Digital Transformation, Data Market, and Growth

Sihang Cai

September 10, 2024

Abstract

This paper investigates the relationship between digital transformation and the data market, as well as its implications for growth. In an endogenous growth model with expanding product varieties, data is generated as a by-product of consumption and used as a nonrival factor of production. Firms face a one-time fixed cost for the transformation. Contingent upon the existence of a data market, digital transformation raises growth rates, and there is sustainable digital transformation in the economy. Digital transformation only temporarily crowds out entry when accompanied by a data market, but it does not address insufficient entry resulting from markups. The study also examines an extended model that incorporates firm heterogeneity, suggesting that firms entering the data economy exhibit increased market power relative to the rest of the economy.

1 Introduction

In the wake of extensive progress and rapid advancement in information technology, data has emerged as a key input in various economic activities. It has become a crucial factor driving growth and innovation and reshaping all sectors. However, a significant portion of economic activities still remains to be digitized, and the adaption of data-dependent technologies remains below potential. Report shows that 14% of firms on average have used big data analytics in 2022, only 8% of firms used AI in 2023, while large firms are five and six times more likely to adopt big data analytics and AI (OECD 2024).

The supply of data, despite its vastness, is significantly constrained by challenges associated with collecting, processing, and analyzing large datasets. Concerns also arise regarding the quality and reliability of data. These limitations hinder firms' ability to fully harness the power of data for extracting valuable insights and making informed decisions. Therefore, it becomes crucial to understand underlying tradeoffs, identify effective policies, alleviate numerous obstacles, and foster coordinated efforts for businesses and policymakers to drive further growth.

Various studies have characterized the macroeconomic implications of data used in production in recent years. Literature has focused on how data contributes to technological change and therefore leads to growth, drawing upon learning by doing effects (Arrow 1962) and endogenous growth models (Romer 1990; Grossman and Helpman 1991; Aghion and Howitt 1992; Jones 1995). Data enters into the production function as a nonrival technology and brings increasing returns to scale, so the economy has a long run growth rate that depends on the importance of data (Jones and Tonetti 2020; Cong, Xie, and Zhang 2021; Cong et al. 2022). At the same time, concerns for consumer privacy and business stealing limit the extent to which an economy can benefit from the nonrivalry of data, a setup distinct from the free access to stock of knowledge as in Romer (1990).

This literature often defines data as a by-product of economic activities. In other words, there is no private cost associated with data production and its social cost falls upon other agents in the economy through creative destruction or business stealing. The marginal cost of data generation may be small, but in practice, firms often face significant barriers or friction when they decide to undergo digital transformation and become capable of generating and using data. These barriers can be financial, organizational, or managerial and resemble fixed costs in nature. The omission of such costs will result in a model where the fundamental tradeoff between data generation and final goods production is absent from firms' problems.

In the literature on the economics of data, it has been noted that different purposes of data usage can have distinct implications for economic growth. Specifically, there is data used in research and idea creation which can enhance long-run growth as in the aforementioned studies. On the other hand, there is data used solely for information purposes that improves prediction accuracy, such as business transaction data, which exhibits diminishing returns and no long-term growth effects (Farboodi and Veldkamp 2022; Li and Xu 2024). In particular, Farboodi and Veldkamp (2022) present a data economy model with an emphasis on the latter. Their model also incorporates heterogeneity with two types of firm data productivity and allows for a closer look into transition dynamics. Despite assuming the data productivity parameter as exogenously given, they point out that a more realistic assumption would allow firms to make investments in improving data productivity driven by market incentives.

In order to model a firm's decision on digital transformation when facing fixed costs, I draw upon the literature in monopolistic competition. Dixit and Stiglitz (1977) first parameterize monopolistic competition with a constant elasticity of substitution (CES) function. Melitz (2003), building upon Hopenhayn (1992), develops a dynamic entry model with heterogeneous firms under monopolistic competition in a trade setting. The framework of monopolistic competition is relevant in studying the economics of data, because it allows firms to have the monopoly power over the variety they produce, resulting in endogenous entry decisions and enabling them to earn monopolist profits as compensation for any incurred fixed costs.

In this paper, I examine the economics of data with a specific focus on digital transformation, by utilizing an endogenous growth model that incorporates expanding product varieties. Unlike the literature that assumes the economy has already undergone complete digital transformation, this model takes into account the absence of data production at the outset and allows firms to determine whether or not they should undergo transformation. The model heavily relies on Jones and Tonetti (2020), but it differs in how firms' decisions are characterized when they face a one-time fixed cost. Furthermore, this paper examines the role of a data market in facilitating the digital transformation. The motivation is that data markets and prospects for data aggregation may enhance firms' value after digital transformation, thereby providing a greater incentive for them to bear the fixed costs and embark on this process.

The key findings of the paper show that digital transformation will lead to higher growth rates only in the presence of a data market. Additionally, implementing a data market increases the likelihood of the economy reaching the threshold and undergoing digital transformation, and sustainable digital transformation depends on having a data market. The impact of digital transformation becomes more pronounced with a higher importance of data or a reduced cost of entry.

The two main sources of suboptimality in the equilibrium allocations are insufficient entry resulting from monopolistic markups and insufficient entry due to increasing returns from using data in production, respectively. Digital transformation mainly contributes to the latter, and with a data market, this effect gradually diminishes as the economy grows. When data becomes more important in production, the economy will experience more crowdout of entry through digital transformation if there is no data market.

In the extension, we introduce heterogeneity in firms' ability to use data effectively, allowing for a more detailed examination of firm dynamics. When a data market is present, firms undergoing digital transformation not only have larger sizes but also experience higher rates of return, suggesting that digital transformation can result in increased market power.

The rest of the paper is organized as follows. Section 2 describes the model and derives the equilibrium outcomes in allocations with and without a data market. Section 3 provides a numerical analysis of the model and discuss the implications. Section 4 investigates an extension involving firms with heterogeneous data productivity. Section 5 concludes and discusses implications for polices related to digital transformation and the data economy.

2 The Model

In the basic model, suppose all firms are symmetric and make decisions on whether to pay a one-time fixed cost for digital transformation. If a firm chooses to transform, it will generate data as a by-product and also use data to enhance productivity. Otherwise, firms produce without any involvement of data.

There are two market scenarios to consider. In one case, there is no data market, so each digitally transformed firm can only use data generated by itself. In another case, there is a data market where all data can be aggregated into a bundle, and each firm is able to utilize the entire data bundle. The data bundle is nonrival as in related literature. The specific mechanism of the data market is not addressed in this paper.

For each scenario, I will describe the competitive equilibrium along a balanced growth path and characterize the conditions for digital transformation to occur. Then, I will compare two scenarios to examine how the data market affects digital transformation and growth.

2.1 The Basic Model

Assume the final goods sector to be in monopolistic competition, where each firm produces one variety, and the total output is aggregated with a CES function. The production function for each firm is

$$\begin{aligned} y &= Al \text{ if not digitally transformed,} \\ y' &= \gamma d^{\eta} l' \text{ if digitally transformed,} \end{aligned}$$

where A is the productivity before digital transformation, while γd^{η} is the productivity when producing with data. η is the general importance of data and assume $\eta \in (0,1)$. γ is the productivity associated with data, or "data savviness" as in Farboodi and Veldkamp (2022), which reflects firms ability to use data. d denotes the data accessible to the firm.

Let $d = y'$ without a data market, meaning data is generated as a by-product of the consumption of the variety and is for internal use only. With a data market, data can be shared and aggregated in a CES fashion with elasticity of substitution $\epsilon \in (1, \infty)$. In particular,

$$d = D = \left(\int_0^{\lambda N} y'^{\frac{\epsilon-1}{\epsilon}} di \right)^{\frac{\epsilon}{\epsilon-1}} = \left(\int_0^{\lambda N} (\gamma D^{\eta} l')^{\frac{\epsilon-1}{\epsilon}} \, di \right)^{\frac{\epsilon}{\epsilon-1}}, \tag{1}$$

where N is the total number of varieties and $\lambda \in [0,1]$ denotes the proportion opting for digital transformation. The law of motion for N is

$$\dot{N}_t = \frac{L_{et}}{\chi} N_t = \frac{\bar{L} - L_{pt}}{\chi} N_t$$

where L_e denotes the total labor devoted to create new varieties, L_p represents the total labor used in final goods production, and $\bar{L}$ is the total labor available in the economy. Assume the total population remains constant at $\bar{L}$.

2.2 No Data Market

Consider the scenario without a data market first. Each firm only uses its own data. No data is shared or aggregated, and $d = y'$ in the production function for each firm that opts for digital transformation.

Household Problem. The problem for the representative consumer is

$$U = \max_{\{C_t\}} \int_0^\infty e^{-\rho t} u(C_t) dt,$$

subject to

$$\begin{aligned} C_t &= \left(\int_0^{N_t} c_{it}^{\frac{\sigma-1}{\sigma}} di \right)^{\frac{\sigma}{\sigma-1}}, \\ \dot{a_t} &= r_t a_t + w_t \bar{L} - \int_0^{N_t} p_{it} c_{it} di, \\ \lim_{t\to\infty} & \quad a_t e^{-\int_o^t r_\tau d\tau} = 0, \\ a(0) &= a_0, \end{aligned}$$

where $\sigma \in (1, \infty)$ is the elasticity of substitution, and $u(C_t) = \ln C_t$. Define $P_t = \left(\int_0^{N_t} p_{it}^{1-\sigma} di \right)^{\frac{1}{1-\sigma}}$ as the price index and normalize P_t to be 1 for all t. Solving the utility maximization problem,we obtain the demand function for each variety i

$$c_{it} = \left(\frac{p_{it}}{P_t} \right)^{-\sigma} C_t, \tag{2}$$

and the growth rate for consumption

$$\frac{\dot{C_t}}{C_t} = r_t - \rho. \tag{3}$$

With market clearing for each variety of final goods, $c_{it} = y_{it}$ for all i and t. In the following analysis, we will omit the subscripts i and t for the sake of conciseness.

Firm Problem. Each firm is a monopoly for the variety it produces.

The decision making process is as follows. At the beginning of each period, firms decide whether to undergo digital transformation. If so, the firms pay the one-time fixed cost $f > 0$ and transform. The production takes place afterwards.

From the consumer problem we have the following inverse demand function for each

variety

$$p = \left(\frac{y}{Y}\right)^{-\frac{1}{\sigma}}, \tag{4}$$

where Y denotes the total output. At each time t, the profit maximization for each firm thus becomes

$$\begin{aligned} \pi &= \max_{l} py - wl = \max_{l} Y^{\frac{1}{\sigma}} y^{\frac{\sigma-1}{\sigma}} - wl \text{ if not d.t.,} \\ \pi' &= \max_{l'} p'y' - wl' = \max_{l'} Y^{\frac{1}{\sigma}} y'^{\frac{\sigma-1}{\sigma}} - wl' \text{ if d.t.,} \end{aligned} \tag{5}$$

where $y' = \gamma d^{\eta} l' = (\gamma l')^{\frac{1}{1-\eta}}$, demonstrating increasing returns associated with data following digital transformation. Assume the total number of varieties N is large, so each firm takes N and the aggregate output Y as given.

