Data Economy to the Rescue of Our Empty Planet?*

Qing Chang† Lin William Cong‡ Longtian Zhang§

December 16, 2024

Abstract

Many medium- and high-income countries are riddled with aging and declining populations. Against the backdrop of the rapid emergence of big data, digitization, and AI, we explore how the development of the data economy interacts with these demographic challenges and has the potential to mitigate or even reverse these troubling trends. Specifically, we construct an endogenous growth model with endogenous fertility, incorporating data into the accumulation of human capital. We identify a powerful reinforcement mechanism: a larger population generates more data, which, due to its nonrival nature, accelerates human capital accumulation and enhances consumers' productivity. This, in turn, frees up time and financial resources, facilitating higher rates of reproduction and population growth. Unlike Jones (2022), the development of the data economy can drive high population and economic growth without the need for distortionary policy measures.

Keywords: Population decline, digital economy, endogenous growth, human capital accumulation.

JEL Classification: I12, J11, and O41

*All the authors contribute equally to this work. The authors are especially grateful to Charles I. Jones for his detailed feedback and invaluable guidance. Chang is grateful for the financial support from the "Shuimu Scholar" project at Tsinghua University, Cong is grateful for the financial support from Ripple's University Blockchain Research Initiative (UBRI) and the FinTech at Cornell Initiative, and Zhang is grateful for the financial support from the National Natural Science Foundation of China (Grant No. 72303261). The contents of this publication are solely the responsibility of the authors. Send correspondence to Cong at will.cong@cornell.edu.

†Institute of Economics, School of Social Sciences, Tsinghua University.

‡Cornell University SC Johnson College of Business (Johnson) and NBER.

§School of International Trade and Economics, Central University of Finance and Economics.

1. Introduction

Many countries around the globe, including Bulgaria, Germany, Greece, Italy, Japan, South Korea, and Ukraine, face unprecedented demographic challenges related to population aging and decline, often driven by the younger generation's increasing reluctance or inability to dedicate time and resources to child-rearing. For instance, official statistics from China reported a negative population growth of approximately -0.21% between 2022 and 2023, marking a critical turning point in the nation's demographic trajectory. Maintaining positive population growth has long been recognized as essential for sustained economic development (e.g., Romer, 1990; Grossman and Helpman, 1991; Aghion and Howitt, 1992; Jones, 1995). A recent study by Jones (2022) introduces a growth model with an endogenous fertility rate to underscore the adverse outcomes of insufficient pro-breeding policies (the "empty planet" scenario) and highlight the necessary interventions to achieve positive population growth (the "expanding cosmos" scenario). With an exogenous negative population growth rate, Boikos et al. (2023) demonstrate that human capital accumulation can mitigate the negative effects of depopulation. Similarly, others emphasize the challenges posed by depopulation from the perspectives of exhaustible resources and education (e.g., Mino and Sasaki, 2023; Strulik, 2024).

None of these studies offer potential solutions with endogenous population growth or examine the role of data in the economy, a gap this paper aims to fill. As an illustration of the importance of data, the emergence of generative AI since late 2022, exemplified by ChatGPT and similar technologies, has transformed the ways we learn and work.[1] At the core of these advancements lies the use of vast datasets. These technologies excel by leveraging rich and diverse data collected from various production sectors and daily life. This data-driven approach not only enhances the functionality of AI but also stimulates

[1]For a detailed introduction to these new technologies, see https://future.com/alphafold-gpt-3-and-how-to-augment-intelligence-with-ai/. For further studies on the effects of ChatGPT on education and human capital, see Kumar et al. (2024) and Giordano et al. (2024).

economic growth by enabling smarter, more efficient business models and workflows.[2] As AI becomes increasingly integrated into the global economy, it is set to transform industries and redefine economic paradigms, establishing itself as a pivotal element of modern economic development.

In light of the rapid development of data technologies, this paper synthesizes the data-driven human capital accumulation framework outlined in Chang et al. (2024) with the analyses of endogenous population growth rates discussed in Jones (2022) (referred to as the "J model" in subsequent discussions). Drawing on the differentiation of data types in the aforementioned studies, we specifically focus on "producer data," as described in Farboodi and Veldkamp (2022) and Xie and Zhang (2023), which refers to data generated as byproducts during the production process.[3] Our model features representative consumers who accumulate human capital through time investment and data usage, data intermediaries responsible for data allocation, and a production sector that generates data alongside final goods for consumption. We first propose a benchmark model without data to illustrate the dynamics of endogenous fertility decisions within a human capital framework, contrasting it with the innovation-driven approach depicted in the J model. Subsequently, we integrate producer data into the economy, using this factor as a substitute for education time in human capital accumulation to develop a general model. We then analyze and compare the outcomes in competitive equilibrium and optimal allocation with those of the J model, revealing a significant externality in population decision-making. Finally, we examine both the steady-state (balanced growth path, BGP) and the transition dynamics, shedding light on the stability of the equilibria.

[2]Some argue that the rise of these technologies may displace human capital. However, although the technology itself is not directly related to population size, the refinement and optimization of these technologies fundamentally depend on data. The more people use these technologies, the more data are generated, improving the quality of training. This paper emphasizes a production-related, population-dependent factor as an input to human capital, thereby promoting both population growth and efficiency improvement.

[3]In Farboodi and Veldkamp (2022), the authors discuss data generated from production, which is directly related to output levels and free from privacy concerns. Xie and Zhang (2023) examine various types of data and compare their differing effects on growth. Chang et al. (2024) incorporate producer data into a human capital framework, providing the rationale for this model setting. In Section 3.1, we further discuss the selection of this type of data.

Our main findings are twofold. First, we find that the incorporation of data positively influences the population growth rate, thereby enhancing overall economic welfare. The nonrival nature of data makes it an exceptionally efficient input for human capital accumulation. Comparative statics show that total welfare and population growth rates increase as the importance of data rises, both in the competitive equilibrium and under optimal allocation. Second, our findings suggest that a social planner only needs to integrate data into the human capital accumulation process to steer the economy toward optimal allocation, achieving a positive and substantial population growth rate. This result contrasts with the outcomes of the J model, where the author suggests that the social planner must implement additional intervention policies to raise the population growth rate to a sufficiently high level for the economy to converge to the "expanding cosmos" scenario.[4] The distinction arises from the different steady states in the optimal allocation of our general model: one state exhibits a low and unstable population growth rate—lower than that observed in the competitive equilibrium of the benchmark model—while the other maintains a positive, high population growth rate and is saddle-path stable. Thus, the economy embarks on a transition path toward the high-growth steady state as soon as the social planner moves from the competitive equilibrium to the optimal allocation. Furthermore, the time series analysis of the transition dynamics reveals considerable stability along the path toward the steady state with a high population growth rate, reinforcing the stability of this state.

The integration of the data economy with human capital represents a significant innovation in how people learn and work. The emergence of these new technologies and phenomena urges us to reconsider the potential of the data economy to counter the prevailing trend of population decline. Furthermore, it is essential to assess whether the utilization of data can enhance the efficiency of human capital accumulation and, in turn, drive economic growth. In earlier studies on data, such as Jones and Tonetti (2020), Cong et al. (2021), and Cong et al. (2022), the authors emphasize that the most critical attribute of data is its nonrival na-

[4]The policy interventions discussed in this paper refer only to policies beyond those required to transition the economy from a competitive equilibrium to the optimal allocation, consistent with the discussion in Jones (2022).

ture—data can be used by one individual without diminishing its availability for others. In the context of this paper, an interesting reinforcement effect emerges: a larger population generates more data, facilitating faster human capital accumulation and enhancing consumers' productivity, thanks to the nonrival nature of data; the resulting time savings and increased financial resources—indirectly through real wage growth—enable further reproduction and promote population growth.

To our knowledge, we are the first to jointly consider the long-run effects of data, endogenous fertility, and human capital accumulation on an economy facing population decline, building on the foundational work of Jones and Tonetti (2020), Farboodi and Veldkamp (2022), and Cong et al. (2021). Recent studies also focus on analyzing the nonrival nature of data and underscore the competition among digital platforms or intermediaries (Ichihashi, 2021a,b; Yang, 2022). In terms of the deployment of digital technologies, Jones (2024) and Acemoglu and Lensman (2024) explore the potential opportunities and challenges in this new era, while Beraja et al. (2023) examine micro-level data to reveal their positive impacts on innovation, privacy, and growth.

This paper also contributes to the literature investigating the impacts of digital technologies on human capital and employment. Following the seminal work of Acemoglu and Restrepo (2018) and Acemoglu and Restrepo (2020) on the multifaceted effects of AI and automation on employment structures, Studies such as Prettner and Strulik (2020) explore the connections between automation, higher education enrollment, and innovation within an R&D-driven growth model. Meanwhile, Lu (2021) attempts to integrate AI into the human capital accumulation framework but falls short of elucidating the implications of digital technologies on population growth. Finally, Chang et al. (2024) successfully establishes a link between data and human capital accumulation in a semi-endogenous growth model. Our study adds by incorporating the decision to raise children into the framework.

The remainder of this paper is organized as follows. Section 2 introduces the benchmark model without data to illustrate the preliminary results on endogenous fertility decisions within our framework. Section 3 expands this to include a general model with data, providing

comparative statics to explore fundamental results concerning the multiple solutions in optimal allocation. Section 4 then examines the transition dynamics of the general model, detailing both the phase diagram and the time series analysis of these dynamics. Finally, Section 5 provides the conclusions.

2. Benchmark Economy Without Data

Within the framework of human capital accumulation, we explore how this type of growth model responds to population declines. Prior to discussing the role of data, we analyze a conventional and simplified model of human capital accumulation that includes endogenous population growth. This model serves as the benchmark for our study, providing a basis for comparison with the general model, which incorporates data usage and is introduced in the subsequent section. This approach allows us to clearly delineate the impact of data on the dynamics of human capital accumulation.

2.1 Model Setup

This model only includes representative consumers and the production of final goods. Each consumer is endowed with one unit of time that can be allocated to one of three activities: producing children, education, and producing final goods. The decision to produce children is made endogenously, leading to a variable rate of population growth.

2.1.1 Population Growth Rate

There exist homogeneous consumers with a population of L_t at time t, which grows at a variable rate n_t and has an initial population of $L_0 = 100$. The population growth rate is endogenously determined by two factors: the fertility rate and the death rate. Assuming each consumer allocates l_t units of time to childbearing, and considering the death rate as constant in developed countries experiencing negative population growth—where life expectancy has nearly reached the natural limit for humans—we derive the following population growth

equation:

$$n_t = b(l_t) - \delta_0, \tag{1}$$

where $b(\cdot)$ represents the fertility rate, and $\delta_0 > 0$ denotes the death rate. For simplicity, like in the J model, we assume the fertility rate follows a linear function, i.e., $b(l_t) = \bar{b} l_t$, where $\bar{b} > 0$ is a constant reflecting the efficiency of child-rearing activities.

Given the population growth rate defined above, we have the evolution of population as:

$$\dot{L}_t = n_t L_t. \tag{2}$$

Unlike related papers such as Jones and Tonetti (2020) and Cong et al. (2021), consumers in this model consider n_t endogenously rather than treating it as an exogenous parameter. This approach constitutes one of the main contributions of this paper.

2.1.2 Representative Consumers

In this economy, a representative consumer lives in an indefinite and continuous environment. This consumer faces the following lifetime utility maximization problem:

$$\max_{c_t, L_t} \int_0^\infty e^{-(\bar{\rho}+\delta_0)t} \frac{(L_t^{\varepsilon} c_t)^{1-\gamma}}{1-\gamma} \mathrm{d}t, \tag{3}$$

where $\bar{\rho} > 0$ is the subjective discount rate of consumption, $\varepsilon > 0$ denotes the importance of descendants of the people born at time t, and $\gamma \geq 1$ is the inverse of the intertemporal elasticity of substitution of consumption. Notice that the representative consumers account for their death rates, thus δ_0 is added to $\bar{\rho}$ as a discount on lifetime utility. This utility framework, derived from Barro and Becker (1989) and Jones and Schoonbroodt (2010), is referred to as "generalized Barro-Becker preferences" and is used to represent attitudes towards children when studying population growth.

The consumer's budget constraint is shown as follows:

$$c_t = (1 - l_t - u_t) w_t h_t, \tag{4}$$

where u_t is the time allocated to education, so $(1-l_t-u_t)$ is the time allocated to the production of final goods, which can be remunerated.[5] Finally, w_t is the wage rate, and h_t is the average level of human capital stock, so $w_t h_t$ together represents the effective wage that one can earn from employment.

The consumer also accumulates human capital in the following form:

$$\dot{h}_t = \eta h_t^{\zeta} u_t - \delta_0 h_t, \tag{5}$$

where η is the efficiency term of accumulating human capital, and ζ represents the spillover effect of human capital. This specification comes from Lucas (2015), which indicates that the spillover effect is significant and should be considered in human capital accumulation. We can interpret this process as education, where consumers spend time learning and accumulate their human capital to achieve higher effective wage rates. Finally, note that the death rate is also considered here, implying that consumers' human capital stock will diminish upon their death, so a portion of human capital is deducted from its total stock in each period.

2.1.3 Production of Final Goods

The production sector follows a simple setting within the human capital framework. There exists a representative producer who solely employs labor for production, and the productivity of labor is enhanced by the human capital stock of the employees. The production function is shown as:

$$Y_t = (1 - l_t - u_t) h_t^{\sigma_p} L_t,$$

[5]The variables u_t and l_t can also be viewed as the labor shares of the whole economy allocated to human capital accumulation and child production, respectively. In this paper, we assume that consumers cannot receive wages from these two activities.

where $(1 - l_t - u_t)$ is the labor allocated to final good production, and $\sigma_P > 0$ represents the augmentation effect of human capital in this process. Additionally, we denote $y_t = Y_t/L_t$ as the per capita level of output for the following derivations.

2.2 Solutions and Discussions

We solve this model along a BGP, which requires all the variables to grow at some constant rates and can be viewed as the steady state of the economy. Obviously, the labor shares of u_t and l_t are both constant along the BGP (denoted as u^* and l^*, respectively).[6] As a result, (5) can be reformulated into the following form:

$$g_h^* = \left(\frac{\dot{h}_t}{h_t}\right)^* = \eta h_t^{\zeta-1} u^* - \delta_0. \tag{6}$$

Since the growth rate g_h^* and the labor share u^* are both constants, the per capita human capital stock h_t should also be constant here (denoted as h^*).

In this subsection, we first derive the results in a competitive equilibrium, and then provide the optimal allocations for comparison. Our discussions on this benchmark model primarily focus on whether the population growth rate is positive or negative under different scenarios, and on determining the values of some important parameters for further discussions on the general model presented in Section 3.

2.2.1 Competitive Equilibrium in the Benchmark Model

Based on the above analysis, we have the following proposition.

Proposition 1. *Along a BGP, the allocations of population growth rate n_t and labor share in education u_t in equilibrium are jointly determined by the following two functions:*

$$u_1(n_t) = \frac{\delta_0}{\bar{b}} \cdot \frac{-n_t + \bar{b} - \delta_0}{\varepsilon(\gamma - 1)n_t + \bar{\rho} + (3 - \zeta)\delta_0}, \tag{7}$$

[6]The key reason is that these variables are ratios within the range of $[0, 1]$. We can find similar proofs in relating papers like Cong et al. (2021) and Chang et al. (2024).

and

$$u_2(n_t) = -\frac{1}{\varepsilon \bar{b}} \left[\varepsilon \gamma n_t + \bar{\rho} + (1+\varepsilon)\,\delta_0\right] + 1. \tag{8}$$

The equilibrium levels are denoted as n^* and u^*, respectively. Given these results, the equilibrium levels of other variables are all constants, which are given as:

$$l^* = \frac{1}{\bar{b}}\left(n^* + \delta_0\right), \qquad h^* = \left(\frac{\delta_0}{\eta u^*}\right)^{\frac{1}{\zeta - 1}},$$
$$w^* = (h^*)^{\sigma_P - 1}, \qquad c^* = y^* = (1 - l^* - u^*)\, w^* h^*.$$

The proof is shown in Appendix A.1. By combining equations (7) and (8), we derive a quadratic equation with one unknown variable n_t, which may lead to at most two solutions. From equation (8), we know that the slope of this linear function is $-\gamma/\bar{b} < 0$, and its intercept when $u_t = 0$ is

$$n_{\text{intercept}} = \frac{\varepsilon \bar{b} - \rho - (1+\varepsilon)\delta_0}{\varepsilon \gamma},$$

which is very close to zero since the importance of descendants ε and the efficiency of raising children $\bar{b}$ do not take very large values (see the calibration in Section 3.5.1) in reality. As a result, the equilibrium level of the population growth rate n^* usually takes a negative value in this benchmark model, which is consistent with the results shown in Jones (2022) but slightly different from those shown in Boikos et al. (2023). For clarity, we show the two functions that determine the equilibrium in Figure 1.

We can also discuss a special scenario to make the equilibrium simpler, where the reverse of elasticity of intertemporal substitution $\gamma = 1$. In this scenario, (7) reduces to a linear function, and the solutions of the model can be uniquely determined as

$$u^* = \frac{\delta_0(\delta_0 + \bar{\rho})}{\varepsilon \bar{b}[\bar{\rho} + (2-\zeta)\delta_0]},$$

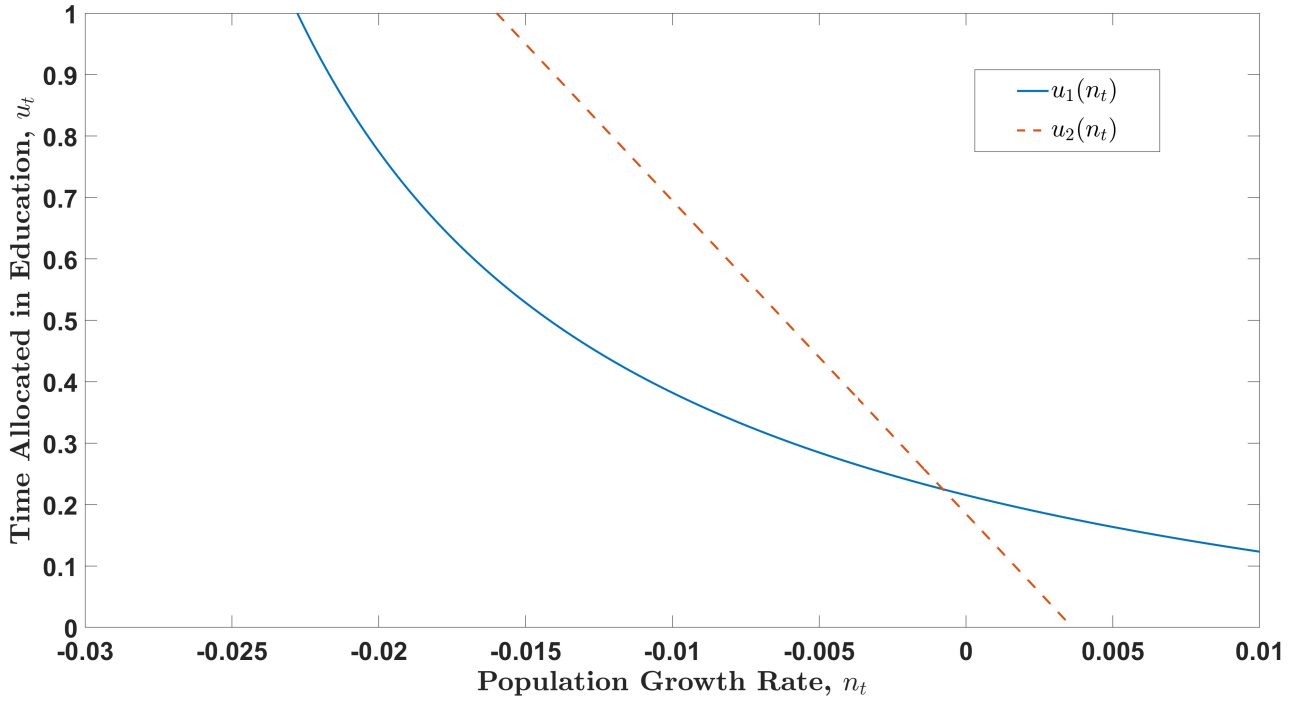


Figure 1: Time allocated in education u^* and population growth rate n^* without data (competitive equilibrium)

Note. This figure shows the time allocated in education and population growth rate in a competitive equilibrium in the benchmark model without data. The function $u_1(n_t)$ is hyperbolic when $\gamma > 1$, but only the right branch lies within the range of $u \in [0, 1]$, given the standard values of other parameters. The function $u_2(n_t)$ is linear. The intersection of these two functions represents the equilibrium levels of n^* and u^* in this benchmark model.

and

$$n^* = -\frac{\delta_0(\delta_0 + \bar{\rho})}{\varepsilon[\bar{\rho} + (2-\zeta)\delta_0]} - \left(1 + \frac{1}{\varepsilon}\right)\delta_0 - \frac{\bar{\rho}}{\varepsilon} + \bar{b}. \tag{9}$$

In this scenario, we can almost confirm that n^* is negative since $\bar{b}$ is typically not very large by definition. The equilibrium in the benchmark model can be viewed as the best approximation of reality. As a result, we set $\gamma = 1$ in the following discussions and calibrate the parameters according to this model.

