruction). Figure 10 illustrates the interaction between quality growth and variety growth: both increase with population growth. However, the direct, linear effect of population growth dominates the indirect, nonlinear effect through quality growth. If the population growth rate is zero, the direct effect disappears, but the indirect effect remains. Additionally, if we assume lower elasticity of substitution for

original data, the indirect effect becomes more pronounced.

Most researchers in endogenous growth theory assume no, or at most very limited, knowledge spillovers between quality and variety R&D (e.g., Aghion and Howitt, 1998; Dinopulos and Thompson, 1998; Howitt, 1999; Peretto, 1998; Peretto and Smulders, 1998). Li (2000) combines quality growth with variety growth by introducing knowledge spillovers in both R&D sectors. Unlike Li (2000), we introduce nonrival data into both innovation activities, connecting quality and variety innovations in a different way.

[Insert Figure 10 here.]

7 Conclusions

In this paper, we develop an endogenous growth model that treats data as innovations, where big data promotes quality-enhancing innovations by increasing the possibility and magnitude of successful quality innovations, and variety-creating innovations through effective labor forces as a fixed cost. The social planner mitigates the welfare cost of monopolistic pricing and internalizes externalities present in the decentralized economy, resulting in an optimal growth rate that is larger than the equilibrium growth rates under both data property right regimes. Data property rights significantly influence equilibrium growth and steady-state welfare. The comparisons for growth and welfare under the two property rights largely hinge on several key model parameters: the weight for privacy, the frequency for creative destruction, and the elastiticy of substitution among goods. Introducing non-rival data in both variety and quality innovations, our model combines variety growth and quality growth in the long run.

Online appendix: Proofs

7.1 Appendix 8.1 Proof of Proposition 1

Proof of Proposition 1. The stochastic optimal control problem faced by the social planner is summarized as follows:

$$\max E_0 \left[\int_{t=0}^{+\infty} e^{-(\rho-n)t} \left(\ln c\left(t\right) - \frac{\kappa}{2} d\left(t\right)^2 \right) dt \right],$$

subject to

$$c(t) \equiv \left[\int_{i=0}^{m(t)} q(i,t)^{\frac{1}{\theta}} c(i,t)^{\frac{\theta-1}{\theta}} di\right]^{\frac{\theta}{\theta-1}}, d(t) \equiv \int_{i=0}^{m(t)} d(i,t) di = \int_{i=0}^{m(t)} s(i,t) c(i,t) di,$$

$$dq(i,t) = \begin{cases} \gamma Q(t) D(t)^{\phi_2}, l_v(i,t)^b D(t)^{\phi_1} dt \\ 0, 1 - l_v(i,t)^b D(t)^{\phi_1} dt \end{cases}, \dot{m}(t) = \frac{l_m(t)}{l_h(t)}, \mu = l_h(t)^{1-\psi} D(t)^{\psi},$$

$$D(t) = \chi\left[m(t)^{-1/\eta} \int_{i=0}^{m(t)} q(i,t)^{1/\eta} (s(i,t) c(i,t) l(t))^{(\eta-1)/\eta} di\right]^{\eta/(\eta-1)}, \eta > 1, \quad (72)$$

$$l(t) = \int_{i=0}^{m(t)} c(i,t) l(t) di + \int_{i=0}^{m(t)} l_v(i,t) di + l_m(t).$$

Define the value function $V(m(t), \{q(i,t), i \in [0, m(t)]\})$. The Bellman equation is written as

$$\begin{aligned} & (\rho - n) V \\ = & \max_{\{c,s,l_v\}} \left\{ \begin{array}{c} \ln c(t) - \frac{\kappa}{2} d(t)^2 + V_m \left[l(t) - \int_{i=0}^{m(t)} c(i,t) l(t) di + \int_{i=0}^{m(t)} l_v(i,t) di\right] \mu^{-\frac{1}{1-\psi}} D(t)^{\frac{\psi}{1-\psi}} + \\ \int_{i=0}^{m(t)} l_v(i,t)^b D(t)^{\phi_1} \left[V\left(m, q(i,t) + \gamma Q(t) D(t)^{\phi_2}, q(k,t)_{k \neq i}\right) - V\left(m, q(i,t), q(k,t)_{k \neq i}\right)\right] di \end{array} \right. \end{aligned}$$

The first-order necessary conditions for $c(i,t), s(i,t), l_v(i,t)$ are

$$\begin{aligned} 0 = & \left[\int_{i=0}^{m(t)} q(i,t)^{\frac{1}{\theta}} c(i,t)^{\frac{\theta-1}{\theta}} di\right]^{-1} q(i,t)^{\frac{1}{\theta}} c(i,t)^{-\frac{1}{\theta}} - \kappa d(t) s(i,t) - V_m \mu^{-\frac{1}{1-\psi}} D(t)^{\frac{\psi}{1-\psi}} + \quad (73) \\ & \left[\begin{array}{c} V_m \left(l(t) - \int_{i=0}^{m(t)} c(i,t) l(t) di - \int_{i=0}^{m(t)} l_v(i,t) di\right) \mu^{-\frac{1}{1-\psi}} \frac{\psi}{1-\psi} D(t)^{\frac{\psi}{1-\psi}-1} \\ + \phi_1 D(t)^{\phi_1 - 1} \int_{j=0}^{m(t)} l_v(j,t)^b \left[V\left(q(j,t) + \gamma Q(t) D(t)^{\phi_2}\right) - V(q(j,t))\right] dj + \\ \gamma Q(t) \phi_2 D(t)^{\phi_1 + \phi_2 - 1} \int_{j=0}^{m(t)} l_v(j,t)^b V_{q(j,t) + \gamma Q(t) D(t)^{\phi_2}}\left(q(j,t) + \gamma Q(t) D(t)^{\phi_2}\right) dj \end{array}\right] \frac{\partial D(t)}{\partial c(i,t)}, \end{aligned}$$

$$\begin{aligned} 0 = & -\kappa d(t) c(i,t) + \quad (74) \\ & \left[\begin{array}{c} V_m \left(l(t) - \int_{i=0}^{m(t)} c(i,t) l(t) di - \int_{i=0}^{m(t)} l_v(i,t) di\right) \mu^{-\frac{1}{1-\psi}} \frac{\psi}{1-\psi} D(t)^{\frac{\psi}{1-\psi}-1} \\ + \phi_1 D(t)^{\phi_1 - 1} \int_{j=0}^{m(t)} l_v(j,t)^b \left[V\left(q(j,t) + \gamma Q(t) D(t)^{\phi_2}\right) - V(q(j,t))\right] dj + \\ \gamma Q(t) \phi_2 D(t)^{\phi_1 + \phi_2 - 1} \int_{j=0}^{m(t)} l_v(j,t)^b V_{q(j,t) + \gamma Q(t) D(t)^{\phi_2}}\left(q(j,t) + \gamma Q(t) D(t)^{\phi_2}\right) dj \end{array}\right] \frac{\partial D(t)}{\partial s(i,t)}, \end{aligned}$$

$$0 = -V_m \mu^{-\frac{1}{1-\psi}} \frac{\psi}{1-\psi} D(t)^{\frac{\psi}{1-\psi}} + b l_v(j,t)^{b-1} \left[V\left(q(i,t) + \gamma Q(t) D(t)^{\phi_2}\right) - V(q(i,t))\right], \quad (75)$$

where

$$\frac{\partial D(t)}{\partial c(i,t)} = \chi \frac{\eta}{\eta-1} \left[\Xi\right]^{\frac{\eta}{\eta-1}-1} m(t)^{-\frac{1}{\eta}} q(i,t)^{\frac{1}{\eta}} \left(s(i,t)\, c(i,t)\, l(t)\right)^{\frac{\eta-1}{\eta}} s(i,t)\, l(t),$$

$$\frac{\partial D(t)}{\partial s(i,t)} = \chi \frac{\eta}{\eta-1} \left[\Xi\right]^{\frac{\eta}{\eta-1}-1} m(t)^{-\frac{1}{\eta}} q(i,t)^{\frac{1}{\eta}} \left(s(i,t)\, c(i,t)\, l(t)\right)^{\frac{\eta-1}{\eta}} c(i,t)\, l(t),$$

$$\Xi \equiv m(t)^{-1/\eta} \int_{i=0}^{m(t)} q(i,t)^{1/\eta} \left(l(t)\, d(i,t)\right)^{(\eta-1)/\eta} di.$$

Combining (73) and (74) leads to

$$c(i,t) = \left[\int_{i=0}^{m(t)} q(i,t)^{\frac{1}{\theta}} c(i,t)^{\frac{\theta-1}{\theta}} di\right]^{-\theta} \left(V_m \mu^{-\frac{1}{1-\psi}} D(t)^{\frac{\psi}{1-\psi}}\right)^{-\theta} q(i,t) \equiv \Gamma(t)\, q(i,t). \quad (76)$$

Plugging (76) in $c(t)$ gives rise to

$$c(t) = \Gamma(t)\, m(t)^{\frac{\theta}{\theta-1}} Q(t)^{\frac{\theta}{\theta-1}}. \quad (77)$$

Putting (76) in $l_p(t)$ gives us

$$l_p(t) = \int_{i=0}^{m(t)} c(i,t)\, l(t)\, di = \Gamma(t)\, l(t)\, m(t)\, Q(t). \quad (78)$$

Combining (77) and (78) turns out to

$$c(t) = l_p(t)\, l(t)^{-1} m(t)^{\frac{1}{\theta-1}} Q(t)^{\frac{1}{\theta-1}}. \quad (79)$$

Dividing both sides of (74) by $c(i,t)$ and using (76), we obtain

$$\frac{\kappa d(t)}{\chi \left[\Xi\right]^{\frac{\eta}{\eta-1}-1} m(t)^{-\frac{1}{\eta}} \Gamma^{-\frac{1}{\eta}} s(i,t)^{-\frac{1}{\eta}} l(t)} = \begin{bmatrix} V_m l_m(t)\, \mu^{-\frac{1}{1-\psi}} \frac{\psi}{1-\psi} D(t)^{\frac{\psi}{1-\psi}-1} + \\ \phi_1 D(t)^{\phi_1-1} \int_{j=0}^{m(t)} l_v(j,t)^b \left[V\left(q(j,t) + \gamma Q D^{\phi_2}\right) - V\left(q(j,t)\right)\right] dj \\ +\gamma Q(t)\, \phi_2 D(t)^{\phi_1+\phi_2-1} \int_{j=0}^{m(t)} l_v(j,t)^b\, V_{q+\gamma Q D^{\phi_2}}\left(q + \gamma Q D^{\phi_2}\right) dj \end{bmatrix},$$

which implies that $s(i,t) = s(t)$ is independent of i. Then, substituting (78) into both $d(t)$ and $D(t)$ gives us

$$d(t) = l_p(t)\, l(t)^{-1} s(t), D(t) = \chi s(t)\, l_p(t)\, Q(t)^{\frac{1}{\eta-1}}. \quad (80)$$

Therefore, there is no aggregate uncertainty in the economy, and the optimal solution is also symmetric. For convenience, we convert the original stochastic problem into an equivalent deterministic problem as follows:

$$\max_{\{l_p,l_v,s,m,Q\}} \int_{t=0}^{\infty} e^{-(\rho-n)t} \left[\ln \left(l_p(t)\, l(t)^{-1} Q(t)^{\frac{1}{\theta-1}} m(t)^{\frac{1}{\theta-1}} \right) - \frac{\kappa}{2} l_p(t)^2\, l(t)^{-2}\, s(t)^2 \right] dt,$$

subject to

$$\dot{Q}(t) = \gamma Q(t)\, l_v(t)^b \left(\chi s(t)\, l_p(t)\, Q(t)^{\frac{1}{\eta}} \right)^{\phi_1+\phi_2}, \tag{81}$$

$$\dot{m}(t) = \left(l(t) - l_v(t)\, m(t) - l_p(t) \right) \mu^{-\frac{1}{1-\psi}} \left(\chi s(t)\, l_p(t)\, Q(t)^{\frac{1}{\eta}} \right)^{\frac{\psi}{1-\psi}}. \tag{82}$$

To solve the above problem, we define the Hamitonian:

$$H = e^{-(\rho-n)t} \left[\ln \left(l_p(t)\, l(t)^{-1} Q(t)^{\frac{1}{\theta-1}} m(t)^{\frac{1}{\theta-1}} \right) - \frac{\kappa}{2} l_p(t)^2\, l(t)^{-2}\, s(t)^2 \right] + \lambda_1(t)\, \dot{Q}(t) + \lambda_2(t)\, \dot{m}(t),$$

where $\lambda_1(t)$ and $\lambda_2(t)$ are two Hamiltonian multipliers. The first-order necessary conditions for $l_p(t)$, $l_v(t)$, $s(t)$, $Q(t)$, and $m(t)$ are

$$e^{-(\rho-n)t} \left(1 - \kappa l_p(t)^2\, l(t)^{-2}\, s(t)^2 \right) + \lambda_1(t)(\phi_1+\phi_2)\, \dot{Q}(t) + \lambda_2(t)\, \dot{m}(t) \left(\frac{\psi}{1-\psi} - \frac{l_p(t)}{l_m(t)} \right) = 0, \tag{83}$$

$$b\lambda_1(t)\, \dot{Q}(t) = \lambda_2(t)\, \dot{m}(t) \frac{l_v(t)\, m(t)}{l(t) - l_v(t)\, m(t) - l_p(t)}, \tag{84}$$

$$-e^{-(\rho-n)t} \kappa l_p(t)^2\, l(t)^{-2}\, s(t)^2 + \lambda_1(t)(\phi_1+\phi_2)\, \dot{Q}(t) + \lambda_2(t)\, \dot{m}(t) \frac{\psi}{1-\psi} = 0, \tag{85}$$

$$e^{-(\rho-n)t} \frac{1}{\theta-1} + \left(1 + \frac{\phi_1+\phi_2}{\eta-1} \right) \lambda_1(t)\, \dot{Q}(t) + \frac{\psi}{1-\psi} \frac{1}{\eta-1} \lambda_2(t)\, \dot{m}(t) = -\lambda_1(t)\, \dot{Q}(t), \tag{86}$$

$$e^{-(\rho-n)t} \frac{1}{\theta-1} - \lambda_2(t)\, \dot{m}(t) \frac{l_v(t)\, m(t)}{l(t) - l_v(t)\, m(t) - l_p(t)} = -\lambda_2(t)\, \dot{m}(t). \tag{87}$$