From the first order conditions, the equilibrium wage rate w satisfies

$$\begin{aligned} w &= \frac{\sigma-1}{\sigma} Y^{\frac{1}{\sigma}} A^{\frac{\sigma-1}{\sigma}} l^{-\frac{1}{\sigma}}, \\ &= \frac{\sigma-1}{\sigma(1-\eta)} Y^{\frac{1}{\sigma}} \gamma^{\frac{\sigma-1}{\sigma(1-\eta)}} l'^{\frac{\sigma\eta-1}{\sigma(1-\eta)}}. \end{aligned} \tag{6}$$

The second equality comes from the competitive labor market.

Assume $\sigma\eta < 1$, the positive monopolist profit flow for each firm is

$$\begin{aligned} \pi &= \frac{1}{\sigma-1} wl, \\ \pi' &= \frac{1-\sigma\eta}{\sigma-1} wl'. \end{aligned} \tag{7}$$

The value of each firm at time t is therefore

$$\begin{aligned} V_t &= \int_t^{t'} e^{-\int_t^\tau r_s ds} \pi_\tau d\tau + \int_{t'}^{\infty} e^{-\int_t^\tau r_s ds} \pi'_\tau d\tau - e^{-\int_t^{t'} r_s ds} f, \\ V'_t &= \int_t^{\infty} e^{-\int_t^\tau r_s ds} \pi'_\tau d\tau - f, \end{aligned} \tag{8}$$

where f is the one-time fixed cost a firm has to invest when it decides to undergo digital transformation. The first equation demonstrates the value of firms that choose not to change today, but they may transform later at $t' > t$. The value functions can also be written in the form of Hamilton-Jacobi-Bellman equations, which give us no arbitrage conditions

$$\begin{aligned} rV &= \dot{V} + \pi, \\ r(V' + f) &= \dot{V}' + \pi'. \end{aligned} \tag{9}$$

When $V' > V$, a firm would achieve a higher value by investing in digital transformation. Meanwhile, there should be no profitable one-shot deviation that would delay a firm's transformation when all other firms are transforming. This means $\int_t^{t+\varepsilon} e^{-\int_t^\tau r_s ds}\pi'_\tau d\tau - f > \int_t^{t+\varepsilon} e^{-\int_t^\tau r_s ds}\pi_\tau d\tau - e^{-\int_t^{t+\varepsilon} r_s ds} f$ as $\varepsilon \to 0$, i.e., $\frac{1}{r}(\pi' - \pi) > f$. Therefore, in equilibrium, a firm will undergo digital transformation only when both inequalities are satisfied. The characteristics of the threshold will be discussed in more detail later.

The free entry conditions of the monopolistic competition model require firms to have intertemporal zero profit, that is,

$$\begin{aligned} (1-\lambda)V\dot{N} &= (1-\nu)wL_e, \\ \lambda V'\dot{N} &= \nu wL_e, \end{aligned}$$

where $\nu \in [0,1]$ denotes the share of production labor employed by firms opting for digital transformation. With symmetry, all firms would make the same choice: either all or none would undergo digital transformation, i.e. $\lambda = 0$ *or* 1 (see Appendix A Lemma 6 for the proof).

The free entry conditions become

$$\begin{aligned} V\dot{N} &= wL_e, \\ V'\dot{N} &= w'L'_e \end{aligned} \tag{10}$$

Competitive Equilibrium. An equilibrium in this economy consists of quantities $\{C_t, c_{it}, a_t\}$ that solve the household problem, $\{y_{it}, p_{it}, l_{it}, d_{it,}\}$ that solve the firm problem, $\{N_t, L_{pt}, L_{et}\}$ as determined by free entry, and factor prices $\{r_t, w_t\}$ that are consistent with asset market clearing $a_t = \int_0^{N_t} V_{it} di$ and labor market clearing $\bar{L} = L_{pt} + L_{et} = \int_0^{N_t} l_{it} di + L_{et}$. In addition, the equilibrium satisfies final goods market clearing $Y_t = C_t$.

The equilibrium wage rate is determined by free entry conditions and the law of motion for N,

$$\begin{aligned} w &= \frac{VN}{\chi}, \\ w' &= \frac{V'N'}{\chi}, \end{aligned} \tag{11}$$

which implies $g_w = g_V + g_N$ or $g_{w'} = g_{V'} + g_{N'}$ along the balanced growth path.

For the equilibrium where $\lambda = 0$, there is no digital transformation in the economy, so $l = \frac{L_p}{N}$. The aggregate output Y is given by

$$Y = N^{\frac{\sigma}{\sigma-1}} Al = N^{\frac{1}{\sigma-1}} AL_p, \tag{12}$$

which implies $g_Y = \frac{1}{\sigma-1} g_N$. According to the first no arbitrage condition in (9), $\frac{\dot{V}}{V} + \frac{\pi}{V} = r$ and $g_w - g_N + \frac{1}{\sigma-1}\frac{L_p}{\chi} = r$. Recall that $g_Y = g_C = r - \rho$, and $g_N = \frac{L_e}{\chi}$. In addition, it can be proved that $g_w = g_Y$.

Along a balanced growth path with constant interest rate, and with the assumption $\bar{L} > (\sigma-1)\rho\chi$, the amount of labor employed in the creation of new varieties and the growth rate are given by

$$L_e^0 = \frac{\bar{L} - (\sigma-1)\rho\chi}{\sigma}, \tag{13}$$

$$g_Y^0 = \frac{\bar{L} - (\sigma-1)\rho\chi}{(\sigma-1)\sigma\chi}, \tag{14}$$

when there is no digital transformation.

For the equilibrium where $\lambda = 1$, all firms undergo digital transformation and $l' = \frac{L'_p}{N'}$. The aggregate output Y' is given by

$$Y' = N'^{\frac{\sigma}{\sigma-1}} (\gamma l')^{\frac{1}{1-\eta}} = N'^{\frac{1-\sigma\eta}{(\sigma-1)(1-\eta)}} (\gamma L'_p)^{\frac{1}{1-\eta}}, \tag{15}$$

which implies $g_{Y'} = \frac{1-\sigma\eta}{(\sigma-1)(1-\eta)} g_{N'}$. According to the second no arbitrage condition in (9), $\frac{\dot{V}'}{V'} + \frac{\pi'}{V'} = r' + \frac{r'f}{V'}$, which implies $g_{w'} - g_{N'} + \frac{1-\sigma\eta}{\sigma-1}\frac{L'_p}{\chi} = r' + \frac{r'f}{V'}$. Again, it can be shown that $g_{w'} = g_{Y'} = r' - \rho$ and $g_{N'} = \frac{L'_e}{\chi}$. The amount of labor employed in the creation of new varieties and the balanced growth rate are given as

$$L_e^{0\prime} = \frac{(1-\sigma\eta)\bar{L} - (\sigma-1)(\rho + \frac{r'f}{V'})\chi}{\sigma(1-\eta)}, \tag{16}$$

$$g_{Y'}^0 = \frac{(1-\sigma\eta)^2\bar{L} - (1-\sigma\eta)(\sigma-1)(\rho + \frac{r'f}{V'})\chi}{\sigma(\sigma-1)(1-\eta)^2\chi}, \tag{17}$$

with the assumption that $(1-\sigma\eta)\bar{L}$ is sufficiently large to ensure the existence of positive growth.

Proposition 1. (Equilibrium without a Data Market) *Along a balanced growth path, as N'_t grows large, there will be digital transformation only if $\sigma + \eta < 2$ and $(\chi - \frac{(1-\eta)^2(2-\sigma)}{(2-\sigma-\eta)})(\bar{L} - (\sigma-1)\rho\chi) - \sigma\eta\bar{L} > 0$. As a result, the economy will face slower growth. Otherwise, there will be no digital transformation in the long run.*

Proof. Recall that all firms in the economy undergo digital transformation when $V' > V$ and $\frac{1}{r'}(\pi' - \pi) > f$. From no arbitrage conditions, $V' > V$ is equivalent to $\frac{1}{r'}(\dot{V}' + \pi') - \frac{1}{r}(\dot{V} + \pi) - f > 0$, where $\dot{V}' = g_{V'}V'$ and $\dot{V} = g_V V$. The threshold for digital transformation is given by $\frac{1}{r'}(\pi' - \pi)f > 0$ if $\frac{\dot{V}'}{r'} - \frac{\dot{V}}{r} + \frac{\pi}{r'} - \frac{\pi}{r} > 0$, otherwise it follows $V' - V > 0$. If $g_{V'} > g_V$, V' grows faster than V does, the threshold converges to $\frac{1}{r'}(\pi' - \pi)f > 0$ as the economy grows,

and the economy will inevitably undergo digital transformation. If $g_{V'} < g_V$, the threshold converges to $V' - V > 0$, and in the long term there will be no digital transformation.

Since $g_{V'} = g_{Y'} - g_{N'} = \frac{2-\sigma-\eta}{(\sigma-1)(1-\eta)} \frac{L_e^{0\prime}}{\chi}$ and $g_V = g_Y - g_N = \frac{2-\sigma}{\sigma-1} \frac{L_e^0}{\chi}$, $g_{V'} > g_V$ holds when $(\chi - \frac{(1-\eta)^2(2-\sigma)}{(2-\sigma-\eta)})(\bar{L} - (\sigma-1)\rho\chi) - \sigma\eta\bar{L} > (\sigma-1)\chi^2 \frac{r'f}{V'}$.

If $\sigma+\eta < 2$, then $g_{V'} > 0$, $V' \to \infty$, and $\frac{r'f}{V'} \to 0$. $g_{V'} > g_V$ will hold if $(\chi - \frac{(1-\eta)^2(2-\sigma)}{(2-\sigma-\eta)})(\bar{L} - (\sigma-1)\rho\chi) - \sigma\eta\bar{L} > 0$. As N_t' becomes large, equations (16) and (17) converges to

$$L_e^{0\prime} \to \frac{(1-\sigma\eta)\bar{L} - (\sigma-1)\rho\chi}{\sigma(1-\eta)}, \tag{18}$$

$$g_{Y'}^0 \to \frac{(1-\sigma\eta)^2\bar{L} - (1-\sigma\eta)(\sigma-1)\rho\chi}{\sigma(\sigma-1)(1-\eta)^2\chi}. \tag{19}$$

Compared with outcomes if the economy stays in the status quo ((13) and (14)), $L_e^{0\prime} \geq L_e^0$ only if $\eta(\bar{L} + \rho\chi) + \frac{\chi r'f}{V'} < 0$, which contradicts the model assumptions. Hence, $L_e^{0\prime} < L_e^0$. Similarly, $g_{Y'}^0 \geq g_Y^0$ only if $\bar{L} + \rho\chi + c \leq 0$, where $c = \frac{(1-\sigma\eta)\chi r'f}{(2-\sigma\eta-\eta)\eta V'} > 0$. Again this contradicts the model assumptions. Hence, $g_Y^{0\prime} < g_Y^0$.Therefore, without first implementing a data market, the economy will face slower growth as a result of digital transformation at the end of the day.

If $\sigma + \eta \geq 2$, then $g_{V'} \leq 0$, $V' \to 0$, $\frac{r'f}{V'} \to \infty$, there will inevitably be a point where $g_{V'} < g_V$. Ultimately, this leads to $V' < V$ and the absence of data in the economy. □

The proposition above shows that without a data market, for a wide range of parameter values, digital transformation cannot sustain in the economy. Even in cases where firms have incentives for digital transformation, the overall economic outcomes will be inferior to the status quo. If there is digital transformation, $L_e^{0\prime} < L_e^0$ suggests a crowd-out of labor from new firm entry due to increasing returns from data used in production, while $g_Y^{0\prime} < g_Y^0$ implies that the economy would face a slower growth rate if firms choose to pursue digital transformation, indicating the potential risk of economies falling into a growth trap in the absence of a data market.