2.2.2 Optimal Allocation in the Benchmark Model

We further discuss this benchmark model in the context of optimal allocations, where the social planner considers the total welfare of the entire economy. Thus, the maximization

problem should be slightly different from that in the equilibrium allocation, which is shown as follows:[7]

$$\max_{c_t, L_t} \int_0^\infty L_t e^{-(\bar{\rho}+\delta_0)t} \frac{(L_t^\varepsilon c_t)^{1-\gamma}}{1-\gamma} \mathrm{d}t,$$

subject to,

$$c_t = (1 - l_t - u_t) h_t^{\sigma_P}, \tag{10}$$

$$\dot{L}_t = [b(l_t) - \delta_0] L_t,$$

$$\dot{h}_t = \eta h_t^\zeta u_t - \delta_0 h_t.$$

We use the subscript "S" to denote the results derived in this scenario. Then, we have the following proposition.

Proposition 2. *Along a BGP, the optimal allocations of population growth rate n_t and labor share in education u_t are jointly determined by the following two equations:*

$$u_{S,1}(n_t) = \frac{\sigma_P \delta_0}{\bar{b}} \cdot \frac{-n_t + \bar{b} - \delta_0}{[\varepsilon(\gamma - 1) - 1] n_t + \bar{\rho} + (2 + \sigma_P - \zeta)\delta_0}, \tag{11}$$

and

$$u_{S,2}(n_t) = -\frac{1}{\varepsilon \bar{b}} [(\varepsilon\gamma - 1)n_t + \rho + (1 + \varepsilon)\delta_0] + 1. \tag{12}$$

The solutions are denoted as n_S^* and u_S^*, respectively. Given these results, the optimal levels of l_S^* and h_S^* are given similarly as those shown in Proposition 1, and c_S^* and y_S^* are shown as

$$c_S^* = y_S^* = (1 - l_S^* - u_S^*)(h_S^*)^{\sigma_P}.$$

The proof is shown in Appendix A.2. Similar to the scenario of competitive equilibrium, the optimal allocation is also determined by the intersections of a hyperbolic function (11) and a linear function (12). However, the difference lies in the slope of the linear function,

[7]Notice that we add an L_t term to the utility in this scenario. This addition does not affect the results of the model when the growth of L_t is set as exogenous, as in some semi-endogenous growth models. However, the results will change in our model since we consider population growth to be endogenous.

which now becomes $(1-\varepsilon\gamma)/(\varepsilon\bar{b})$. The parameter ε can be viewed as the relative importance of the consumption of descendants of the consumer, which is usually smaller than 1 since current consumption is typically considered more important. Thus, as we can see in Figure 2, the slope of (11) becomes positive, leading to multiple potential intersections in the optimal allocation.

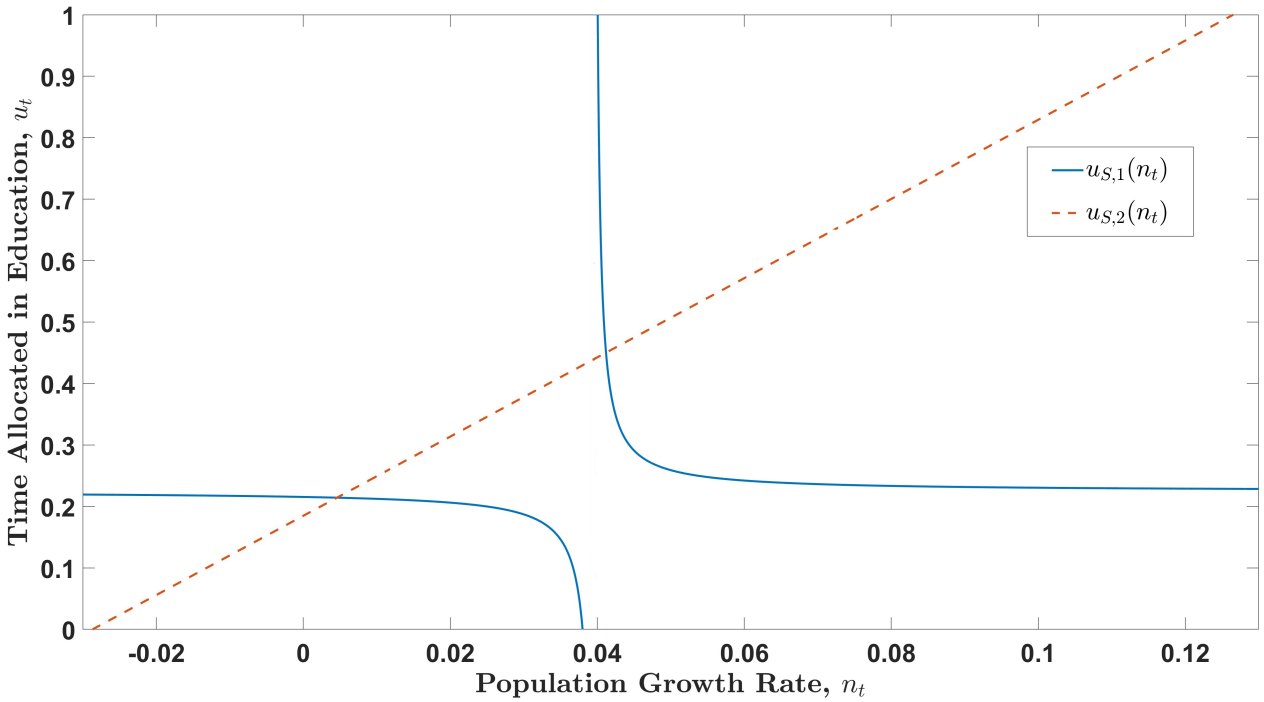


Figure 2: Time allocated in education u_S^* and population growth rate n_S^* without data (optimal allocation)

Note. This figure shows the time allocated in education and population growth rate in the optimal allocation in the benchmark model without data. The function $u_{S,1}(n_t)$ is hyperbolic, while $u_{S,2}(n_t)$ is linear. The intersections represent the equilibrium levels of n_S^* and u_S^* in this benchmark model.

However, we can further derive that the intersection with higher n_S^* and u_S^* is not feasible in the model. From (11), the left and right branches are divided by the line (given that $\gamma = 1$)

$$n_t = \bar{\rho} + (2 + \sigma_P - \zeta)\delta_0.$$

Notice that the time allocated to raising children l_t should not exceed 1, so the maximum value of n_t in this model should be $n_t \leq \bar{b} - \delta_0$ from the definition shown in equation (1).

Given the standard values of the parameters (shown in Table 2 in Section 3.5.1), we can prove that the only solution to this model is derived by the intersection with lower n_S^* and u_S^*. From Figure 1 we see that, the optimal population growth rate n_S^* is much larger than that in the competitive equilibrium, and can be even positive given some appropriate values of parameters.

Our benchmark model shares some similarities in model settings with those in the J model, but we switch to the framework of human capital accumulation instead of the idea-based growth model. There are also some differences between our model and that presented in Boikos et al. (2023), although both consider the effect of human capital accumulation in an economy with a negative population growth rate. Our model asserts that the population growth rate is determined endogenously and introduces human capital accumulation to demonstrate that the "empty planet" result cannot be fully mitigated merely by adding this labor augmentation effect.[8] There must be a stronger incentive for consumers to devote their time to raising children, hence the introduction of the data economy, which will be discussed in the next section.

3. General Digital Economy with Data

In this section, we introduce a general model with data based on the aforementioned benchmark model. Some of the notations are reused in this section to make the model more explicit. The data generation process is detailed in Farboodi and Veldkamp (2022) and Xie and Zhang (2023), which we refer to as "producer data". We innovate to incorporate the model on human capital accumulation with endogenous decisions on raising children, leading to varying population growth rates under different scenarios. We will first outline the economic environment and discuss in detail the deviations from the benchmark model, and then derive the results in competitive equilibrium and optimal allocations, respectively.

[8] Although we observe a positive population growth rate in the optimal allocations, this result is driven by the social planner who prioritizes total welfare. This outcome is similar to the middle steady state shown in the J model.

3.1 Economic Environment

The economic environment of the general model we mainly focus on in this paper is shown in Table 1, and for comparison, we also present the economic environment of the benchmark model without data. The benchmark model is a special scenario of our general model with $\theta = 0$, where the only distinction lies in the inclusion of a data sector in the latter.

The determination of the population growth rate remains the same as in the benchmark model, where (1) and (2) still hold in this model. In addition to the three activities that consumers can allocate their time to, they can also devote time to data generation, with the corresponding ratio denoted as s_t. The utility function is still given by (1), but the consumer's budget constraint changes to the following form:

$$c_t + p_{D,t} D_t = (1 - l_t - u_t) w_t h_t + \pi_t, \tag{13}$$

where D_t is the data purchased by the consumer and $p_{D,t}$ is the corresponding price. π_t is the profit from final good producer, which is owned by the consumers. Other variables are defined similarly to those in the benchmark model.

Table 1: Economic Environment

Meaning	Equation (general model)	Equation (benchmark model)
Population growth	$\frac{\dot{L}_t}{L_t} = n_t = b(l_t) - \delta_0$	Same as left
Generation 0 utility	$U_0 = \int_0^\infty e^{-(\bar{\rho}+\delta_0)t} u(c_t, L_t) \mathrm{d}t$, $L_0 = 1$	Same as left
Flow utility	$u(c_t, L_t) = \frac{(L_t^\xi c_t)^{1-\gamma}}{1-\gamma}$	Same as left
Per capita consumption	$c_t = y_t = Y_t / L_t$	Same as left
Human capital accumulation	$\dot{h}_t = \eta h_t^\zeta D_t^\theta u_t^{1-\theta} - \delta_0 h_t$	$\dot{h}_t = \eta h_t^\zeta u_t - \delta_0 h_t$
Production function	$Y_t = (1 - l_t - u_t - s_t) h_t^{\sigma_P} L_t$	$Y_t = (1 - l_t - u_t) h_t^{\sigma_P} L_t$
Data generation function	$D_t = y_t s_t h_t^{\sigma_D} L_t$	—

Data are generated as a byproduct in the production process and are owned by the firms

(this will be discussed in the following section). Consumers use data to substitute part of the time allocated to education, u_t, to accumulate human capital in the following form:

$$\boxed{\dot{h}_t = \eta h_t^{\zeta} D_t^{\theta} u_t^{1-\theta} - \delta_0 h_t,} \tag{14}$$

where θ represents the importance of data in human capital accumulation. Note that due to the nonrival nature of data, each person in the economy has full access to the entire dataset available, so we use D_t instead of D_t/L_t in this equation.[9]

The production sector differs significantly from the benchmark model. Besides final goods, firms also generate data as a byproduct of production. Specifically, the production function is now modified to:

$$Y_t = (1 - l_t - u_t - s_t) h_t^{\sigma_P} L_t,$$

where $(1 - l_t - u_t - s_t)$ represents the labor allocated to final goods production. The data generation function is given by:

$$\boxed{D_t = y_t s_t h_t^{\sigma_D} L_t,} \tag{15}$$

where σ_D represents the augmentation effect of human capital in data generation, and s_t denotes the labor allocated to data generation. Following Xie and Zhang (2023), per capita output y_t here serves merely as a reference for data generation, and is not one of the input factors. Thus, output is not used in this process.[10] D_t represents the total provision of data in the economy, which can be sold to consumers in a nonrival manner.

[9]One might argue that not everyone in society has access to the entire dataset. However, it is reasonable to assume that consumers can learn beyond their personal experiences. For instance, while there may be some variation across schools, students generally use the same textbooks, leading to a largely uniform educational experience. Furthermore, in today's information-driven society, access to information has become increasingly democratized through the internet, as traditional barriers are being dismantled. Therefore, we find it reasonable to adopt this assumption, which is a key element of our model.

[10]In Xie and Zhang (2023), data are divided into three different types based on their source: consumer data, producer data, and nature data. In Chang et al. (2024), these different types of data are compared within the framework of human capital accumulation, showing that producer data are most suitable for this framework. The key reason is that data are used by consumers in this model, so data should be generated by other agents, such as producers. Moreover, learning from production by employing producer data is more realistic in this context.

Data Intermediary. We follow the settings used in Jones and Tonetti (2020) to model the data intermediary. The nonrival nature of data makes the market for this special factor non-competitive. Thus, we assume that there is a monopolistic data intermediary that collects data from the producer and then sells it to each consumer. This data intermediary is subject to free entry, so it earns zero profit. Thus, the profit maximization problem of the data intermediary is:

$$\max_{p_{D,t}, D_t^d} \; p_{D,t} D_t L_t - p_{b,D,t} D_t^d,$$

subject to

$$D_t \leq D_t^d,$$

where $p_{b,D,t}$ is the price at which data is bought from the producer, and D_t^d is the demand for data. Clearly, we have the following result:

$$D_t = D_t^d, \quad \text{and} \quad p_{b,D,t} = p_{D,t} L_t. \tag{16}$$

This result provides the main incentive for consumers to devote more time to raising children (and thus leads to a higher population growth rate) in this model. From (16), we see that the larger the population, the cheaper the data consumers will get to fuel human capital accumulation.

3.2 Competitive Equilibrium

Unlike the benchmark model, the growth rates of the key variables in this model are no longer zero. Through derivations, we arrive at the following proposition.

Proposition 3. *In the competitive equilibrium, along a BGP, the growth rates of the key variables are given by:*

$$g_h^* = \Gamma n^*, \quad g_c^* = g_y^* = \sigma_P \Gamma n^*, \quad g_D^* = \frac{1-\zeta}{\theta}\Gamma n^*,$$
$$g_{p_D}^* = \frac{\zeta + \theta\sigma_P - 1}{\theta}\Gamma n^*, \quad \textit{and} \quad g_w^* = (\sigma_P - 1)\Gamma n^*,$$

where the parameter Γ *is derived as*

$$\Gamma \equiv \frac{\theta}{1 - \zeta - \theta(\sigma_P + \sigma_D)}. \tag{17}$$

Meanwhile, the population growth rate n^* *and the labor share allocated to education* u^* *are endogenously determined by the following equations:*

$$u_{D,1}(n_t) = \frac{1}{\bar{b}} \cdot \frac{\left(\bar{b} - n_t - \delta_0\right)(1-\theta)(\Gamma n_t + \delta_0)}{\{\varepsilon(\gamma - 1) + \Gamma\left[\gamma\sigma_P - \sigma_P + 2 - \theta - \zeta\right]\} n_t + \bar{\rho} + (3 - \theta - \zeta)\delta_0}. \tag{18}$$

and

$$u_{D,2}(n_t) = -\frac{1-\theta}{\varepsilon\bar{b}(1+\theta)}\left\{\left[(\gamma - 1)(\varepsilon + \Gamma\sigma_P) + \varepsilon\right] n_t - \varepsilon\bar{b} + \bar{\rho} + (1+\varepsilon)\delta_0\right\}. \tag{19}$$

Given the equilibrium values of u_D^* *and* n_D^**, the labor share allocated to raising children* l_D^* *and data production* s_D^* *are derived by the following equations, respectively:*

$$l_D^* = \frac{n_D^* + \delta_0}{\bar{b}},$$

and

$$s_D^* = \frac{\theta}{1-\theta}u_D^*. \tag{20}$$

The proof is shown in Appendix B.1. Unlike the benchmark model, we can no longer derive two linear functions even if we set $\gamma = 1$. This means that an explicit solution cannot be derived in this scenario. Given the same parameter values used in Figure 1 (except that $\gamma = 1$ here), we plot the functions of $u_{D,1}(n_t)$ and $u_{D,2}(n_t)$ together in Figure 3. Since the slope of the linear function $u_2(n_t)$ is negative, and only the right branch of the hyperbolic function $u_1(n_t)$ is within the range of $u \in [0, 1]$, there is only one feasible solution to this

model. Numerically, this solution still gives a negative population growth rate n_D^*, but it is larger than that derived in the benchmark model.

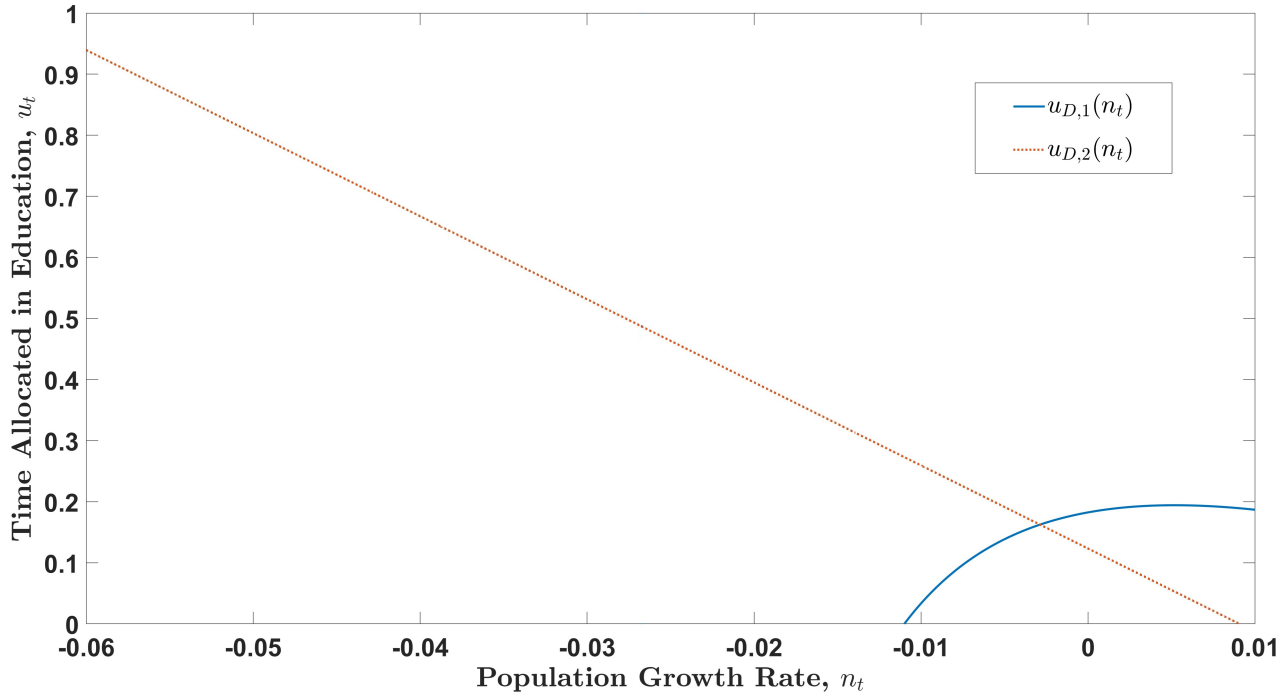


Figure 3: Time allocated in education u_D^* and population growth rate n_D^* in the general model (competitive equilibrium)

Note. This figure shows the time allocated to education and the population growth rate in a competitive equilibrium in the general model. $u_1(n_t)$ is a hyperbolic function, but only its right branch lies within the range of $u \in [0, 1]$ given the standard values of other parameters. $u_2(n_t)$ is a linear function. The intersection represents the equilibrium levels of n_D^* and u_D^* in this general model.

Data work as a substitute for labor allocated to education in the human capital accumulation process. Although the generation of this new factor still requires labor, there is also an augmentation effect from the output level, which makes the usage of this factor more efficient (see s_t in the data generation function (15)). Additionally, the larger the population, the cheaper the data consumers can get from the producer. As a result, we see that the population growth rate is pushed upward when data are introduced into the model. This shows a potential way to mitigate the "empty planet" results discussed in Jones (2022).

We can further show the BGP levels of other key variables explicitly, leading to the following proposition. The proof can also be found in Appendix B.1. These results will be

used in the transition dynamics discussed in Section 4.

Proposition 4. *Given the BGP growth rates of the variables and the labor shares in the competitive equilibrium shown in Proposition 3, the BGP levels of the key variables are derived as follows:*

$$h_D^* = \left[\frac{\Gamma n^* + \delta_0}{\eta(1-l^*-u^*-s^*)^\theta (s^*)^\theta (u^*)^{1-\theta}}\right]^{\frac{1}{\zeta-1+\theta(\sigma_P+\sigma_D)}} L_t^{\frac{\theta}{1-\zeta-\theta(\sigma_P+\sigma_D)}} \equiv \tilde{h}^* L_t^{\frac{\theta}{1-\zeta-\theta(\sigma_P+\sigma_D)}}, \tag{21}$$

$$y_D^* = (1-l^*-u^*-s^*)(h^*)^{\sigma_P} \equiv \tilde{y}^* L_t^{\frac{\sigma_P\theta}{1-\zeta-\theta(\sigma_P+\sigma_D)}},$$

$$D^* = (1-l^*-u^*-s^*)s^*(h^*)^{\sigma_P+\sigma_D} L_t \equiv \tilde{D}^* L_t^{\frac{1-\zeta}{1-\zeta-\theta(\sigma_P+\sigma_D)}},$$

$$p_D^* = \frac{1}{1-l^*-u^*-2s^*}(h^*)^{-\sigma_D} L_t^{-1} \equiv \tilde{p}_D^* L_t^{-1-\frac{\sigma_D\theta}{1-\zeta-\theta(\sigma_P+\sigma_D)}},$$

$$w_D^* = \frac{(1-\theta)(1-l^*-u^*-s^*)s^*}{\theta(1-l^*-u^*-2s^*)u^*}(h^*)^{\sigma_P-1} \equiv \tilde{w}^* L_t^{\frac{(\sigma_P-1)\theta}{1-\zeta-\theta(\sigma_P+\sigma_D)}},$$

and

$$c_D^* = \frac{(1-l^*-u^*-s^*)[(1-\theta)(1-l^*)-u^*]s^*}{\theta(1-l^*-u^*-2s^*)u^*}(h^*)^{\sigma_P} \equiv \tilde{c}^* L_t^{\frac{\sigma_P\theta}{1-\zeta-\theta(\sigma_P+\sigma_D)}} = y_D^*.$$

The levels of the variables shown in Proposition 4 can all be divided into two parts: the constant part, which is represented by tilde symbols, and the growing part, which is related to population growth. We will further study the transition dynamics of these variables in Section 4.