Putting (85) in (83) and using (84) lead to

$$e^{-(\rho-n)t} = \lambda_2(t)\, \dot{m}(t) \frac{l_p(t)}{l(t) - l_v(t)\, m(t) - l_p(t)} = \frac{b\lambda_1(t)\, \dot{Q}(t) l_p(t)}{l_v(t)\, m(t)}, \tag{88}$$

which is

$$e^{-(\rho-n)t} = \lambda_2(t)\, g_m(t)\, \frac{l_p(t)\, m(t)}{l(t) - l_v(t)\, m(t) - l_p(t)} = \frac{b\lambda_1(t)\, g_Q(t)\, l_p(t)\, Q(t)}{l_v(t)\, m(t)}. \tag{89}$$

Putting (84) and (88) in (86) and dividing both sides by $\lambda_1(t) Q(t)$, we obtain

$$\left[\frac{b}{\theta-1}\frac{l_p(t)}{l_v(t)\, m(t)} + 1 + \frac{\phi_1+\phi_2}{\eta-1} + \frac{\psi}{1-\psi}\frac{b}{\eta-1}\frac{l(t) - l_v(t)\, m(t) - l_p(t)}{l_p(t)\, m(t)}\right] g_Q(t) = -g_{\lambda_1}(t). \tag{90}$$

We conjecture that (i) the BGP growth rates of $m(t)$, $Q(t)$, $\lambda_1(t)$, and $\lambda_2(t)$ are constant, i.e., $g_m(t) = g_m$, $g_Q(t) = g_Q$, $g_{\lambda_1}(t) = g_{\lambda_1}$, $g_{\lambda_2}(t) = g_{\lambda 2}$; and (ii) the labor employment shares $L_p(t) \equiv l_p(t)/l(t)$, $L_v(t) \equiv l_v(t)\, m(t)/l(t)$, and $L_m(t) \equiv l_m(t)/l(t)$ are constant, namely, $L_p(t) = L_p$, $L_v(t) = L_v$, $L_m(t) = L_m$. Thus, on the BGP, the labor employment ratios between production and horizontal/vertical innovation, $\frac{l_p(t)}{l_m(t)}$ and $\frac{l_p(t)}{l_v(t)m(t)}$, are also constant. Taking time derivatives on both sides of equation (89), we know that on the BGP,

$$-(\rho - n) = g_{\lambda 2} + g_m = g_{\lambda 1} + g_Q. \tag{91}$$

On the BGP, equation (90) is changed as

$$\left[\frac{b}{\theta-1}\frac{l_p(t)}{l_v(t)\, m(t)} + 1 + \frac{\phi_1+\phi_2}{\eta-1} + \frac{\psi}{1-\psi}\frac{b}{\eta-1}\frac{l(t) - l_v(t)\, m(t) - l_p(t)}{l_p(t)\, m(t)}\right] g_Q = -g_{\lambda_1}. \tag{92}$$

Putting (92) in (91), we solve for

$$g_Q = \frac{\rho - n}{\frac{b}{\theta-1}\frac{l_p(t)}{l_v(t)m(t)} + \frac{\phi_1+\phi_2}{\eta-1} + \frac{\psi}{1-\psi}\frac{b}{\eta-1}\frac{l(t)-l_v(t)m(t)-l_p(t)}{l_v(t)m(t)}}. \tag{93}$$

Substituting (88) into (87), dividing both sides of the resulting equation by $\lambda_2(t)\, m(t)$, and using equation (90), we know that

$$g_m = \frac{\rho - n}{\frac{1}{\theta-1}\frac{l_p(t)}{l(t)-l_v(t)m(t)-l_p(t)} - \frac{l_v(t)m(t)}{l(t)-l_v(t)m(t)-l_p(t)} - 1}. \tag{94}$$

Equation (81) tells that at any time t,

$$g_Q(t) = \gamma l_v(t)^b\, D(t)^{\phi_1+\phi_2} = \gamma \left(\frac{l_v(t)\, m(t)}{l(t)}\right)^b m(t)^{-b}\, l(t)^b\, D(t)^{\phi_1+\phi_2}, \tag{95}$$

which implies that on the BGP,

$$n - g_m + \frac{\phi_1 + \phi_2}{b} g_D = 0. \tag{96}$$

Equation (82) tells that at any time t,

$$g_m(t) = \frac{l(t) - l_v(t) m(t) - l_p(t)}{l(t)} l(t) m(t)^{-1} \mu^{-\frac{1}{1-\psi}} D(t)^{\frac{\psi}{1-\psi}}, \tag{97}$$

which implies that on the BGP,

$$n - g_m + \frac{\psi}{1-\psi} g_D = 0. \tag{98}$$

The Knife edge condition (i.e., $\frac{\phi_1+\phi_2}{b} = \frac{\psi}{1-\psi}$) establishes that equations (96) and (98) are the same.

Putting (88) in (85) establishes that $s(t)$ is constant on the BGP, namely,

$$s(t) = \sqrt{\frac{\frac{\phi_1+\phi_2}{b}\frac{l_v(t)m(t)}{l_p(t)} + \frac{\psi}{1-\psi}\frac{l(t)-l_v(t)m(t)-l_p(t)}{l_p(t)}}{\kappa\left(\frac{l_p(t)}{l(t)}\right)^2}} = s. \tag{99}$$

Equation (72) tells that $D(t) = \chi \frac{l_p(t)}{l(t)} l(t) s(t) Q(t)^{\frac{1}{\eta-1}}$, which establishes that on the BGP,

$$g_D = n + \frac{1}{\eta - 1} g_Q = \frac{b}{\phi_1 + \phi_2} (g_m - n). \tag{100}$$

Then, we know that

$$g_m = \left(1 + \frac{\phi_1 + \phi_2}{b}\right) n + \frac{\phi_1 + \phi_2}{b} \frac{1}{\eta - 1} g_Q \equiv z_1 n + z_2 g_Q, \tag{101}$$

where

$$z_1 \equiv 1 + \frac{\phi_1 + \phi_2}{b}, z_2 \equiv \frac{\phi_1 + \phi_2}{b} \frac{1}{\eta - 1}.$$

Substituting (97) into (95), we have that at any time t,

$$g_Q(t)^{\frac{1}{b}} = \gamma^{\frac{1}{b}} \mu^{\frac{1}{1-\psi}} \frac{l_v(t) m(t)}{l(t) - l_v(t) m(t) - l_p(t)} g_m(t).$$

Thus, on the BGP, we know that

$$g_Q^{\frac{1}{b}} = \gamma^{\frac{1}{b}} \mu^{\frac{1}{1-\psi}} \frac{l_v(t)\, m(t)}{l(t) - l_v(t)\, m(t) - l_p(t)} g_m. \tag{102}$$

Define

$$x(t) \equiv \frac{l_p(t)}{l_m(t)} = \frac{L_p(t)}{L_m(t)}, y(t) \equiv \frac{l_v(t)\, m(t)}{l_m(t)} = \frac{L_v(t)}{L_m(t)}.$$

On the BGP, (93), (94), (101) and (102) turn out to an equation system for (x, y, g_m, g_Q):

$$\begin{aligned}
(\rho - n) g_Q^{-1} &= \frac{b}{\theta - 1}\frac{x}{y} + \frac{\phi_1 + \phi_2}{\eta - 1} + \frac{\psi}{1-\psi}\frac{b}{\eta - 1}\frac{b}{y}, && (103)\\
(\rho - n) g_m^{-1} &= \frac{1}{\theta - 1}x - y - 1, && (104)\\
g_m &= z_1 n + z_2 g_Q, && (105)\\
g_Q^{\frac{1}{b}} &= \gamma^{\frac{1}{b}} \mu^{\frac{1}{1-\psi}} y g_m. && (106)
\end{aligned}$$

Substituting (104), (105), and (106) into (103) one by one, we obtain the equation pinning down g_Q:

$$\frac{(\rho - n) g_Q^{\frac{1}{b}-1} - \left(b + \frac{\phi_1+\phi_2}{\eta-1}\right) g_Q^{\frac{1}{b}}}{(\rho - n) b + \left(b + \frac{\phi_1+\phi_2}{\eta-1}\right)(z_1 n + z_2 g_Q)} = \gamma^{\frac{1}{b}} \mu^{\frac{1}{1-\psi}}. \tag{107}$$

By substitution, we have that

$$\begin{aligned}
g_m &= z_1 n + z_2 g_Q,\\
y &= \gamma^{-\frac{1}{b}} \mu^{-\frac{1}{1-\psi}} g_Q^{\frac{1}{b}} (z_1 n + z_2 g_Q)^{-1},\\
x &= (\theta - 1)\left[\left(\rho - n + \gamma^{-\frac{1}{b}} \mu^{-\frac{1}{1-\psi}} g_Q^{\frac{1}{b}}\right)(z_1 n + z_2 g_Q)^{-1} + 1\right].
\end{aligned}$$

Using $x = L_p/L_m$, $y = L_v/L_m$, and $L_p + L_v + L_m = 1$, we solve for

$$L_p = \frac{x}{x + y + 1}, L_v = \frac{y}{x + y + 1}, L_m = \frac{1}{x + y + 1}. \tag{108}$$

Using (96) and (105), we solve for

$$g_D = \frac{b}{\phi_1 + \phi_2}\left[(z_1 - 1) n + z_2 g_Q\right]. \tag{109}$$

From (79), we solve for

$$g_c = \frac{1}{\theta - 1}\left[z_1 n + (z_2 + 1)\, g_Q\right]. \tag{110}$$

Using (99), we have that

$$s = \sqrt{\frac{\phi_1 + \phi_2}{b}\frac{1 - L_p}{L_p}\kappa^{-1} L_p^{-2}}. \tag{111}$$

Plugging (111) in (80) gives rise to

$$d = \sqrt{\frac{\phi_1 + \phi_2}{b\kappa}\frac{1 - L_p}{L_p}}. \tag{112}$$

Using (108) and $l_p(t) = (\theta - 1)\, e(t)\, l(t)/\theta$, we obtain

$$e = \frac{\theta}{\theta - 1} L_p. \tag{113}$$

Thus, we complete the proof of Proposition 1. □

7.2 Appendix 8.2 Proof of Proposition 2

Proof of Proposition 2. First, we solve the households' problems. Each household's problem is summarized as follows:

$$\max_{\{c(i,t),a(t)\}} E_0\left[\int_{t=0}^{+\infty} e^{-(\rho - n)t}\left(\ln c(t) - \frac{\kappa}{2} d(t)^2\right) dt\right],$$

subject to

$$da(t) = \left[(r(t) - n)\, a(t) + w(t) - e(t)\right] dt,$$

$$dm(t) = \dot{m}(t) dt,$$

$$dq(i,t) = \begin{cases} \gamma Q(t)\, d_v(i,t)^{\phi_2}, l_v(i,t)^b\, d_v(i,t)^{\phi_1}\, dt \\ 0, 1 - l_v(i,t)^b\, d_v(i,t)^{\phi_1}\, dt \end{cases},$$

$$c(t) \equiv \left[\int_{i=0}^{m(t)} q(i,t)^{\frac{1}{\theta}}\, c(i,t)^{\frac{\theta - 1}{\theta}}\, di\right]^{\frac{\theta}{\theta - 1}}, e(t) \equiv \int_{i=0}^{m(t)} p(i,t)\, c(i,t)\, di.$$

Define the value function $J(t, a(t), m(t), \{q(i,t), i \in [0, m(t)]\})$. The Bellman equation is

written as follows:

$$(\rho - n) J = \max_{\{c(i,t)\}} \left\{ \begin{array}{l} \ln c(t) - \frac{\kappa}{2} d(t)^2 + J_t + J_a \left[(r(t) - n) a(t) + w(t) - e(t) \right] + J_m \dot{m}(t) \\ + \int_{i=0}^{m(t)} l_v(i,t)^b d_v(i,t)^{\phi_1} \left[J\left(q(i,t) + \gamma Q(t) d_v(i,t)^{\phi_2} \right) - J(q(i,t)) \right] \end{array} \right\}.$$

The first-order condition for $c(i,t)$ is

$$\left(\int_{i=0}^{m(t)} q(i,t)^{\frac{1}{\theta}} c(i,t)^{\frac{\theta-1}{\theta}} di \right)^{-1} q(i,t)^{\frac{1}{\theta}} c(i,t)^{\frac{\theta-1}{\theta}} = J_a p(i,t) c(i,t). \tag{114}$$

Integrating (114) with respect to i gives rise to

$$1 = J_a e(t). \tag{115}$$

Taking powers θ and $(\theta - 1)$ in turn on both sides of (114) gives us

$$\left[\int_{i=0}^{m(t)} q(i,t)^{\frac{1}{\theta}} c(i,t)^{\frac{\theta-1}{\theta}} di \right]^{-\theta} q(i,t) J_a^{-\theta} p(i,t)^{1-\theta} = c(i,t), \tag{116}$$

$$\left[\int_{i=0}^{m(t)} q(i,t)^{\frac{1}{\theta}} c(i,t)^{\frac{\theta-1}{\theta}} di \right]^{1-\theta} q(i,t) J_a^{1-\theta} p(i,t)^{1-\theta} = q(i,t)^{\frac{1}{\theta}} c(i,t)^{\frac{\theta-1}{\theta}},$$

respectively. Integrating the second equation leads to

$$\left[\int_{i=0}^{m(t)} q(i,t)^{\frac{1}{\theta}} c(i,t)^{\frac{\theta-1}{\theta}} di \right]^{-\theta} J_a^{1-\theta} \int_{i=0}^{m(t)} q(i,t) p(i,t)^{1-\theta} di = 1. \tag{117}$$