Threshold for Digital Transformation in Transitional Dynamics. The threshold for digital transformation in equilibrium can be characterized by $\underline{V}^0$, a function of V' and V', where $V' = \frac{\sigma-1}{\sigma(1-\eta)} \chi \gamma^{\frac{\sigma-1}{\sigma(1-\eta)}} Y^{\frac{1}{\sigma}} N'^{\frac{1-\sigma}{\sigma(1-\eta)}} L_p'^{\frac{\sigma\eta-1}{\sigma(1-\eta)}}$ and $V = \frac{\sigma-1}{\sigma} \chi A^{\frac{\sigma-1}{\sigma}} Y^{\frac{1}{\sigma}} N^{\frac{1-\sigma}{\sigma}} L_p^{-\frac{1}{\sigma}}$. The economy will undergo digital transformation whenever $\underline{V}^0 > 0$.

From the Firm Problem and free entry conditions, $\pi' = \frac{1-\sigma\eta}{(\sigma-1)\chi}(\bar{L} - L_e^{0\prime})V'$ and $\pi = \frac{1}{(\sigma-1)\chi}(\bar{L} - L_e^0)V$. Moreover, $r' = \frac{(1-\sigma\eta)^2\bar{L} + (\sigma(1-\eta)^2 + (1-\sigma\eta))(\sigma-1)\rho\chi}{(\sigma-1)\chi} \cdot \frac{V'}{\sigma(1-\eta)^2V' + (1-\sigma\eta)f}$ and $r = \frac{\bar{L} + (\sigma-1)^2\rho\chi}{\sigma(\sigma-1)\chi}$. Thus, $\frac{1}{r'}(\pi' - \pi) > f$ is equivalent to $\sigma(1-\eta)(\bar{L} + \rho\chi)(V' - V - \eta(\sigma V' - V))V' + f(\sigma(2-\sigma-\eta)\rho\chi V' - (1-\sigma\eta)(\bar{L} + \rho\chi)V) > 0$.

Therefore, when there is no data market, the threshold for digital transformation is given by

$$\underline{V}^0 = \begin{cases} V' - V & \text{if } \frac{\dot{V}'}{r'} - \frac{\dot{V}}{r} + \frac{\pi}{r'} - \frac{\pi}{r} < 0 \\ \sigma(1-\eta)(\bar{L}+\rho\chi)(V'-V-\eta(\sigma V'-V))V' + & \\ f(\sigma(2-\sigma-\eta)\rho\chi V' - (1-\sigma\eta)(\bar{L}+\rho\chi)V) & \text{o.w.} \end{cases}. \tag{20}$$

2.3 With Data Market

Now consider the scenario with a data market, where all data can be shared and aggregated into a nonrival bundle denoted as $d = D$. The production function for a firm after digital transformation becomes $y' = \gamma D^{\eta} l'$, with D defined by equation (1).

Household Problem. The household problem remains unchanged as in the case without a data market.

Firm Problem. According to the first order conditions from profit maximization problems, the equilibrium wage rate satisfies

$$\begin{aligned} w &= \frac{\sigma-1}{\sigma} Y^{\frac{1}{\sigma}} A^{\frac{\sigma-1}{\sigma}} l^{-\frac{1}{\sigma}}, \\ &= \frac{\sigma-1}{\sigma} Y^{\frac{1}{\sigma}} \left(\gamma D^{\eta}\right)^{\frac{\sigma-1}{\sigma}} l'^{-\frac{1}{\sigma}}. \end{aligned} \tag{21}$$

The positive monopolist profit flow for each firm is now

$$\begin{aligned} \pi &= \frac{1}{\sigma-1} wl, \\ \pi' &= \frac{1}{\sigma-1} wl'. \end{aligned} \tag{22}$$

The value of each firm and the no arbitrary conditions stay the same as

$$V_t = \int_t^{t'} e^{-\int_t^{\tau} r_s ds} \pi_\tau d\tau + \int_{t'}^{\infty} e^{-\int_t^{\tau} r_s ds} \pi'_\tau d\tau - e^{-\int_t^{t'} r_s ds} f, \tag{23}$$

$$V'_t = \int_t^{\infty} e^{-\int_t^{\tau} r_s ds} \pi'_\tau d\tau - f, \tag{24}$$

and

$$\begin{aligned} rV &= \dot{V} + \pi, \\ r(V' + f) &= \dot{V}' + \pi'. \end{aligned} \tag{25}$$

Just like before, an equilibrium with digital transformation exists only when both $V' > V$ and $\frac{1}{r'}(\pi' - \pi) > f$ are satisfied.

The free entry conditions of the monopolistic competition model require firms to have intertemporal zero profit. As in the previous case, either all or none of the firms would undergo digital transformation, i.e. $\lambda = 0$ *or* 1 (Appendix A Lemma 6). The free entry conditions become

$$\begin{aligned} V\dot{N} &= wL_e, \\ V'\dot{N} &= w'L'_e \end{aligned} \tag{26}$$

Competitive Equilibrium. An equilibrium in this economy consists of quantities $\{C_t, c_{it}, a_t\}$ that solve the household problem, $\{y_{it}, p_{it}, l_{it}, D_{t,}\}$ that solve the firm problem, $\{N_t, L_{pt}, L_{et}\}$ as determined by free entry, and $\{r_t, w_t\}$ that are consistent with asset market clearing $a_t = \int_0^{N_t} V_{it} di$ and labor market clearing $\bar{L} = L_{pt} + L_{et} = \int_0^{N_t} l_{it} di + L_{et}$. In addition, the equilibrium satisfies final goods market clearing $Y_t = C_t$ and data market clearing $D_t = \left(\int_0^{N_t} d_{it}^{\frac{\epsilon-1}{\epsilon}} di\right)^{\frac{\epsilon}{\epsilon-1}}$.

Proposition 2. (Equilibrium with a Data Market) *Along a balanced growth path, as N'_t grows large, the economy will undergo digital transformation and experience a higher growth rate, except in cases of coordination failure.*

Proof. Notice that, with the data market, there exists an equilibrium where all firms produce without using data, as the status quo. There will be no digital transformation whatsoever, i.e. $\lambda = 0$, since no firm can profit from deviation, as the aggregate bundle D is always 0 in the data market. This equilibrium can be interpreted as a case of coordination failure.

In the equilibrium where $\lambda = 0$, there is no digital transformation in the economy. The labor allocation and growth rate along the balanced growth path remain unchanged compared to the previous scenario

$$L_e^1 = \frac{\bar{L} - (\sigma - 1)\rho\chi}{\sigma}, \tag{27}$$

$$g_Y^1 = \frac{\bar{L} - (\sigma - 1)\rho\chi}{(\sigma - 1)\sigma\chi}. \tag{28}$$

In the equilibrium with $\lambda = 1$, all firms undergo digital transformation and $l' = \frac{L'_p}{N'}$. The

aggregate data bundle D is given by

$$D = N'^{\frac{\epsilon}{\epsilon-1}}\gamma D^{\eta} l' = N'^{\frac{1}{\epsilon-1}\frac{1}{1-\eta}}(\gamma L'_p)^{\frac{1}{1-\eta}}, \tag{29}$$

which implies $g_D = \frac{1}{\epsilon-1}\frac{1}{1-\eta}g_{N'}$. The aggregate output Y' is given by

$$Y' = N'^{\frac{\sigma}{\sigma-1}}\gamma D^{\eta} l' = N'^{\frac{1}{\sigma-1}}\gamma D^{\eta} L'_p, \tag{30}$$

which implies $g_{Y'} = \frac{1}{\sigma-1}g_{N'} + \eta g_D$. According to the second no arbitrage condition in (25), $\frac{\dot{V'}}{V'} + \frac{\pi'}{V'} = r' + \frac{r'f}{V'}$, which implies $g_{w'} - g_{N'} + \frac{1}{\sigma-1}\frac{L'_p}{\chi} = r' + \frac{r'f}{V'}$. Again, it can be shown that $g_{w'} = g_{Y'}$. The amount of labor employed in the creation of new varieties and the balanced growth rate are given as

$$L_e^{1\prime} = \frac{\bar{L} - (\sigma-1)(\rho + \frac{r'f}{V'})\chi}{\sigma}, \tag{31}$$

$$g^1_{Y'} = \left(\frac{1}{\sigma-1} + \frac{1}{\epsilon-1}\frac{\eta}{1-\eta}\right)\frac{\bar{L} - (\sigma-1)(\rho + \frac{r'f}{V'})\chi}{\sigma\chi}. \tag{32}$$

Consider the case with $\epsilon = \frac{1}{1-\eta}$ as an example, we have $g^1_{Y'} = \frac{\sigma}{\sigma-1}\frac{L_e^{1\prime}}{\chi}$, which implies that $g^1_{Y'} > g^1_Y$ when $\frac{r'f}{V'} < \frac{\bar{L}-(\sigma-1)\rho\chi}{\sigma}$. In addition, there is $g_{V'} > 0$, and $\lim_{t\to\infty}\frac{r'f}{V'} = 0$.

As a result, as N'_t grows large, the equilibrium allocation with a data market under digital transformation converges to

$$L_e^{1\prime} \to \frac{\bar{L} - (\sigma-1)\rho\chi}{\sigma}, \tag{33}$$

$$g^1_{Y'} \to \frac{\bar{L} - (\sigma-1)\rho\chi}{(\sigma-1)\chi}. \tag{34}$$

The results indicate that $L_e^{1\prime} \leq L_e^1$, suggesting a crowd-out of labor from entry as the use of data increases productivity and firm sizes, although the crowd-out effect gradually diminishes as the economy grows. Also, $g^1_{Y'} > g^1_Y$ when $V' > \frac{(\bar{L}+(\sigma-1)^2\rho\chi)f}{(\sigma-1)\bar{L}+(\sigma-1)^2\rho\chi}$, indicating that, with a data market, the economy will experience faster growth following digital transformation as the economy grows.

Nonetheless, note that for $\epsilon \to \infty$, the aggregation of data becomes $D = (\gamma L_p)^{\frac{1}{1-\eta}}$, which remains constant and $g_D = 0$ on the balanced growth path. In other words, when data generated from different varieties are perfect substitutes and the data aggregation is linear, data has no contribution to the growth rate and $g^{1\prime}_Y < g^1_Y$. In this particular case, if digital transformation occurs, it will hinder entry and harm economic growth as a whole. □

From BGP outcomes, we can observe that productivity parameters A, γ, and the fixed cost f do not manifest in either the growth rate or labor allocation. However, they do affect

the level of output and, more relevantly, the timing of digital transformation. For a given initial state of an economy, digital transformation occurs sooner if A is smaller, γ is larger, or f is lower.

Comparing equations ((33) and (34)) with the ones in the scenario of no data market ((18) and (19)), we observe the following. After digital transformation, the equilibrium share of labor devoted to the creation of new varieties is lower in the economy without a data market by $\sigma\eta\bar{L}$. The long-term growth rate following digital transformation is generally higher with a data market compared to one without. Overall, a data market allows for a positive economic impact resulting from digital transformation.

Threshold for Digital Transformation in Transitional Dynamics. Define the threshold with a data market as $\underline{V}^1$. The economy will undergo digital transformation whenever $\underline{V}^1 > 0$.

In the equilibrium where firms always choose to remain in the status quo as a result of coordination failure, $\underline{V}^1$ can be defined as 0 and there will be no digital transformation.