3.3 Optimal Allocations

Similar to the benchmark model, we also discuss the optimal allocations in the general model. Specifically, the social planner's problem is show as follows.

$$\max_{c_t, L_t} \int_0^\infty L_t e^{-(\bar{\rho}+\delta_0)t} \frac{(L_t^\varepsilon c_t)^{1-\gamma}}{1-\gamma} \mathrm{d}t,$$

subject to,

$$c_t = (1 - l_t - u_t - s_t)h_t^{\sigma_P}, \tag{22}$$

$$D_t = (1 - l_t - u_t - s_t)s_t h_t^{\sigma_P + \sigma_D} L_t,$$

$$\dot{L}_t = [b(l_t) - \delta_0] L_t,$$

$$\dot{h}_t = \eta h_t^{\zeta} D_t^{\theta} u_t^{1-\theta} - \delta_0 h_t.$$

Solving this problem, we have the following proposition.

Proposition 5. *In the optimal allocations, along a BGP, the growth rates of the key variables are all the same as those shown in Proposition 3. Moreover, the population growth rate and the labor share allocated in education are endogenously determined by the following equations:*

$$u_{DS,1}(n_t) = \frac{\sigma_P}{\bar{b}} \cdot \frac{\left(\bar{b} - n_t - \delta_0\right)(1-\theta)(\Gamma n_t + \delta_0)}{[\varepsilon(\gamma - 1) + \Gamma(\gamma\sigma_P + 1 - \sigma_D\theta - \zeta) - 1]\, n_t + \bar{\rho} + (2 + \sigma_P - \sigma_D\theta - \zeta)\delta_0}, \tag{23}$$

and

$$u_{DS,2}(n_t) = -\frac{1-\theta}{\bar{b}[\varepsilon(1+\theta) - \theta]} \left\{[(\gamma - 1)(\varepsilon + \Gamma\sigma_P) + \varepsilon - 1]\, n_t - \bar{b}\varepsilon + \bar{\rho} + (1+\varepsilon)\delta_0\right\}. \tag{24}$$

The solutions are denoted as n_{DS}^* and u_{DS}^*, respectively. Given these results, the labor share allocated in raising children l_{DS}^* and data production s_{DS}^* can be derived similarly to those shown in Proposition 3. The levels of other variables can also be derived similarly to those shown in Proposition 4, except that we only have h_{DS}^*, D_S^*, y_{DS}^*, and c_{DS}^* here.

The proof is shown in Appendix B.2. Similar to the scenario shown in Proposition 2, the slope of $u_{DS,2}(n_t)$ is positive since we usually assume $\varepsilon < 1$. In Figure 4, we see that there are two intersections between the functions $u_{DS,1}(n_t)$ and $u_{DS,2}(n_t)$. Both of these solutions to the model are feasible: the lower solution has a negative n_{DS}^*, while the higher one is positive. This is similar to the multiple steady states in the optimal allocation shown in the

J model, where the former solution corresponds to the "empty planet" result and the latter one corresponds to the "expanding cosmos" result. We will further discuss these multiple solutions in the next subsection given the parameter values.

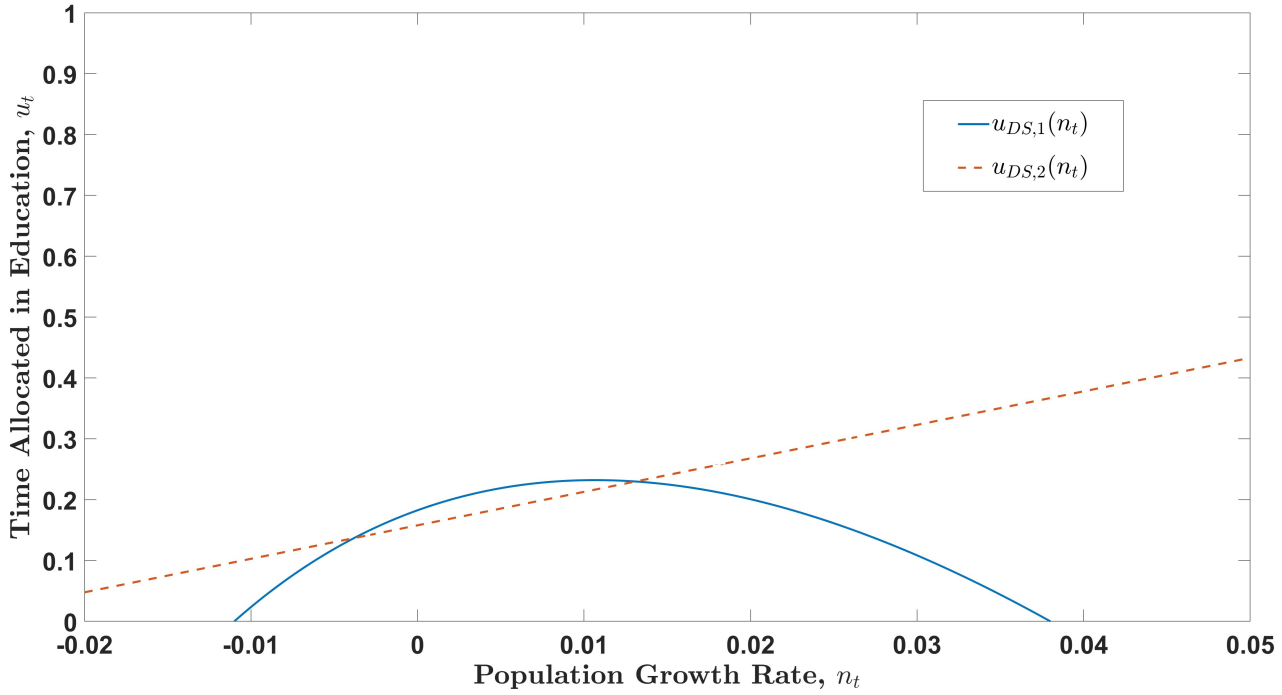


Figure 4: Time allocated in education u^* and population growth rate n^* in the general model (optimal allocation)

Note. This figure shows the time allocated to education and the population growth rate in the optimal allocation in the general model. $u_{DS,1}(n_t)$ is a hyperbolic function, but only its right branch lies within the range of $u \in [0, 1]$ given the standard values of other parameters. $u_{DS,2}(n_t)$ is a linear function. The two intersections represent the multiple solutions of n^*_{DS} and u^*_{DS} in this general model.

3.4 Inefficiency in the Equilibrium

The comparison between the competitive equilibrium and optimal allocations highlights distinct differences, primarily stemming from inefficiencies in the equilibrium. In this model, three main sources of inefficiency arise. Two are present in both the benchmark and general models, while the third is unique to the general model. This third inefficiency, introduced by the inclusion of data, allows us to isolate and evaluate the impact of the additional settings in the general model. Understanding these inefficiencies is critical before proceeding to the

quantitative analysis, as they play a fundamental role in shaping the dynamics of the model.

The first inefficiency arises from the human capital externality, a well-established concept in growth models, particularly those based on Lucas (1988). In such models, consumers accumulate human capital by investing time and other resources, yet the overall productivity of the economy depends on the average level of human capital. This creates a spillover effect across consumers of different types, as each consumer's productivity benefits from the collective human capital stock. This externality is a fundamental inefficiency introduced when applying the human capital framework to the model.

The second inefficiency stems from population externality, which arises due to differences in utility structures between the two scenarios, as discussed in Barro and Becker (1989). In this model, we employ "generalized Barro-Becker preferences" for the utility function, which explicitly accounts for the total population level. This introduces a key distinction between the optimization objectives in the competitive equilibrium, where consumers maximize their own utility, and in the optimal allocations, where the social planner seeks to maximize the total sum of utilities in the economy. Specifically, the planner's objective is to maximize $L_t u(c_t, L_t)$, rather than just $u(c_t, L_t)$ as consumers do. While this difference may be negligible in models where L_t is absent from the utility function (or say, L_t is exogenous in those models), it plays a significant role here. As demonstrated in the numerical solutions section, the total utility differs substantially between the two scenarios, underscoring the importance of this externality in shaping the model's outcomes.

Finally, and most distinctively in our framework, is the nonrival nature of data—a hallmark of the digital economy. Nonrivalry has been extensively explored in seminal papers examining the impact of data, influencing the economy in various ways. For instance, data can be nonrival among different intermediate producers, across time periods, or even among countries (Jones and Tonetti, 2020; Cong et al., 2021; Chang et al., 2023). In this paper, we introduce a novel dimension of nonrivalry, demonstrating that data can be utilized without depletion by different consumers during human capital accumulation, as discussed in equation (14). This unique nonrival property is pivotal to our findings, particularly when

illustrating long-run growth within the human capital framework—a result not observed in the benchmark model. In the following sections, we will show how this aspect of data nonrivalry significantly influences the model's long-term dynamics.

3.5 Numerical Solutions and Discussions

Given all the models shown above, we now turn to presenting some numerical solutions and comparing the results of these four scenarios: the competitive equilibrium and optimal allocation of the benchmark model without data, and those two scenarios in the general model with data. We first perform calibrations to estimate the values of the key parameters, and then compare population growth rate, labor allocations, and welfare levels among the different scenarios. Additionally, we provide some implications on the labor shares when the importance of data, θ, changes.

3.5.1 Calibrations

Table 2 shows the estimated values of the parameters. We choose these values to make our model more realistic, but the general results remain robust across a range of alternative values.

Table 2: Parameters for Studying the Model

Parameters	Meaning	Value	Comment
$\bar{b}$	Efficiency of raising children	0.049	Estimated from benchmark model
δ_0	Death rate	1/90	Life expectancy 90 years
$\bar{\rho}$	Subjective discount rate	1/90	Discounting from mortality $\bar{\rho} = \delta_0$
ε	Importance of descendants	0.718	Estimated from benchmark model
γ	Inverse of the intertemporal elasticity of substitution	1	Simplifies analysis
ζ	Spillover effect of human capital	0.4	Lucas (1988)
θ	Importance of data in human capital accumulation	0.2	Deduced from DeJong and Ingram (2001)
η	Efficiency of human capital accumulation	1	Standard
σ_P	Efficiency of human capital in final good production	1	Normalization
σ_D	Efficiency of human capital in data generation	1	Normalization

Our calibrations are divided into three parts. First, since it is not easy to measure human capital directly, it is convenient to normalize $\sigma_P = 1$ and $\sigma_D = 1$ to avoid introducing further augmentation or reduction effects when these two parameters are not equal to one. Meanwhile, η is also standardized to 1, which is a common practice when determining the efficiency term in the innovation possibility frontier in idea-driven models. It is also important to note that we set $\gamma = 1$ as this simplifies the derivation of the models.

Second, we follow the methods shown in Jones (2022) to estimate some parameters from real data. It is convenient to set the death rate δ_0 to $1/90 \approx 1.1\%$, implying a life expectancy of approximately 90 years. The discount rate $\bar{\rho}$ is set equal to δ_0 since we assume most of the discounting arises from mortality. Next, we estimate the equilibrium levels of two variables, n^* and l^*, based on real data. We assume that people typically spend 1/8 of their lifetime raising children (i.e., $l^* = 1/8$), and the average population growth rate in advanced economies like the OECD, Japan, and the U.S. is approximately $n^* = -0.5\%$. Then, from (1), we derive that

$$\bar{b} = \frac{n^* + \delta_0}{l^*} = 0.049.$$

For the estimation of ε, we calculate from (9) to achieve $n^* = -0.5\%$ and find that $\varepsilon = 0.718$. Our estimations are all derived from the equilibrium results in the benchmark model, which we believe is the scenario closest to reality with a declining population. After determining these parameters, we use them in the optimal allocation and the general model to demonstrate the ideal results from introducing data into the economy.

Finally, the two parameters ζ and θ, which are related to human capital accumulation, should also be estimated. Following Lucas (1988), who built an endogenous growth model, the spillover effect of human capital is estimated to be approximately 0.4. On the other hand, the estimation of θ comes from DeJong and Ingram (2001), where they estimate that the contribution of education to human capital accumulation is approximately 0.8. Additionally, from the calibrations shown in Chang et al. (2024), using the Human Capital Index (HCI) provided by the World Development Indicators database of the World Bank, they find that

the importance of data in some advanced economies is approximately 0.26. Combining these two estimations and considering that other factors may affect human capital accumulation, we set $\theta = 0.2$. In the following analysis, we present the comparative statics for this parameter to provide further insights into its effects on the model. We also examine the robustness of our results by considering parameter values of 0.15 and 0.25 when comparing welfare across different scenarios. It should also be noted that θ should be within $\theta \leq (1-\zeta)/(\sigma_P + \sigma_D) = 0.3$ from (17), as we need the BGP growth rates to be positive when the population growth rate becomes positive.

3.5.2 Comparative Statics on Data Importance

The most important and newly introduced parameter in the general model is the data importance θ, which represents the contribution of data relative to education time in human capital accumulation. The larger the value of θ, the higher the proportion of data that substitutes education time, thereby creating a greater incentive to raise more children for more data. The comparative statics with respect to θ are divided into two parts: we first show the labor shares allocated to the four sectors, and then show the output growth rates in both the competitive equilibrium and the optimal allocation. Note that the subscript "*alloc*" represents the corresponding results derived in different scenarios of the two models we discuss in this paper.

Figure 5 shows the comparative statics on labor shares. Since there is only one feasible solution when θ is relatively small, we only show the time allocations in the high state of the optimal allocation to make the change smoother.[11] From the two panels, we see that the time allocated to data generation is zero when $\theta = 0$, which corresponds to the benchmark model. Additionally, we observe an increasing share of time allocated to raising children

[11]When θ is relatively small, the only feasible solution corresponds to the high state, which features a positive population growth rate. As θ increases, although the low state emerges, the population growth rate in the high state rises to an even higher level. From the discussions on transition dynamics in Section 4, we know that the economy will never reach the low state if it starts from the competitive equilibrium. Thus, we demonstrate how the introduction of data promotes population growth and further enhances the welfare level in the economy.

and data generation in both scenarios as θ increases. This illustrates the effect of data on population growth: as data becomes more important, it becomes more efficient to use more data in human capital accumulation, leading to a higher population growth rate to promote data generation. Moreover, as the population growth rate increases, the total volume of data also grows, which consequently reduces the time allocated to education in human capital accumulation. This creates a feedback loop between data volume and population scale. However, a key difference between the competitive equilibrium and the optimal allocation is that, in the latter scenario, people devote much more time to raising children, and the time allocated to final goods production is crowded out as θ increases. In the competitive equilibrium, the time allocated to education is crowded out. The reason lies in the different weights that consumers and the social planner place on the total population. It is more desirable in the optimal allocation to keep the time allocated to education nearly unchanged to ensure the economy's future growth.

In addition, Figure 6 shows the output growth rate and total welfare along a BGP as θ changes. Similar to the analysis shown above, we only show the results derived in the competitive equilibrium and the optimal allocation in the high state for comparison. We observe that the output growth rate in the former scenario decreases as θ increases. This is due to the negative population growth rate in the competitive equilibrium we derived, as the growth rates of other variables largely depend on this rate, similar to other semi-endogenous models first proposed in Jones (1995). However, when we focus on the total welfare levels of the two scenarios, we see increasing patterns as θ increases, which highlights the positive effect of data. As the contribution of this new factor becomes more significant, human capital accumulates with higher efficiency, leading to a higher overall welfare level for society. This figure demonstrates that, since the utility function in our model does not solely depend on consumption, the output growth rate alone cannot fully capture the improvement brought by data. Therefore, it is more reasonable to compare the discounted sum of welfare levels over all periods instead of focusing solely on the growth rate.

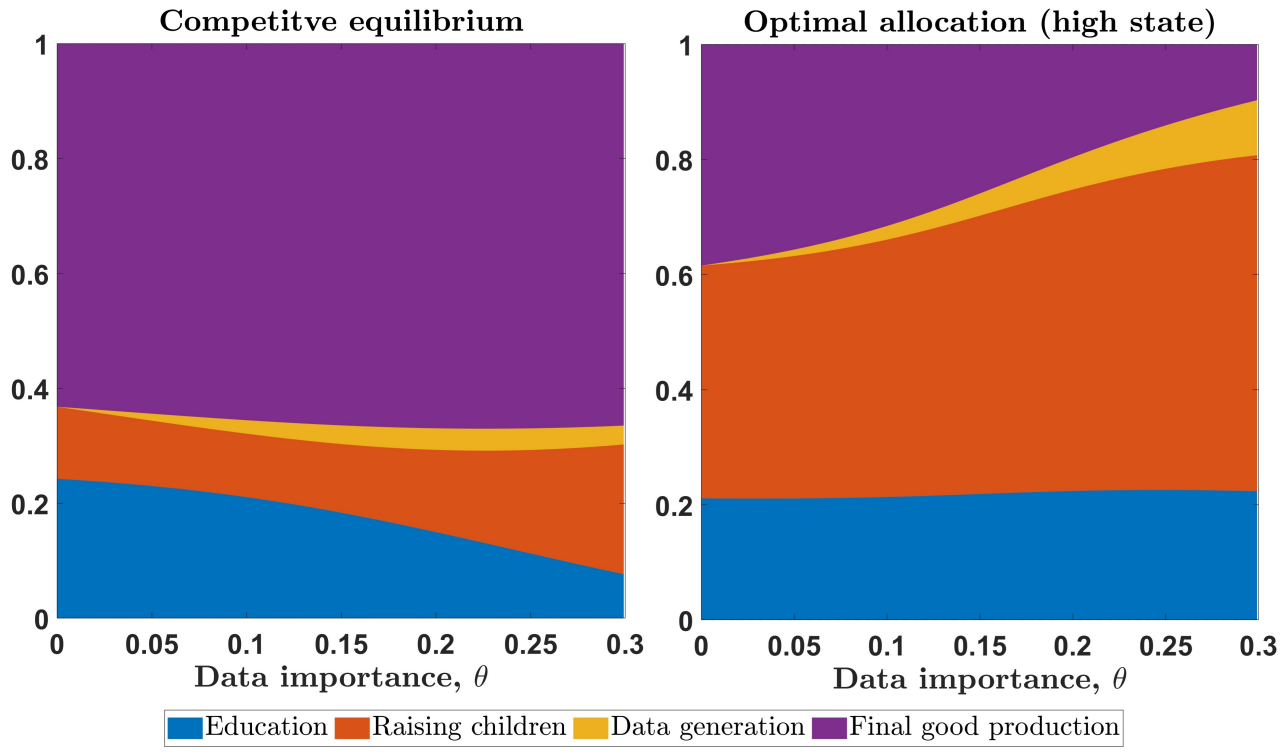


Figure 5: Time allocated in different activities in the general model

Note. These figures show the time allocated to the four activities in the general model: education (bottom and blue area), raising children (lower center and orange area), data generation (higher center and yellow area), and final goods production (upper and purple area).

3.5.3 Comparisons of Different Scenarios

In this section, we calculate six key variables in the four different scenarios to show the effect of data: population growth rate n^*_{alloc}, time allocated to education u^*_{alloc}, time allocated to raising children l^*_{alloc}, output growth rate $(g^*_y)_{alloc}$, total welfare U^*_{alloc}, and average welfare $\bar{U}^*_{alloc}$. Here, the subscript "alloc" represents the corresponding variables derived in different scenarios. The first four variables are derived directly from Propositions 1 to 5, while the last two variables can be derived from the utility function (3), i.e.,

$$U^*_{alloc} = \frac{\varepsilon n^*_{alloc} + (g^*_c)_{alloc}}{(\bar{\rho} + \delta_0 - n^*_{alloc})^2} + \frac{\ln\left(L_0^{\varepsilon}\tilde{c}^*_{alloc}\right)}{\bar{\rho} + \delta_0 - n^*_{alloc}}, \tag{25}$$

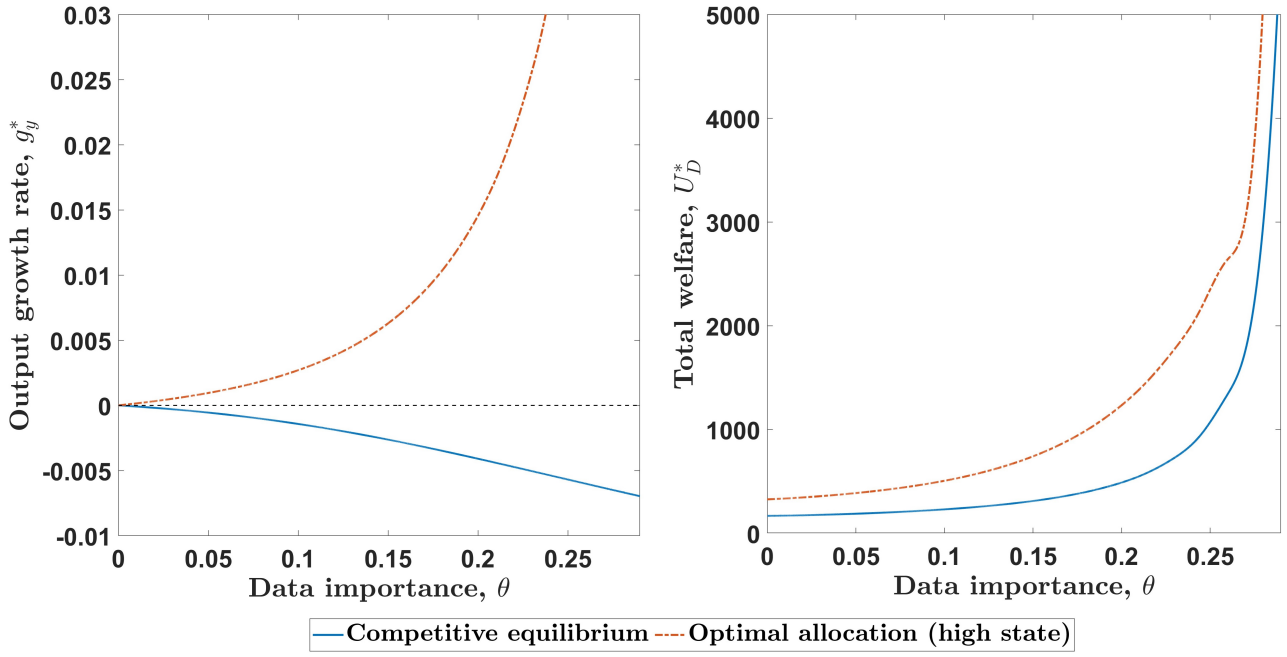


Figure 6: Comparative statics of the general model

Note. These figures show the comparative statics of the output growth rate and total welfare with respect to data importance θ in the general model. The blue solid lines represent the results derived in the competitive equilibrium, and the orange dotted lines represent the results derived in the optimal allocation.

and

$$\bar{U}^*_{alloc} = \frac{\varepsilon n^*_{alloc} + (g^*_c)_{alloc}}{(\bar{\rho} + \delta_0)^2} + \frac{\ln\left(L_0^{\varepsilon}\tilde{c}^*_{alloc}\right)}{\bar{\rho} + \delta_0}. \tag{26}$$

The derivation can be found in Appendix C.1. Note that (25) and (26) only exist when $\gamma = 1$ (the immediate utility function changes to a logarithmic form). The results for the benchmark value with $\theta = 0.2$ are presented in Table 3. Additionally, we provide robustness checks for other values of θ in Table C.1 in Appendix C.2. For clarity, we also show the results derived in the J model for comparison.