Combining (115) and (116) and using (115) and (117) give rise to

$$\left[\int_{i=0}^{m(t)} q(i,t)^{\frac{1}{\theta}} c(i,t)^{\frac{\theta-1}{\theta}} di \right]^{-\theta} q(i,t) e(t)^{\theta} p(i,t)^{-\theta} = c(i,t),$$

$$\left[\int_{i=0}^{m(t)} q(i,t)^{\frac{1}{\theta}} c(i,t)^{\frac{\theta-1}{\theta}} di \right]^{-\theta} e(t)^{\theta-1} \int_{i=0}^{m(t)} q(i,t) p(i,t)^{1-\theta} di = 1.$$

Using the above two equations, we know that the optimal demand for good i is

$$c(i,t) = \frac{q(i,t) p(i,t)^{-\theta} e(t)}{\int_{i=0}^{m(t)} q(i,t) p(i,t)^{1-\theta} di}. \tag{118}$$

Using the Envelope theorem with respect to $a(t)$, we have that

$$(\rho - n) J_a = \begin{pmatrix} J_{ta} + J_{aa}\dot{a}(t) + J_a(r(t) - n) + J_{ma}\dot{m}(t) + \\ \int_{i=0}^{m(t)} l_v(i,t)^b d_v(i,t)^{\phi_1} \left[J_a\left(q(i,t) + \gamma Q(t) d_v(i,t)^{\phi_2}\right) - J_a(q(i,t)) \right] \end{pmatrix}.$$

Taking the time derivatives on both sides of $J_a = J_a(t, a(t), m(t), \{q(i,t), i \in [0, m(t)]\})$, we obtain

$$\frac{dJ_a}{dt} = J_{at} + J_{aa}\dot{a}(t) + J_{am}\dot{m}(t) + \int_{i=0}^{m(t)} l_v(i,t)^b d_v(i,t)^{\phi_1} \left[J_a\left(q(i,t) + \gamma Q(t) d_v(i,t)^{\phi_2}\right) - J_a(q(i,t)) \right].$$

Combining the above two equations and using (115), we obtain the Euler equation

$$g_e(t) \equiv \frac{\dot{e}(t)}{e(t)} = r(t) - \rho. \tag{119}$$

Second, we solve the production problem of the final goods. The owner of variety i's profit-maximizing problem is:

$$\max_{\{p(i,t), c(i,t)\}} [p(i,t) - 1] c(i,t) l(t),$$

subject to the market demand for good i, $c(i,t) l(t)$, where $c(i,t)$ satisfies (118). Putting (118) in the above problem and solving the first-order condition wrt $p(i,t)$ lead to the monopolistic pricing rule

$$p(i,t) = \frac{\theta}{\theta - 1} = p(t), \tag{120}$$

and the optimal profits:

$$\pi(i,t) = \frac{1}{\theta} \frac{q(i,t)}{Q(t)} \frac{e(t) l(t)}{m(t)} \equiv \frac{1}{\theta} \alpha(i,t) \lambda(t).$$

Plugging (120) in (118) gives rise to

$$c(i,t) = \alpha(i,t) \frac{e(t)}{m(t)} \frac{\theta - 1}{\theta}. \tag{121}$$

Third, we solve the quality-enhancing innovators' problems. Using production profits and revenues from data sales, the owner of variety i hires labor, $l_v(i,t)$, and purchases integrated

data, $d_v(i,t)$, to improve product quality. Specifically, each variety solves the following problem:

$$J(\alpha(i,t),t) \equiv \max_{\{l_v,d_v,s,p_d\}} E_t \left[\int_{u=t}^{+\infty} e^{-\int_{\tau=t}^{u} r(\tau)d\tau} \left(\begin{array}{c} (1-\delta(s(i,u))) \frac{\alpha(i,u)\lambda(u)}{\theta} - l_v(i,u) - \\ p_D^v(u) d_v(i,u) + p_d(i,u) s(i,u) c(i,u) l(u) \end{array} \right) du \right],$$

subject to the evolutionary equation of its relative quality

$$d\alpha(i,t) = -\alpha(i,t) g_Q(t) dt + \frac{1}{Q(t)} dq(i,t), \tag{122}$$

and the demand function for the original data,

$$p_d(i,t) = \left(p_D^v(t) m(t) + p_D^h(t) \dot{m(t)} \right) \chi [\Xi]^{\frac{1}{\eta-1}} m(t)^{-\frac{1}{\eta}} q(i,t)^{\frac{1}{\eta}} (l(t) c(i,t) s(i,t))^{-\frac{1}{\eta}}, \tag{123}$$

where

$$\Xi \equiv m(t)^{-1/\eta} \int_{i=0}^{m(t)} q(i,t)^{1/\eta} (l(t) d(i,t))^{(\eta-1)/\eta} di,$$

and $\delta(s(i,t)) = 0.5\delta_0 s(i,t)^2$ stands for the profit loss of data sharing for variety i. To solve the above problem, we write down the Bellman equation:

$$0 = \max_{\{l_v,d_v,s\}} \left\{ \begin{array}{c} e^{-\int_{u=0}^{t} r(u)du} \left(\begin{array}{c} (1-\delta(s(i,t))) \pi(i,t) - l_v(i,t) - \\ p_D^v(t) d_v(i,t) + p_d(i,t) s(i,t) c(i,t) l(t) \end{array} \right) \\ -J_\alpha(\alpha(i,t),t) \alpha(i,t) g_Q(t) + J_t(\alpha(i,t),t) + \\ l_v(i,t)^b d_v(i,t)^{\phi_1} \left[J\left(\alpha(i,t) + \gamma d_v(i,t)^{\phi_2}\right) - J(\alpha(i,t)) \right] \end{array} \right\}.$$

This is a nonlinear differential equation with jump components, which cannot be solved directly. Therefore, we will restrict our attention to balanced growth solutions, where all variables are either constant or growing at constant rates. This will yield an autonomous problem that can be solved. Similar to Dinopoulos and Thompson (1998), on the balanced growth path, we assume that $J(\alpha(i),t) = e^{-rt} V(\alpha(i), \lambda(t), p_D(t))$. To solve for the BGP, we set $J(\alpha(i),t) = e^{-rt} V(\alpha(i), \lambda(t), p_D(t))$ and substitute it into the above equation, leading to

$$0=\max_{\{l_v,d_v,s,p_d\}}\left\{\begin{array}{c}\frac{1-\delta(s(i))}{\theta}\alpha\left(i\right)\lambda\left(t\right)-l_v\left(i,t\right)-p_D^v\left(t\right)d_v\left(i,t\right)+p_d\left(i,t\right)s\left(i,t\right)c\left(i,t\right)l\left(t\right)-\\ \left(\rho+g_e\right)V+V_\lambda\left(g_e+n-g_m\right)\lambda\left(t\right)+V_{p_D^v}g_{p_D^v}p_D^v\left(t\right)-\alpha\left(i\right)g_QV_\alpha+\\ l_v\left(i,t\right)^b d_v\left(i,t\right)^{\phi_1}\left[V\left(\alpha\left(i\right)+\gamma d_v\left(i,t\right)^{\phi_2},\lambda,p_D^v\left(t\right)\right)-V\left(\alpha\left(i\right),\lambda,p_D^v\left(t\right)\right)\right]\end{array}\right\}. \tag{124}$$

The first-order necessary conditions with respect to $l_v\left(i,t\right)$, $d_v\left(i,t\right)$, and $s\left(i,t\right)$ are:

$$1=bl_v\left(i,t\right)^{b-1}d_v\left(i,t\right)^{\phi_1}\left[V\left(\alpha\left(i\right)+\gamma d_v\left(i,t\right)^{\phi_2},\lambda\left(t\right),p_D^v\left(t\right)\right)-V\left(\alpha\left(i\right),\lambda\left(t\right),p_D^v\left(t\right)\right)\right], \tag{125}$$

$$p_D^v\left(t\right)=\left\{\begin{array}{c}\phi_1 l_v\left(i,t\right)^b d_v\left(i,t\right)^{\phi_1-1}\left[V\left(\alpha\left(i\right)+\gamma d_v\left(i,t\right)^{\phi_2}\right)-V\left(\alpha\left(i\right)\right)\right]\\ +\gamma\phi_2 l_v\left(i,t\right)^b d_v\left(i,t\right)^{\phi_1+\phi_2-1}V_\alpha\left(\alpha\left(i\right)+\gamma d_v\left(i,t\right)^{\phi_2}\right)\end{array}\right\}, \tag{126}$$

$$\delta_0 s\left(i,t\right)\frac{\alpha\left(i\right)\lambda}{\theta}=\frac{\eta-1}{\eta}c\left(i,t\right)l\left(t\right)p_d\left(i,t\right). \tag{127}$$

Putting (121) and (123) in (127) leads to

$$s\left(i,t\right)^{1+\frac{1}{\eta}}=\delta_0^{-1}\left(p_D^v\left(t\right)m+p_D^h\left(t\right)\dot{m}\right)\chi\left[\Xi\right]^{\frac{1}{\eta-1}}m\left(t\right)^{-\frac{1}{\eta}}\frac{\eta-1}{\eta}\left(\frac{e\left(t\right)l\left(t\right)}{m\left(t\right)Q\left(t\right)}\frac{\theta-1}{\theta}\right)^{-\frac{1}{\eta}}\left(\theta-1\right)=s\left(t\right)^{1+1/\eta}, \tag{128}$$

which shows that all varieties sell the same share of original data, namely, $s\left(i,t\right)=s\left(t\right)$. Thus, we have that

$$s\left(t\right)=\delta_0^{-1}\left(p_D^v\left(t\right)m+p_D^h\left(t\right)\dot{m}\right)\chi Q\left(t\right)^{\frac{1}{\eta-1}}\frac{\eta-1}{\eta}\left(\theta-1\right). \tag{129}$$

Using $s\left(i,t\right)=s\left(t\right)$ in equation (123), we find that $p_d\left(i,t\right)$ is also independent of i, namely,

$$p_d\left(i,t\right)=\left(p_D^v\left(t\right)m+p_D^h\left(t\right)\dot{m}\right)\chi Q\left(t\right)^{\frac{1}{\eta-1}}=p_d\left(t\right). \tag{130}$$

Combining (129) and (130) gives

$$s\left(t\right)=\delta_0^{-1}\frac{\eta-1}{\eta}\left(\theta-1\right)p_d\left(t\right). \tag{131}$$

We conjecture that the value function takes the following form:

$$V\left(\alpha\left(i\right),\lambda\left(t\right),p_D^v\left(t\right)\right)=y_1\alpha\left(i\right)\lambda\left(t\right)+y_2p_D^v\left(t\right)^{x_1}\lambda\left(t\right)^{x_2}. \tag{132}$$

Then, the first-order necessary conditions become:

$$l_v(i,t) = \left[b\gamma y_1 \lambda(t) d_v(i,t)^{\phi_1+\phi_2}\right]^{\frac{1}{1-b}} (= l_v(t)), \tag{133}$$

$$d_v(i,t) = \left[\frac{\phi_1+\phi_2}{b} p_D^v(t)^{-1} (b\gamma y_1 \lambda(t))^{\frac{1}{1-b}}\right]^{\frac{1-b}{1-b-(\phi_1+\phi_2)}} = d_v(t), \tag{134}$$

which establish that all varieties employ the same levels of labor and data for quality-enhancing innovations.

Substituting equations (132)-(134) into the Bellman equation, we determine the undetermined coefficients as follows:

$$x_1 = -\frac{\phi_1+\phi_2}{1-b-(\phi_1+\phi_2)}, x_2 = \frac{1}{1-b-(\phi_1+\phi_2)}, \tag{135}$$

$$y_1 = \frac{1+\delta_0^{-1}\left(1-\frac{1}{\eta}\right) p_d^2 (\theta-1)^2 \left(1-2^{-1}\left(1-\eta^{-1}\right)\right)}{\theta(\rho+g_e-g_\lambda+g_Q)}, y_2 = \frac{\frac{1-b-(\phi_1+\phi_2)}{b} (\gamma b y_1)^{x_2} \left(\frac{\phi_1+\phi_2}{b}\right)^{-x_1}}{\rho+g_e-x_2 g_\lambda - x_1 g_{p_D^v}}. \tag{136}$$

Thus, we obtain the solution to the Bellman equation (124).