In the equilibrium where firms undergo digital transformation, $V' - V > 0$ and $\frac{1}{r'}(\pi' - \pi) - f > 0$, where $V = \frac{\sigma-1}{\sigma}\chi A^{\frac{\sigma-1}{\sigma}} Y^{\frac{1}{\sigma}} N^{\frac{1-\sigma}{\sigma}} L_p^{-\frac{1}{\sigma}}$, $V' = \frac{\sigma-1}{\sigma}\chi\gamma^{\frac{\sigma-1}{\sigma}} D^{\frac{\eta(\sigma-1)}{\sigma}} Y'^{\frac{1}{\sigma}} N'^{\frac{1-\sigma}{\sigma}} L_p'^{-\frac{1}{\sigma}}$, $\pi' = \frac{1}{(\sigma-1)\chi}(\bar{L} - L_e^{1\prime})V'$, $\pi = \frac{1}{(\sigma-1)\chi}(\bar{L} - L_e^1)V$, $r' = \frac{\bar{L}V'}{(\sigma-1)\chi(V'+f)}$ and $r = \frac{\bar{L}+(\sigma-1)^2\rho\chi}{\sigma(\sigma-1)\chi}$. Therefore, $\frac{1}{r'}(\pi' - \pi) - f > 0$ is equivalent to $V' - V - \frac{\bar{L}fV}{(\bar{L}+\rho\chi)V'+\rho\chi f} > 0$. Notice that the left hand side is smaller than $V' - V$, so the threshold is given by $\frac{1}{r'}(\pi' - \pi) - f > 0$.

Hence, the threshold for digital transformation with the presence of a data market is given by

$$\underline{V}^1 = \begin{cases} 0 & \text{if there is a coordination failure} \\ \frac{1}{r'}(\pi' - \pi) - f & \text{o.w.} \end{cases}. \tag{35}$$

Since $g_{\pi'} = g_{V'} = \frac{2-\sigma}{\sigma-1}g'_N + \eta g_D = \frac{1}{\sigma-1}g'_N$, and $g_\pi = g_V = \frac{2-\sigma}{\sigma-1}g_N$, it will always be the case that $g_{\pi'} > g_\pi$ as the economy grows. That is, the profit flow and the value of the firm grow faster if it undergoes digital transformation than if it does not. As a result, $\underline{V}^1 = \frac{1}{r'}(\pi' - \pi) - f$ converges to ∞, leading to inevitable digital transformation except in cases of coordination failure.

Compared to the scenario without a data market, the implementation of a data market increases the likelihood of the economy embarking on digital transformation. Another implication is that opening government data may serve as a solution to prevent coordination failure by ensuring that the aggregate data bundle $D > 0$, thus pushing the economy towards an equilibrium with digital transformation in the presence of a data market.

2.4 Discussion

From previous sections, the suboptimality in equilibrium allocations mainly arises from two sources. One is insufficient entry due to the presence of markups, as in a standard monopolistic competition framework. The other source, more specific to the model for economics of data, is insufficient entry due to increasing returns from using data in production.

By examining how labor allocation changes with digital transformation, we can infer the inefficiencies in each scenario.

Proposition 3. (Labor Allocation and Digital Transformation) *Whenever digital transformation occurs, there will be a permanent decrease in the BGP share of labor devoted to variety creation without a data market, or a temporary decrease with a data market.*

Proof. By comparing equations (13) and (18) for the scenario without a data market, and equations (27) and (33) for the scenario with a data market. □

We have shown that in the scenario with no data market, digital transformation, whenever occurs, will lead to a strictly lower share of labor devoted to variety creation with $L_e^{0\prime} < L_e^0$. This suggests that there could be an increase in supoptimality by crowding out labor from entry. With the presence of a data market, the inefficiency in labor allocation becomes temporarily more severe with $L_e^{1\prime} < L_e^1$, but as the economy grows along the balanced growth path, $L_e^{1\prime} \to L_e^1$, and the additional inefficiency imposed by digital transformation gradually disappears. In either scenario, digital transformation does not alleviate the issue of insufficient entry caused by monopolist markups.

Also, it would be useful to see how labor allocations respond to changes in the parameter representing the importance of data η. In practical terms, an economy characterized by a higher η may demonstrate a larger contribution of the data economy to GDP due to superior data quality or more advanced algorithms. Without a data market, equation (18) reveals that economies with a lower η will experience less suboptimality induced by data. When there is a data market, however, equation (33) indicates that the suboptimality driven by data is not affected by η; instead, it depends on the size of the economy or the stage of growth.

3 Numerical Analysis

To facilitate the investigation of outcomes on the balanced growth path and their implications, we provide numerical examples in this section. Firstly, we present growth rates and labor outcomes given a set of parameter values for different allocation mechanisms, aiming to interpret the relationship among these results. Additionally, we explore how the outcomes may vary with possible changes in parameters such as the importance of data and the

difficulty of entry. Furthermore, we compare the growth rates before and after digital transformation to evaluate its impact, while identifying sources of inefficiency and misallocation in the equilibrium allocations.

The parameter values we have selected are commonly used in closely related studies. As shown in Table 1, we set the rate of time preference ρ, elasticity of substitution for goods σ, importance of data η and productivity parameters A and γ at the same level as in Jones and Tonetti (2020). The population level $\bar{L}$ is assumed to be fixed at 1 as in Cong et al. (2022). Specifically for the model setup in this paper, we choose the elasticity of substitution for data to be equal to $\frac{1}{1-\eta}$ for the illustration, and consider a parameter χ that measures the difficulty or cost of firm entry with a value of 10.

Table 1: Parameter Values

Description	Parameter	Value
Rate of time preference	ρ	0.025
Elasticity of substitution (goods)	σ	4
Elasticity of substitution (data)	ϵ	$\frac{50}{47}$
Importance of data	η	0.06
Difficulty of entry	χ	8
Population level	$\bar{L}$	1
Productivity without data	A	1
Data productivity	γ	1

Note: Baseline parameters for the numerical example. The example sets $\epsilon = \frac{1}{1-\eta}$.

3.1 Baseline Results

Table 2 shows the growth rates and shares of labor in new variety creation for each allocation: the equilibrium with a data market in place, and the equilibrium without a data market.

Table 2: Numerical Example

Allocation	Digital transformation (Y/N)	Growth before d.t. g	Share of labor in entry before d.t. L_e
Data market	Y	0.0167	0.1000
No data market	N	0.0042	0.1000

In this example, the optimal allocation have the highest growth rates and the greatest share of labor in variety creation, as predicted by the model. In the optimal allocation with a data market in place, digital transformation can boost the growth from a rate of 0.83 percent to 10.67 percent, with a substantial increase in the share of labor devoted to variety

creation from 25 percent of the total labor to 80.05 percent. The planner will opt for digital transformation immediately if it it can lead to a higher growth rate, which would greatly benefit the economy.

In the presence of a data market, the economy has a status quo growth rate of 0.42 percent. Digital transformation will boost growth to a rate of 1.67 percent. Labor allocation remains unchanged despite digital transformation. The share of labor devoted to variety creation stays constant at 10 percent. Firms decide to undergo digital transformation as soon as the increase in profit is sufficient to cover the interest for the one-time fixed cost. The inclusion of data in production can lead to a substantially higher growth rate, but it does not improve insufficient entry resulting from monopolistic competition, nor does it exacerbate it.

Without the data market, the economy would fare worse after undergoing digital transformation, with both the growth rate and share of labor in entry converge to zero for the given parameter values. Digital transformation may or may not occur depending on the initial conditions, but ultimately all firms will transition away from it. Therefore, the economy will experience a long-term growth rate of 0.42 percent, with 10 percent of the labor force employed in firm entry.

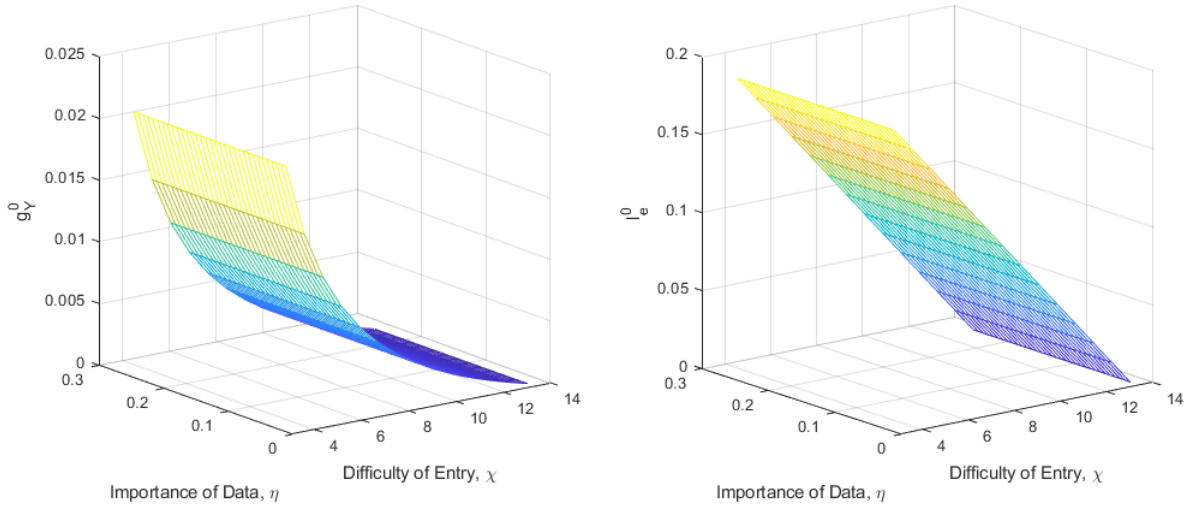
3.2 Digital Transformation and Growth

In this exercise, we explore robustness to varying parameter values. Specifically, we examine the outcomes resulting from a range of values for parameters such as data importance η and the difficulty of introducing new varieties into the economy χ. Policies can sometimes have a greater influence on these parameters than on others. Specifically, a policy that reduces entry barriers can be associated with a decrease in χ, and an improvement in data quality or the advancement of algorithms can result in an increase in η. Figure 1 presents the growth rates and shares of labor in entry under different scenarios.

Without a data market, there is no digital transformation in the economy in the long run. Both the growth rate and the share of labor devoted to entry decrease in χ and remain unaffected by changes in η within the given parameter ranges (Panel (a) of Figure 1). For a wider range of potential parameter values, it can be observed from equations (18) and (19) that $\frac{\mathrm{d}L_e^{0\prime}}{\mathrm{d}\eta} \leq 0$, $\frac{\mathrm{d}g_Y^{0\prime}}{\mathrm{d}\eta} \leq 0$, $\frac{\mathrm{d}L_e^{0\prime}}{\mathrm{d}\chi} < 0$ and $\frac{\mathrm{d}g_Y^{0\prime}}{\mathrm{d}\chi} < 0$. In any case, positive growth only occurs within a narrow range of parameter values since the growth rate decreases rapidly with both η and χ.