By comparing the results in the benchmark model and the general model shown in the table, we can easily observe the improvements caused by introducing data into the economy. In the competitive equilibrium, both the population growth rate and the welfare levels (total and average) increase when we substitute data for labor in human capital accumulation. It should also be noted that although the output growth rate changes from zero to a negative

Table 3: Comparisons of Different Scenarios ($\theta = 0.2$)

Variables	Benchmark model		General model			J model		
	CE	OA	CE	OA		CE	OA	
				Low state	High state		Low state	High state
n^*_{alloc}	-0.5%	0.87%	-0.41%	-0.80%	1.46%	-0.5%	0.26%	1.16%
l^*_{alloc}	1/8	0.40	0.14	0.06	0.52	1/8	0.28	0.46
u^*_{alloc}	0.24	0.21	0.15	0.07	0.22	—	—	—
$(g^*_y)_{alloc}$	0	0	-0.41%	-0.80%	1.46%	0	0.10%	0.43%
U^*_{alloc}	167.39	326.73	488.57	437.29	1223.43	—	—	—
$\bar{U}^*_{alloc}$	203.47	190.54	576.35	587.54	329.20	—	—	—

Note: This table shows the results of six key variables in the four different scenarios we discuss in this paper, in addition to the corresponding results shown in the J model. The columns labeled "CE" represent the results derived in competitive equilibria, while the columns labeled "OA" represent the results derived in optimal allocations. Note that there are multiple steady states in both our general model and the J model.

level, welfare is still improved since the initial level of consumption $\tilde{c}^*_{alloc}$ increases significantly. Additionally, due to the decreasing population in some scenarios, the average welfare levels become larger than the total levels. These average levels even become smaller when we turn to optimal allocations where there are positive population growth rates. As a result, it is more reasonable to compare the total welfare improvement instead of the average levels, since population L_t is also endogenous in our model.

Moreover, given the same results derived in the competitive equilibrium, our model shows distinct differences in optimal allocations compared with the J model. In our benchmark model, the BGP growth rates of all variables are always zero, while in the J model, those growth rates become positive in optimal allocations. This is due to the human capital accumulation framework we use in this paper. However, when it comes to the general model with data, these growth rates become non-zero and more aligned with reality. Finally, we find an important result when comparing the optimal allocations between the general model and the J model: the population growth rate in the low state (the "empty planet" state in the J model) is much lower and negative in the general model, while in the high state (the

"expanding cosmos" state in the J model), this rate becomes higher than that in the J model. If we can avoid the economy being trapped in the low state and ensure it remains in the high state, the introduction of data will move the economy to a more desirable steady state with larger population and output growth rates. We will discuss the stability of these multiple states in Section 4 to provide further implications for our model.

4. Transition Dynamics

Given the multiple steady states derived in the optimal allocation in the general model, it is important to discuss their stability to determine which state is more likely to be reached in the long run. In this section, we analyze the evolution of the general model before the economy reaches the BGP, providing a further investigation of our model. Specifically, we focus on the optimal allocations and leave the transitions in the competitive equilibrium to the Appendix D.1 as supplementary material. We first derive the differential equations of the model and then illustrate the evolution of the economy through a phase diagram and the changes of several key variables.

4.1 Methodology

Following Jones (2016), Cong et al. (2021), and Jones (2022), we use the "reverse shooting" method to derive the transition dynamics numerically. This method, first proposed by Judd (1998), has been used in several papers to show the dynamics of a system given an initial state and the equations of motion.

In Appendix D, we derive two systems of equations to describe the transition dynamics in competitive equilibrium and optimal allocation, respectively. These systems contain three "state-like" variables that can uniquely represent the evolution of the economy, and other variables can be derived from them. Moreover, since the economy grows at a constant rate along a BGP in the long run, some variables should be transformed into the "detrended" form to derive a constant steady state along the BGP. Specifically, in this model we only

need to "detrend" the average level of human capital stock h_t into $\tilde{h}_t = h_t L_t^{-\Gamma}$, in addition to the time allocated to education u_t and population growth rate n_t to form the "state-like" variables. Then, the systems can be represented by three equations with respect to $\dot{\tilde{h}}_t$, $\dot{u}_t$, and $\dot{n}_t$, and the steady state of $\tilde{h}_t$ can be derived by (21), while the other two variables are derived from solving the equations shown in Proposition 5. In the following proposition, we show the differential equations in the optimal allocation.

Proposition 6. *In the optimal allocation of the general model, the transition dynamics of the economy can be represented by the following three differential equations. Given $\tilde{h}_t$, u_t, and n_t, its change is shown as*

$$\dot{\tilde{h}}_t = \eta \tilde{h}_t^{\zeta+\theta(\sigma_P+\sigma_D)} \left(1 - \frac{n_t + \delta_0}{\bar{b}} - \frac{1}{1-\theta} u_t\right)^{\theta} \left(\frac{\theta}{1-\theta}\right)^{\theta} u_t - (\delta_0 + \Gamma n_t)\, \tilde{h}_t.$$

Then, given $\tilde{h}_t$, $\dot{\tilde{h}}_t$, u_t, and n_t, the differential equations with respect to $\dot{u}_t$ and $\dot{n}_t$ are derived from the following two linear equations:

$$\begin{aligned}
&-(\gamma+\theta)\frac{\frac{1}{\bar{b}}\dot{n}_t + \frac{1}{1-\theta}\dot{u}_t}{1 - \frac{1}{\bar{b}}(n_t+\delta_0) - \frac{1}{1-\theta}u_t} - \frac{\theta}{\bar{b}} \cdot \frac{u_t\dot{n}_t - (n_t+\delta_0)\dot{u}_t}{\left[(1-\theta)\left(1 - \frac{n_t+\delta_0}{\bar{b}}\right) - (1+\theta)u_t\right]\left(1 - \frac{n_t+\delta_0}{\bar{b}} - \frac{1}{1-\theta}u_t\right)} \\
&= \left[\sigma_P(1-\theta)\left(\frac{1 - \frac{1}{\bar{b}}(n_t+\delta_0)}{u_t} - \frac{1}{1-\theta}\right) + \sigma_D\theta + \zeta\right]\left(\frac{\dot{h}_t}{h_t} + \delta_0\right) - [(\gamma+\theta-1)\sigma_P + \theta\sigma_D + \zeta]\frac{\dot{h}_t}{h_t} - \\
&[\varepsilon(\gamma-1) - 1 + \theta]n_t - \bar{\rho} - 2\delta_0, \qquad (27)
\end{aligned}$$

and

$$\begin{aligned}
&-\gamma\frac{\frac{1}{\bar{b}}\dot{n}_t + \frac{1}{1-\theta}\dot{u}_t}{1 - \frac{1}{\bar{b}}(n_t+\delta_0) - \frac{1}{1-\theta}u_t} + \frac{\frac{1}{\bar{b}}\dot{n}_t + \theta\left(1 - \frac{n_t+\delta_0}{\bar{b}}\right)\dot{u}_t}{\left[(1-\theta)\left(1 - \frac{n_t+\delta_0}{\bar{b}}\right) - (1+\theta)u_t\right]\left[1 - \frac{n_t+\delta_0}{\bar{b}} - \frac{1}{1-\theta}u_t\right]} \\
&= \bar{b}\left[\varepsilon\left(1 - \frac{n_t+\delta_0}{\bar{b}} - \frac{1+\theta}{1-\theta}u_t\right) + \frac{\theta}{1-\theta}u_t\right] - \sigma_P(\gamma-1)\frac{\dot{h}_t}{h_t} - [\varepsilon(\gamma-1) - 1]n_t - \bar{\rho} - \delta_0, \qquad (28)
\end{aligned}$$

where the growth rate of human capital $\dot{h}_t/h_t = \dot{\tilde{h}}_t/\tilde{h}_t + \Gamma n_t$.

4.2 Phase Diagram and Policy Implications

Applying the parameters calibrated in Table 2, we first draw the phase diagram for the optimal allocation to show the stability of the multiple steady states. Considering that there are three "state-like" variables in this economy, it will be complicated and not easy to observe the transitions if we draw a 3-dimensional diagram. As a result, for clarity, we only derive a 2-dimensional phase diagram by keeping the "detrended" human capital stock unchanged, which means we only discuss the path that is mapped to the surface derived by $\dot{\tilde{h}}_t = 0$, to show the relationship between the other two variables.

The derivation is shown in Appendix D.2. Given $\dot{\tilde{h}}_t = 0$ (and thus $\dot{h}_t/h_t = \Gamma n_t$), the curves $\dot{u}_t = 0$ and $\dot{n}_t = 0$ are derived implicitly from two equations. Moreover, from the implicit functions, we find that there may be multiple solutions to u_t given the value of n_t.[12] Considering this, we only choose the minimal value of u_t to draw the phase diagram given a certain point of n_t. Figure 7 shows the diagram. Besides the $\dot{u}_t = 0$ and $\dot{n}_t = 0$ curves, we also present the transition path of the economy and the arrows that indicate the dynamics.[13]

From the figure, we first see that there are two intersections between the $\dot{u}_t = 0$ and $\dot{n}_t = 0$ curves. Although it seems that the $\dot{n}_t = 0$ curve is not continuous within the range of the diagram, we can still divide the diagram into eight different areas according to the dynamics of the model. Clearly, the steady state with lower u_t and n_t is not stable, while the one with higher u_t and n_t remains stable around it.[14] Once the economy starts from a transition state with higher u_t and n_t, it will converge to the high state along the transition path and remain stable in this state. This is similar to the results derived in the J model, where there are two

[12] In the phase diagram, we let n_t be the horizontal axis and u_t be the vertical axis. This is also the case if we switch the two axes since there will also be multiple solutions to n_t given the value of u_t.

[13] In Appendix D.1, we show the differential equations describing the dynamics of the competitive equilibrium, draw the phase diagram for this scenario, and find that the unique solution of the economy is also saddle-path stable.

[14] It should be noted that in a 3-dimensional phase diagram in $(\tilde{h}, n, u)$ space, the lower steady state has three positive eigenvalues, indicating instability, while the higher steady state has one negative eigenvalue along the n_t axis, indicating saddle-path stability. Since our main focus is the dynamics of population growth rate changes, we restrict our discussion to the plane of $\dot{\tilde{h}}_t = 0$ and provide the corresponding 2-dimensional phase diagram here to show the transitions of the economy.

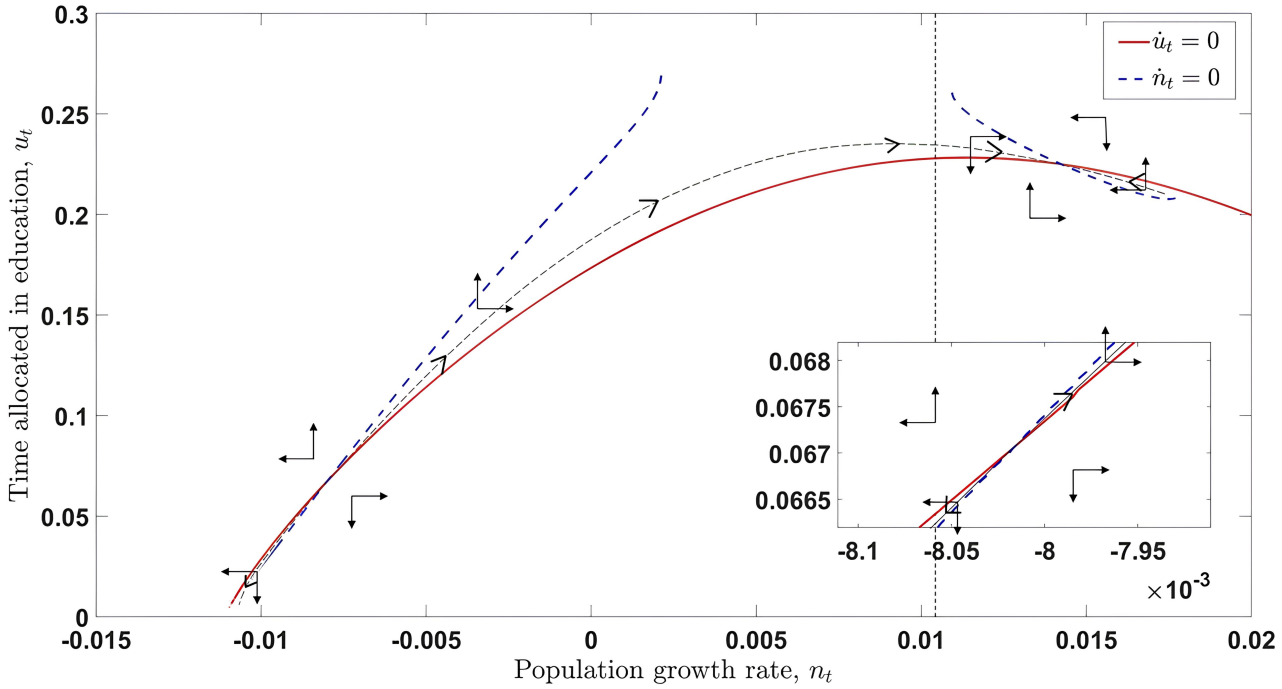


Figure 7: Phase diagram in the optimal allocation of the general model

Note. This figure shows the phase diagram in the optimal allocation of the general model. The figure is in (n, u) space and is based on the differential equations (27) and (28) given that $\dot{\tilde{h}}_t = 0$. The red solid line shows the $\dot{u}_t = 0$ curve, and the blue dashed line shows the $\dot{n}_t = 0$ curve. Arrows indicate the dynamics in different areas of the diagram, and the green dashed line represents the path of the economy. The dynamics with respect to u_t are divided by a black vertical dashed line, which intersects with the highest point of the $\dot{u}_t = 0$ curve. We also show in a subfigure the dynamics around the low steady state for clarity.

steady states with one being stable and high. However, combining the numerical results shown in Table 3, the conclusion differs. In that model, the low state in the optimal allocation is still higher than that in the competitive equilibrium. This suggests that policies should be introduced to encourage population growth before the economy converges to the high state. However, in our model, the low state in the optimal allocation is much lower than that derived in the competitive equilibrium.[15] Thus, there is no need to introduce further policies that may cause additional distortions to push the population growth rate upward to move the economy to the high state. In other words, the economy will move along the path to the

[15]The population growth rate in the low state of the optimal allocation in the general model is about -0.8%, which is much lower than that derived in the competitive equilibrium in the benchmark model (about -0.5%) and in the general model (about -0.41%).

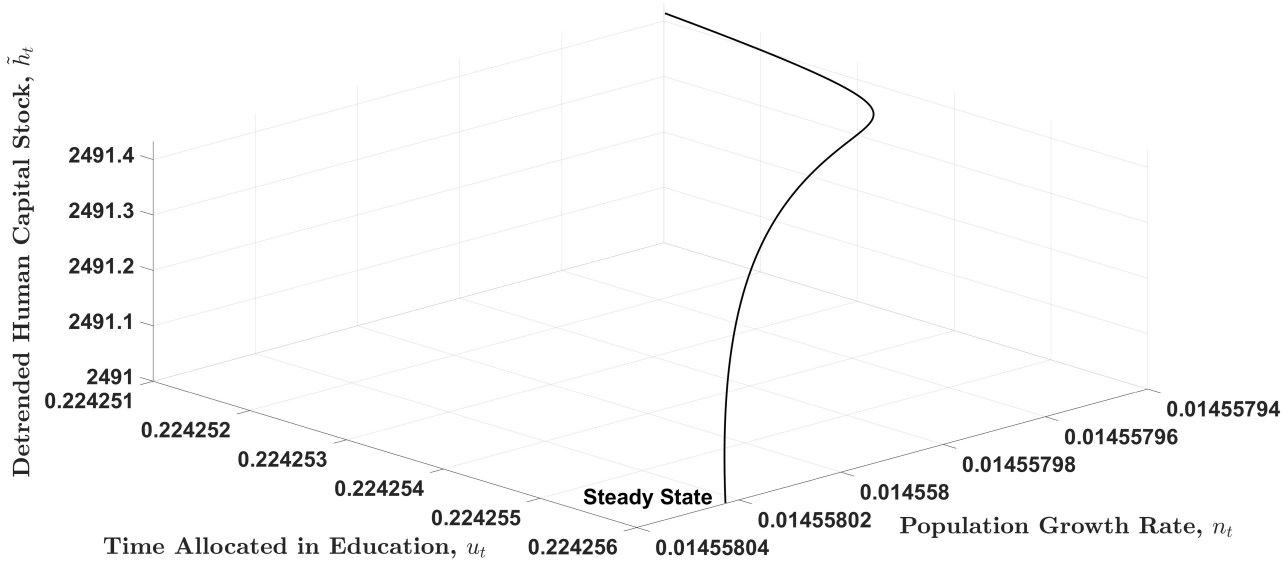
high state after we introduce data into human capital accumulation and place the economy under the social planner's planning. This explains how data can reform the development of the economy by providing a new channel to encourage population growth and thus improve total welfare.

4.3 The Economics of Transition Dynamics

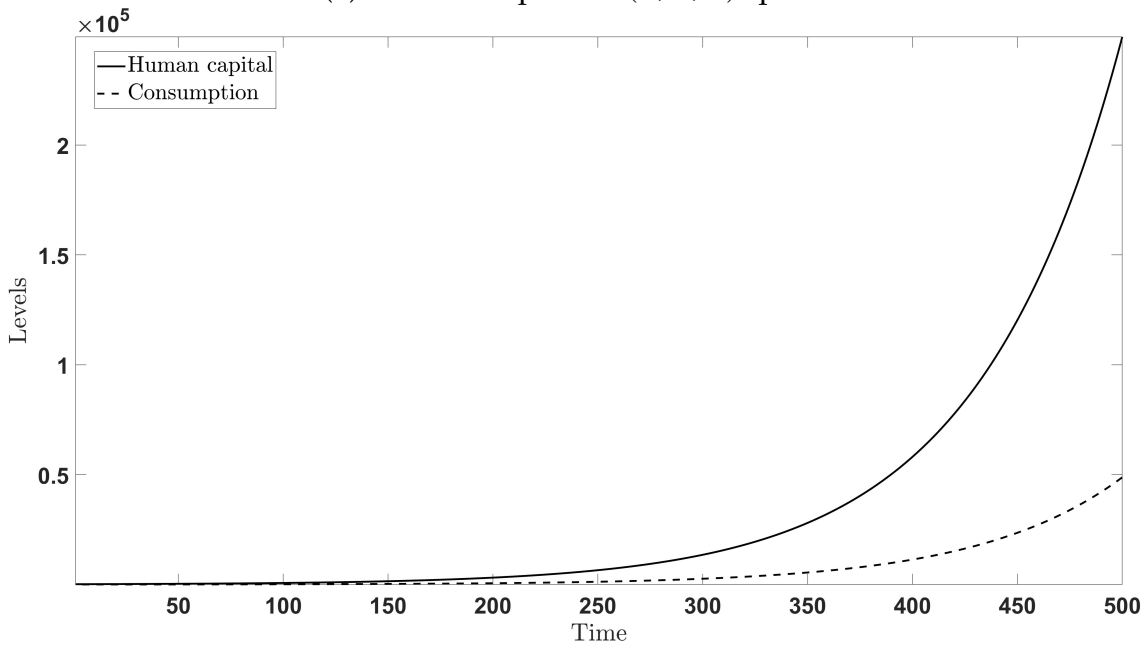
Besides the phase diagram, it is also important to show the transition dynamics to see how the economy moves towards its steady state. Our analysis includes two parts, which are shown in Figure 8. First, we show the transition path of the optimal allocation in $(n, u, \tilde{h})$ space. The path is derived from the three differential equations in Proposition 6. We use the "reverse shooting" method by setting the ending point of the economy at the steady state with higher u_t and n_t and then push the economy backward 500 periods before the final state.[16] The figure shows a very slow change in all of the key variables, although we see all of the variables increase before reaching the steady state. This also proves that the high steady state we discussed in the last subsection is stable.