Fourth, we solve the problem of variety-expanding innovators. Construct the Lagrangian:

$$\mathcal{L} = l_h(t) + p_D^h(t) d_h(t) + \vartheta\left[\mu - d_h(t)^\psi l_h(t)^{1-\psi}\right],$$

where ϑ is the Lagrangian multiplier. The first-order necessary conditions are

$$1 = \vartheta d_h(t)^\psi (1-\psi) l_h(t)^{-\psi}, p_D^h(t) = \vartheta\psi d_h(t)^{\psi-1} l_h(t)^{1-\psi}.$$

Eliminating ϑ in the above two equations gives rise to

$$p_D^h(t) = \frac{\psi}{1-\psi}\frac{l_h(t)}{d_h(t)}. \tag{137}$$

Combining (137) with $\mu = d_h(t)^\psi l_h(t)^{1-\psi}$, we solve for the demand functions for the two inputs:

$$l_h(t) = \mu\left(\frac{1-\psi}{\psi}\right)^\psi p_D^h(t)^\psi, d_h(t) = \mu\left(\frac{\psi}{1-\psi}\right)^{1-\psi} p_D^h(t)^{\psi-1}. \tag{138}$$

We assume that new entrants obtain the profit flows from all existing varieties who suffering

from creative destruction. Using the linearity of V wrt α and (138), we have a zero-profit condition for variety-expanding R&D, namely,

$$\begin{aligned} V\left(1,\lambda\left(t\right),p_{D}^{v}\left(t\right)\right) &= \frac{1}{1-\psi}\mu^{\frac{1}{1-\psi}}d_{h}\left(t\right)^{-\frac{\psi}{1-\psi}}-\frac{\int_{i=0}^{m(t)}\delta\left(s\left(i,t\right)\right)\frac{1}{\theta}\alpha\left(i,t\right)\lambda\left(t\right)di}{\dot{m}\left(t\right)} \\ &= \frac{1}{1-\psi}l_{h}\left(t\right)-\frac{\int_{i=0}^{m(t)}\delta\left(s\left(i,t\right)\right)\frac{1}{\theta}\alpha\left(i,t\right)\lambda\left(t\right)di}{\dot{m}\left(t\right)} \\ &= \frac{1}{\psi}d_{h}\left(t\right)p_{D}^{h}\left(t\right)-\frac{\int_{i=0}^{m(t)}\delta\left(s\left(i,t\right)\right)\frac{1}{\theta}\alpha\left(i,t\right)\lambda\left(t\right)di}{\dot{m}\left(t\right)}. \end{aligned} \tag{139}$$

Fifth, the competitive data intermediary sector solves the following optimization problem:

$$\max_{\{d(i,t)\}}\left(p_{D}^{v}\left(t\right)m+p_{D}^{h}\left(t\right)\dot{m}\right)D\left(t\right)-\int_{i=0}^{m(t)}p_{d}\left(i,t\right)d\left(i,t\right)l\left(t\right)di,$$

subject to the data-integrating technology

$$D\left(t\right)=\chi\left[m\left(t\right)^{-1/\eta}\int_{i=0}^{m(t)}q\left(i,t\right)^{1/\eta}\left(l\left(t\right)d\left(i,t\right)\right)^{(\eta-1)/\eta}di\right]^{\eta/(\eta-1)},\eta>1. \tag{140}$$

Solving the FOC w.r.t $d\left(i,t\right)$ leads to the demand function for the original data

$$\begin{aligned} p_{d}\left(i,t\right) &= \left(p_{D}^{v}\left(t\right)m+p_{D}^{h}\left(t\right)\dot{m}\right)\chi\left[\Xi\right]^{\frac{1}{\eta-1}}m\left(t\right)^{-\frac{1}{\eta}}q\left(i,t\right)^{\frac{1}{\eta}}\left(d\left(i,t\right)l\left(t\right)\right)^{-\frac{1}{\eta}} \\ &= \left(p_{D}^{v}\left(t\right)m+p_{D}^{h}\left(t\right)\dot{m}\right)\chi Q\left(t\right)^{\frac{1}{\eta-1}}=p_{d}\left(t\right). \end{aligned} \tag{141}$$

Free entry leads to zero profit in the data intermediary sector, namely,

$$\left(p_{D}^{v}\left(t\right)m+p_{D}^{h}\left(t\right)\dot{m}\right)D\left(t\right)=p_{d}\left(t\right)l\left(t\right)s\left(t\right)e\left(t\right)\frac{\theta-1}{\theta}. \tag{142}$$

Sixth, we impose the market-clearing conditions for labor and capital markets, namely,

$$l\left(t\right)=\int_{i=0}^{m(t)}c\left(i,t\right)l\left(t\right)di+\int_{i=0}^{m(t)}l_{v}\left(i,t\right)di+\dot{m}\left(t\right)l_{h}\left(t\right)=\frac{\theta-1}{\theta}e\left(t\right)l\left(t\right)+l_{v}\left(t\right)m\left(t\right)+\dot{m}\left(t\right)l_{h}\left(t\right), \tag{143}$$

$$a\left(t\right)l\left(t\right)=\int_{i=0}^{m(t)}V\left(\alpha\left(i\right),\lambda\left(t\right),p_{D}^{v}\left(t\right)\right)di=m\left(t\right)V\left(1,\lambda\left(t\right),p_{D}^{v}\left(t\right)\right). \tag{144}$$

Finally, we solve for the stationary equilibrium (or BGP), on which (g_Q, g_m, g_λ, g_D, g_{p_D}, p_d, r, e, s, d, L_p, L_v, L_m) are constant. Similar to Dinopoulos and Thompson (1998), the growth rate of average quality level at any time t is given by

$$g_Q(t) \equiv \frac{\dot{Q}(t)}{Q(t)} = \gamma l_v(t)^b D(t)^{\phi_1+\phi_2}. \tag{145}$$

Putting (133) in (145) tells that on the BGP,

$$\lambda(t) = \gamma^{-\frac{1}{b}} b^{-1} g_Q^{\frac{1-b}{b}} y_1^{-1} D(t)^{-\frac{\phi_1+\phi_2}{b}}, \tag{146}$$

and thus,

$$g_\lambda = -\frac{\phi_1+\phi_2}{b} g_D, \tag{147}$$

since g_Q is constant on the BGP.

Using the definitions of labor shares and substituting (121), (133) and (146) into them, we find that along the BGP,

$$L_p = \frac{\theta-1}{\theta} e, L_v = eby_1 g_Q, L_m = 1 - L_p - L_v. \tag{148}$$

Substituting (133) and (146) into (134) leads to

$$p_D^v(t) = \frac{\phi_1+\phi_2}{b} \gamma^{-\frac{1}{b}} g_Q^{\frac{1}{b}} D(t)^{-\left(1+\frac{\phi_1+\phi_2}{b}\right)}. \tag{149}$$

Putting (146) and (149) in the value function (132) gives rise to

$$V(\alpha(i), \lambda(t), p_D^v(t)) = \gamma^{-\frac{1}{b}} b^{-1} g_Q^{\frac{1-b}{b}} D(t)^{-\frac{\phi_1+\phi_2}{b}} \alpha(i) + \frac{1-b-(\phi_1+\phi_2)}{b(\rho+g_e-x_2 g_\lambda - x_1 g_{p_D})} \gamma^{-\frac{1}{b}} g_Q^{\frac{1}{b}} D(t)^{-\frac{\phi_1+\phi_2}{b}}. \tag{150}$$

Putting (146) in (139) gives us

$$V(1, \lambda(t), p_D^v(t)) = \frac{1}{1-\psi} \mu^{\frac{1}{1-\psi}} D(t)^{-\frac{\psi}{1-\psi}} - \frac{1}{\theta} {g_m}^{-1} \delta(s) \gamma^{-\frac{1}{b}} b^{-1} y_1^{-1} g_Q^{\frac{1-b}{b}} D(t)^{-\frac{\phi_1+\phi_2}{b}}. \tag{151}$$

Taking $\alpha(i) = 1$ in equation (150), using the knife-edge condition, and combining it with

(151), we obtain

$$\gamma^{-\frac{1}{b}} b^{-1} g_Q^{\frac{1-b}{b}} + \frac{1-b-(\phi_1+\phi_2)}{b\left(\rho+g_e-x_2 g_\lambda - x_1 g_{p_D^v}\right)} \gamma^{-\frac{1}{b}} g_Q^{\frac{1}{b}} = \frac{1}{1-\psi} \mu^{\frac{1}{1-\psi}} - \frac{1}{\theta} {g_m}^{-1} \delta(s) \gamma^{-\frac{1}{b}} b^{-1} y_1^{-1} g_Q^{\frac{1-b}{b}}. \tag{152}$$

Combining (147) and (149), we know that on the BGP,

$$g_{p_D^v} = -\left(1+\frac{\phi_1+\phi_2}{b}\right) g_D = \left(1+\frac{b}{\phi_1+\phi_2}\right) g_\lambda. \tag{153}$$

Substituting (121) into (140) leads to

$$D(t) = \chi \frac{\theta-1}{\theta} s(t) e(t) l(t) Q(t)^{\frac{1}{\eta-1}}, \tag{154}$$

which implies that on the BGP,

$$g_D = n + \frac{1}{\eta-1} g_Q. \tag{155}$$

Using (153) and (155), we have

$$g_m = \left(1+\frac{\phi_1+\phi_2}{b}\right) n + \frac{\phi_1+\phi_2}{b} \frac{1}{\eta-1} g_Q \equiv z_1 n + z_2 g_Q. \tag{156}$$

Using (153) and (156), we have that

$$\rho + g_e - x_2 g_\lambda - x_1 g_{p_D^v} = \rho + (z_1 - 1) n + z_2 g_Q. \tag{157}$$

From (145) and by definition, we know that

$$g_Q^{\frac{1}{b}} = \gamma^{\frac{1}{b}} L_v l(t) m(t)^{-1} D(t)^{\frac{\phi_1+\phi_2}{b}}, g_m = g_Q^{\frac{1}{b}} \gamma^{-\frac{1}{b}} \mu^{-\frac{1}{1-\psi}} \frac{1-L_v-L_p}{L_v}, \tag{158}$$

which imply that

$$L_v = \frac{1-\frac{\theta-1}{\theta} e}{1+(z_1 n + z_2 g_Q) g_Q^{-\frac{1}{b}} \gamma^{\frac{1}{b}} \mu^{\frac{1}{1-\psi}}}. \tag{159}$$

Combining (147) and (159) leads to

$$eby_1 g_Q = \frac{1 - \frac{\theta-1}{\theta} e}{1 + (z_1 n + z_2 g_Q) g_Q^{-\frac{1}{b}} \gamma^{\frac{1}{b}} \mu^{\frac{1}{1-\psi}}}. \tag{160}$$

Plugging (160) in $\frac{1}{\theta} {g_m}^{-1} \delta(s) \gamma^{-\frac{1}{b}} b^{-1} y_1^{-1} g_Q^{\frac{1-b}{b}}$ gives us

$$\frac{1}{\theta} {g_m}^{-1} \delta(s) \gamma^{-\frac{1}{b}} b^{-1} y_1^{-1} g_Q^{\frac{1-b}{b}} = \frac{1}{2} \frac{\eta - 1}{\eta} \frac{\phi_1 + \phi_2}{b} \left(g_m^{-1} g_Q^{\frac{1}{b}} \gamma^{-\frac{1}{b}} + \mu^{\frac{1}{1-\psi}} \right). \tag{161}$$

Putting (157) and (161) in (152) yields us the key equation pinning down g_Q:

$$\gamma^{-\frac{1}{b}} b^{-1} g_Q^{\frac{1-b}{b}} + \frac{\frac{1-b-(\phi_1+\phi_2)}{b} \gamma^{-\frac{1}{b}} g_Q^{\frac{1}{b}}}{(\rho + (z_1 - 1) n + z_2 g_Q)} = \frac{\mu^{\frac{1}{1-\psi}}}{1-\psi} - \frac{\eta - 1}{2\eta} \frac{\phi_1 + \phi_2}{b} \left(g_m^{-1} g_Q^{\frac{1}{b}} \gamma^{-\frac{1}{b}} + \mu^{\frac{1}{1-\psi}} \right). \tag{162}$$

Using (134), (137), and (141), we know that

$$p_d = \frac{\phi_1 + \phi_2}{b} \left(1 - e \frac{\theta - 1}{\theta} \right) \frac{\theta}{\theta - 1} s^{-1} e^{-1}. \tag{163}$$

Combining (131) and (163) gives rise to

$$s^2 = \delta_0^{-1} (\theta - 1) \frac{\eta - 1}{\eta} \frac{\phi_1 + \phi_2}{b} \left(1 - e \frac{\theta - 1}{\theta} \right) \left(e \frac{\theta - 1}{\theta} \right)^{-1}. \tag{164}$$

Using (160), (163), and (164), we obtain the equation pinning dow the steady state value of e:

$$beg_Q \frac{1 + (\theta - 1) \left[1 - \frac{1}{2} \left(1 - \frac{1}{\eta} \right) \right] \frac{\phi_1 + \phi_2}{b} \left(1 - \frac{\theta-1}{\theta} e \right) \left(\frac{\theta-1}{\theta} e \right)^{-1}}{\theta \left[\rho + (z_1 - 1) n + (z_1 + 1) g_Q \right]} = \frac{1 - \frac{\theta-1}{\theta} e}{1 + (z_1 n + z_2 g_Q) g_Q^{-1/b} \gamma^{1/b} \mu^{1/(1-\psi)}}.$$

By substition, we have the steady state expressions for other variables listed in Proposition 1.