With the existence of a data market, the growth rate and the share of labor in entry both decrease in χ, while the growth rate increases in η (Panel (b) of Figure 1). Recall that equations (33) and (34) from the previous section imply that $\frac{\mathrm{d}L_e^{1\prime}}{\mathrm{d}\chi} < 0$, $\frac{\mathrm{d}g_Y^{1\prime}}{\mathrm{d}\chi} < 0$, $\frac{\mathrm{d}g_Y^{1\prime}}{\mathrm{d}\eta} > 0$ and $\frac{\mathrm{d}L_e^{1\prime}}{\mathrm{d}\eta} = 0$. As data becomes more important in the production of goods, the use of nonrival data bundles leads to faster growth in the competitive equilibrium. In the mean time, it has

(a) No data market, status quo



(b) With a data market, after digital transformation

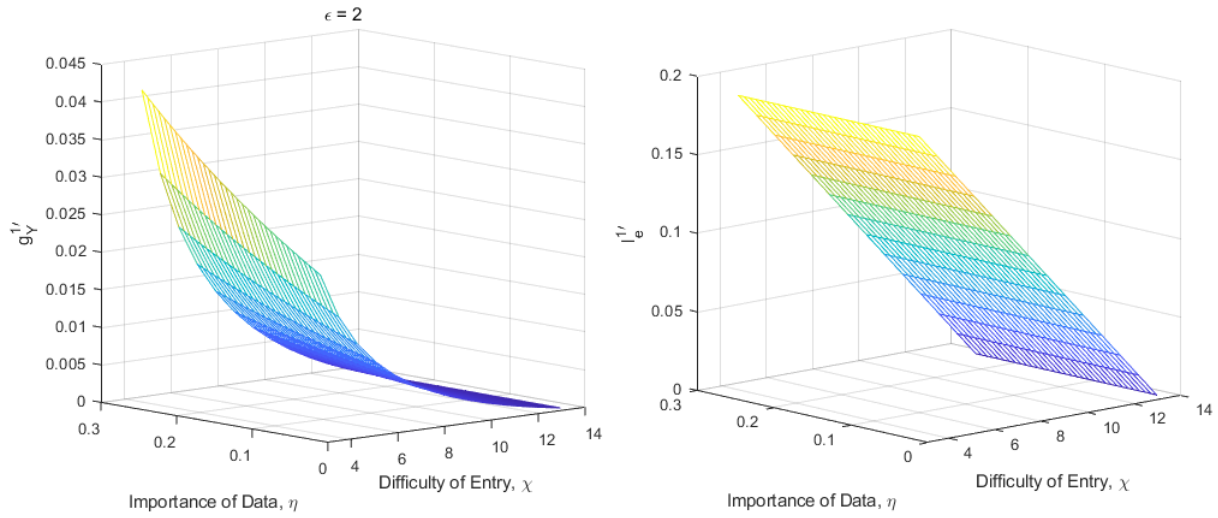


Figure 1: Growth Rate and Share of Labor in Entry

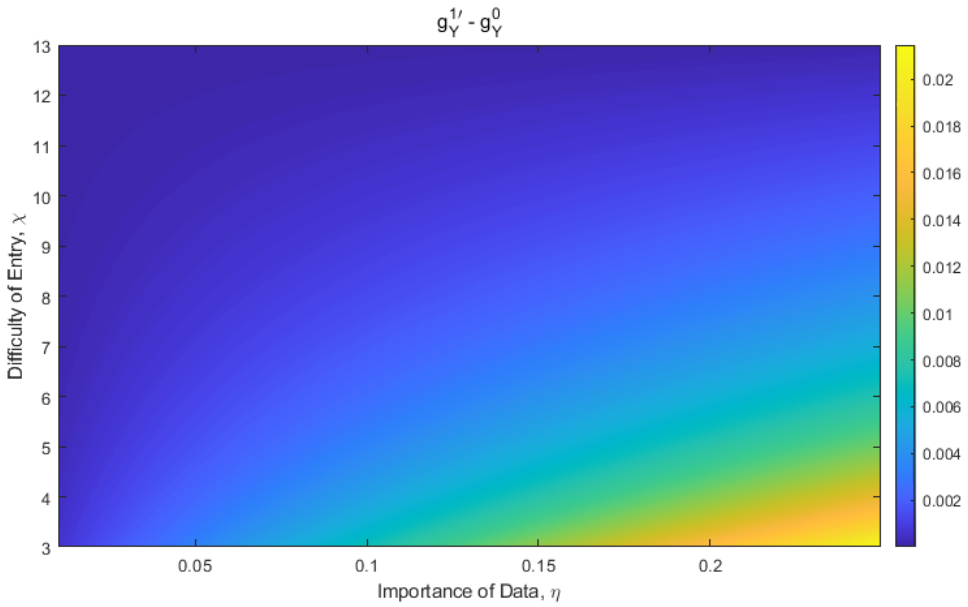


Figure 2: Comparison of Growth Rates with and without Data Market and Digital Transformation

no effect on the creation of new varieties.

Within the chosen range of parameter values, the growth rate exhibits a stronger positive correlation with a decrease in the difficulty of entry χ. In particular, a smaller χ can greatly enhance growth rates due to increasing returns from employing a larger amount of labor in creating new varieties. Here the elasticity of substitution of data is set to $\epsilon = 2$. However, the observations would still hold if we choose different values for ϵ. Appendix Figure 3 provides further explorations with $\epsilon = \frac{1}{1-\eta}$ and $\epsilon \to \infty$.

By examining the differences in BGP growth rates with and without data market, and therefore with and without digital transformation, we can assess the impact of data market and digital transformation. Figure 2 illustrates the difference in growth rates in the equilibrium allocation with and without a data market. Given the reasonable range of parameter values, the equilibrium always exhibits a faster growth rate following digital transformation in the presence of a data market. The impact from data market and digital transformation becomes larger with a higher η and lower χ.

4 Heterogeneous Firms

The model has so far assumed firm homogeneity, implying that all firms in the economy would undergo digital transformation simultaneously. However, this section introduces an extension with heterogeneous firms, allowing for the analysis of cases where some firms undergo transformation while others do not. Specifically, we introduce heterogeneity in firms'

productivity when using data. In this setting, each firm decides whether or not to digitally transform based on its type and market conditions. As a result, the economy will be divided into two parts: one consisting of firms that choose to undergo digital transformation, known as "the data economy", and another composed of firms that continue producing without utilizing data.

The production function for the firm producing variety i is

$$\begin{aligned} y_i &= A_i l_i \text{ if not digitally transformed,} \\ y_i' &= \gamma_i d_i^{\eta} l_i' \text{ if digitally transformed,} \end{aligned}$$

where i denotes a variety. A_i denotes the firm-specific productivity when producing without data, while γ_i represents the firm-specific productivity associated with data, which reflects firm's ability to utilize data.

Let $A_i = A$ for all i for the sake of tractability, and suppose γ_i is uniformly distributed between $[\underline{\gamma}, \overline{\gamma}]$. Each firm draws its data productivity type before the decision to undergo digital transformation, and the type does not change over time. Under the assumption of complete information and permanent types, the decision will feature a cutoff property. Let γ^* denote the cutoff. Firms with data productivity $\gamma_i \geq \gamma^*$ will choose to transform. The proportion of firms opting for digital transformation at each t is defined as

$$\lambda = 1 - F(\gamma^*) = \frac{\overline{\gamma} - \gamma^*}{\overline{\gamma} - \underline{\gamma}}. \tag{36}$$

Let $d_i = y_i'$, i.e. data is generated as a by-product of the consumption of the variety. Re-label each variety so that $\gamma_i \geq \gamma_j$ for $i > j$. With the data market, data can be shared and aggregated in a CES fashion with elasticity of substitution $\epsilon \in (1, \infty)$. In particular,

$$d = D = \left(\int_0^{\lambda N} \int_{\gamma^*}^{\overline{\gamma}} {y_i'}^{\frac{\epsilon-1}{\epsilon}} dF(\gamma_i) di \right)^{\frac{\epsilon}{\epsilon-1}} = \left(\int_0^{\lambda N} \int_{\gamma^*}^{\overline{\gamma}} (\gamma_i d_i^{\eta} l_i')^{\frac{\epsilon-1}{\epsilon}} dF(\gamma_i) di \right)^{\frac{\epsilon}{\epsilon-1}}, \tag{37}$$

where N is the total number of varieties, λ denotes the fraction of firms undergoing digital transformation and D denotes the total amount of data available in the economy. The law of motion for N is

$$\dot{N}_t = \frac{L_{et}}{\chi} N_t = \frac{\bar{L} - L_{pt}}{\chi} N_t.$$

4.1 No Data Market

Consider the scenario with no data market. The Household Problem is the same as in the basic model.

Firm Problem. Suppose the fixed cost for digital transformation is constant across firms. The decision making process remains unchanged.

The inverse demand function for each variety is

$$p_i = \left(\frac{y_i}{Y}\right)^{-\frac{1}{\sigma}}. \tag{38}$$

From the first order conditions of the firm problem and the competitive labor market assumption, we can establish the following relationships between the profit-maximizing input choices of firm i before and after digital transformation, as well as between two firms $i \neq j$,

$$\begin{aligned} \frac{l_i}{{l_i'}^{\frac{1-\sigma\eta}{1-\eta}}} &= (1-\eta)^{\sigma} A^{\sigma-1} {\gamma_i}^{-\frac{\sigma-1}{1-\eta}}, & (39)\\ \frac{l_i'}{l_j'} &= (\frac{\gamma_i}{\gamma_j})^{\frac{\sigma-1}{1-\sigma\eta}}. & (40) \end{aligned}$$

Define $\nu \in [0,1]$ as the share of production labor employed in the data economy. With symmetry in productivity prior to digital transformation, $l_i = l = \frac{(1-\nu)L_p}{(1-\lambda)N}$ and $l_i' = \left(\frac{(1-\nu)L_p}{(1-\lambda)N}\right)^{\frac{1-\eta}{1-\sigma\eta}} (1-\eta)^{-\frac{\sigma(1-\eta)}{1-\sigma\eta}} A^{-\frac{(\sigma-1)(1-\eta)}{1-\sigma\eta}} \gamma_i^{\frac{\sigma-1}{1-\sigma\eta}}$. Firm i would have a larger size through digital transformation as long as $l_i > (1-\eta)^{\frac{\sigma(1-\eta)}{(\sigma-1)\eta}} A^{\frac{1-\eta}{\eta}} {\gamma_i}^{-\frac{1}{\eta}}$.

Assume $\sigma\eta < 1$, the monopolist profit flow for firm i is

$$\begin{aligned} \pi_i &= \frac{1}{\sigma-1}\frac{(1-\nu)L_p}{(1-\lambda)N} w, \\ \pi_i' &= \frac{1-\sigma\eta}{\sigma-1}\left(\frac{(1-\nu)L_p}{(1-\lambda)N}\right)^{\frac{1-\eta}{1-\sigma\eta}} (1-\eta)^{-\frac{\sigma(1-\eta)}{1-\sigma\eta}} A^{-\frac{(\sigma-1)(1-\eta)}{1-\sigma\eta}} \gamma_i^{\frac{\sigma-1}{1-\sigma\eta}} w. \end{aligned} \tag{41}$$

Let $L_{pd} = \nu L_p$ be the labor employed in the data economy, $L_{pn} = (1-\nu)L_p$ be the labor employed in the rest of production, and $L_{pn} + L_{pd} = L_p$, then

$$\nu = \frac{L_{pd}}{L_p} = \frac{L_{pd}}{L_{pn}+L_{pd}} = \frac{\lambda {l_i'}^{\frac{\eta(\sigma-1)}{1-\sigma\eta}}}{(1-\lambda)(1-\eta)^{\frac{\sigma(1-\eta)}{1-\sigma\eta}} A^{\frac{(\sigma-1)(1-\eta)}{1-\sigma\eta}} \widetilde{\gamma} + \lambda {l_i'}^{\frac{\eta(\sigma-1)}{1-\sigma\eta}}}, \tag{42}$$

where $\widetilde{\gamma} = \int_{\gamma^*}^{\overline{\gamma}} \gamma_i^{\frac{\sigma-1}{1-\sigma\eta}} dF(\gamma_i) = \frac{1-\sigma\eta}{\sigma(1-\eta)} \frac{\overline{\gamma}^{\frac{\sigma(1-\eta)}{1-\sigma\eta}} - {\gamma^*}^{\frac{\sigma(1-\eta)}{1-\sigma\eta}}}{\overline{\gamma}-\gamma^*}$.