Second, we recover the "detrended" variables and show their changes along the transition path. The human capital stock h_t is calculated from its "detrended" variable $\tilde{h}_t$ with $h_t = \tilde{h}_t L_t^{\Gamma}$, while per capita consumption c_t is derived from the resource constraint (22). Moreover, we set $L(T) = 100$ as the population in the last period and calculate the population in each period using the endogenous n_t derived in the model. Considering that $\tilde{h}_t$ does not change significantly along the path, the increasing pattern in Figure 8(b) mainly comes from the growth of population since we derive a positive n_t in this scenario. Consumption c_t has the same growth rate as h_t (recall that $\sigma_P = 1$), so the difference between the growth of these two variables comes from their "detrended" parts, $\tilde{c}_t$ and $\tilde{h}_t$. Thus, this figure shows that the "detrended" part of the variables is also important in determining their values, making the discussions on welfare level in Section 3.5.3 more significant, rather than just comparing

[16] This transition path shows how the economy converges to the high steady state. Following Jones (2022), we achieve this by letting the starting point deviate slightly from the steady state, determined by the negative eigenvalue and the corresponding eigenvector around the steady state.



(a) Transition path in $(n, u, \tilde{h})$ space



(b) Time series of human capital and consumption

Figure 8: Transition dynamics of the optimal allocation in general model

Note. The upper panel shows the transition path of the economy in a 3-dimensional graph. The economy starts from a state with higher $\tilde{h}_t$ and lower u_t and n_t, and converges to the steady state. The lower panel shows the time series of two key variables during the transition. The solid line represents human capital, and the dashed line represents consumption.

the growth rates in different scenarios.

5. Conclusion

We build an endogenous growth model with endogenous fertility decisions and a data factor entering the process of human capital accumulation, to emphasize the role of data in the economy and on population growth. We find a crucial reinforcement effect: as the population grows larger, it generates more data, which, due to its non-rival nature, accelerates human capital accumulation and boosts consumers' productivity. This, in turn, frees up time and increases financial resources (indirectly through real wage increases), thus enabling more reproduction and further population growth. Additionally, comparative statics indicate that both total welfare and population growth rate increase with the rising importance of data, whether in a decentralized economy or under optimal allocation.

In contrast to Jones (2022), our analysis of multiple steady states in optimal allocations and their transition dynamics reveals that the development of a data economy can achieve high population and economic growth without relying on distortionary policy interventions. Our study contributes to the literature by extending the usage of data factors and data-related technologies into a wider range to fuel economic growth. Our model thus completes the theory of fertility decisions and economic growth, and adds to the emerging studies on the effects of data and their usage in various aspects, including production, data generation, and human capital accumulation.

References

Acemoglu, D. and T. Lensman (2024). Regulating Transformative Technologies. *American Economic Review: Insights 6*(3), 359–376.

Acemoglu, D. and P. Restrepo (2018). The Race between Man and Machine: Implications of

Technology for Growth, Factor Shares, and Employment. *American Economic Review 108*(6), 1488–1542.

Acemoglu, D. and P. Restrepo (2020). Robots and Jobs: Evidence from U.S. Labor Markets. *Journal of Political Economy 128*(6), 2188–2244.

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

Barro, R. J. and G. S. Becker (1989). Fertility Choice in a Model of Economic Growth. *Econometrica 57*(2), 481–501.

Beraja, M., D. Y. Yang, and N. Yuchtman (2023). Data-Intensive Innovation and the State: Evidence from AI Firms in China. *Review of Economic Studies 90*(4), 1701–1723.

Boikos, S., A. Bucci, and T. N. Sequeira (2023). May Human Capital Rescue the Empty Planet? *Economics Letters 232*, 111352.

Chang, Q., L. W. Cong, L. Wang, and L. Zhang (2023). Production, Trade, and Cross-Border Data Flows. NBER Working Papers 31416, National Bureau of Economic Research.

Chang, Q., M. Wu, and L. Zhang (2024). Endogenous Growth and Human Capital Accumulation in a Data Economy. *Structural Change and Economic Dynamics 69*, 298–312.

Cong, L. W., W. Wei, D. Xie, and L. Zhang (2022). Endogenous Growth Under Multiple Uses of Data. *Journal of Economic Dynamics and Control 141*, 104395.

Cong, L. W., D. Xie, and L. Zhang (2021). Knowledge Accumulation, Privacy, and Growth in a Data Economy. *Management Science 67*(10), 6480–6492.

DeJong, D. N. and B. F. Ingram (2001). The Cyclical Behavior of Skill Acquisition. *Review of Economic Dynamics 4*(3), 536–561.

Farboodi, M. and L. Veldkamp (2022). A Model of the Data Economy. NBER Working Papers 28427, National Bureau of Economic Research.

Giordano, V., I. Spada, F. Chiarello, and G. Fantoni (2024). The Impact of ChatGPT on Human Skills: A Quantitative Study on Twitter Data. *Technological Forecasting and Social Change 203*, 123389.

Grossman, G. M. and E. Helpman (1991). *Innovation and Growth in the Global Economy*. Cambridge: MIT Press.

Ichihashi, S. (2021a). Competing Data Intermediaries. *RAND Journal of Economics 52*(3), 515–537.

Ichihashi, S. (2021b). The Economics of Data Externalities. *Journal of Economic Theory 196*, 105316.

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

Jones, C. I. (2016). Life and Growth. *Journal of Political Economy 124*(2), 539–578.

Jones, C. I. (2022). The End of Economic Growth? Unintended Consequences of a Declining Population. *American Economic Review 112*(11), 3489–3527.

Jones, C. I. (2024). The AI Dilemma: Growth versus Existential Risk. *American Economic Review: Insights 6*(4), 575–590.

Jones, C. I. and C. Tonetti (2020). Nonrivalry and the Economics of Data. *American Economic Review 110*(9), 2819–2858.

Jones, L. E. and A. Schoonbroodt (2010). Complements Versus Substitutes and Trends in Fertility Choice in Dynastic Models. *Econometrica 51*(3), 671–699.

Judd, K. L. (1998). *Numerical Method in Economics*. Cambridge: MIT Press.

Kumar, S., P. Rao, S. Singhania, S. Verma, and M. Kheterpal (2024). Will Artificial Intelligence Drive the Advancements in Higher Education? A Tri-Phased Exploration. *Technological Forecasting and Social Change 201*, 123258.

Lu, C.-H. (2021). The Impact of Artificial Intelligence on Economic Growth and Welfare. *Journal of Macroeconomics 69*, 103342.

Lucas, R. E. (1988). On the Mechanics of Economic Development. *Journal of Monetary Economics 22*, 3–42.

Lucas, R. E. (2015). Human Capital and Growth. *American Economic Review 105*(5), 85–88.

Mino, K. and H. Sasaki (2023). Long-Run Consequences of Population Decline in an Economy with Exhaustible Resources. *Economic Modelling 121*, 106212.

Prettner, K. and H. Strulik (2020). Innovation, Automation, and Inequality: Policy Challenges in the Race Against the Machine. *Journal of Monetary Economics 116*, 249–265.

Romer, P. M. (1990). Endogenous Technological Change. *Journal of Political Economy 98*(5), S71–S102.

Strulik, H. (2024). Higher Education and the Income-Fertility Nexus. *Journal of Population Economics 37*, 35.

Xie, D. and L. Zhang (2023). A Generalized Model of Growth in the Data Economy. Available at SSRN 4033576.

Yang, K. H. (2022). Selling Consumer Data for Profit: Optimal Market-Segmentation Design and Its Consequences. *American Economic Review 112*(4), 1364–1393.

Appendix

A. Proofs of the Benchmark Model Along a BGP

A.1 Proof of Proposition 1: Competitive Equilibrium

The current-value Hamiltonian for the consumer's utility maximization problem is given by

$$\mathcal{H} = \frac{(L_t^{\varepsilon} c_t)^{1-\gamma}}{1-\gamma} + \nu_t \left[(1-l_t-u_t)w_t h_t - c_t\right] + \lambda_{L,t}\left[b(l_t)-\delta_0\right]L_t + \lambda_{h,t}\left(\eta h_t^{\zeta} u_t - \delta_0 h_t\right). \quad \text{(A.1)}$$

In this equation, the co-state variables $\lambda_{L,t}$ and $\lambda_{h,t}$ capture the shadow prices of an extra unit of population and human capital, respectively, to maximize the representative consumer's utility. The variable ν_t denotes the shadow price of the budget constraint.

The FOCs of L_t, c_t, l_t, u_t, and h_t are shown as follows:

$$\frac{\partial \mathcal{H}}{\partial L_t} = \varepsilon L_t^{\varepsilon(1-\gamma)-1} c_t^{1-\gamma} + \lambda_{L,t} n_t = -\dot{\lambda}_{L,t} + (\bar{\rho}+\delta_0)\lambda_{L,t}, \quad \text{(A.2)}$$

$$\frac{\partial \mathcal{H}}{\partial c_t} = L_t^{\varepsilon(1-\gamma)} c_t^{-\gamma} - \nu_t = 0, \quad \text{(A.3)}$$

$$\frac{\partial \mathcal{H}}{\partial l_t} = -\nu_t w_t h_t + \lambda_{L,t} b'(l_t) L_t = 0, \quad \text{(A.4)}$$

$$\frac{\partial \mathcal{H}}{\partial u_t} = -\nu_t w_t h_t + \eta \lambda_{h,t} h_t^{\zeta} = 0, \quad \text{(A.5)}$$

$$\frac{\partial \mathcal{H}}{\partial h_t} = \nu_t (1-l_t-u_t) w_t + \lambda_{h,t}\left(\eta\zeta h_t^{\zeta-1} u_t - \delta_0\right) = -\dot{\lambda}_{h,t} + (\bar{\rho}+\delta_0)\lambda_{h,t}. \quad \text{(A.6)}$$

First, from (A.3), we have

$$\nu_t = L_t^{\varepsilon(1-\gamma)} c_t^{-\gamma}. \quad \text{(A.7)}$$

Substituting (A.7) into (A.4), (A.5), and (A.6), we obtain

$$L_t^{\varepsilon(1-\gamma)-1} c_t^{-\gamma} w_t h_t = \lambda_{L,t} \bar{b}, \tag{A.8}$$

$$L_t^{\varepsilon(1-\gamma)} c_t^{-\gamma} w_t = \eta \lambda_{h,t} h_t^{\zeta-1}, \tag{A.9}$$

$$L_t^{\varepsilon(1-\gamma)} c_t^{-\gamma} (1 - l_t - u_t) w_t + \lambda_{h,t} \left(\eta \zeta h_t^{\zeta-1} u_t - \delta_0 \right) = -\dot{\lambda}_{h,t} + (\bar{\rho} + \delta_0) \lambda_{h,t}. \tag{A.10}$$

Combining (A.9) and (A.10), we obtain

$$(1 - l_t - u_t + \zeta u_t)\, \eta h_t^{\zeta-1} = \varepsilon(\gamma - 1) n_t + \gamma \frac{\dot{c}_t}{c_t} - \frac{\dot{w}_t}{w_t} - (1 - \zeta) \frac{\dot{h}_t}{h_t} + \bar{\rho} + 2\delta_0. \tag{A.11}$$

Plugging (A.8) into (A.2), we have

$$\frac{\varepsilon \bar{b} c_t}{w_t h_t} = \varepsilon(\gamma - 1) n_t + \gamma \frac{\dot{c}_t}{c_t} - \frac{\dot{w}_t}{w_t} - \frac{\dot{h}_t}{h_t} + \bar{\rho} + \delta_0. \tag{A.12}$$

Finally, consider the profit maximization problem of the producer, and based on the conclusions derived from the problem of the data intermediary, we write the problem as:

$$\max_{(l_t + u_t)} (1 - l_t - u_t) h_t^{\sigma_P} L_t - w_t h_t (1 - l_t - u_t) L_t. \tag{A.13}$$

The FOC with respect to $(l_t + u_t)$ is

$$h_t^{\sigma_P - 1} = w_t. \tag{A.14}$$

We now turn to solving the equations. Since the growth rate of human capital g_h is zero along a BGP, we derive that g_w is also zero from (A.14), and so is g_c from (A.12).

Plugging in the BGP growth rates of h_t, c_t, and w_t, from (A.11) we have

$$\begin{aligned}
& \left(\frac{1 - l_t}{u_t} - 1 + \zeta \right) \eta h_t^{\zeta-1} u_t = \varepsilon(\gamma - 1) n_t + \bar{\rho} + 2\delta_0 \\
\Rightarrow \quad & \left(\frac{1 - l_t}{u_t} - 1 + \zeta \right) \delta_0 = \varepsilon(\gamma - 1) n_t + \bar{\rho} + 2\delta_0 \\
\Rightarrow \quad & \frac{1 - l_t}{u_t} = \frac{\varepsilon(\gamma - 1)}{\delta_0} n_t + \frac{\bar{\rho}}{\delta_0} + 3 - \zeta \\
\Rightarrow \quad & u_t = \frac{\delta_0}{\bar{b}} \cdot \frac{-n_t + \bar{b} - \delta_0}{\varepsilon(\gamma - 1) n_t + \bar{\rho} + (3 - \zeta)\delta_0}.
\end{aligned} \tag{A.15}$$

From the budget constraint, we have

$$\frac{c_t}{w_t h_t} = \frac{(1 - l_t - u_t) w_t h_t}{w_t h_t} = 1 - l_t - u_t. \tag{A.16}$$

Thus, (A.12) can be rewritten as

$$\begin{aligned} & \varepsilon \bar{b}(1 - l_t - u_t) = \varepsilon(\gamma - 1) n_t + \bar{\rho} + \delta_0 \\ \Rightarrow \quad & u_t = -\frac{1}{\bar{b}} \left[\gamma n_t + \left(1 + \frac{1}{\varepsilon}\right) \delta_0 + \frac{\bar{\rho}}{\varepsilon} \right] + 1. \end{aligned} \tag{A.17}$$

Thus, the equilibrium allocation of n_t and u_t can be derived by solving equations (A.15) and (A.17), which are denoted by n^* and u^*, respectively.

A.2 Proof of Proposition 2: Optimal Allocation

The current-value Hamiltonian equation for the consumer's utility maximization problem is

$$\mathcal{H} = \frac{(L_t^{\varepsilon} c_t)^{1-\gamma}}{1-\gamma} + \nu_t \left[(1 - l_t - u_t) h_t^{\sigma_P} - c_t \right] + \lambda_{L,t} \left[b(l_t) - \delta_0 \right] L_t + \lambda_{h,t} \left(\eta h_t^{\zeta} u_t - \delta_0 h_t \right). \tag{A.18}$$

Here, $\lambda_{L,t}$, $\lambda_{h,t}$, and ν_t defined similarly as in Appendix A.1.

The FOCs of L_t, c_t, l_t, u_t, and h_t are shown as follows:

$$\frac{\partial \mathcal{H}}{\partial L_t} = \varepsilon L_t^{\varepsilon(1-\gamma)-1} c_t^{1-\gamma} + \lambda_{L,t} \left[b(l_t) - \delta_0 \right] = -\dot{\lambda}_{L,t} + (\bar{\rho} + \delta_0 - n_t) \lambda_{L,t}, \tag{A.19}$$

$$\frac{\partial \mathcal{H}}{\partial c_t} = L_t^{\varepsilon(1-\gamma)} c_t^{-\gamma} - \nu_t = 0, \tag{A.20}$$

$$\frac{\partial \mathcal{H}}{\partial l_t} = -\nu_t h_t^{\sigma_P} + \lambda_{L,t} b'(l_t) L_t = 0, \tag{A.21}$$

$$\frac{\partial \mathcal{H}}{\partial u_t} = -\nu_t h_t^{\sigma_P} + \eta \lambda_{h,t} h_t^{\zeta} = 0, \tag{A.22}$$

$$\frac{\partial \mathcal{H}}{\partial h_t} = \nu_t (1 - l_t - u_t) \sigma_P h_t^{\sigma_P - 1} + \lambda_{h,t} \left(\eta \zeta h_t^{\zeta - 1} u_t - \delta_0 \right) = -\dot{\lambda}_{h,t} + (\bar{\rho} + \delta_0 - n_t) \lambda_{h,t}. \tag{A.23}$$

First, from (A.20), we have

$$v_t = L_t^{\varepsilon(1-\gamma)} c_t^{-\gamma} \tag{A.24}$$

Substituting (A.24) into (A.21), (A.22), and (A.23), we obtain

$$L_t^{\varepsilon(1-\gamma)-1} c_t^{-\gamma} h_t^{\sigma_P} = \lambda_{L,t} \bar{b}, \tag{A.25}$$

$$L_t^{\varepsilon(1-\gamma)} c_t^{-\gamma} h_t^{\sigma_P} = \eta \lambda_{h,t} h_t^{\zeta}, \tag{A.26}$$

$$L_t^{\varepsilon(1-\gamma)} c_t^{-\gamma} (1 - l_t - u_t) \sigma_P h_t^{\sigma_P - 1} + \lambda_{h,t} \left(\eta \zeta h_t^{\zeta-1} u_t - \delta_0 \right) = -\dot{\lambda}_{h,t} + (\bar{\rho} + \delta_0 - n_t) \lambda_{h,t}. \tag{A.27}$$

Combining (A.26) and (A.27), we have

$$[(1 - l_t - u_t)\sigma_P + \zeta u_t] \eta h_t^{\zeta-1} = [\varepsilon(\gamma - 1) - 1] n_t + \gamma \frac{\dot{c}_t}{c_t} + (\zeta - \sigma_P) \frac{\dot{h}_t}{h_t} + \bar{\rho} + 2\delta_0. \tag{A.28}$$

Considering that the growth rate of human capital g_h is zero along a BGP, g_c is also zero from (A.29). Plugging (A.25) into (A.19), we have

$$\frac{\varepsilon \bar{b} c_t}{h_t^{\sigma_P}} = [\varepsilon(\gamma - 1) - 1] n_t + \bar{\rho} + \delta_0. \tag{A.29}$$

We now turn to solving the equations. Plugging in the BGP growth rates of h_t and c_t, from (A.28) we have

$$\begin{aligned}
& \left[\left(\frac{1 - l_t}{u_t} - 1 \right) \sigma_P + \zeta \right] \eta h_t^{\zeta-1} u_t = [\varepsilon(\gamma - 1) - 1] n_t + \bar{\rho} + 2\delta_0 \\
\Rightarrow \quad & \left[\left(\frac{1 - l_t}{u_t} - 1 \right) \sigma_P + \zeta \right] \delta_0 = [\varepsilon(\gamma - 1) - 1] n_t + \bar{\rho} + 2\delta_0 \\
\Rightarrow \quad & u_t = \frac{1 - \frac{1}{\bar{b}}(n_t + \delta_0)}{1 + \frac{1}{\sigma_P} \left\{ \frac{[\varepsilon(\gamma-1)-1] n_t + \bar{\rho} + 2\delta_0}{\delta_0} - \zeta \right\}} \\
\Rightarrow \quad & u_t = \frac{\sigma_P \delta_0}{\bar{b}} \cdot \frac{-n_t + \bar{b} - \delta_0}{[\varepsilon(\gamma - 1) - 1] n_t + \bar{\rho} + (2 + \sigma_P - \zeta)\delta_0}.
\end{aligned} \tag{A.30}$$

From the resource constraint (10), we have

$$\frac{c_t}{h_t^{\sigma_P}} = \frac{(1 - l_t - u_t) h_t^{\sigma_P}}{h_t^{\sigma_P}} = 1 - l_t - u_t. \tag{A.31}$$

Thus, (A.29) can be rewritten as

$$\begin{aligned} & \varepsilon\bar{b}(1-l_t-u_t) = [\varepsilon(\gamma-1)-1]n_t + \bar{\rho} + \delta_0 \\ \Rightarrow \quad & u_t = -\frac{1}{\varepsilon\bar{b}}\left[(\varepsilon\gamma-1)n_t + \bar{\rho} + (1+\varepsilon)\delta_0\right] + 1. \end{aligned} \tag{A.32}$$

Thus, the equilibrium allocation of n_t and u_t can be derived by solving equations (A.30) and (A.32), which are denoted by n_S^* and u_S^*, respectively.

B. Proofs of the General Model Along a BGP

B.1 Proof of Propositions 3 and 4: Competitive Equilibrium

The current-value Hamiltonian equation for the consumer's utility maximization problem is

$$\begin{aligned} \mathcal{H} = & \frac{(L_t^{\varepsilon} c_t)^{1-\gamma}}{1-\gamma} + \nu_t\left[(1-l_t-u_t)w_t h_t + \pi_t - p_{D,t}D_t - c_t\right] + \\ & \lambda_{L,t}\left[b(l_t)-\delta_0\right]L_t + \lambda_{h,t}\left(\eta h_t^{\zeta} D_t^{\theta} u_t^{1-\theta} - \delta_0 h_t\right). \end{aligned} \tag{B.1}$$

Here, the co-state variables $\lambda_{L,t}$, $\lambda_{h,t}$, and ν_t are defined similarly as in Appendix A.1.