Now we prove that $p_D^v(t)$ and $p_D^h(t)$ have the same BGP growth rates, i.e., $g_{p_D^v} = g_{p_D^h}$. Putting (133), (146), and (149) in $L_v \equiv l_v(t) m(t) / l(t)$ in turn leads to

$$L_v = \left(\frac{\phi_1 + \phi_2}{b} \right)^{-1} D(t) p_D^v(t) m(t) l(t)^{-1}. \tag{165}$$

Plugging (137) in $L_m \equiv l_h(t)\dot{m}(t)/l(t)$ in turn gives rise to

$$L_m = \frac{1-\psi}{\psi}\dot{m}(t)D(t)p_D^h(t)l(t)^{-1}. \tag{166}$$

Combining (165) and (166) and using (149) and (158) in turn, we have that

$$p_D^h(t) = \frac{\phi_1+\phi_2}{b}\mu^{\frac{1}{1-\psi}}D(t)^{-\left(1+\frac{\phi_1+\phi_2}{b}\right)}. \tag{167}$$

Taking the time derivatives on both sides of (167) and using (153), we obtain that

$$g_{p_D^h} = -\left(1+\frac{\phi_1+\phi_2}{b}\right)g_D = g_{p_D^v}.$$

Thus, we complete the proof of Proposition 2. □

7.3 Appendix 8.3 Proof of Proposition 3

Proof of Proposition 3. First, we solve the households' problems. Each household's problem is summarized as follows:

$$\max_{\{c(i,t),a(t)\}} E_0\left[\int_{t=0}^{+\infty} e^{-(\rho-n)t}\left(\ln c(t) - \frac{\kappa}{2}d(t)^2\right)dt\right],$$

subject to

$$da(t) = \left[(r(t)-n)a(t) + w(t) - e(t) + \int_{i=0}^{m(t)} p_d(i,t)d(i,t)\right]dt,$$

$$dm(t) = \dot{m}(t)dt,$$

$$dq(i,t) = \begin{cases} \gamma Q(t)d_v(i,t)^{\phi_2}, l_v(i,t)^b d_v(i,t)^{\phi_1}dt \\ 0, 1-l_v(i,t)^b d_v(i,t)^{\phi_1}dt \end{cases},$$

$$c(t) \equiv \left[\int_{i=0}^{m(t)} q(i,t)^{\frac{1}{\theta}} c(i,t)^{\frac{\theta-1}{\theta}}di\right]^{\frac{\theta}{\theta-1}}, e(t) \equiv \int_{i=0}^{m(t)} p(i,t)c(i,t)di.$$

Define the value function $J(t, a(t), m(t), \{q(i,t), i \in [0, m(t)]\})$. The Bellman equation is

written as follows:

$$(\rho-n)J=\max_{\{c(i,t)\}}\left\{\begin{array}{l}\ln c(t)-\frac{\kappa}{2}d(t)^2+J_t+J_a\left[(r(t)-n)a(t)+w(t)+\int_{i=0}^{m(t)}p_d(i,t)d(i,t)-e(t)\right]\\+J_m\dot{m}(t)+\int_{i=0}^{m(t)}l_v(i,t)^b d_v(i,t)^{\phi_1}\left[J\left(q(i,t)+\gamma Q(t)d_v(i,t)^{\phi_2}\right)-J(q(i,t))\right]\end{array}\right\}.$$

The first order condition for $c(i,t)$ is

$$\left(\int_{i=0}^{m(t)}q(i,t)^{\frac{1}{\theta}}c(i,t)^{\frac{\theta-1}{\theta}}di\right)^{-1}q(i,t)^{\frac{1}{\theta}}c(i,t)^{\frac{\theta-1}{\theta}}=J_a p(i,t)c(i,t). \tag{168}$$

Integrating (168) with respect to i gives rise to

$$1=J_a e(t). \tag{169}$$

Taking powers θ and $(\theta-1)$ on both sides of (168) gives us

$$\left[\int_{i=0}^{m(t)}q(i,t)^{\frac{1}{\theta}}c(i,t)^{\frac{\theta-1}{\theta}}di\right]^{-\theta}q(i,t)J_a^{-\theta}p(i,t)^{1-\theta}=c(i,t), \tag{170}$$

$$\left[\int_{i=0}^{m(t)}q(i,t)^{\frac{1}{\theta}}c(i,t)^{\frac{\theta-1}{\theta}}di\right]^{1-\theta}q(i,t)J_a^{1-\theta}p(i,t)^{1-\theta}=q(i,t)^{\frac{1}{\theta}}c(i,t)^{\frac{\theta-1}{\theta}}.$$

Integrating the above second equation leads to

$$\left[\int_{i=0}^{m(t)}q(i,t)^{\frac{1}{\theta}}c(i,t)^{\frac{\theta-1}{\theta}}di\right]^{-\theta}J_a^{1-\theta}\int_{i=0}^{m(t)}q(i,t)p(i,t)^{1-\theta}di=1. \tag{171}$$

Combining (169) and (170) and using (169) and (171) give rise to

$$\left[\int_{i=0}^{m(t)}q(i,t)^{\frac{1}{\theta}}c(i,t)^{\frac{\theta-1}{\theta}}di\right]^{-\theta}q(i,t)e(t)^{\theta}p(i,t)^{-\theta}=c(i,t),$$

$$\left[\int_{i=0}^{m(t)}q(i,t)^{\frac{1}{\theta}}c(i,t)^{\frac{\theta-1}{\theta}}di\right]^{-\theta}e(t)^{\theta-1}\int_{i=0}^{m(t)}q(i,t)p(i,t)^{1-\theta}di=1.$$

Using the above two equations, we know that

$$c(i,t)=\frac{q(i,t)p(i,t)^{-\theta}e(t)}{\int_{i=0}^{m(t)}q(i,t)p(i,t)^{1-\theta}di}. \tag{172}$$

The first order condition with respect to $d(i,t)$ gives us

$$\kappa d(t) = J_a p_d(i,t).$$

Plugging (169) in the above equation leads to

$$p_d(i,t) = \kappa d(t) e(t) = p_d(t), \tag{173}$$

which implies that the equilibrium prices of all original data are the same.

Using the Envelope theorem with respect to $a(t)$, we have that

$$(\rho - n) J_a = \begin{pmatrix} J_{ta} + J_{aa}\dot{a}(t) + J_a(r(t) - n) + J_{ma}\dot{m}(t) + \\ \int_{i=0}^{m(t)} l_v(i,t)^b d_v(i,t)^{\phi_1} \left[J_a\left(q(i,t) + \gamma Q(t) d_v(i,t)^{\phi_2} \right) - J_a(q(i,t)) \right] \end{pmatrix}.$$

Taking the time derivatives wrt t on both sides of $J_a = J_a(t, a(t), m(t), \{q(i,t), i \in [0, m(t)]\})$, we obtain

$$\frac{dJ_a}{dt} = J_{at} + J_{aa}\dot{a}(t) + J_{am}\dot{m}(t) + \int_{i=0}^{m(t)} l_v(i,t)^b d_v(i,t)^{\phi_1} \left[J_a\left(q(i,t) + \gamma Q(t) d_v(i,t)^{\phi_2} \right) - J_a(q(i,t)) \right].$$

Combining the above two equations and using (169) gives rise to

$$g_e(t) \equiv \frac{\dot{e}(t)}{e(t)} = r(t) - \rho. \tag{174}$$

Second, we examine the production of each variety. The owner of variety i solves the following profit-maximizing problem:

$$\max_{\{p(i,t), c(i,t)\}} [p(i,t) - 1] c(i,t) l(t),$$

subject to the market demand for variety i, $c(i,t) l(t)$, where $c(i,t)$ satisfies (172). Putting (172) in the above problem and solving the first-order condition with respect to $p(i,t)$ lead to the monopolistic pricing rule:

$$p(i,t) = \frac{\theta}{\theta - 1} = p(t), \tag{175}$$

which implies that all varieties set the same monopoly price. Plugging (175) in (172) and the

objective function of variety i, we obtain:

$$c(i,t) = \frac{\theta-1}{\theta}\frac{\alpha(i,t)\,e(t)}{m(t)}, \tag{176}$$

$$\pi(i,t) = \frac{1}{\theta}\frac{q(i,t)}{Q(t)}\frac{e(t)\,l(t)}{m(t)} \equiv \frac{1}{\theta}\alpha(i,t)\,\lambda(t), \tag{177}$$

which establish that the discrepancies of both optimal demands and profits among different varieties hinge their relative quality levels.

Third, we address the problems faced by quality-enhancing innovators. Using the production profits and the revenues from selling data, the owner of product line i hires a labor force, $l_v(i,t)$, and purchases the integrated data, $d_v(i,t)$, to improve product quality. Specifically, any variety i solves the following problem:

$$J(\alpha(i,t),t) \equiv \max_{\{l_v,d_v,s,p_d\}} E_t\left[\int_{u=t}^{+\infty} e^{-\int_{\tau=t}^{u} r(\tau)d\tau}\left((1-\delta(s(i,u)))\frac{\alpha(i,u)\,\lambda(u)}{\theta} - l_v(i,u) - p_D^v(u)\,d_v(i,u)\right)du\right],$$

subject to the evolutionary equation of its relative quality

$$d\alpha(i,t) = -\alpha(i,t)\,g_Q(t)\,dt + \frac{1}{Q(t)}dq(i,t), \tag{178}$$

where $\delta(s(i,t)) = 0.5\delta_0 s(i,t)^2$ represents the profit losses of data sharing for variety i. To solve the above problem, we write down the Bellman equation:

$$0 = \max_{\{l_v,d_v,s\}}\left\{\begin{array}{c} e^{-\int_{u=0}^{t} r(u)du}\left((1-\delta(s(i,t)))\frac{\alpha(i,t)\lambda(t)}{\theta} - l_v(i,t) - p_D^v(t)\,d_v(i,t)\right) \\ -J_\alpha(\alpha(i,t),t)\,\alpha(i,t)\,g_Q(t) + J_t(\alpha(i,t),t) + \\ l_v(i,t)^b\,d_v(i,t)^{\phi_1}\left[J\left(\alpha(i,t)+\gamma d_v(i,t)^{\phi_2}\right) - J(\alpha(i,t))\right] \end{array}\right\}.$$

To solve for the BGP, setting $J(\alpha(i),t) = e^{-rt}V(\alpha(i),\lambda(t),p_D^v(t))$ and plugging it in the above equation lead to

$$0 = \max_{\{l_v,d_v,s,p_d\}}\left\{\begin{array}{c} \frac{1-\delta(s(i))}{\theta}\alpha(i)\,\lambda(t) - l_v(i,t) - p_D^v(t)\,d_v(i,t) - (\rho+g_e)\,V \\ +V_\lambda(g_e+n-g_m)\,\lambda(t) + V_{p_D^v}g_{p_D^v}p_D^v(t) - \alpha(i)\,g_Q V_\alpha + \\ l_v(i,t)^b\,d_v(i,t)^{\phi_1}\left[V\left(\alpha(i)+\gamma d_v(i,t)^{\phi_2},\lambda,p_D^v,p_d(i,t)\right) - V(\alpha(i),\lambda,p_D^v,p_d(i,t))\right] \end{array}\right\}. \tag{179}$$

The first order necessary conditions w.r.t $l_v(i,t)$ and $d_v(i,t)$ are:

$$1 = bl_v(i,t)^{b-1} d_v(i,t)^{\phi_1} \left[V\left(\alpha(i) + \gamma d_v(i,t)^{\phi_2}\right) - V(\alpha(i)) \right], \tag{180}$$

$$p_D(t) d_v(i,t) = \left\{ \begin{array}{l} \phi_1 l_v(i,t)^b d_v(i,t)^{\phi_1} \left[V\left(\alpha(i) + \gamma d_v(i,t)^{\phi_2}\right) - V(\alpha(i)) \right] \\ +\gamma\phi_2 l_v(i,t)^b d_v(i,t)^{\phi_1+\phi_2} V_{\alpha+\gamma d_v{}^{\phi_2}}\left(\alpha(i) + \gamma d_v(i,t)^{\phi_2}\right) \end{array} \right\}. \tag{181}$$

Conjecture that the value function has the following form:

$$V(\alpha(i), \lambda(t), p_D^v(t)) = y_1 \alpha(i) \lambda(t) + y_2 p_D^v(t)^{x_1} \lambda(t)^{x_2}.$$

The first order necessary conditions are changed as

$$l_v(i,t) = \left[b\gamma y_1 \lambda(t) d_v(i,t)^{\phi_1+\phi_2} \right]^{\frac{1}{1-b}}, \tag{182}$$

$$p_D^v(t) d_v(i,t) = \frac{\phi_1+\phi_2}{b} \left[b\gamma y_1 \lambda(t) d_v(i,t)^{\phi_1+\phi_2} \right]^{\frac{1}{1-b}} = \frac{\phi_1+\phi_2}{b} l_v(i,t). \tag{183}$$

Substituting (182) and (183) into the Bellman equation (179) and equating the coefficients and exponents, we obtain

$$x_1 = -\frac{\phi_1+\phi_2}{1-b-(\phi_1+\phi_2)}, x_2 = \frac{1}{1-b-(\phi_1+\phi_2)},$$

$$y_1 = \frac{1-\delta(s(i))}{\theta(\rho - n + g_m + g_Q)}, y_2 = \frac{\frac{1-b-(\phi_1+\phi_2)}{b} (\gamma b y_1)^{x_2} \left(\frac{\phi_1+\phi_2}{b}\right)^{-x_1}}{\rho + g_e - x_2 g_\lambda - x_1 g_{p_D^v}}.$$

Fourth, we solve the problem of variety-expanding innovators. Construct the Lagrangian:

$$\mathcal{L} = l_h(t) + p_D^h(t) d_h(t) + \widehat{\vartheta} \left[\mu - d_h(t)^{\psi} l_h(t)^{1-\psi} \right],$$

where $\widehat{\vartheta}$ is the Lagrangian multiplier. The first order necessary conditions are

$$1 = \widehat{\vartheta} d_h(t)^{\psi} (1-\psi) l_h(t)^{-\psi}, p_D^h(t) = \widehat{\vartheta}\psi d_h(t)^{\psi-1} l_h(t)^{1-\psi}.$$

Eliminating $\widehat{\vartheta}$ in the above two equations gives rise to

$$p_D^h(t) = \frac{\psi}{1-\psi}\frac{l_h(t)}{d_h(t)}. \tag{184}$$

Combining (184) with $\mu = d_h(t)^{\psi} l_h(t)^{1-\psi}$, we solve for the demand functions for two inputs:

$$l_h(t) = \mu\left(\frac{1-\psi}{\psi}\right)^{\psi} p_D^h(t)^{\psi}, d_h(t) = \mu\left(\frac{\psi}{1-\psi}\right)^{1-\psi} p_D^h(t)^{\psi-1}. \tag{185}$$

Similar to the situation with firms owning data, we assume that new entrants receive profit flows from all existing varieties affected by creative destruction. Utilizing the linearity of V with respect to α and equation (185), we obtain a zero-profit condition for variety-expanding R&D, specifically,

$$\begin{aligned} V(1,\lambda(t),p_D^v(t)) &= \frac{1}{1-\psi}\mu^{\frac{1}{1-\psi}} d_h(t)^{-\frac{\psi}{1-\psi}} - \frac{\int_{i=0}^{m(t)} \delta(s(i,t))\frac{1}{\theta}\alpha(i,t)\lambda(t)\,di}{\dot{m}(t)} \\ &= \frac{1}{1-\psi} l_h(t) - \frac{\int_{i=0}^{m(t)} \delta(s(i,t))\frac{1}{\theta}\alpha(i,t)\lambda(t)\,di}{\dot{m}(t)} \\ &= \frac{1}{\psi} d_h(t) p_D^h(t) - \frac{\int_{i=0}^{m(t)} \delta(s(i,t))\frac{1}{\theta}\alpha(i,t)\lambda(t)\,di}{\dot{m}(t)}. \end{aligned} \tag{186}$$