At time t, the value of firm i in the status quo, V, and the value of i if it undergoes

digital transformation, V', are given by

$$\begin{aligned} V_{it} &= \int_t^{t'} e^{-\int_t^\tau r_s ds} \pi_{i\tau} d\tau + \int_{t'}^{\infty} e^{-\int_t^\tau r_s ds} \pi'_{i\tau} d\tau - e^{-\int_t^{t'} r_s ds} f, \\ V'_{it} &= \int_t^{\infty} e^{-\int_t^\tau r_s ds} \pi'_{i\tau} d\tau - f, \end{aligned} \tag{43}$$

where f is the one-time fixed cost. No arbitrage conditions are given by the HJB equations

$$\begin{aligned} rV_i &= \dot{V}_i + \pi_i, \\ r(V'_i + f) &= \dot{V}'_i + \pi'_i. \end{aligned} \tag{44}$$

In equilibrium, firm i would choose to transform when $V'_i > V_i$, which implies $\frac{1}{r}(\pi'_i - \pi_i + \dot{V}'_i - \dot{V}_i) - f > 0$. Moreover, there should be no profitable one-shot deviations. That is, $\int_t^{t+\varepsilon} e^{-\int_t^\tau r_s ds} \pi'_{i\tau} d\tau - f > \int_t^{t+\varepsilon} e^{-\int_t^\tau r_s ds} \pi_{i\tau} d\tau - e^{-\int_t^{t+\varepsilon} r_s ds} f$ as $\varepsilon \to 0$, i.e., $\frac{1}{r}(\pi'_i - \pi_i) - f > 0$. Depending on the state of the economy, if $\dot{V}'_i - \dot{V}_i$, the cutoff for digital transformation is given by $\frac{1}{r}(\pi'_i - \pi_i) - f > 0$. Otherwise, firm i chooses to transform if $\frac{1}{r}(\pi'_i - \pi_i + \dot{V}'_i - \dot{V}_i) - f > 0$. We will discuss the cutoff in more detail later.

The free entry conditions, which ensure intertemporal zero profit for each part of the economy, are

$$\begin{aligned} (1-\lambda) V \dot{N} &= (1-\nu) w L_e, \\ \lambda \tilde{V}' \dot{N} &= \nu w L_e, \end{aligned} \tag{45}$$

where V is the average value of firms before digital transformation, and $\tilde{V}' = \int_{\gamma^*}^{\overline{\gamma}} V'_i dF(\gamma_i)$ represents the average value of firms in the data economy.

Competitive equilibrium. An equilibrium in this economy consists of quantities $\{C_t, c_{it}, a_t\}$ that solve the household problem, $\{y_{it}, p_{it}, l_{it}, d_{it}, \lambda_t\}$ that solve the firm problem, $\{N_t, L_{pt}, L_{et}, \nu_t\}$ as determined by the free entry conditions, and $\{r_t, w_t\}$ that are consistent with asset market clearing $a_t = \int_0^{\lambda_t N_t} \int_{\gamma_t^*}^{\overline{\gamma}} V'_{it} dF(\gamma_i) di + \int_{\lambda_t N_t}^{N_t} \int_{\underline{\gamma}}^{\gamma_t^*} V_{it} dF(\gamma_i) di$, labor market clearing $\bar{L} = L_{et} + L_{pt}$ and $L_{pt} = \int_0^{\lambda_t N_t} \int_{\gamma_t^*}^{\overline{\gamma}} l'_{it} dF(\gamma_i) di + \int_{\lambda_t N_t}^{N_t} \int_{\underline{\gamma}}^{\gamma_t^*} l_{it} dF(\gamma_i) di$.

Proposition 4. (Equilibrium with No Data Market and Heterogeneous Firms) *Along a balanced growth path, as N_t grows large, there will be no data economy. The share of labor in entry, share of labor in the data economy, and growth rate are given by*

$$
\begin{aligned}
L_e^0 &= \frac{\bar{L}-(\sigma-1)\rho\chi}{\sigma}, && (46)\\
L_{pd}^0 &\to 0, && (47)\\
g_Y^0 &\to (\frac{1}{\sigma-1}-\frac{\sigma\eta}{1-\sigma\eta})\frac{\bar{L}-(\sigma-1)\rho\chi}{\sigma\chi}. && (48)
\end{aligned}
$$

Proof. From the free entry conditions and the law of motion for N, we can derive the equilibrium wage rate

$$w=\frac{(1-\lambda)VN}{(1-\nu)\chi}=\frac{\lambda}{\nu}\frac{\tilde{V}'N}{\chi}, \tag{49}$$

and the relationship between $\tilde{V}'$ and V is given by

$$\frac{\tilde{V}'}{V}=\frac{(1-\lambda)\nu}{(1-\nu)\lambda}. \tag{50}$$

Notice that equation (49) and $\tilde{V}'>V$ implies $(1-\lambda)\nu/(1-\nu)\lambda>1$. This indicates that $\nu>\lambda$ and $1-\nu<1-\lambda$, meaning a larger fraction of labor employed in the data economy than the fraction of firms in the data economy. It also suggests that on average, firms in the data economy are larger in size.

The monopolist profit can now be expressed in terms of the firm value

$$
\begin{aligned}
\pi_i &= \frac{L_pV}{(\sigma-1)\chi}, && (51)\\
\pi_i' &= \frac{1-\sigma\eta}{\sigma-1}\left(\frac{1-\nu}{(1-\lambda)N}\right)^{\frac{\eta(\sigma-1)}{1-\sigma\eta}}(1-\eta)^{-\frac{\sigma(1-\eta)}{1-\sigma\eta}}A^{-\frac{(\sigma-1)(1-\eta)}{1-\sigma\eta}}\gamma_i^{\frac{\sigma-1}{1-\sigma\eta}}L_p^{\frac{1-\eta}{1-\sigma\eta}}V.
\end{aligned}
$$

This implies $g_{\pi'}=g_{V'}=g_V-\frac{\eta(\sigma-1)}{1-\sigma\eta}g_N$ for any firm in the data economy and $g_\pi=g_V$ for the rest of the economy on a balanced growth path. Since $g_{V'}<g_V$, $\dot{V}_i'=g_{V'}V_i'$ and $\dot{V}_i=g_VV_i$, $\dot{V}_i'$ may be greater or less than $\dot{V}_i$ when $V_i'>V_i$. If the initial conditions imply $\dot{V}_i'>\dot{V}_i$, any firm with $\frac{1}{r}(\pi_i'-\pi_i)-f>0$ will choose to undergo digital transformation. However, since V_i' grows at a slower rate than V_i, there will eventually be $\dot{V}_i'<\dot{V}_i$. As a result, the cutoff point thus becomes $\frac{1}{r}(\pi_i'-\pi_i+\dot{V}_i'-\dot{V}_i)-f>0$, or equivalently $V_i'-V_i>0$ thereafter. Nonetheless, there will also be $V_i'<V_i$. In other words, as the economy grows, firms have less incentive for digital transformation and more of them will opt to produce without utilizing any data.

The aggregate output Y is given by

$$\begin{aligned} Y &= \left(\int_0^{\lambda N}\int_{\gamma^*}^{\overline{\gamma}}(\gamma_i l_i')^{\frac{\sigma-1}{\sigma(1-\eta)}}dF(\gamma_i)di + \int_{\lambda N}^{N}(Al_i)^{\frac{\sigma-1}{\sigma}}di\right)^{\frac{\sigma}{\sigma-1}} \\ &= a_1 L_p N^{\frac{1}{\sigma-1}}(1+a_2\left(\frac{L_p}{N}\right)^{\frac{\eta(\sigma-1)}{1-\sigma\eta}})^{\frac{\sigma}{\sigma-1}} \end{aligned} \tag{52}$$

where $a_1 = (1-\nu)(1-\lambda)^{\frac{1}{\sigma-1}}A$, $a_2 = \lambda(1-\lambda)^{-\frac{1-\eta}{1-\sigma\eta}}(1-\nu)^{\frac{\eta(\sigma-1)}{1-\sigma\eta}}(1-\eta)^{-\frac{\sigma-1}{1-\sigma\eta}}A^{-\frac{(\sigma-1)(1-\eta)}{1-\sigma\eta}}\widetilde{\gamma}$, with $\widetilde{\gamma}$ as defined above. This implies that $g_Y \to (\frac{1}{\sigma-1}-\frac{\sigma\eta}{1-\sigma\eta})g_N$.

The first no arbitrage condition in (44) implies $\frac{\dot{V}}{V}+\frac{\pi}{V}=r$, which suggests $g_w - g_N + \frac{L_p}{(\sigma-1)\chi} = r$. Recall that $g_Y = g_C = r-\rho$ from the representative household problem, $g_N = \frac{L_e}{\chi}$ from the law of motion for N, and $g_w = g_Y$. Along the balanced growth path, as N_t grows large, the share of labor in entry, share of labor in the data economy, and growth rate are given by $L_e^0 = \frac{\bar{L}-(\sigma-1)\rho\chi}{\sigma}$, $L_{pd}^0 = \frac{\nu(\sigma-1)(\bar{L}+\rho\chi)}{\sigma}$, and $g_Y^0 \to (\frac{1}{\sigma-1}-\frac{\sigma\eta}{1-\sigma\eta})\frac{\bar{L}-(\sigma-1)\rho\chi}{\sigma\chi}$, where $\nu = \frac{\lambda l_i'^{\frac{\eta(\sigma-1)}{1-\sigma\eta}}}{(1-\lambda)(1-\eta)^{\frac{\sigma(1-\eta)}{1-\sigma\eta}}A^{\frac{(\sigma-1)(1-\eta)}{1-\sigma\eta}}\widetilde{\gamma}+\lambda l_i'^{\frac{\eta(\sigma-1)}{1-\sigma\eta}}}$ with $\widetilde{\gamma} = \frac{1-\sigma\eta}{\sigma(1-\eta)}\frac{\overline{\gamma}^{\frac{\sigma(1-\eta)}{1-\sigma\eta}}-\gamma^{*\frac{\sigma(1-\eta)}{1-\sigma\eta}}}{\overline{\gamma}-\gamma^*}$.

Notice that as the economy grows, each firm becomes smaller in size as $l_i' \to 0$, resulting in $\nu \to 0$. The amount of labor employed in the data economy $L_{pd} \to 0$, and there will be no data economy in the long term with $\lambda \to 0$, which is consistent with the observation from cutoff analysis. □

4.2 With Data Market

Now consider the scenario with a data market. Suppose the elasticity of substitution of data $\epsilon = \frac{1}{1-\eta}$.