The FOCs of L_t, c_t, l_t, u_t, h_t, and D_t are shown as follows:

$$\frac{\partial \mathcal{H}}{\partial L_t} = \varepsilon L_t^{\varepsilon(1-\gamma)-1} c_t^{1-\gamma} + \lambda_{L,t}\left[b(l_t)-\delta_0\right] = -\dot{\lambda}_{L,t} + (\bar{\rho}+\delta_0)\lambda_{L,t}, \tag{B.2}$$

$$\frac{\partial \mathcal{H}}{\partial c_t} = L_t^{\varepsilon(1-\gamma)} c_t^{-\gamma} - \nu_t = 0, \tag{B.3}$$

$$\frac{\partial \mathcal{H}}{\partial l_t} = -\nu_t w_t h_t + \lambda_{L,t} b'(l_t) L_t = 0, \tag{B.4}$$

$$\frac{\partial \mathcal{H}}{\partial u_t} = -\nu_t w_t h_t + (1-\theta)\eta\lambda_{h,t} h_t^{\zeta} D_t^{\theta} u_t^{-\theta} = 0, \tag{B.5}$$

$$\frac{\partial \mathcal{H}}{\partial h_t} = \nu_t(1-l_t-u_t)w_t + \lambda_{h,t}\left(\eta\zeta h_t^{\zeta-1} D_t^{\theta} u_t^{1-\theta} - \delta_0\right) = -\dot{\lambda}_{h,t} + (\bar{\rho}+\delta_0)\lambda_{h,t}, \tag{B.6}$$

$$\frac{\partial \mathcal{H}}{\partial D_t} = -\nu_t p_{D,t} + \theta\eta\lambda_{h,t} h_t^{\zeta} D_t^{\theta-1} u_t^{1-\theta} = 0. \tag{B.7}$$

First, from (B.3), we have

$$\nu_t = L_t^{\varepsilon(1-\gamma)} c_t^{-\gamma}. \tag{B.8}$$

Substituting (B.8) into (B.4), (B.5), (B.6), and (B.7), we obtain

$$L_t^{\varepsilon(1-\gamma)} c_t^{-\gamma} w_t h_t = \lambda_{L,t} \bar{b} L_t, \tag{B.9}$$

$$L_t^{\varepsilon(1-\gamma)} c_t^{-\gamma} w_t = (1-\theta)\eta\lambda_{h,t} h_t^{\zeta-1} D_t^{\theta} u_t^{-\theta}, \tag{B.10}$$

$$L_t^{\varepsilon(1-\gamma)} c_t^{-\gamma} (1-l_t-u_t) w_t + \lambda_{h,t}\left(\eta\zeta h_t^{\zeta-1} D_t^{\theta} u_t^{1-\theta} - \delta_0\right) = -\dot{\lambda}_{h,t} + (\bar{\rho}+\delta_0)\lambda_{h,t}, \tag{B.11}$$

$$L_t^{\varepsilon(1-\gamma)} c_t^{-\gamma} p_{D,t} = \theta\eta\lambda_{h,t} h_t^{\zeta} D_t^{\theta-1} u_t^{1-\theta}. \tag{B.12}$$

Combining (B.10) and (B.12), we have

$$(1-\theta) p_{D,t} D_t = \theta w_t u_t h_t. \tag{B.13}$$

Then, combining (B.10) and (B.11), we obtain

$$\frac{(1-\theta)\eta(1-l_t-u_t)+\eta\zeta u_t}{h_t^{1-\zeta} D_t^{-\theta} u_t^{\theta}} = \varepsilon(\gamma-1) n_t + \gamma\frac{\dot{c}_t}{c_t} - \frac{\dot{w}_t}{w_t} - (1-\zeta)\frac{\dot{h}_t}{h_t} + \theta\frac{\dot{D}_t}{D_t} + \bar{\rho} + 2\delta_0. \tag{B.14}$$

Considering that the right-hand side of (B.14) should be constant along a BGP, its left-hand side should also be constant, i.e.,

$$(1-\zeta)\frac{\dot{h}_t}{h_t} = \theta\frac{\dot{D}_t}{D_t}. \tag{B.15}$$

Then, (B.14) can be further written as

$$\left[(1-\theta)\left(\frac{1-l_t}{u_t} - 1\right) + \zeta\right]\eta h_t^{\zeta-1} D_t^{\theta} u_t^{1-\theta} = \varepsilon(\gamma-1) n_t + \gamma\frac{\dot{c}_t}{c_t} - \frac{\dot{w}_t}{w_t} + \bar{\rho} + 2\delta_0. \tag{B.16}$$

Similarly, combining (B.11) and (B.12), we have

$$\left[\frac{\theta(1-l_t-u_t)w_t h_t}{p_{D,t}D_t}+\zeta\right]\eta h_t^{\zeta-1}D_t^{\theta}u_t^{1-\theta}=\varepsilon(\gamma-1)n_t+\gamma\frac{\dot{c}_t}{c_t}-\frac{\dot{p}_{D,t}}{p_{D,t}}+\zeta\frac{\dot{h}_t}{h_t}+(\theta-1)\frac{\dot{D}_t}{D_t}+\bar{\rho}+2\delta_0. \tag{B.17}$$

Thus, we have

$$\frac{\dot{p}_{D,t}}{p_{D,t}}+\frac{\dot{D}_t}{D_t}=\frac{\dot{h}_t}{h_t}+\frac{\dot{w}_t}{w_t}. \tag{B.18}$$

Plugging (B.9) into (B.2), we have

$$\frac{\varepsilon\bar{b}c_t}{w_t h_t}=\varepsilon(\gamma-1)n_t+\gamma\frac{\dot{c}_t}{c_t}-\frac{\dot{w}_t}{w_t}-\frac{\dot{h}_t}{h_t}+\bar{\rho}+\delta_0. \tag{B.19}$$

Along a BGP, the growth rates of c_t, w_t, and h_t should all be constant, so the left-hand side of (B.19) should also be constant. As a result, we have

$$\frac{\dot{c}_t}{c_t}=\frac{\dot{w}_t}{w_t}+\frac{\dot{h}_t}{h_t}. \tag{B.20}$$

Then, (B.19) can be further written as

$$\frac{\varepsilon\bar{b}c_t}{w_t h_t}=\varepsilon(\gamma-1)n_t+(\gamma-1)\frac{\dot{c}_t}{c_t}+\bar{\rho}+\delta_0. \tag{B.21}$$

Finally, consider the profit maximization problem of the producer, and taking into account the conclusions derived from the problem of the data intermediary, we write the problem as:

$$\max_{(l_t+u_t),s_t}(1-l_t-u_t-s_t)h_t^{\sigma_P}L_t+p_{D,t}L_t(1-l_t-u_t-s_t)h_t^{\sigma_P+\sigma_D}s_tL_t-w_t h_t(1-l_t-u_t)L_t. \tag{B.22}$$

The FOCs with respect to (l_t+u_t) and s_t are

$$h_t^{\sigma_P-1}+p_{D,t}h_t^{\sigma_P+\sigma_D-1}L_t s_t=w_t, \tag{B.23}$$

$$p_{D,t}h_t^{\sigma_D}(1-l_t-u_t-2s_t)L_t=1. \tag{B.24}$$

From (B.24), we have

$$s_t = \frac{1}{2}\left(1 - l_t - u_t - \frac{1}{p_{D,t} h_t^{\sigma_D} L_t}\right). \tag{B.25}$$

From (B.25) we can see that the right-hand side of the equation should be constant along a BGP, then,

$$\frac{\dot{p}_{D,t}}{p_{D,t}} + \sigma_D \frac{\dot{h}_t}{h_t} + n_t = 0. \tag{B.26}$$

Now, we derive the BGP level of s_t. Combining (B.13) with (B.23), and plugging in the definition of D_t, we have

$$\begin{aligned}
& (1-\theta) p_{D,t}(1 - l_t - u_t - s_t) s_t h_t^{\sigma_D + \sigma_P} L_t = \theta u_t (h_t^{\sigma_P} + p_{D,t} h_t^{\sigma_D + \sigma_P} s_t L_t) \\
\Rightarrow \quad & \left[(1-\theta)(1 - l_t - u_t - s_t) - \theta u_t\right] p_{D,t} s_t h_t^{\sigma_D} L_t = \theta u_t \\
\Rightarrow \quad & p_{D,t} h_t^{\sigma_D} L_t = \frac{\theta u_t}{s_t\left[(1-\theta)(1 - l_t - u_t - s_t) - \theta u_t\right]}.
\end{aligned} \tag{B.27}$$

Plugging (B.27) into (B.24), we have

$$s_t = \frac{1}{2}\left[1 - l_t - u_t + \frac{s_t}{\theta}\left(1 - (1-\theta)\frac{1 - l_t - s_t}{u_t}\right)\right]. \tag{B.28}$$

Then, the BGP level of s_t can be derived from (B.28) given the levels of u_t and l_t. We can easily derive that $s_t = 1 - l_t - u_t$, which is the upper limit of s_t due to the labor market clearing condition, is one of the solutions to the equation. This solution will lead to zero labor share allocated in the production of final goods, which is obviously not feasible. As a result, there is only one solution for s_t, which is

$$s_t = \frac{\theta}{1-\theta} u_t. \tag{B.29}$$

We still want to derive the net profit of the producer, which may be non-zero in this model. Substituting (B.23) and (B.24) back into the profit maximization problem (B.22), we derive the per capita profit as

$$\begin{aligned}
\pi_t = & \frac{1}{L_t}\Big[(1 - l_t - u_t - s_t) h_t^{\sigma_P} L_t + (1 + p_{D,t} h_t^{\sigma_D} s_t L_t) h_t^{\sigma_P} s_t L_t - \\
& h_t (1 - l_t - u_t) L_t (h_t^{\sigma_P - 1} + p_{D,t} h_t^{\sigma_P + \sigma_D - 1} L_t s_t)\Big]
\end{aligned}$$

$$
\begin{aligned}
=& p_{D,t} h_t^{\sigma_P+\sigma_D} s_t^2 L_t - p_D h_t^{\sigma_P+\sigma_D} (1-l_t-u_t) s_t L_t \\
=& - p_{D,t} h_t^{\sigma_P+\sigma_D} s_t (1-l_t-u_t-s_t) L_t = -p_{D,t} D_t .
\end{aligned} \tag{B.30}
$$

As a result, the consumers should pay a subsidy to the producer using their incomes.

We now turn to solving the equations. First, from the data generation function (15), we have the BGP growth rate of data as:

$$
g_D = \frac{\dot{D}_t}{D_t} = (\sigma_P + \sigma_D) g_h + n^* . \tag{B.31}
$$

Meanwhile, (B.15) can be rewritten as

$$
\begin{aligned}
(1-\zeta) g_h = \theta g_D &= \theta \left[(\sigma_P + \sigma_D) g_h + n^* \right] \\
\Rightarrow \left[1 - \zeta - \theta(\sigma_P + \sigma_D) \right] g_h &= \theta n^* \\
g_h &= \left[\frac{\theta}{1 - \zeta - \theta(\sigma_P + \sigma_D)} \right] n^* ,
\end{aligned} \tag{B.32}
$$

and g_D can be rewritten as

$$
g_D = (\sigma_P + \sigma_D) g_h + n^* = \left[\frac{1-\zeta}{1 - \zeta - \theta(\sigma_P + \sigma_D)} \right] n^* . \tag{B.33}
$$

From (B.26), we have the BGP growth rate of data price as

$$
g_{p_D} = -\sigma_D g_h - n_t = \left[\frac{\zeta + \theta\sigma_P - 1}{1 - \zeta - \theta(\sigma_P + \sigma_D)} \right] n^* . \tag{B.34}
$$

Then, from (B.18), we have the BGP growth rate of wage rate as

$$
g_w = g_{p_D} + g_D - g_h = \left[\frac{\theta(\sigma_P - 1)}{1 - \zeta - \theta(\sigma_P + \sigma_D)} \right] n^* . \tag{B.35}
$$

From the budget constraint (13), we know that the BGP growth rate of $p_{D,t} D_t$ and c_t

should be the same. Thus, we have the BGP growth rate of consumption as

$$g_c = g_{p_D} + g_D = \left[\frac{\sigma_P \theta}{1-\zeta-\theta(\sigma_P+\sigma_D)}\right] n^* \tag{B.36}$$

Define

$$\Gamma \equiv \frac{\theta}{1-\zeta-\theta(\sigma_P+\sigma_D)}, \tag{B.37}$$

So we have $g_c = \Gamma\sigma_P n_t$, $g_h = \Gamma n_t$, and $g_w = \Gamma(\sigma_P - 1)n_t$.

Plugging in the BGP growth rates of h_t, c_t, and w_t, from (B.16) we have

$$\begin{aligned}
&\left[(1-\theta)\left(\frac{1-l_t}{u_t}-1\right)+\zeta\right](g_h+\delta_0) = \varepsilon(\gamma-1)n_t + \gamma g_c - g_w + \bar{\rho} + 2\delta_0 \\
\Rightarrow\quad &\left[(1-\theta)\left(\frac{1-l_t}{u_t}-1\right)+\zeta\right](\Gamma n_t+\delta_0) = \varepsilon(\gamma-1)n_t + \gamma\Gamma\sigma_P n_t - \Gamma(\sigma_P-1)n_t + \bar{\rho} + 2\delta_0 \\
\Rightarrow\quad &\frac{1-l_t}{u_t} = \frac{\left[\varepsilon(\gamma-1)+\Gamma\left(\gamma\sigma_P-(\sigma_P-1)+(1-\theta-\zeta)\right)\right]n_t + \bar{\rho} + (3-\theta-\zeta)\delta_0}{(1-\theta)(\Gamma n_t+\delta_0)} \\
\Rightarrow\quad &u_t = \frac{1}{\bar{b}} \cdot \frac{\left(\bar{b}-n_t-\delta_0\right)(1-\theta)(\Gamma n_t+\delta_0)}{\left\{\varepsilon(\gamma-1)+\Gamma\left[\gamma\sigma_P-\sigma_P+2-\theta-\zeta\right]\right\}n_t + \bar{\rho} + (3-\theta-\zeta)\delta_0}.
\end{aligned} \tag{B.38}$$

From the budget constraint (13) and (B.13), we have

$$\frac{c_t}{w_t h_t} = \frac{(1-l_t-u_t)w_t h_t + \pi_t - p_{D,t}D_t}{w_t h_t} = \frac{(1-l_t-u_t)w_t h_t - 2p_{D,t}D_t}{w_t h_t} = 1 - l_t - \frac{1+\theta}{1-\theta}u_t. \tag{B.39}$$

Thus, (B.21) can be rewritten as

$$\begin{aligned}
&\frac{\varepsilon\bar{b}c_t}{w_t h_t} = \varepsilon(\gamma-1)n_t + (\gamma-1)\Gamma\sigma_P n_t + \bar{\rho} + \delta_0 \\
\Rightarrow\quad &\varepsilon\bar{b}\left(1-l_t-\frac{1+\theta}{1-\theta}u_t\right) = (\gamma-1)(\varepsilon+\Gamma\sigma_P)n_t + \bar{\rho} + \delta_0 \\
\Rightarrow\quad &u_t = -\frac{1-\theta}{\varepsilon\bar{b}(1+\theta)}\left\{\left[(\gamma-1)(\varepsilon+\Gamma\sigma_P)+\varepsilon\right]n_t - \varepsilon\bar{b} + \bar{\rho} + (1+\varepsilon)\delta_0\right\}.
\end{aligned} \tag{B.40}$$

Combining (B.38) and (B.40), we derive the BGP levels of u_D^* and n_D^*. Moreover, given u_D^* and l_D^* (derived from n_D^*), we can derive s_D^* by solving (B.28).

We next derive the BGP levels of the other key variables. From (14), we have

$$
\begin{aligned}
& g_h^* + \delta_0 = \eta (h^*)^{\zeta-1} (D^*)^{\theta} (u^*)^{1-\theta} \\
\Rightarrow \quad & g_h^* + \delta_0 = \eta (h^*)^{\zeta-1+\theta(\sigma_P+\sigma_D)} (1-l^*-u^*-s^*)^{\theta} (s^*)^{\theta} (u^*)^{1-\theta} L_t^{\theta} \\
\Rightarrow \quad & h^* = \left[\frac{\Gamma n^* + \delta_0}{\eta (1-l^*-u^*-s^*)^{\theta} (s^*)^{\theta} (u^*)^{1-\theta}} \right]^{\frac{1}{\zeta-1+\theta(\sigma_P+\sigma_D)}} L_t^{\frac{\theta}{1-\zeta-\theta(\sigma_P+\sigma_D)}} \\
\Rightarrow \quad & h^* = \left[\frac{\Gamma n^* + \delta_0}{\eta (1-l^*-u^*-s^*)^{\theta} (s^*)^{\theta} (u^*)^{1-\theta} L_0^{\theta}} \right]^{\frac{1}{\zeta-1+\theta(\sigma_P+\sigma_D)}} e^{g_h^* t} \equiv \tilde{h}^* e^{\Gamma n^* t}.
\end{aligned} \tag{B.41}
$$

We also have

$$
y^* = (1-l^*-u^*-s^*)(h^*)^{\sigma_P} \equiv \tilde{y}^* e^{\sigma_P \Gamma n^* t}, \tag{B.42}
$$

and

$$
D^* = y^* s^* (h^*)^{\sigma_D} L_t = (1-l^*-u^*-s^*) s^* (h^*)^{\sigma_P+\sigma_D} L_t \equiv \tilde{D}^* e^{[1+(\sigma_P+\sigma_D)\Gamma] n^* t}. \tag{B.43}
$$

From (B.24), we have

$$
p_D^* = \frac{1}{1-l^*-u^*-2s^*} (h^*)^{-\sigma_D} L_t^{-1} \equiv \tilde{p}_D^* e^{(-\sigma_D \Gamma - 1) n^* t}. \tag{B.44}
$$

From (B.13), we have

$$
w^* = \frac{(1-\theta) p_D^* D^*}{\theta u^* h^*} = \frac{(1-\theta)(1-l^*-u^*-s^*) s^*}{\theta (1-l^*-u^*-2s^*) u^*} (h^*)^{\sigma_P - 1} \equiv \tilde{w}^* e^{(\sigma_P - 1)\Gamma n^* t}. \tag{B.45}
$$

From (B.39), we have

$$
c^* = \left(1 - l^* - \frac{1+\theta}{1-\theta} u^* \right) w^* h^* = \frac{(1-l^*-u^*-s^*)[(1-\theta)(1-l^*) - (1+\theta)u^*] s^*}{\theta (1-l^*-u^*-2s^*) u^*} (h^*)^{\sigma_P}. \tag{B.46}
$$

Obviously, by plugging $s_t = \theta u_t / (1-\theta)$ into (B.46), we can derive that

$$
c^* = y^* = (1-l^*-u^*-s^*)(h^*)^{\sigma_P} \equiv \tilde{c}^* e^{\sigma_P \Gamma n^* t}. \tag{B.47}
$$

B.2 Proof of Proposition 5: Optimal Allocations

The current-value Hamiltonian equation for the consumer's utility maximization problem is

$$\begin{aligned}\mathcal{H} =&\frac{(L_t^{\varepsilon} c_t)^{1-\gamma}}{1-\gamma} + \nu_t \left[(1-l_t-u_t-s_t)h_t^{\sigma_P} - c_t\right] + \mu_t \left[(1-l_t-u_t-s_t)h_t^{\sigma_P+\sigma_D} s_t L_t - D_t\right] \\ &\lambda_{L,t}\left[b(l_t)-\delta_0\right]L_t + \lambda_{h,t}\left(\eta h_t^{\zeta} D_t^{\theta} u_t^{1-\theta} - \delta_0 h_t\right). \end{aligned} \tag{B.48}$$

The FOCs of L_t, c_t, l_t, u_t, s_t, h_t, and D_t are shown as follows:

$$\begin{aligned}\frac{\partial \mathcal{H}}{\partial L_t} =&\varepsilon L_t^{\varepsilon(1-\gamma)-1} c_t^{1-\gamma} + \mu_t(1-l_t-u_t-s_t)h_t^{\sigma_P+\sigma_D} s_t + \lambda_{L,t}\left[b(l_t)-\delta_0\right] \\ =&-\dot{\lambda}_{L,t} + (\bar{\rho}+\delta_0-n_t)\lambda_{L,t}, \end{aligned} \tag{B.49}$$

$$\frac{\partial \mathcal{H}}{\partial c_t} = L_t^{\varepsilon(1-\gamma)} c_t^{-\gamma} - \nu_t = 0, \tag{B.50}$$

$$\frac{\partial \mathcal{H}}{\partial l_t} = -\nu_t h_t^{\sigma_P} - \mu h_t^{\sigma_P+\sigma_D} s_t L_t + \lambda_{L,t} b'(l_t) L_t = 0, \tag{B.51}$$

$$\frac{\partial \mathcal{H}}{\partial u_t} = -\nu_t h_t^{\sigma_P} - \mu h_t^{\sigma_P+\sigma_D} s_t L_t + (1-\theta)\eta\lambda_{h,t} h_t^{\zeta} D_t^{\theta} u_t^{-\theta} = 0, \tag{B.52}$$

$$\frac{\partial \mathcal{H}}{\partial s_t} = -\nu_t h_t^{\sigma_P} + \mu(1-l_t-u_t-2s_t)h_t^{\sigma_P+\sigma_D} L_t = 0, \tag{B.53}$$

$$\begin{aligned}\frac{\partial \mathcal{H}}{\partial h_t} =&\sigma_P \nu_t (1-l_t-u_t-s_t)h_t^{\sigma_P-1} + (\sigma_P+\sigma_D)\mu_t(1-l_t-u_t-s_t)s_t L_t h_t^{\sigma_P+\sigma_D-1} + \\ &\lambda_{h,t}\left(\eta\zeta h_t^{\zeta-1} D_t^{\theta} u_t^{1-\theta} - \delta_0\right) = -\dot{\lambda}_{h,t} + (\bar{\rho}+\delta_0-n_t)\lambda_{h,t}, \end{aligned} \tag{B.54}$$

$$\frac{\partial \mathcal{H}}{\partial D_t} = -\mu_t + \theta\eta\lambda_{h,t} h_t^{\zeta} D_t^{\theta-1} u_t^{1-\theta} = 0. \tag{B.55}$$