Five, a competitive data intermediary sector solves the following optimization problem:

$$\max_{\{d(i,t)\}} \left(p_D^v(t) m(t) + p_D^h(t) \dot{m}(t)\right) D(t) - \int_{i=0}^{m(t)} p_d(i,t) d(i,t) l(t)\,di,$$

subject to the data-integrating technology

$$D(t) = \chi\left[m(t)^{-1/\eta}\int_{i=0}^{m(t)} q(i,t)^{1/\eta} (l(t) d(i,t))^{(\eta-1)/\eta}\,di\right]^{\eta/(\eta-1)}, \eta > 1. \tag{187}$$

Solving the FOC w.r.t $d(i,t)$ leads to the demand function for the original data

$$p_d(i,t) = \left(p_D^v(t) m(t) + p_D^h(t)\dot{m}(t)\right)\chi[\Xi]^{\frac{1}{\eta-1}} m(t)^{-\frac{1}{\eta}} q(i,t)^{\frac{1}{\eta}} (d(i,t) l(t))^{-\frac{1}{\eta}}, \tag{188}$$

where

$$\Xi \equiv m(t)^{-1/\eta}\int_{i=0}^{m(t)} q(i,t)^{1/\eta} (l(t) d(i,t))^{(\eta-1)/\eta}\,di.$$

Plugging $d(i,t) = s(i,t)\, c(i,t)$, (176), and (173) in (188) leads to

$$p_d(i,t) = \left(p_D^v(t)\, m(t) + p_D^h(t)\, \dot{m}(t)\right) \chi\, [\Xi]^{\frac{1}{\eta-1}} \left(\frac{\theta-1}{\theta} l(t)\, s(i,t)\, e(t)\, Q(t)\right)^{-\frac{1}{\eta}} = p_d(t) = \kappa d(t)\, e(t)\,, \tag{189}$$

which implies that $s(i,t)$ is independent of i, namley, $s(i,t) = s(t)$. Thus, y_1, y_2, $l_v(i,t)$, and $d_v(i,t)$ are independent of i. Substituting (183), (186), and (189) into the zero-profit condition of the data intermediary sector gives rise to

$$\kappa d(t)^2\, e(t)\, l(t) = \frac{\theta-1}{\theta} l_v(t)\, m(t) + \frac{\psi}{1-\psi} \dot{m}(t) l_h(t)\,. \tag{190}$$

Six, we impose the market-clearing conditions for both labor and capital markets, namely,

$$l(t) = \int_{i=0}^{m(t)} c(i,t)\, l(t)\, di + \int_{i=0}^{m(t)} l_v(i,t)\, di + \dot{m}(t) l_h(t) = \frac{\theta-1}{\theta} e(t)\, l(t) + l_v(t)\, m(t) + \dot{m}(t) l_h(t)\,, \tag{191}$$

$$a(t)\, l(t) = \int_{i=0}^{m(t)} V\left(\alpha(i), \lambda(t), p_D^v(t)\right) di = m(t)\, V\left(1, \lambda(t), p_D^v(t)\right). \tag{192}$$

Finally, we solve for the stationary equilibrium (or the balanced growth path), in which g_Q, g_m, g_λ, g_D, $g_{p_D^v}$, p_d, r, e, s, d, L_p, L_v, L_m are constant. Given the symmetry in the equilibrium behaviors of all varieties, $dq(i,t)$ and $dQ(t)$ share the same distribution. Specifically,

$$g_Q(t) \equiv \frac{\dot{Q}(t)}{Q(t)} = \gamma l_v(t)^b\, D(t)^{\phi_1+\phi_2}\,. \tag{193}$$

Putting (182) in (193) tells that on the BGP,

$$\lambda(t) = \gamma^{-\frac{1}{b}} b^{-1} g_Q^{\frac{1-b}{b}} y_1^{-1} D(t)^{-\frac{\phi_1+\phi_2}{b}}\,, \tag{194}$$

and thus,

$$g_\lambda = -\frac{\phi_1+\phi_2}{b} g_D. \tag{195}$$

Using the definitions of labor shares and substituting (176), (182) and (194) into them, we know that, on the BGP,

$$L_p = \frac{\theta-1}{\theta} e, L_v = eby_1 g_Q, L_m = 1 - L_p - L_v. \tag{196}$$

Putting (190) in (191) leads to

$$l(t)=\left(\frac{\theta-1}{\theta}+\frac{1-\psi}{\psi}\kappa d^{2}\right)el(t),$$

which implies that on the BGP, the total original data sold by anyone are constant, namely,

$$d=\sqrt{\left(e^{-1}-\frac{\theta-1}{\theta}\right)\frac{\psi}{1-\psi}\kappa^{-1}}. \tag{197}$$

Putting (172) in the definition of $d(t)=\int_{i=0}^{m}d(i,t)\,di$ gives rise to

$$d(t)=\frac{\theta-1}{\theta}s(t)\,e(t). \tag{198}$$

Combining (197) and (198) leads to

$$s(t)=\frac{\theta}{\theta-1}\frac{d}{e}=s. \tag{199}$$

Plugging (172) and (199) in (187) gives us

$$D(t)=\chi se\frac{\theta-1}{\theta}l(t)\,Q(t)^{\frac{1}{\eta-1}}, \tag{200}$$

which implies that on the BGP,

$$g_{D}=n+\frac{1}{\eta-1}g_{Q}. \tag{201}$$

Substituting (194) into (183) leads to

$$p_{D}(t)=\frac{\phi_{1}+\phi_{2}}{b}\gamma^{-\frac{1}{b}}g_{Q}^{\frac{1}{b}}D(t)^{-\left(1+\frac{\phi_{1}+\phi_{2}}{b}\right)}, \tag{202}$$

which implies that on the BGP,

$$g_{P_{D}^{v}}=-\left(1+\frac{\phi_{1}+\phi_{2}}{b}\right)g_{D}=\left(1+\frac{b}{\phi_{1}+\phi_{2}}\right)g_{\lambda}=\left(1+\frac{b}{\phi_{1}+\phi_{2}}\right)(n-g_{m}). \tag{203}$$

Combining (201) and (203) gives rise to

$$g_{m}=\left(1+\frac{b}{\phi_{1}+\phi_{2}}\right)n+\frac{b}{\phi_{1}+\phi_{2}}\frac{1}{\eta-1}g_{Q}\equiv z_{1}n+z_{2}g_{Q}. \tag{204}$$

Putting (183), (194), and $\alpha(i)=1$ in the conjectured value function gives rise to

$$V\left(1,\lambda\left(t\right),p_{D}^{v}\left(t\right)\right)=\gamma^{-\frac{1}{b}}b^{-1}g_{Q}^{\frac{1-b}{b}}D\left(t\right)^{-\frac{\phi_{1}+\phi_{2}}{b}}+\frac{1-b-\left(\phi_{1}+\phi_{2}\right)}{b\left(\rho+g_{e}-x_{2}g_{\lambda}-x_{1}g_{p_{D}^{v}}\right)}\gamma^{-\frac{1}{b}}g_{Q}^{\frac{1}{b}}D\left(t\right)^{-\frac{\phi_{1}+\phi_{2}}{b}}. \tag{205}$$

Combining (186) in (205) and using the knife-edge condition give us

$$\gamma^{-\frac{1}{b}}b^{-1}g_{Q}^{\frac{1-b}{b}}+\frac{1-b-\left(\phi_{1}+\phi_{2}\right)}{b\left(\rho+g_{e}-x_{2}g_{\lambda}-x_{1}g_{p_{D}^{v}}\right)}\gamma^{-\frac{1}{b}}g_{Q}^{\frac{1}{b}}=\frac{\mu^{\frac{1}{1-\psi}}}{1-\psi}-\frac{\delta\left(s\right)}{\theta}{g_{m}}^{-1}\gamma^{-\frac{1}{b}}b^{-1}{y_{1}}^{-1}g_{Q}^{\frac{1-b}{b}}. \tag{206}$$

Using (184) and (185), we know that

$$g_{Q}^{\frac{1}{b}}=\gamma^{\frac{1}{b}}L_{v}l\left(t\right)m\left(t\right)^{-1}D\left(t\right)^{\frac{\phi_{1}+\phi_{2}}{b}},g_{m}=L_{m}l\left(t\right)m\left(t\right)^{-1}\mu^{-\frac{1}{1-\psi}}D\left(t\right)^{\frac{\psi}{1-\psi}}, \tag{207}$$

which together imply that

$$L_{v}=\frac{1-\frac{\theta-1}{\theta}e}{1+g_{Q}^{-\frac{1}{b}}\gamma^{\frac{1}{b}}\mu^{\frac{1}{1-\psi}}\left(z_{1}n+z_{2}g_{Q}\right)}. \tag{208}$$

Combining (196) and (208) leads to

$$eby_{1}g_{Q}=\frac{1-\frac{\theta-1}{\theta}e}{1+g_{Q}^{-\frac{1}{b}}\gamma^{\frac{1}{b}}\mu^{\frac{1}{1-\psi}}\left(z_{1}n+z_{2}g_{Q}\right)}. \tag{209}$$

Putting (203), (204), (197), (199) and (209) in (206) and (209) yields us the equations determining g_Q and e:

$$\left(\frac{1}{bg_{Q}}+\frac{1-b-\left(\phi_{1}+\phi_{2}\right)}{b\left[\rho+\left(z_{1}-1\right)n+z_{2}g_{Q}\right]}\right)\gamma^{-\frac{1}{b}}g_{Q}^{\frac{1}{b}}=\frac{\mu^{\frac{1}{1-\psi}}}{\left(1-\psi\right)}-\frac{\delta_{0}}{2}\left(\frac{\theta e^{-1}}{\theta-1}\right)^{2}\frac{\psi\kappa^{-1}}{1-\psi}\left(\mu^{\frac{1}{1-\psi}}+\frac{\gamma^{-\frac{1}{b}}g_{Q}^{\frac{1}{b}}}{z_{1}n+z_{2}g_{Q}}\right), \tag{210}$$

$$bg_{Q}e\frac{1-\frac{\delta_{0}}{2}\left(\frac{\theta e^{-1}}{\theta-1}\right)^{2}\left(e^{-1}-\frac{\theta-1}{\theta}\right)\frac{\psi\kappa^{-1}}{1-\psi}}{\theta\left[\rho+\left(z_{1}-1\right)n+\left(z_{2}+1\right)g_{Q}\right]}=\frac{\left(1-\frac{\theta-1}{\theta}e\right)}{1+\gamma^{\frac{1}{b}}g_{Q}^{-\frac{1}{b}}\mu^{\frac{1}{1-\psi}}\left(z_{1}n+z_{2}g_{Q}\right)}. \tag{211}$$

By substition, we obtain the steady-state expressions for the remaining variables, as listed in Proposition 3.

Similar to the proof of Proposition 2, we can demonstrate that $p_D^v(t)$ and $p_D^h(t)$ share the same growth rates, i.e., $g_{p_D^v}=g_{p_D^v}=g_{p_D}$. This concludes the proof of Proposition 3. □

7.4 Appendix 8.4: Modelling strategies of growth theories with data

The growth theories involving data in the literature employ different modelling strategies. We summarize these strategies as follows:

Table 7: Modelling strategies for growth theories involving data

Literature	Data modelling strategy
Farboodi & Veldkamp (2019)	$a_{it} = E\left(\theta_t \middle\vert \mathcal{I}_{it}\right), \mathcal{I}_{it} = \left\{\left\{s^a_{i\tau}\right\}_{\tau=0}^{t-1}, \left\{\left\{s_{i\tau m}\right\}_{m=1}^{n_{i\tau}}\right\}_{\tau=1}^{t-1}\right\}$
(Data as prediction)	$s^a_{it-1} = \theta_{t-1} + \epsilon_{ait-1}, s_{it-1m} = \theta_t + \epsilon_{it-1m}$
Jones&Tonetti (2020)	$Y_{it} = D^{\eta}_{it} L_{it}, Y_t = \left(\int_{i=0}^{N_t} Y_{it}^{\frac{\sigma-1}{\sigma}} di\right)^{\frac{\sigma}{\sigma-1}}$
(Data as production)	
Cong et al (2021)	$\dot{N}(t) = \eta N(t)^{\zeta} \left(\varphi(t) L(t)\right)^{\xi} L_R(t)^{1-\xi}$
(Data as variety innovation)	$Y(t) = L_E(t)^{\beta} \int_{\upsilon=0}^{N(t)} x(\upsilon, t)^{1-\beta} d\upsilon$
Cong et al (2022)	$\dot{N}(t) = \varepsilon N(t) L_R(t)^{1-\xi} L_R(t)^{\xi}$
(Data as variety innovation and production)	$Y(\upsilon, t) = L_E(\upsilon, t) D(\upsilon, t)^{\eta}$
	$Y(t) = \left(\int_{\upsilon=0}^{N(t)} Y(\upsilon, t)^{\frac{\gamma-1}{\gamma}} d\upsilon\right)^{\frac{\gamma}{\gamma-1}}$
Our Model	$dq(i,t) = \begin{cases} \gamma Q(t) D(t)^{\phi_2}, l_v(i,t)^b D(t)^{\phi_1} dt \\ 0, 1 - l_v(i,t)^b D(t)^{\phi_1} dt \end{cases}$
(Data as innovation (quality and variety)	$g_Q(t) = \gamma l_v(i,t)^b D(t)^{\phi_1+\phi_2},$
	$\dot{m}(t) = l_m(t) \mu^{-\frac{1}{1-\psi}} D(t)^{\frac{\psi}{1-\psi}}$