Firm Problem. From the first order conditions and the competitive labor market assumption, the labor input choices are given by

$$\frac{l_i}{l_i'} = (\frac{A}{\gamma_i D^\eta})^{\sigma-1}, \tag{53}$$

$$\frac{l_i'}{l_j'} = (\frac{\gamma_i}{\gamma_j})^{\sigma-1}. \tag{54}$$

With symmetry prior to digital transformation, $l_i = l = \frac{(1-\nu)L_p}{(1-\lambda)N}$ and $l_i' = \frac{(1-\nu)L_p}{(1-\lambda)N}(\frac{\gamma_i D^\eta}{A})^{\sigma-1}$, where ν denotes the share of production labor employed in the data economy. Firms can achieve a larger size after digital transformation whenever $D^\eta > A\gamma_i^{-1}$.

The profit flow for a firm in or not in the data economy is now given by

$$\begin{aligned}\pi_i &= \frac{1}{\sigma-1}\frac{(1-\nu)L_p}{(1-\lambda)N}w,\\ \pi_i' &= \frac{1}{\sigma-1}\frac{(1-\nu)L_p}{(1-\lambda)N}(\frac{\gamma_i D^\eta}{A})^{\sigma-1}w.\end{aligned} \tag{55}$$

It can be shown that

$$\nu = \frac{L_{pd}}{L_p} = \frac{L_{pd}}{L_{pn}+L_{pd}} = \frac{\lambda\widetilde{\gamma}D^{\eta(\sigma-1)}}{(1-\lambda)A^{\sigma-1}+\lambda\widetilde{\gamma}D^{\eta(\sigma-1)}}, \tag{56}$$

where L_{pd} is the labor employed in the data economy, L_{pn} is the labor employed in the rest of production, $L_{pn}+L_{pd}=L_p$, and $\widetilde{\gamma}=\int_{\gamma^*}^{\overline{\gamma}}\gamma_i^{\sigma-1}dF(\gamma_i)=(\overline{\gamma}^\sigma-\gamma^{*\sigma})/(\overline{\gamma}-\gamma^*)\sigma$.

The value of firm i and no arbitrage conditions are given by

$$\begin{aligned}V_{it} &= \int_t^{t'} e^{-\int_t^\tau r_s ds}\pi_{i\tau}d\tau + \int_{t'}^{\infty} e^{-\int_t^\tau r_s ds}\pi_{i\tau}'d\tau - e^{-\int_t^{t'} r_s ds}f,\\ V_{it}' &= \int_t^{\infty} e^{-\int_t^\tau r_s ds}\pi_{i\tau}'d\tau - f,\end{aligned} \tag{57}$$

and

$$\begin{aligned}rV_i &= \dot{V}_i + \pi_i,\\ r(V_i'+f) &= \dot{V}_i' + \pi_i'.\end{aligned} \tag{58}$$

In equilibrium, firm i would choose to transform when $V_i' > V_i$, which implies $\frac{1}{r}(\pi_i' - \pi_i + \dot{V}_i' - \dot{V}_i) - f > 0$. Moreover, there should be no profitable one-shot deviations. That is, $\int_t^{t+\varepsilon} e^{-\int_t^\tau r_s ds}\pi_{i\tau}'d\tau - f > \int_t^{t+\varepsilon} e^{-\int_t^\tau r_s ds}\pi_{i\tau}d\tau - e^{-\int_t^{t+\varepsilon} r_s ds}f$ as $\varepsilon\to 0$, i.e., $\frac{1}{r}(\pi_i'-\pi_i)-f>0$. Since $\dot{V}_i' > \dot{V}_i$, as we will show later, firm i will choose to undergo digital transformation as soon as $\frac{1}{r}(\pi_i'-\pi_i)-f>0$. For firms at the cutoff γ^*,

$$\frac{w}{r(\sigma-1)}(l_i'-l_i)-f=0. \tag{59}$$

Hence, we can solve the equation below for γ^*,

$$\gamma^* = ((\sigma-1)rf\frac{(1-\lambda)N}{w(1-\nu)L_p}+1)^{\frac{1}{\sigma-1}}AD^{-\eta}. \tag{60}$$

This, in turn, determines λ, the overall level of digital transformation in the economy.

The free entry conditions are

$$\begin{aligned}(1-\lambda)V\dot{N} &= (1-\nu)wL_e, \\ \lambda\tilde{V}'\dot{N} &= \nu wL_e, \end{aligned} \tag{61}$$

where V is the average value of firms without digital transformation, and $\tilde{V}' = \int_{\gamma^*}^{\overline{\gamma}} V_i' dF(\gamma_i)$ represents the average value of firms in the data economy.

Competitive equilibrium. An equilibrium in this economy consists of quantities $\{C_t, c_{it}, a_t\}$ that solve the household problem, $\{y_{it}, p_{it}, l_{it}, D_t, \lambda_t\}$ that solve the firm problem, $\{N_t, L_{pt}, L_{et}, \nu_t\}$ as determined by the free entry conditions, and $\{r_t, w_t\}$ that are consistent with asset market clearing $a_t = \int_0^{\lambda_t N_t} \int_{\gamma_t^*}^{\overline{\gamma}} V_{it}' dF(\gamma_i) di + \int_{\lambda_t N_t}^{N_t} \int_{\underline{\gamma}}^{\gamma_t^*} V_{it} dF(\gamma_i) di$, labor market clearing $\bar{L} = L_{et} + L_{pt}$ and $L_{pt} = \int_0^{\lambda_t N_t} \int_{\gamma_t^*}^{\overline{\gamma}} l_{it}' dF(\gamma_i) di + \int_{\lambda_t N_t}^{N_t} \int_{\underline{\gamma}}^{\gamma_t^*} l_{it} dF(\gamma_i) di$.

Proposition 5. (Equilibrium with a Data Market and Heterogeneous Firms) *Along a balanced growth path, as N_t grows large, the economy will undergo a complete digital transformation. The share of labor in entry, share of labor in the data economy, and growth rate are given by*

$$L_e^1 = \frac{\bar{L} - (\sigma-1)\rho\chi}{\sigma}, \tag{62}$$

$$L_{pd}^1 \to \frac{(\sigma-1)(\bar{L}+\rho\chi)}{\sigma}, \tag{63}$$

$$g_Y^1 \to (\frac{1}{\sigma-1} + \frac{\sigma(1-\eta)}{1-\sigma\eta}) \frac{\bar{L} - (\sigma-1)\rho\chi}{\sigma\chi}, \tag{64}$$

where $L_{pd}^1 \to L_p^1 = \bar{L} - L_e^1$. In particular, the data economy will feature larger firms and experience faster growth rates.

For the equilibrium with coordination failure, the balanced growth path outcomes are characterized by

$$L_e = \frac{\bar{L} - (\sigma-1)\rho\chi}{\sigma}, \tag{65}$$

$$L_{pd} = 0, \tag{66}$$

$$g_Y = \frac{\bar{L} - (\sigma-1)\rho\chi}{(\sigma-1)\sigma\chi}. \tag{67}$$

Proof. Like in the basic model, when there is a data market, there always exists an equlibrium where no firm undergoes digital transformation. In this case, $\lambda = 0$ and $l_i = \frac{L_p}{N}$ for all i, the share of labor in entry, share of labor in the data economy, and growth rate are given by (65), (66) and (67).

The equilibrium wage rate is

$$w = \frac{(1-\lambda)VN}{(1-\nu)\chi} = \frac{\lambda}{\nu}\frac{\tilde{V}'N}{\chi}, \tag{68}$$

where $\tilde{V}' > V$ implies $(1-\lambda)\nu/(1-\nu)\lambda > 1$ and $\nu > \lambda$. This means that firms in the data economy are larger in size.

The monopolist profit in terms of the firm value becomes

$$\begin{aligned} \pi_i &= \frac{L_p V}{(\sigma-1)\chi}, \\ \pi_i' &= \frac{L_p V}{(\sigma-1)\chi}(\frac{\gamma_i D^\eta}{A})^{\sigma-1}. \end{aligned} \tag{69}$$

This implies $g_{\pi'} = g_{V'} = g_V + \eta(\sigma-1)g_D$ for any firm in the data economy and $g_\pi = g_V$ for the rest of the economy on a balanced growth path. Since $g_{V'} > g_V$, $\dot{V}_i' = g_{V'}V_i'$ and $\dot{V}_i = g_V V_i$, there is $\dot{V}_i' > \dot{V}_i$when $V_i' > V_i$.

With an elasticity of substitution of data $\epsilon = \frac{1}{1-\eta}$, in the competitive equilibrium, the aggregate, nonrival data bundle is given by

$$\begin{aligned} D &= \left(\int_0^{\lambda N}\int_{\gamma^*}^{\overline{\gamma}}\left(\gamma_i D^\eta \frac{(1-\nu)L_p}{(1-\lambda)N}\left(\frac{\gamma_i D^\eta}{A}\right)^{\sigma-1}\right)^\eta dF(\gamma_i)di\right)^{\frac{1}{\eta}} \\ &= \left(\frac{(1-\nu)L_p}{(1-\lambda)N}A^{1-\sigma}\widetilde{\gamma}'\right)^{\frac{1}{1-\sigma\eta}} \end{aligned} \tag{70}$$

where $\widetilde{\gamma}' = \left(\int_{\gamma^*}^{\overline{\gamma}}\gamma_i^{\sigma\eta}dF(\gamma_i)\right)^{\frac{1}{\eta}} = ((\overline{\gamma}^{1+\sigma\eta} - \gamma^{*1+\sigma\eta})/(\overline{\gamma}-\gamma^*)(1+\sigma\eta))^{\frac{1}{\eta}}$. Therefore, $g_D = \frac{1-\eta}{\eta(1-\sigma\eta)}g_N$. Since $\eta g_D > g_N$, this suggests that the cutoff γ^* as described by (60) will decrease as the economy grows. It also suggests that the growth rate of profits in the data economy is higher than that in the rest of the economy by a level of $\eta(\sigma-1)g_D$ which is equivalent to $\frac{(\sigma-1)(1-\eta)}{1-\sigma\eta}g_N$.

The aggregate output Y is given by

$$\begin{aligned} Y &= \left(\int_0^{\lambda N}\int_{\gamma^*}^{\overline{\gamma}}(\gamma_i D^\eta l_i')^{\frac{\sigma-1}{\sigma}}dF(\gamma_i)di + \int_{\lambda N}^{N}(Al_i)^{\frac{\sigma-1}{\sigma}}di\right)^{\frac{\sigma}{\sigma-1}} \\ &= \frac{1-\nu}{1-\lambda}AL_pN^{\frac{1}{\sigma-1}}(1-\lambda+\lambda A^{1-\sigma}\widetilde{\gamma}D^{\eta(\sigma-1)})^{\frac{\sigma}{\sigma-1}} \end{aligned} \tag{71}$$

where $\widetilde{\gamma}$ is defined above. This implies that g_Y converges to $\frac{1}{\sigma-1}g_N + \eta\sigma g_D = (\frac{1}{\sigma-1} + \frac{\sigma(1-\eta)}{1-\sigma\eta})g_N$ asymptotically.