First, from (B.50) and (B.55), we have

$$\nu_t = L_t^{\varepsilon(1-\gamma)} c_t^{-\gamma} \tag{B.56}$$

and

$$\mu_t = \theta\eta\lambda_{h,t} h_t^{\zeta} D_t^{\theta-1} u_t^{1-\theta}. \tag{B.57}$$

Substituting (B.56) and (B.57) into (B.49), (B.51), (B.52), (B.53), and (B.54), respectively, we have

$$\varepsilon L_t^{\varepsilon(1-\gamma)-1} c_t^{1-\gamma} + \theta\eta\lambda_{h,t} h_t^{\zeta} D_t^{\theta} u_t^{1-\theta} L_t^{-1} + \lambda_{L,t}\left[b(l_t) - \delta_0\right] = -\dot{\lambda}_{L,t} + (\bar{\rho} + \delta_0 - n_t)\lambda_{L,t}, \tag{B.58}$$

$$L_t^{\varepsilon(1-\gamma)} c_t^{-\gamma} h_t^{\sigma_P} + \theta\eta\lambda_{h,t} h_t^{\zeta+\sigma_P+\sigma_D} D_t^{\theta-1} u_t^{1-\theta} s_t L_t = \lambda_{L,t}\bar{b} L_t, \tag{B.59}$$

$$L_t^{\varepsilon(1-\gamma)} c_t^{-\gamma} = \eta\lambda_{h,t} h_t^{\zeta-\sigma_P} D_t^{\theta} u_t^{-\theta}\left(1 - \theta - \theta\frac{u_t}{1 - l_t - u_t - s_t}\right), \tag{B.60}$$

$$L_t^{\varepsilon(1-\gamma)-1} c_t^{-\gamma} = \theta\eta\lambda_{h,t} h_t^{\zeta+\sigma_D} D_t^{\theta-1} u_t^{1-\theta}(1 - l_t - u_t - 2s_t), \tag{B.61}$$

$$\sigma_P L_t^{\varepsilon(1-\gamma)} c_t^{-\gamma}(1-l_t-u_t-s_t) h_t^{\sigma_P-1} + \lambda_{h,t}\left\{\eta h_t^{\zeta-1} D_t^{\theta} u_t^{1-\theta}\left[(\sigma_P + \sigma_D)\theta + \zeta\right] - \delta_0\right\} = -\dot{\lambda}_{h,t} + (\bar{\rho} + \delta_0 - n_t)\lambda_{h,t}. \tag{B.62}$$

Combining (B.60) and (B.61), we have

$$\begin{aligned} & L_t = \frac{D_t\left(1 - \theta - \theta\frac{u_t}{1-l_t-u_t-s_t}\right)}{\theta h_t^{\sigma_P+\sigma_D} u_t(1 - l_t - u_t - 2s_t)} \\ \Rightarrow \quad & \theta h_t^{\sigma_P+\sigma_D} u_t(1 - l_t - u_t - 2s_t) L_t = (1 - l_t - u_t - s_t) h_t^{\sigma_P+\sigma_D} s_t L_t\left(1 - \theta - \theta\frac{u_t}{1 - l_t - u_t - s_t}\right) \\ \Rightarrow \quad & s_t(1-\theta)(1 - l_t - s_t) = u_t\left[s_t + \theta(1 - l_t - u_t - 2s_t)\right] \end{aligned} \tag{B.63}$$

Equation (B.63) is exactly the same as (B.28), thus we have the only one feasible solution to s_t as:

$$s_t = \frac{\theta}{1-\theta} u_t. \tag{B.64}$$

Combining (B.60) and (B.62), we have

$$\begin{aligned} \eta h_t^{\zeta-1} D_t^{\theta} u_t^{1-\theta}\left[\sigma_P(1-\theta)\left(\frac{1 - l_t - u_t - s_t}{u_t}\right) + \sigma_D\theta + \zeta\right] = & [\varepsilon(\gamma-1) - 1] n_t + \gamma\frac{\dot{c}_t}{c_t} - \\ & (\sigma_P - \zeta)\frac{\dot{h}_t}{h_t} + \theta\frac{\dot{D}_t}{D_t} + \bar{\rho} + 2\delta_0. \end{aligned} \tag{B.65}$$

Considering that the right-hand side of (B.65) should be constant along a BGP, then its left-hand side should also be constant, i.e.,

$$(1-\zeta)\frac{\dot{h}_t}{h_t} = \theta\frac{\dot{D}_t}{D_t}. \tag{B.66}$$

Then, equation (B.65) can be further written as

$$\eta h_t^{\zeta-1} D_t^{\theta} u_t^{1-\theta}\left[\sigma_P(1-\theta)\left(\frac{1-l_t-u_t-s_t}{u_t}\right)+\sigma_D\theta+\zeta\right] = [\varepsilon(\gamma-1)-1]n_t + \gamma\frac{\dot{c}_t}{c_t} + (1-\sigma_P)\frac{\dot{h}_t}{h_t} + \bar{\rho} + 2\delta_0. \tag{B.67}$$

Combining (B.59) and (B.61), we have

$$L_t^{\varepsilon(1-\gamma)-1} c_t^{-\gamma} h_t^{\sigma_P}\left(1+\frac{s_t}{1-l_t-u_t-2s_t}\right) = \lambda_{L,t}\bar{b}. \tag{B.68}$$

Plugging (B.61) and (B.68) into (B.58), we have

$$\bar{b}\left[\varepsilon(1-l_t-u_t-2s_t)+s_t\right] = [\varepsilon(\gamma-1)-1]n_t + \gamma\frac{\dot{c}_t}{c_t} - \sigma_p\frac{\dot{h}_t}{h_t} + \bar{\rho} + \delta_0. \tag{B.69}$$

We now turn to solving the equations. First, from the data generation function (15), we have the BGP growth rate of data as:

$$g_D = \frac{\dot{D}_t}{D_t} = (\sigma_P+\sigma_D)g_h + n^*. \tag{B.70}$$

Meanwhile, (B.66) can be rewritten as

$$\begin{aligned} & (1-\zeta)g_h = \theta g_D = \theta\left[(\sigma_P+\sigma_D)g_h + n^*\right] \\ \Rightarrow\quad & \left[1-\zeta-\theta(\sigma_P+\sigma_D)\right]g_h = \theta n^* \\ \Rightarrow\quad & g_h = \left[\frac{\theta}{1-\zeta-\theta(\sigma_P+\sigma_D)}\right]n^*, \end{aligned} \tag{B.71}$$

and g_D can be rewritten as

$$g_D = (\sigma_P + \sigma_D)g_h + n^* = \left[\frac{1-\zeta}{1-\zeta-\theta(\sigma_P+\sigma_D)}\right] n^*. \tag{B.72}$$

From the resource constraint (22), we know that the BGP growth rate of consumption is

$$g_c = \sigma_P g_h = \left[\frac{\sigma_P \theta}{1-\zeta-\theta(\sigma_P+\sigma_D)}\right] n^* \tag{B.73}$$

Define

$$\Gamma \equiv \frac{\theta}{1-\zeta-\theta(\sigma_P+\sigma_D)}, \tag{B.74}$$

So we have $g_c = \Gamma\sigma_P n_t$ and $g_h = \Gamma n_t$.

Plugging in the BGP growth rates of h_t and c_t, from (B.67) we have

$$\begin{aligned}
&\left[\sigma_P(1-\theta)\left(\frac{1-l_t-u_t-s_t}{u_t}\right)+\sigma_D\theta+\zeta\right](g_h+\delta_0) = [\varepsilon(\gamma-1)-1]n_t + \gamma g_c + (1-\sigma_P)g_h + \bar{\rho} + 2\delta_0 \\
\Rightarrow &\left[\sigma_P(1-\theta)\left(\frac{1-l_t-u_t-s_t}{u_t}\right)+\sigma_D\theta+\zeta\right](\Gamma n_t+\delta_0) = [\varepsilon(\gamma-1)+\gamma\Gamma\sigma_P+\Gamma(1-\sigma_P)-1]\, n_t + \bar{\rho} + 2\delta_0 \\
\Rightarrow &\frac{1-l_t-s_t}{u_t} = 1 + \frac{1}{\sigma_P(1-\theta)}\left[\frac{[\varepsilon(\gamma-1)+\gamma\Gamma\sigma_P+\Gamma(1-\sigma_P)-1]\, n_t}{\Gamma n_t+\delta_0} + \frac{\bar{\rho}+2\delta_0}{\Gamma n_t+\delta_0} - \zeta - \sigma_D\theta\right] \\
\Rightarrow &u_t = \frac{\sigma_P}{\bar{b}} \cdot \frac{\left(\bar{b}-n_t-\delta_0\right)(1-\theta)(\Gamma n_t+\delta_0)}{[\varepsilon(\gamma-1)+\Gamma(\gamma\sigma_P+1-\sigma_D\theta-\zeta)-1]\, n_t + \bar{\rho} + (2+\sigma_P-\sigma_D\theta-\zeta)\delta_0}.
\end{aligned} \tag{B.75}$$

Thus, (B.69) can be rewritten as

$$\begin{aligned}
&\bar{b}\,[\varepsilon(1-l_t-u_t-2s_t)+s_t] = [\varepsilon(\gamma-1)+\gamma\Gamma\sigma_P-\sigma_P\Gamma-1]\, n_t + \bar{\rho} + \delta_0 \\
\Rightarrow\quad &\bar{b}\varepsilon(1-l_t) + \bar{b}\left(\frac{\theta}{1-\theta} - \varepsilon\frac{1+\theta}{1-\theta}\right)u_t = [\varepsilon(\gamma-1)+\gamma\Gamma\sigma_P-\sigma_P\Gamma-1]\, n_t + \bar{\rho} + \delta_0 \\
\Rightarrow\quad &u_t = -\frac{1-\theta}{\bar{b}[\varepsilon(1+\theta)-\theta]}\left\{[(\gamma-1)(\varepsilon+\Gamma\sigma_P)+\varepsilon-1]\, n_t - \bar{b}\varepsilon + \bar{\rho} + (1+\varepsilon)\delta_0\right\}.
\end{aligned} \tag{B.76}$$

Then, combining (B.75) and (B.76), we derive the BGP level of u^*_{DS} and n^*_{DS}. We next derive the BGP levels of the other key variables. From (14), we have

$$g^*_h + \delta_0 = \eta(h^*_{DS})^{\zeta-1}(D^*_{DS})^{\theta}(u^*_{DS})^{1-\theta}$$

$$\Rightarrow \qquad g_h^* + \delta_0 = \eta(h_{DS}^*)^{\zeta-1+\theta(\sigma_P+\sigma_D)}(1-l_{DS}^*-u_{DS}^*-s_{DS}^*)^{\theta}(s_{DS}^*)^{\theta}(u_{DS}^*)^{1-\theta}L_t^{\theta}$$

$$\Rightarrow \qquad h_{DS}^* = \left[\frac{\Gamma n_{DS}^* + \delta_0}{\eta(1-l_{DS}^*-u_{DS}^*-s_{DS}^*)^{\theta}(s_{DS}^*)^{\theta}(u_{DS}^*)^{1-\theta}}\right]^{\frac{1}{\zeta-1+\theta(\sigma_P+\sigma_D)}} L_t^{\frac{\theta}{1-\zeta-\theta(\sigma_P+\sigma_D)}}. \tag{B.77}$$

We also have

$$y_{DS}^* = c_{DS}^* = (1-l_{DS}^*-u_{DS}^*-s_{DS}^*)(h_{DS}^*)^{\sigma_P} \tag{B.78}$$

and

$$D_{DS}^* = y_{DS}^* s_{DS}^* (h_{DS}^*)^{\sigma_D} L_t = (1-l_{DS}^*-u_{DS}^*-s_{DS}^*)s_{DS}^*(h_{DS}^*)^{\sigma_P+\sigma_D} L_t. \tag{B.79}$$

C. Welfare Analysis

C.1 Proofs of Welfare Analysis

After deriving the BGP levels of the key variables, we can obtain the welfare function of the economy by substituting these results into the utility function of the representative consumers. To illustrate, we consider the utility function of a representative consumer in the context of a competitive equilibrium. For simplicity and tractability, we set $\gamma = 1$. Along a BGP, the per capita welfare function is derived as follows:

$$\begin{aligned}
&\int_0^{\infty} e^{-(\bar{\rho}+\delta_0)t}\left(\varepsilon \ln L_t + \ln c_t\right) \mathrm{d}t \\
=&\int_0^{\infty} e^{-(\bar{\rho}+\delta_0)t}\left[\varepsilon \ln\left(L_0 e^{n^* t}\right) + \ln\left(\tilde{c}^* e^{g_c^* t}\right)\right] \mathrm{d}t \\
=&(\varepsilon n^* + g_c^*)\int_0^{\infty} e^{-(\bar{\rho}+\delta_0)t} t \mathrm{d}t + \ln\left(L_0^{\varepsilon}\tilde{c}^*\right)\int_0^{\infty} e^{-(\bar{\rho}+\delta_0)t}\mathrm{d}t \\
=&-\frac{\varepsilon n^* + g_c^*}{\bar{\rho}+\delta_0}\int_0^{\infty} t \mathrm{d}e^{-(\bar{\rho}+\delta_0)t} + \frac{\ln\left(L_0^{\varepsilon}\tilde{c}^*\right)}{\bar{\rho}+\delta_0} \\
=&-\frac{\varepsilon n^* + g_c^*}{\bar{\rho}+\delta_0}\left(t e^{-(\bar{\rho}+\delta_0)t}\Big|_0^{\infty} - \int_0^{\infty} e^{-(\bar{\rho}+\delta_0)t}\mathrm{d}t\right) + \frac{\ln\left(L_0^{\varepsilon}\tilde{c}^*\right)}{\bar{\rho}+\delta_0} \\
=&\frac{\varepsilon n^* + g_c^*}{(\bar{\rho}+\delta_0)^2} + \frac{\ln\left(L_0^{\varepsilon}\tilde{c}^*\right)}{\bar{\rho}+\delta_0}.
\end{aligned} \tag{C.1}$$

Here, $\tilde{c}^*$ comes from the BGP level of consumption (or the constant part of consumption in the general model) in different scenarios. Similarly, the total welfare is derived as

$$\int_0^\infty L_t e^{-(\bar{\rho}+\delta_0)t}\left(\varepsilon \ln L_t + \ln c_t\right) \mathrm{d}t = \frac{\varepsilon n^* + g_c^*}{(\bar{\rho}+\delta_0-n^*)^2} + \frac{\ln\left(L_0^{\varepsilon}\tilde{c}^*\right)}{\bar{\rho}+\delta_0-n^*}. \tag{C.2}$$

C.2 Additional Results on Welfare Analysis

In this subsection, we present additional results from the welfare analysis using robustness values of θ set at 0.15 and 0.25, which are shown in Table C.1. For comparison purposes, we also include the benchmark value of $\theta = 0.2$ in the table below. From the table, we observe that the results remain robust as the parameter θ changes. The high state consistently persists, even though the low state disappears when θ takes smaller values.

Table C.1: Robustness Checks on Comparisons of the General Models

Variables	General model ($\theta = 0.15$)			General model ($\theta = 0.2$)			General model ($\theta = 0.25$)		
	CE	OA		CE	OA		CE	OA	
		Low state	High state		Low state	High state		Low state	High state
n^*_{alloc}	-0.53%	—	1.26%	-0.41%	-0.80%	1.46%	-0.23%	-0.23%	1.62%
l^*_{alloc}	0.12	—	0.48	0.14	0.06	0.52	0.18	0.18	0.56
u^*_{alloc}	0.18	—	0.22	0.15	0.07	0.22	0.11	0.10	0.23
$(g^*_y)_{alloc}$	-0.26%	—	0.63%	-0.41%	-0.80%	1.46	-0.57%	-0.59%	4.06%
U^*_{alloc}	310.82	—	741.16	488.57	437.29	1223.43	1057.39	1029.96	2320.78
$\bar{U}^*_{alloc}$	381.82	—	280.59	576.35	587.54	329.20	1164.30	1137.17	336.66

Note: This table shows the results of six key variables in the general models when θ takes different values. The columns labeled "CE" represent the results derived in competitive equilibria, while the columns labeled "OA" represent the results derived in optimal allocations. Note that there are multiple steady states in the "OA" except the scenario when $\theta = 0.15$.

D. Transition Dynamics

D.1 Competitive Equilibrium in the General Model

According to the per capita level of each variable along a BGP, as shown in (B.41), we denote $\tilde{h}_t = h_t L_t^{-\Gamma}$ as one of the state-like variables. It turns out that $\tilde{h}_t$, u_t, and n_t are constants

along the BGP. Therefore, we will derive the differential equations for these three state-like variables to describe the transition dynamics of the competitive equilibrium.

To begin, differentiating $\tilde{h}_t$ with respect to time t, we obtain

$$\dot{\tilde{h}}_t = \frac{\mathrm{d}(h_t L_t^{-\Gamma})}{\mathrm{d}t} = \dot{h}_t L_t^{-\Gamma} - \Gamma h_t L_t^{-\Gamma} n_t,$$

and

$$\frac{\dot{\tilde{h}}_t}{\tilde{h}_t} = \frac{\dot{h}_t}{h_t} - \Gamma n_t. \tag{D.1}$$

Referring to (14), the above equation can be rearranged as

$$\begin{aligned}
\dot{\tilde{h}}_t =& \left(\eta h_t^{\zeta} D_t^{\theta} u_t^{1-\theta} - \delta_0 h_t\right) L_t^{-\Gamma} - \Gamma h_t L_t^{-\Gamma} n_t \\
=& \eta h_t^{\zeta+\theta(\sigma_P+\sigma_D)} \left(1 - l_t - \frac{1}{1-\theta} u_t\right)^{\theta} \left(\frac{\theta}{1-\theta}\right)^{\theta} u_t L_t^{\theta-\Gamma} - (\delta_0 + \Gamma n_t)\, h_t L_t^{-\Gamma} \\
=& \eta \tilde{h}_t^{\zeta+\theta(\sigma_P+\sigma_D)} \left(1 - \frac{n_t + \delta_0}{\bar{b}} - \frac{1}{1-\theta} u_t\right)^{\theta} \left(\frac{\theta}{1-\theta}\right)^{\theta} u_t - (\delta_0 + \Gamma n_t)\, \tilde{h}_t.
\end{aligned} \tag{D.2}$$

From (15), data D_t can be rewritten as

$$D_t = \left(1 - \frac{n_t + \delta_0}{\bar{b}} - \frac{1}{1-\theta} u_t\right) \frac{\theta}{1-\theta} u_t h_t^{\sigma_P+\sigma_D} L_t. \tag{D.3}$$

Then, its growth rate can be written as

$$\frac{\dot{D}_t}{D_t} = -\frac{\frac{1}{\bar{b}}\dot{n}_t + \frac{1}{1-\theta}\dot{u}_t}{1 - \frac{n_t+\delta_0}{\bar{b}} - \frac{1}{1-\theta} u_t} + \frac{\dot{u}_t}{u_t} + (\sigma_P + \sigma_D)\frac{\dot{h}_t}{h_t} + n_t. \tag{D.4}$$

From (B.24), the price of data can be derived as

$$p_{D,t} = \frac{1}{1 - \frac{1}{\bar{b}}(n_t + \delta_0) - \frac{1+\theta}{1-\theta} u_t} h_t^{-\sigma_D} L_t^{-1} \tag{D.5}$$

Then, its growth rate can be derived as

$$\frac{\dot{p}_{D,t}}{p_{D,t}} = \frac{\frac{1}{\bar{b}}\dot{n}_t + \frac{1+\theta}{1-\theta}\dot{u}_t}{1 - \frac{n_t+\delta_0}{\bar{b}} - \frac{1+\theta}{1-\theta}u_t} - \sigma_D\frac{\dot{h}_t}{h_t} - n_t. \tag{D.6}$$

Thus, we derive the growth rate of wage from (B.13) as

$$\begin{aligned}\frac{\dot{w}_t}{w_t} =& \frac{\dot{p}_{D,t}}{p_{D,t}} + \frac{\dot{D}_t}{D_t} - \frac{\dot{u}_t}{u_t} - \frac{\dot{h}_t}{h_t} \\ =& \frac{\frac{1}{\bar{b}}\dot{n}_t + \frac{1+\theta}{1-\theta}\dot{u}_t}{1 - \frac{n_t+\delta_0}{\bar{b}} - \frac{1+\theta}{1-\theta}u_t} - \sigma_D\frac{\dot{h}_t}{h_t} - n_t - \frac{\frac{1}{\bar{b}}\dot{n}_t + \frac{1}{1-\theta}\dot{u}_t}{1 - \frac{n_t+\delta_0}{\bar{b}} - \frac{1}{1-\theta}u_t} + \frac{\dot{u}_t}{u_t} + (\sigma_P + \sigma_D)\frac{\dot{h}_t}{h_t} + n_t - \frac{\dot{u}_t}{u_t} - \frac{\dot{h}_t}{h_t} \\ =& \frac{\frac{1}{\bar{b}}\dot{n}_t + \frac{1+\theta}{1-\theta}\dot{u}_t}{1 - \frac{n_t+\delta_0}{\bar{b}} - \frac{1+\theta}{1-\theta}u_t} - \frac{\frac{1}{\bar{b}}\dot{n}_t + \frac{1}{1-\theta}\dot{u}_t}{1 - \frac{n_t+\delta_0}{\bar{b}} - \frac{1}{1-\theta}u_t} + (\sigma_P - 1)\frac{\dot{h}_t}{h_t}.\end{aligned} \tag{D.7}$$