7.5 Appendix 8.5: Growth effects of subsidies and taxes

In this appendix, we analyze the growth effects of subsidies and taxes when firms own data. Specifically, we consider the impact of proportional subsidies, $(s^v_l,\ s^v_d)$, on quality-enhancing expenditures, and $\left(s^h_l,\ s^h_d\right)$ on variety-expanding inputs. Suppose the government levies a lump-sum tax on consumers to finance its transfers. Hence, the government's balanced budget constraint is

$$T(t) l(t) = \int_{i=0}^{m(t)} \left[s^v_l l_v(i,t) + s^v_d p^v_D(t) d_v(i,t)\right] di + \dot{m}(t) \left[s^h_l l_h(t) + s^h_d p^h_D(t) d_h(t)\right], \quad (212)$$

where $T(t)$ is the per capita lump-sum tax. Introducing the government into this model alters the representative household's flow budget constraint but leaves its demand function for good i and the Euler equation unchanged. In the objective function for variety i, the quality-enhancing cost $l_v(i,t) + p^v_D(t) d_v(i,t)$ is replaced with the subsidized cost $(1 - s^v_l) l_v(i,t) +$

$(1-s_d^v)\,p_D^v(t)\,d_v(i,t)$. Similarly, in the cost function of variety-expanding innovators, $l_h(t)+d_h(t)\,p_D^h(t)$ is replaced with $\left(1-s_l^h\right)l_h(t)+\left(1-s_d^h\right)d_h(t)\,p_D^h(t)$. Following the same analysis as before, the growth rate of the average quality level in the balanced growth path is implicitly defined by

$$\left[\frac{1-s_l^v}{bg_Q}+\frac{(1-s_l^v)\,(1-b-(\phi_1+\phi_2))}{b\left[\rho+(z_1-1)\,n+z_2g_Q\right]}\right]\gamma^{-\frac{1}{b}}g_Q^{\frac{1}{b}}=\frac{\left(1-s_l^h\right)\mu^{\frac{1}{1-\psi}}}{1-\psi}-\frac{(\eta-1)\,\psi}{2\eta\,(1-\psi)}\left(\frac{\gamma^{-\frac{1}{b}}g_Q^{\frac{1}{b}}}{(z_1n+z_2g_Q)}+\mu^{\frac{1}{1-\psi}}\right). \tag{213}$$

The following result is obvious on inspection of (213):

Proposition A1 *In the balanced growth equilibrium, when firms own data, the growth rate of average quality depends not on the subsidies for data inputs in both innovation activities* $\left(s_d^v,s_d^h\right)$ *, but on the subsidies for labor inputs in these activities* $\left(s_l^v,s_l^h\right)$.

Proof The proof is straightforward and available upon request. □

Proposition A1 establishes that subsidies for data inputs in both innovation activities have no long-run effects on quality growth. This result reflects a knife-edge condition, highlighting a homotheticity property of data's influence on the macroeconomy. Unlike the existing literature, equation (213) shows that both equal and unequal proportionate subsidies for labor inputs in the two innovation activities have persistent effects on growth. Jones (1995) and Young (1998) demonstrate that neither subsidies nor taxes/tariffs affects long-run growth. Dinopoulos and Thompson (1998) argue that only unequal subsidies have long-run effects on quality growth. In our model with big data, data sales induce creative destructions, from which new varieties benefit, introducing a new term in the equation determining g_Q (i.e., the second term in equation (213)). Howitt (1997) also finds subsidies (whether equal or not) have long-run effects in settings with constant returns to labor in quality-improving R&D, whereas in our model, we assume diminishing returns.

To further examine how subsidies for labor inputs in both innnovation activities affect long-run quality growth, we present Figure 11. The figure shows that quality growth increases with subsidies for labor in quality-enhancing innovations, but decreases with subsidies for labor in variety-expanding innovations. Intuitively, higher subsidies for labor in quality-enhancing innovations attract more labor and data towards these activities, thereby boosting quality growth. In constrast, increasing subsidies for labor in variety-expanding innovations diverts labor away from quality-enhancing innovations, thereby dampening quality growth.

[Insert Figure 11 here.]

We now examine the optimal intervention necessary for the equilibirum growth rate to reach the optimal growth rate. By combining equation (213) with the equation determining the optimal growth rate, we derive the following result:

Proposition A2 *The optimal intervention along the balanced growth path satisfies the following formula:*

$$\nu_1 \left(1 - s_l^v\right) + \nu_2 \left(1 - s_l^h\right) = 1, \tag{214}$$

where

$$\begin{aligned}
\nu_1 &\equiv -\frac{\left[\rho + \left(z_1 - 1\right) n + \left(z_2 + 1\right) g_Q - b z_1 g_Q\right]}{b g_Q \left[\rho + \left(z_1 - 1\right) n + z_2 g_Q\right]} v_0 < 0, \\
\nu_2 &\equiv \frac{\left[\left(\rho - n\right) b^{-1} g_Q^{-1} - \left(z_2 + 1\right)\right]}{b g_Q \left[\left(\rho - n\right) + \left(1 + z_2\right)\left(z_1 n + z_2 g_Q\right)\right]} v_0 > 0, \\
\nu_0 &\equiv \frac{2\eta}{\eta - 1} \frac{1 - \psi}{\psi} \frac{\left[\rho - n + \left(1 + z_2\right)\left(z_1 n + z_2 g_Q\right)\right]\left(z_1 n + z_2 g_Q\right)}{\left(\rho - n\right)\left[b^{-1} g_Q^{-1}\left(z_1 n + z_2 g_Q\right) + 1\right]}.
\end{aligned}$$

Proof The proof is straightforward and available upon request. □

References

[1] Aghion, P., Howitt, P., 1998. Endogenous growth theory. Cambridge MA: MIT Press.

[2] Agrawal, A., McHale, J., Oettl, A., 2018. Finding needls in haystacks: artificial intelligence and recombinant growth. In "The Economics of Artificial Intelligence: An Agenda." National Bureau of Economic Research, Inc.

[3] Cockburn, I., Henderson, R., Stern, S., 2018. The impact of artificial intelligence on innovation. In Agrawal, Gans, and Goldfarb Eds. The Economics of Artificial Intelligence: An Agenda. University of Chicago Press.

[4] Cong, L., Wei, W., Xie, D., Zhang, L., 2022. Endogenous growth under multiple uses of data. Journal of Economic Dynamics and Control 141, 104395.

[5] Cong, L., Xie, D., Zhang, L., 2021. Knowledge accumulation, privacy, and growth in a data economy. Management Science 67 (10), 6480-6492.

[6] Dinopoulos, E., Thompson, P., 1998. Schumpeterian growth without scale effects. Journal of Economic Growth 3, 313-335.

[7] Dixit, A.K., Pindyck, R.S., 1994. Investment under uncertainty. Princeton University Press, Princeton, New Jersey.

[8] Farboodi, M., Veldkamp, L., 2019. A model of the data economy. Working paper.

[9] Farboodi, M., Veldkamp, L., 2023. Data and markets. Annual Review of Economics 15, 23-40.

[10] Grossman, G., Helpman, E., 1991. Innovation and growth in the global economy. Cambridge MA: MIT Press.

[11] Howitt, P., 1999. Steady endogenous growth with population and R&D inputs growing. Journal of Political Economy 107(4), 715-730.

[12] Jones, C., 1995. R&D-based models of economic growth. Journal of Political Economy 103, 759-784.

[13] Jones, C., Tonetti, C., 2020. Nonrivalry and the economics of data. American Economic Review 110 (9): 2819-2858.

[14] Kolanovic, M., Krishnamachari, R.T., 2017. Big data and AI strategies: machine learning and alternative data approach to investing. J.P. Morgan Global Quantitative & Derivatives Strategy Report, New York.

[15] Kortum, S., 1997. Research, patenting, and technological change. Econometrica 65, 1389-1419.

[16] Li, C., 2000. Endogenous vs. semi-endogenous growth in a two-R&D-sector model. Economic Journal 110, 109-122.

[17] Martin, K.E., 2020. Ethical issues in the big data industry. In Strategic Information Management: Theory and Practice, ed. RD Galliers, DE Leidner, B Simeonova, pp. 450-471. London: Routledge.

[18] Merton, R., 1971. Optimum consumption and portfolio rules in a continuous-time model. Journal of Economic Theory 3, 373-413.

[19] Nielsen, M., 2012. Reinventing discovery: the new era of networked science. Princeton University Press, Princeton, NJ.

[20] Peretto, P., 1998. Technological change and population growth. Journal of Economic Growth 3, 283-312.

[21] Peretto, P., Smulders, S., 1998. Specialization, knowledge dilution, and scale effects in an IO-based growth model. Mimeo, Duke University.

[22] Romer, P., 1990. Endogenous technological change. Journal of Political Economy 98, 71-102.

[23] Segerstrom, P.S., 1998. Endogenous growth without scale effects. American Economic Review 88, 1290-1310.

[24] Thompson, P., Waldo, D., 1994. Growth and trustified capitalism. Journal of Monetary Economics 34, 445-462.

[25] Thompson, P., 1998. Rationality, rules of thumb, and R&D. Structural Change and Economic Dynamics 10, 321-340.

[26] Young, A., 1998. Growth without scale effects. Journal of Political Economy 106, 41-63.

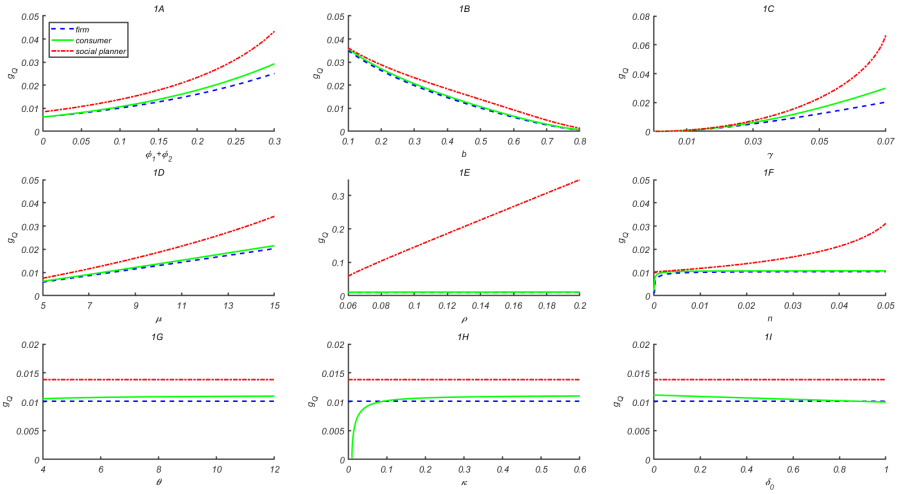

Figure 1: Growth rates under three different allocations. It is shown that the optimal growth rate (the dotted-dashed red line) is larger than the equilibrium growth rate when firms own data (the dashed blue line) and the equilibrium growth rate when consumers own data (the solid green line).

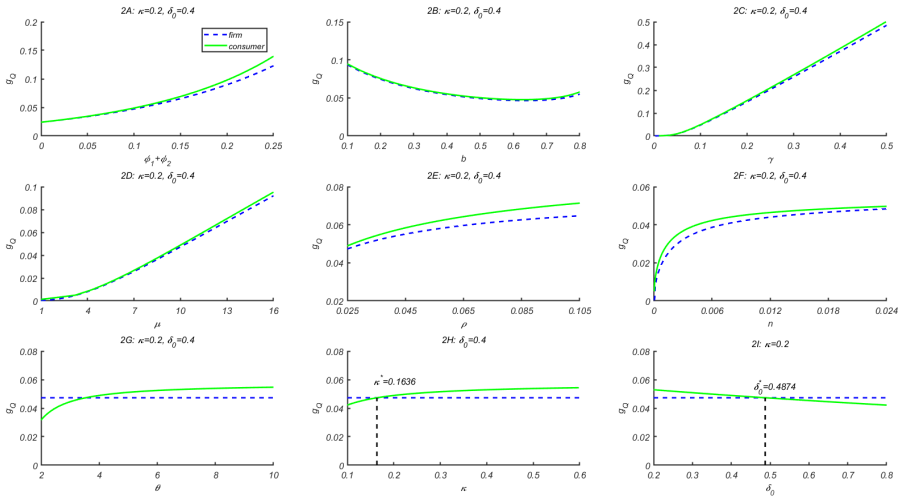

Figure 2: Under benchmark parameter values, if the weight for privacy exceeds its critical value (i.e., $\kappa > \kappa^*$) or the frequency of creative destructon is smaller than its critical value (i.e., $\delta_0 < \delta_0^*$), then quality growth under consumers property rights is larger than that under firm property rights (i.e., $g_Q^c > g_Q^f$).

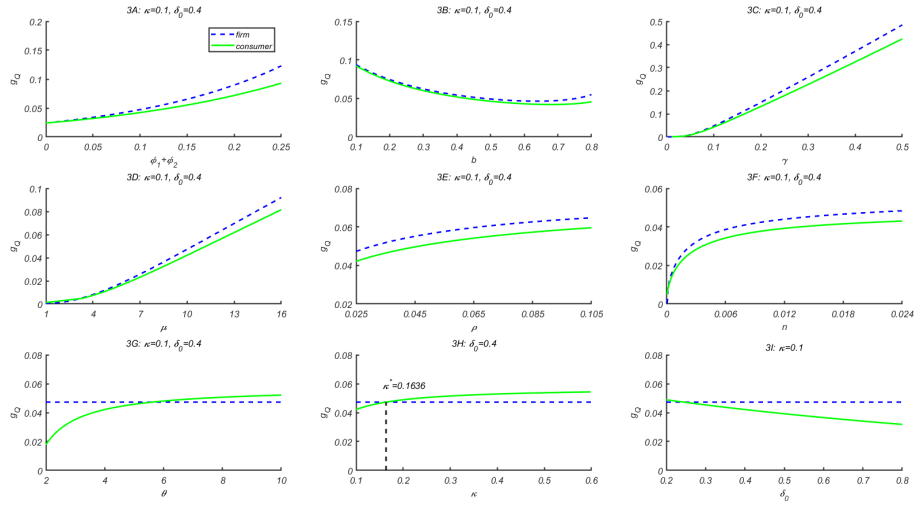

Figure 3: Under benchmark parameter values, if the weight for privacy is smaller than its critical value (i.e., $\kappa < \kappa^*$), then quality growth under consumer property rights is less than that under firms property rights (i.e., $g_Q^c < g_Q^f$).