The first no arbitrage condition in (58) implies $\frac{\dot{V}}{V} + \frac{\pi}{V} = r$, which suggests $g_w - g_N +$

$\frac{L_p}{(\sigma-1)\chi} = r$. Recall that $g_Y = g_C = r - \rho$, $g_N = \frac{L_e}{\chi}$, and $g_w = g_Y$. Along the balanced growth path, as N_t grows large, the share of labor in entry, share of labor in the data economy, and growth rate are given by $L_e^1 = \frac{\bar{L}-(\sigma-1)\rho\chi}{\sigma}$, $L_{pd}^1 = \frac{\nu(\sigma-1)(\bar{L}+\rho\chi)}{\sigma}$, and $g_Y^1 \to (\frac{1}{\sigma-1} + \frac{\sigma(1-\eta)}{1-\sigma\eta})\frac{\bar{L}-(\sigma-1)\rho\chi}{\sigma\chi}$, where $\nu = \frac{(1-\lambda)\widetilde{\gamma}D^{\eta(\sigma-1)}}{\lambda A^{\sigma-1}+(1-\lambda)\widetilde{\gamma}D^{\eta(\sigma-1)}}$ and $\widetilde{\gamma} = (\overline{\gamma}^{\sigma} - \gamma^{*\sigma})/(\overline{\gamma} - \gamma^{*})\sigma$.

Notice that as the economy grows, the amount of data D becomes large causing $\nu \to 1$. The amount of labor employed in the data economy converges to total labor in production $L_{pd} \to L_p$. At the same time, this implies that as the economy grows, it will gradually undergo a complete digital transformation with $\lambda \to 1$. □

Compare equations (62), (63) and (64) with the equilibrium outcomes where coordination failure prevents digital transformation ((65), (66) and (67)), the allocation of labor between entry and production remains unchanged, while the growth rate will increase by a factor of $\frac{\sigma(\sigma-1)(1-\eta)}{1-\sigma\eta}$. Furthermore, the economy grows strictly faster with a data market compared to the allocation without one ((46), (47) and (48)), while the shares of labor in entry remain equal in both scenarios. This suggests that implementing a data market does not improve suboptimality caused by monopolistic profits but rather increases growth rates by facilitating a complete digital transformation in the economy. In addition, firms in the data economy not only have larger sizes but also experience higher rates of return than the rest of the economy, which aligns with the findings in Farboodi et al. (2019) and Eeckhout and Veldkamp (2022) that increasing returns to data may lead to increased market power by data-rich firms.

5 Conclusion

We construct an endogenous growth model with expanding product varieties to investigate the decision regarding digital transformation. Following the transformation, data is generated as a by-product of consumption of each variety and used in production to improve productivity. In particular, firms face a tradeoff between one-time fixed costs and productivity enhancement.

In the basic model, we compare outcomes from the allocations when a data market is in place which aggregates all data into a nonrival bundle, and when every firm uses its own data only. The analytical solutions to the model, as well as the results from numerical analysis, indicate that in the presence of a data market, the economy exhibits accelerated growth rates following digital transformation. However, in the absence of a data market, such transformation can have negative impacts on the growth and cannot sustain over time. Moreover, when data is more important due to higher increasing returns or constituting a higher share of GDP, the impact of digital transformation - both positive and negative - becomes more pronounced. This highlights the need for implementing a data market to facilitate successful digital transformation. Therefore, policymakers should carefully consider the sequence of policies regarding the establishment of a data market and the promotion of

digital transformation. Relatedly, the government's open data initiative has the potential to alleviate coordination failure in the data market.

There are two sources of suboptimality in the equilibrium allocations. One is insufficient entry due to the presence of monopolistic markups, the other is insufficient entry due to increasing returns from data used in production. Without a data market, overall suboptimality is likelty to increase after digital transformation, exacerbated when data has a higher importance. When there is a data market, digital transformation may temporarily contribute to the second source of suboptimality and show no improvement in the first.

In addition, as the entry cost decreases and the importance of data increases, economies with a data market will benefit more from digital transformation. This suggests that policies aimed at reducing barriers to entry, improving data quality, and supporting R&D in algorithms can effectively complement digital transformation.

In the extension, we explore a model with firms that are heterogeneous in their ability to utilize data. Most of the results from the basic model still hold true. Furthermore, in equilibrium with a data market, firms undergoing digital transformation not only have larger sizes but also enjoy higher profit growth rates. This finding is consistent with related studies indicating that digital transformation can result in increased market power for firms using data compared to the rest of the economy.

There are still several unanswered questions, such as the potential outcomes if the model is stochastic, if firms have the ability to choose the extent of digital transformation or amount of data to trade, and if the fixed cost of digital transformation becomes endogenous. With that being said, this paper represents a small step towards an ongoing effort to study the economics of data. Future research can delve deeper into designing data market mechanisms, incorporating digital platforms, and examining firm dynamics in more detail. Alternatively, one can explore how different policies and market structures can facilitate growth, improve efficiency, and achieve better outcomes for various stakeholders in the data economy.

References

Acemoglu, D. (2009). *Introduction to Modern Economic Growth*. Princeton University Press, Princeton.

Aghion, P. and Howitt, P. (1992). A Model of Growth through Creative Destruction. *Econometrica*, 60(2):323–351.

Arrow, K. J. (1962). The economic implications of learning by doing. *The Review of Economic Studies*, 29(3):155–173.

Chamberlin, E. H. (1965). *The theory of monopolistic competition: A Re-orientation of the Theory of Value*, volume 38. Harvard University Press, Cambridge, 8th edition.

Cong, L. W., Wei, W., Xie, D., and Zhang, L. (2022). Endogenous growth under multiple uses of data. *Journal of Economic Dynamics and Control*, 141:104395. Markets and Economies with Information Frictions.

Cong, L. W., Xie, D., and Zhang, L. (2021). Knowledge accumulation, privacy, and growth in a data economy. *Management Science*, 67(10):6480–6492.

Dixit, A. and Stiglitz, J. (1977). Monopolistic competition and optimum product diversity. *American Economic Review*, 67(3):297–308.

Eeckhout, J. and Veldkamp, L. (2022). Data and Market Power. NBER Working Papers 30022, National Bureau of Economic Research.

Farboodi, M., Mihet, R., Philippon, T., and Veldkamp, L. (2019). Big data and firm dynamics. *AEA Papers and Proceedings*, 109:38–42.

Farboodi, M. and Veldkamp, L. (2021). A model of the data economy. Working Paper 28427, National Bureau of Economic Research.

Grossman, G. M. and Helpman, E. (1991). Quality Ladders in the Theory of Growth. *The Review of Economic Studies*, 58(1):43–61.

Hopenhayn, H. A. (1992). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica*, 60(5):1127–1150.

Jones, C. I. (1995). Time Series Tests of Endogenous Growth Models. *The Quarterly Journal of Economics*, 110(2):495–525.

Jones, C. I. and Tonetti, C. (2020). Nonrivalry and the economics of data. *American Economic Review*, 110(9):2819–58.

Li, Z. and Xu, Z. (2024). Production and macroeconomy: A general equilibrium theory with firm digitalization. *Journal of Management World*, 40(9):80–97.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695–1725.

OECD (2024). Oecd digital economy outlook 2024 (volume 1): Embracing the technology frontier. Technical report, Paris.

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, 98(5):S71–S102.

Appendix

A Proofs

Lemma 6. *In equilibrium, all firms would make the same choice: either all or none would undergo digital transformation.*

Proof. Let λ denote the proportion of firms in the economy opting for digital transformation, and ν denote the share of production labor employed by these firms. We want to show that $\lambda = \nu = 0$ *or* 1 in any possible equilibrium.

Suppose there exists an equilibrium with $\lambda \in (0,1)$, which implies $V' = V$. From the free entry conditions $(1-\lambda)V\dot{N} = (1-\nu)wL_e$ and $\lambda V'\dot{N} = \nu wL_e$, so $\frac{V'}{V} = \frac{(1-\lambda)\nu}{\lambda(1-\nu)} = 1$,implying $\lambda = \nu$. This gives $l' = \frac{\lambda L_p}{\nu N} = \frac{(1-\lambda)L_p}{(1-\nu)N} = l$.

For the equilibrium allocation with the data market, from no arbitrage conditions, $rf = \pi' - \pi = \frac{1}{\sigma-1}w(l'-l) = 0$. As a result, $V' = V \implies f = 0$, which leads to a contradiction.

For the equilibrium allocation with no data market, $rf = \pi' - \pi = \frac{1}{\sigma-1}w(l'-l-\sigma\eta l') < 0$. As a result, $V' = V \implies f < 0$, which leads to a contradiction.

Therefore, there does not exist an equilibrium with $\lambda \in (0,1)$ either with a data market or without. □

B Figures

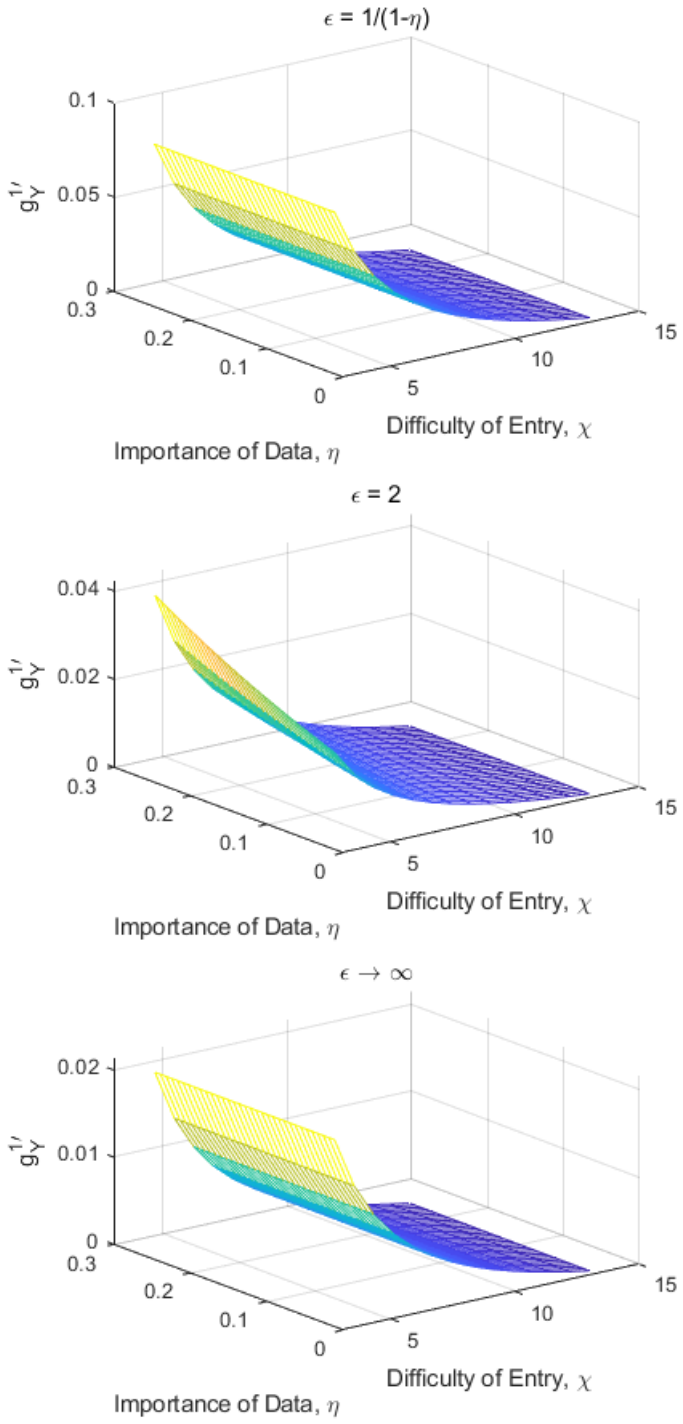


Figure 3: Growth Rates after Digital Transformation with Data Market for Different Values of ϵ