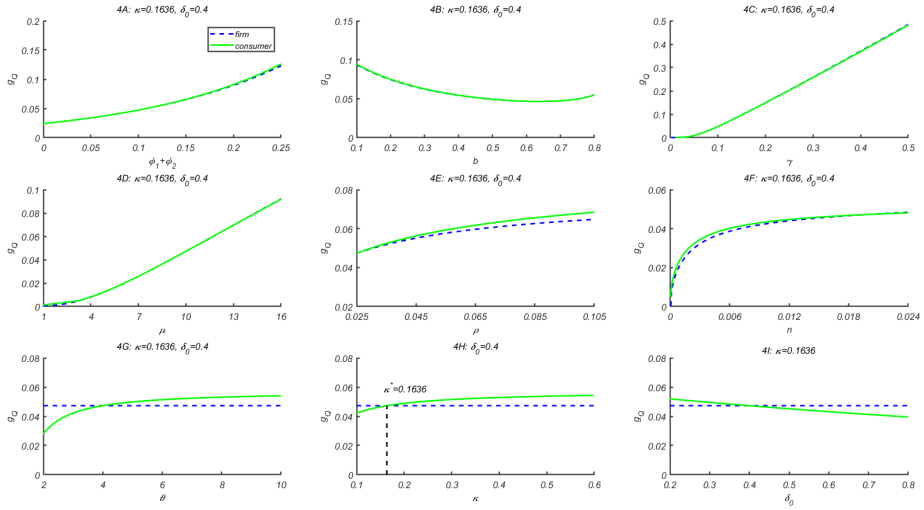

Figure 4: Under benchmark parameter values, if the weight for privacy equals its critical value (i.e., $\kappa = \kappa^*$), then quality growth under the two property right regimes is close to each other (i.e., $g_Q^c \approx g_Q^f$).

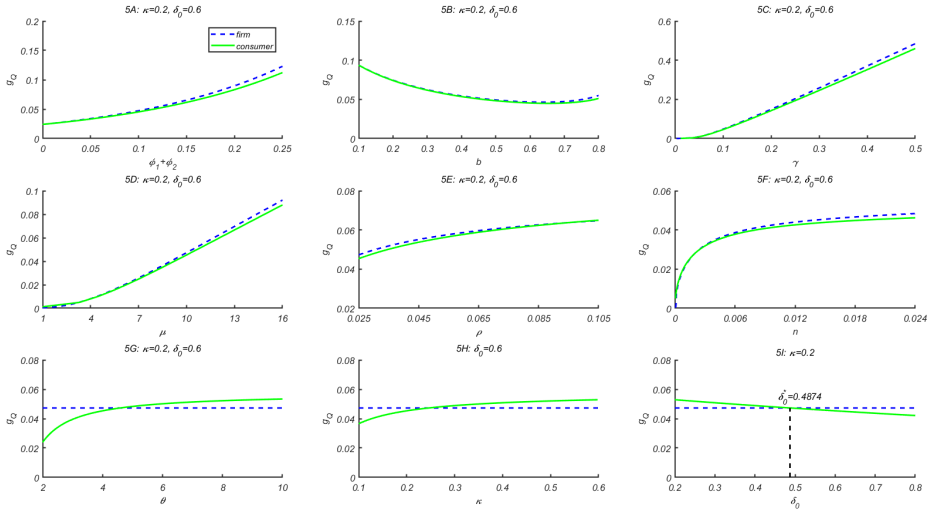

Figure 5: Figure 5. Under benchmark parameter values, if the frequency of creative destructon is smaller than its critical value (i.e., $\delta_0 > \delta_0^*$), then quality growth under consumer property rights is less than that under firm property rights (i.e., $g_Q^c < g_Q^f$).

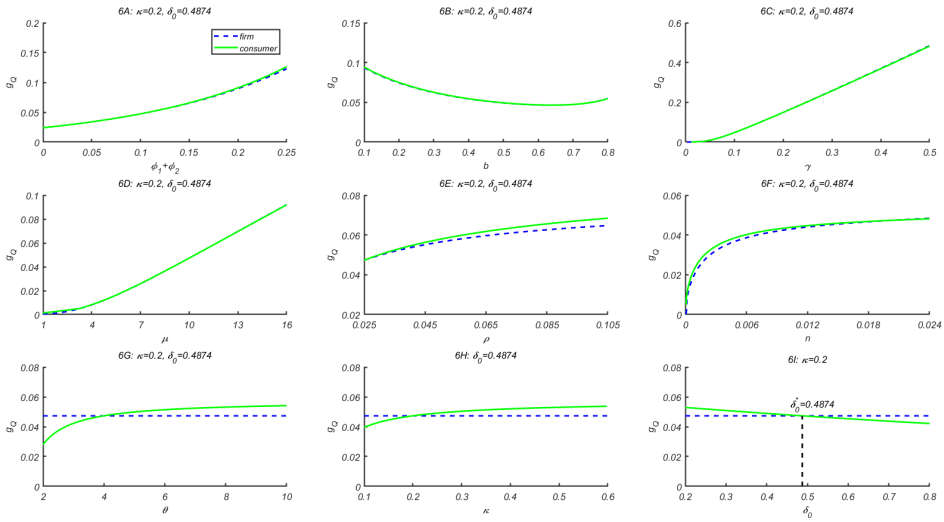

Figure 6: Under benchmark parameter values, if the frequency of creative destructon equals its critical value (i.e., $\delta_0 = \delta_0^*$), then quality growth under the two property right regimes is close to each other (i.e., $g_Q^c \approx g_Q^f$).

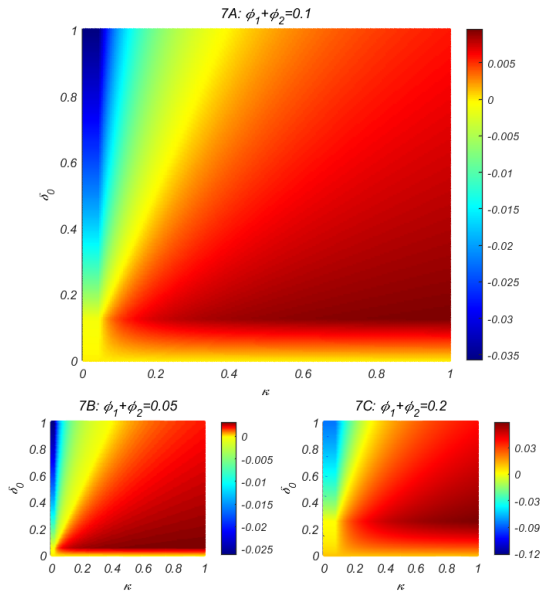

Figure 7: Growth and data property rights. The Consumers Own Data allocation produces higher growth in relatively large parameter ranges, where the difference $g_Q^c - g_Q^f$ is positive. In constrast, the Firms Own Data allocation leads to higher growth in relatively small parameter ranges, where the difference $g_Q^c - g_Q^f$ is negative.

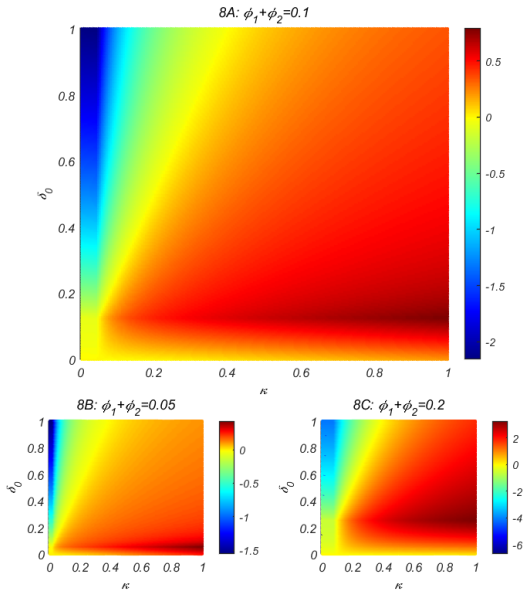

Figure 8: Welfare and data property rights. The Consumers Own Data allocation is superior in relatively large parameter ranges, whereas the Firms Own Data allocation is superior in relatively small parameter ranges.

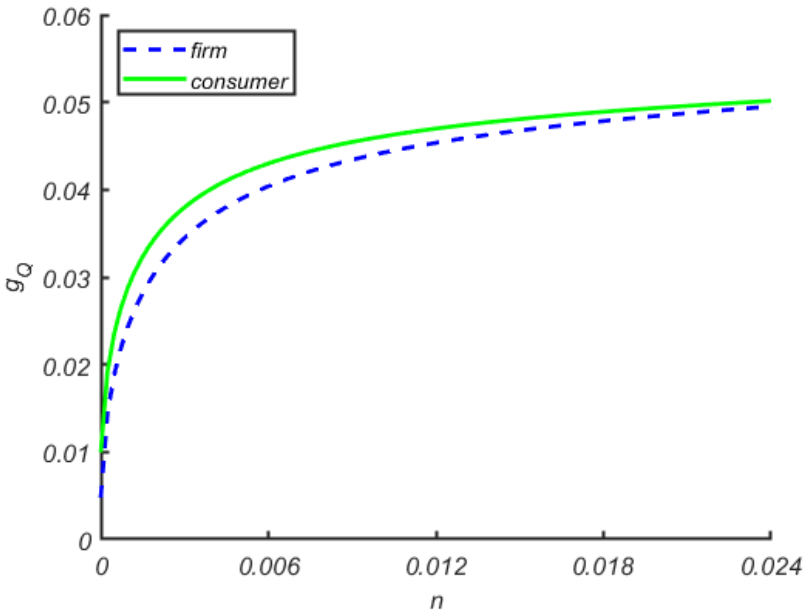

Figure 9: Endogenous growth with population growth. The average quality growth rates under the two data property rights are increasing and a nonlinear funtion of the population growth rate.

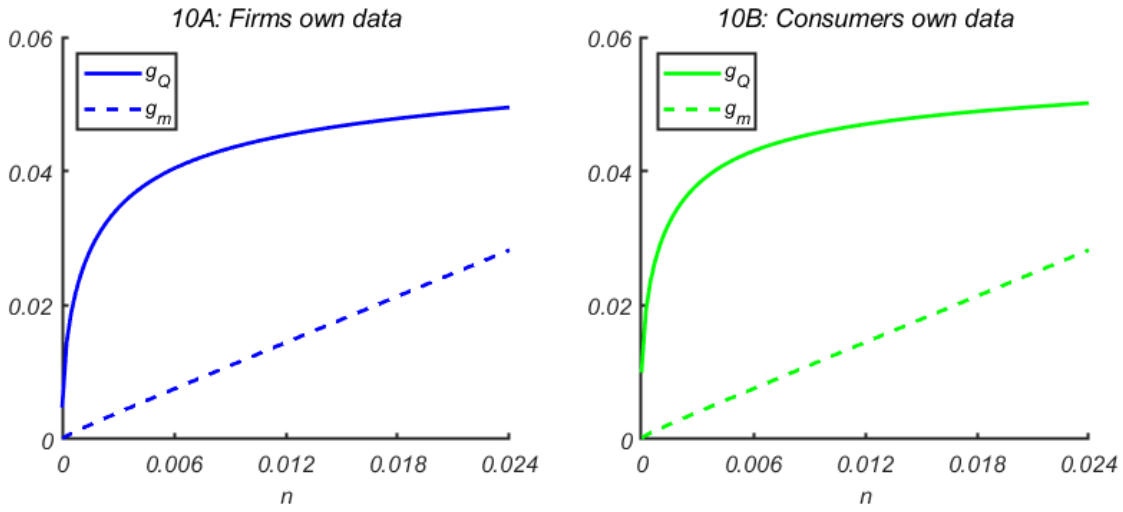

Figure 10: Figure 10: Interactions between quality growth and variety growth. Both quality growth and variety growth increase in population growth. And the direct/linear effect on variety growth of population growth dominates the indirect/nonliear effect.

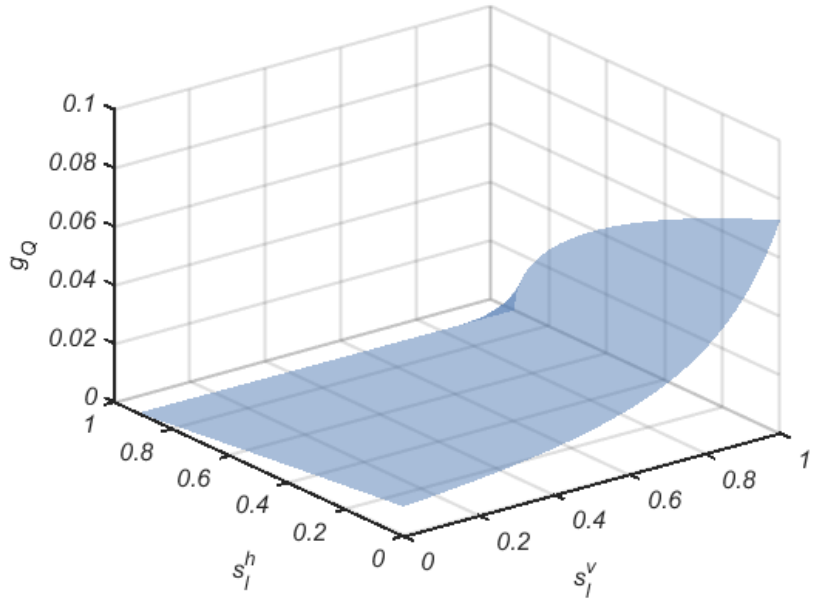

Figure 11: Fiscal policy and growth. The figure shows that quality growth increases with subsidies for labor in quality-enhancing innovations, but decreases with subsidies for labor in variety-expanding innovations.