From (B.39), the growth rate of consumption is derived as

$$\begin{aligned}\frac{\dot{c}_t}{c_t} =& -\frac{\frac{1}{\bar{b}}\dot{n}_t + \frac{1+\theta}{1-\theta}\dot{u}_t}{1 - \frac{n_t+\delta_0}{\bar{b}} - \frac{1+\theta}{1-\theta}u_t} + \frac{\dot{w}_t}{w_t} + \frac{\dot{h}_t}{h_t} \\ =& -\frac{\frac{1}{\bar{b}}\dot{n}_t + \frac{1}{1-\theta}\dot{u}_t}{1 - \frac{n_t+\delta_0}{\bar{b}} - \frac{1}{1-\theta}u_t} + \sigma_P\frac{\dot{h}_t}{h_t}.\end{aligned} \tag{D.8}$$

From (B.14), we have

$$\begin{aligned}&\left[(1-\theta)\left(\frac{\bar{b} - n_t - \delta_0}{\bar{b}u_t} - 1\right) + \zeta\right]\left(\frac{\dot{h}_t}{h_t} + \delta_0\right) \\ =&\varepsilon(\gamma-1)n_t + \gamma\frac{\dot{c}_t}{c_t} - \frac{\dot{w}_t}{w_t} - (1-\zeta)\frac{\dot{h}_t}{h_t} + \theta\frac{\dot{D}_t}{D_t} - \theta\frac{\dot{u}_t}{u_t} + \bar{\rho} + 2\delta_0 \\ =&(1-\gamma-\theta)\frac{\frac{1}{\bar{b}}\dot{n}_t + \frac{1}{1-\theta}\dot{u}_t}{1 - \frac{n_t+\delta_0}{\bar{b}} - \frac{1}{1-\theta}u_t} - \frac{\frac{1}{\bar{b}}\dot{n}_t + \frac{1+\theta}{1-\theta}\dot{u}_t}{1 - \frac{n_t+\delta_0}{\bar{b}} - \frac{1+\theta}{1-\theta}u_t} + [\sigma_P(\gamma-1) + \zeta + \theta(\sigma_P+\sigma_D)]\frac{\dot{h}_t}{h_t} + \\ &[\varepsilon(\gamma-1) + \theta]n_t + \bar{\rho} + 2\delta_0.\end{aligned} \tag{D.9}$$

From (B.19), we have

$$\varepsilon\bar{b}\left[1 - \frac{1}{\bar{b}}(n_t + \delta_0) - \frac{1+\theta}{1-\theta}u_t\right] = \varepsilon(\gamma-1)n_t + \gamma\frac{\dot{c}_t}{c_t} - \frac{\dot{w}_t}{w_t} - \frac{\dot{h}_t}{h_t} + \bar{\rho} + \delta_0$$

$$=(1-\gamma)\frac{\frac{1}{\bar{b}}\dot{n}_t+\frac{1}{1-\theta}\dot{u}_t}{1-\frac{n_t+\delta_0}{\bar{b}}-\frac{1}{1-\theta}u_t}-\frac{\frac{1}{\bar{b}}\dot{n}_t+\frac{1+\theta}{1-\theta}\dot{u}_t}{1-\frac{n_t+\delta_0}{\bar{b}}-\frac{1+\theta}{1-\theta}u_t}+\sigma_P(\gamma-1)\frac{\dot{h}_t}{h_t}+\varepsilon(\gamma-1)n_t+\bar{\rho}+\delta_0. \quad \text{(D.10)}$$

Combine (D.9) and (D.10), we derive the change of u_t and n_t as

$$\dot{u}_t=\frac{\mathcal{CD}-\mathcal{AF}}{\mathcal{BD}-\mathcal{AE}}, \quad \text{and} \quad \dot{n}_t=\frac{\mathcal{BF}-\mathcal{CE}}{\mathcal{BD}-\mathcal{AE}}, \quad \text{(D.11)}$$

where

$$\mathcal{A}=\frac{\frac{1-\gamma-\theta}{\bar{b}}}{1-\frac{n_t+\delta_0}{\bar{b}}-\frac{1}{1-\theta}u_t}-\frac{\frac{1}{\bar{b}}}{1-\frac{n_t+\delta_0}{\bar{b}}-\frac{1+\theta}{1-\theta}u_t}, \quad \text{(D.12)}$$

$$\mathcal{B}=\frac{\frac{1-\gamma-\theta}{1-\theta}}{1-\frac{n_t+\delta_0}{\bar{b}}-\frac{1}{1-\theta}u_t}-\frac{\frac{1+\theta}{1-\theta}}{1-\frac{n_t+\delta_0}{\bar{b}}-\frac{1+\theta}{1-\theta}u_t}, \quad \text{(D.13)}$$

$$\mathcal{C}=\left[(1-\theta)\left(\frac{\bar{b}-n_t-\delta_0}{\bar{b}u_t}-1\right)+\zeta\right]\left(\frac{\dot{h}_t}{h_t}+\delta_0\right)-[\sigma_P(\gamma-1)+\zeta+\theta(\sigma_P+\sigma_D)]\frac{\dot{h}_t}{h_t}-[\varepsilon(\gamma-1)+\theta]n_t-\bar{\rho}-2\delta_0, \quad \text{(D.14)}$$

$$\mathcal{D}=\frac{\frac{1-\gamma}{\bar{b}}}{1-\frac{n_t+\delta_0}{\bar{b}}-\frac{1}{1-\theta}u_t}-\frac{\frac{1}{\bar{b}}}{1-\frac{n_t+\delta_0}{\bar{b}}-\frac{1+\theta}{1-\theta}u_t}, \quad \text{(D.15)}$$

$$\mathcal{E}=\frac{\frac{1-\gamma}{1-\theta}}{1-\frac{n_t+\delta_0}{\bar{b}}-\frac{1}{1-\theta}u_t}-\frac{\frac{1+\theta}{1-\theta}}{1-\frac{n_t+\delta_0}{\bar{b}}-\frac{1+\theta}{1-\theta}u_t}, \quad \text{(D.16)}$$

$$\mathcal{F}=\varepsilon\bar{b}\left[1-\frac{1}{\bar{b}}(n_t+\delta_0)-\frac{1+\theta}{1-\theta}u_t\right]-\sigma_P(\gamma-1)\frac{\dot{h}_t}{h_t}-\varepsilon(\gamma-1)n_t-\bar{\rho}-\delta_0. \quad \text{(D.17)}$$

Based on the differential equations presented above, the phase diagram around the equilibrium point is depicted in Figure D.1. From the figure, it is evident that the two curves $\dot{u}_t=0$ and $\dot{n}_t=0$ are in close proximity to each other, indicating that the equilibrium point exhibits saddle-path stability. The transition path of the economy lies between the two curves, which, for the sake of clarity, is not shown in the figure.

D.2 Optimal Allocation in the General Model

Obviously, the change of $\tilde{h}_t$ can still be presented as (D.2), and the change of D_t can also be denoted as (D.4).

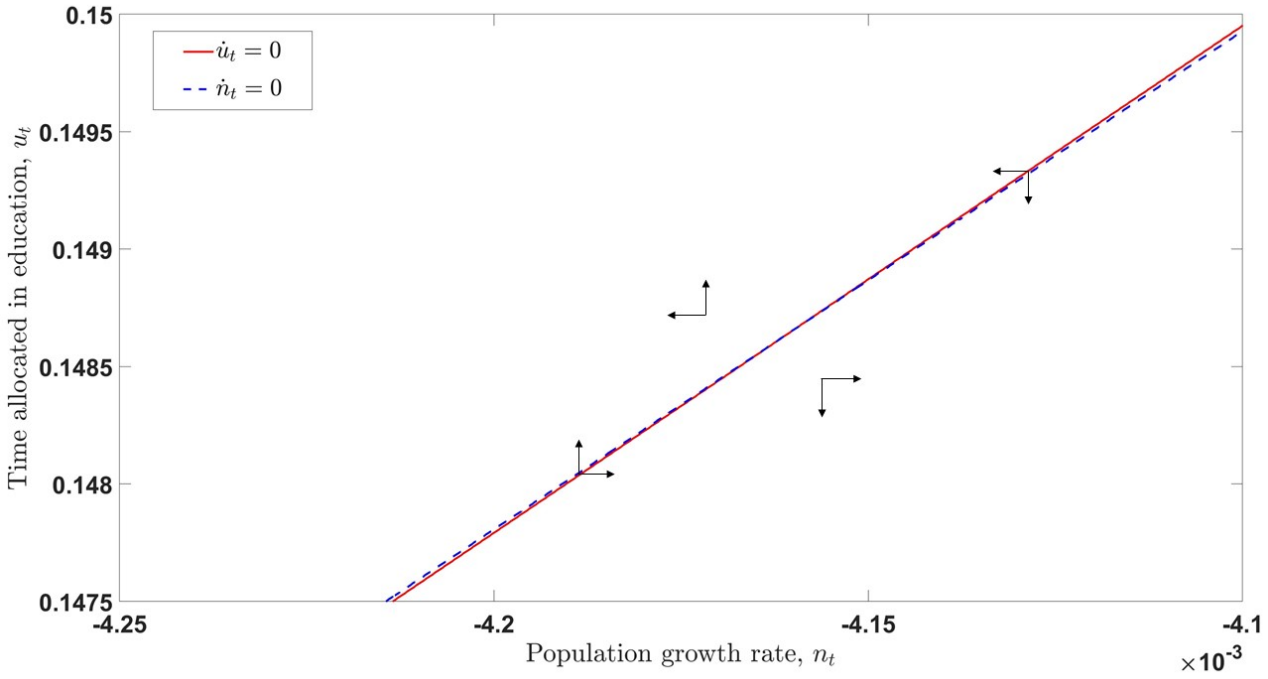


Figure D.1: Phase diagram in the competitive equilibrium of the general model

Note. This figure shows the phase diagram in the competitive equilibrium of the general model. The figure is in (n, u) space and is based on the differential equations (D.9) and (D.10) given that $\dot{\bar{h}}_t = 0$. The red solid line shows the $\dot{u}_t = 0$ curve, and the blue dashed line shows the $\dot{n}_t = 0$ curve. Arrows indicate the dynamics in different areas of the diagram.

We first need to derive the change of the following equation:

$$\Delta_t \equiv \frac{1}{1-\theta-\theta\frac{u_t}{1-l_t-u_t-s_t}} = \frac{1-\frac{n_t+\delta_0}{\bar{b}}-\frac{1}{1-\theta}u_t}{(1-\theta)\left(1-\frac{n_t+\delta_0}{\bar{b}}\right)-(1+\theta)u_t}. \tag{D.18}$$

Thus, we have

$$\begin{aligned}\frac{\mathrm{d}\Delta_t}{\mathrm{d}t} =& \frac{\partial\Delta_t}{\partial n_t}\dot{n}_t + \frac{\partial\Delta_t}{\partial u_t}\dot{u}_t \\ =& \frac{-\frac{1}{\bar{b}}\left[(1-\theta)\left(1-\frac{n_t+\delta_0}{\bar{b}}\right)-(1+\theta)u_t\right]+\frac{1-\theta}{\bar{b}}\left(1-\frac{n_t+\delta_0}{\bar{b}}-\frac{1}{1-\theta}u_t\right)}{\left[(1-\theta)\left(1-\frac{n_t+\delta_0}{\bar{b}}\right)-(1+\theta)u_t\right]^2}\dot{n}_t + \\ & \frac{-\frac{1}{1-\theta}\left[(1-\theta)\left(1-\frac{n_t+\delta_0}{\bar{b}}\right)-(1+\theta)u_t\right]+(1+\theta)\left(1-\frac{n_t+\delta_0}{\bar{b}}-\frac{1}{1-\theta}u_t\right)}{\left[(1-\theta)\left(1-\frac{n_t+\delta_0}{\bar{b}}\right)-(1+\theta)u_t\right]^2}\dot{u}_t\end{aligned}$$

$$=\frac{\theta}{\bar{b}}\cdot\frac{u_t\dot{n}_t-(n_t+\delta_0)\dot{u}_t}{\left[(1-\theta)\left(1-\frac{n_t+\delta_0}{\bar{b}}\right)-(1+\theta)u_t\right]^2}. \tag{D.19}$$

As a result, we have

$$\frac{\dot{\Delta}_t}{\Delta_t}=\frac{\theta}{\bar{b}}\cdot\frac{u_t\dot{n}_t-(n_t+\delta_0)\dot{u}_t}{\left[(1-\theta)\left(1-\frac{n_t+\delta_0}{\bar{b}}\right)-(1+\theta)u_t\right]\left(1-\frac{n_t+\delta_0}{\bar{b}}-\frac{1}{1-\theta}u_t\right)}. \tag{D.20}$$

We also need to derive the growth rate of consumption, which is shown as

$$\frac{\dot{c}_t}{c_t}=-\frac{\frac{1}{\bar{b}}\dot{n}_t+\frac{1}{1-\theta}\dot{u}_t}{1-\frac{1}{\bar{b}}(n_t+\delta_0)-\frac{1}{1-\theta}u_t}+\sigma_P\frac{\dot{h}_t}{h_t}. \tag{D.21}$$

From (B.65), we have

$$\begin{aligned}
&\left[\sigma_P(1-\theta)\left(\frac{1-\frac{1}{\bar{b}}(n_t+\delta_0)}{u_t}-\frac{1}{1-\theta}\right)+\sigma_D\theta+\zeta\right]\left(\frac{\dot{h}_t}{h_t}+\delta_0\right)\\
=&[\varepsilon(\gamma-1)-1]n_t+\gamma\frac{\dot{c}_t}{c_t}-(\sigma_P-\zeta)\frac{\dot{h}_t}{h_t}+\theta\frac{\dot{D}_t}{D_t}-\theta\frac{\dot{u}_t}{u_t}-\frac{\dot{\Delta}_t}{\Delta_t}+\bar{\rho}+2\delta_0\\
=&-(\gamma+\theta)\frac{\frac{1}{\bar{b}}\dot{n}_t+\frac{1}{1-\theta}\dot{u}_t}{1-\frac{1}{\bar{b}}(n_t+\delta_0)-\frac{1}{1-\theta}u_t}-\frac{\theta}{\bar{b}}\cdot\frac{u_t\dot{n}_t-(n_t+\delta_0)\dot{u}_t}{\left[(1-\theta)\left(1-\frac{n_t+\delta_0}{\bar{b}}\right)-(1+\theta)u_t\right]\left(1-\frac{n_t+\delta_0}{\bar{b}}-\frac{1}{1-\theta}u_t\right)}+\\
&[(\gamma+\theta-1)\sigma_P+\theta\sigma_D+\zeta]\frac{\dot{h}_t}{h_t}+[\varepsilon(\gamma-1)-1+\theta]n_t+\bar{\rho}+2\delta_0.
\end{aligned} \tag{D.22}$$

We also need to derive the change of the following equation:

$$\Theta_t\equiv 1+\frac{s_t}{1-l_t-u_t-2s_t}=1+\frac{\theta u_t}{(1-\theta)\left(1-\frac{n_t+\delta_0}{\bar{b}}\right)-(1+\theta)u_t}. \tag{D.23}$$

Thus, we have

$$\begin{aligned}
\frac{\mathrm{d}\Theta_t}{\mathrm{d}t}=&\frac{\partial\Theta_t}{\partial n_t}\dot{n}_t+\frac{\partial\Theta_t}{\partial u_t}\dot{u}_t\\
=&\frac{\frac{1-\theta}{\bar{b}}\dot{n}_t+\left\{\theta\left[(1-\theta)\left(1-\frac{n_t+\delta_0}{\bar{b}}\right)-(1+\theta)u_t\right]+(1+\theta)\theta u_t\right\}\dot{u}_t}{\left[(1-\theta)\left(1-\frac{n_t+\delta_0}{\bar{b}}\right)-(1+\theta)u_t\right]^2}
\end{aligned}$$

$$= \frac{\frac{1-\theta}{\bar{b}}\dot{n}_t + \theta(1-\theta)\left(1-\frac{n_t+\delta_0}{\bar{b}}\right)\dot{u}_t}{\left[(1-\theta)\left(1-\frac{n_t+\delta_0}{\bar{b}}\right)-(1+\theta)u_t\right]^2}. \tag{D.24}$$

As a result, we have

$$\frac{\dot{\Theta}_t}{\Theta_t} = \frac{\frac{1-\theta}{\bar{b}}\dot{n}_t + \theta(1-\theta)\left(1-\frac{n_t+\delta_0}{\bar{b}}\right)\dot{u}_t}{\left[(1-\theta)\left(1-\frac{n_t+\delta_0}{\bar{b}}\right)-(1+\theta)u_t\right]\left[(1-\theta)\left(1-\frac{n_t+\delta_0}{\bar{b}}\right)-u_t\right]}. \tag{D.25}$$

From (B.69), we have

$$\begin{aligned}
&\bar{b}\left[\varepsilon\left(1-\frac{n_t+\delta_0}{\bar{b}}-\frac{1+\theta}{1-\theta}u_t\right)+\frac{\theta}{1-\theta}u_t\right] = [\varepsilon(\gamma-1)-1]n_t + \gamma\frac{\dot{c}_t}{c_t} - \sigma_p\frac{\dot{h}_t}{h_t} + \frac{\dot{\Theta}_t}{\Theta_t} + \bar{\rho} + \delta_0 \\
&= -\gamma\frac{\frac{1}{\bar{b}}\dot{n}_t + \frac{1}{1-\theta}\dot{u}_t}{1-\frac{1}{\bar{b}}(n_t+\delta_0)-\frac{1}{1-\theta}u_t} + \frac{\frac{1}{\bar{b}}\dot{n}_t + \theta\left(1-\frac{n_t+\delta_0}{\bar{b}}\right)\dot{u}_t}{\left[(1-\theta)\left(1-\frac{n_t+\delta_0}{\bar{b}}\right)-(1+\theta)u_t\right]\left[1-\frac{n_t+\delta_0}{\bar{b}}-\frac{1}{1-\theta}u_t\right]} \\
&\quad + \sigma_P(\gamma-1)\frac{\dot{h}_t}{h_t} + [\varepsilon(\gamma-1)-1]n_t + \bar{\rho} + \delta_0.
\end{aligned} \tag{D.26}$$

Then, $\dot{u}_t$ and $\dot{n}_t$ can be derived by solving (D.22) and (D.26), that is,

$$\dot{u}_t = \frac{\mathcal{C}^*\mathcal{D}^* - \mathcal{A}^*\mathcal{F}^*}{\mathcal{B}^*\mathcal{D}^* - \mathcal{A}^*\mathcal{E}^*}, \quad \text{and} \quad \dot{n}_t = \frac{\mathcal{B}^*\mathcal{F}^* - \mathcal{C}^*\mathcal{E}^*}{\mathcal{B}^*\mathcal{D}^* - \mathcal{A}^*\mathcal{E}^*}, \tag{D.27}$$

where

$$\mathcal{A}^* = -\frac{\frac{\gamma+\theta}{\bar{b}}}{1-\frac{n_t+\delta_0}{\bar{b}}-\frac{1}{1-\theta}u_t} - \frac{\frac{\theta}{\bar{b}}u_t}{\left[(1-\theta)\left(1-\frac{n_t+\delta_0}{\bar{b}}\right)-(1+\theta)u_t\right]\left(1-\frac{n_t+\delta_0}{\bar{b}}-\frac{1}{1-\theta}u_t\right)}, \tag{D.28}$$

$$\mathcal{B}^* = -\frac{\frac{\gamma+\theta}{1-\theta}}{1-\frac{n_t+\delta_0}{\bar{b}}-\frac{1}{1-\theta}u_t} + \frac{\frac{\theta(n_t+\delta_0)}{\bar{b}}}{\left[(1-\theta)\left(1-\frac{n_t+\delta_0}{\bar{b}}\right)-(1+\theta)u_t\right]\left(1-\frac{n_t+\delta_0}{\bar{b}}-\frac{1}{1-\theta}u_t\right)}, \tag{D.29}$$

$$\mathcal{C}^* = \left[\sigma_P(1-\theta)\left(\frac{1-\frac{1}{\bar{b}}(n_t+\delta_0)}{u_t}-\frac{1}{1-\theta}\right)+\sigma_D\theta+\zeta\right]\left(\frac{\dot{h}_t}{h_t}+\delta_0\right) - [(\gamma+\theta-1)\sigma_P+\theta\sigma_D+\zeta]\frac{\dot{h}_t}{h_t} -$$

$$[\varepsilon(\gamma-1)-1+\theta]n_t-\bar{\rho}-2\delta_0, \tag{D.30}$$

$$\mathcal{D}^*=-\frac{\frac{\gamma}{\bar{b}}}{1-\frac{n_t+\delta_0}{\bar{b}}-\frac{1}{1-\theta}u_t}+\frac{\frac{1}{\bar{b}}}{\left[(1-\theta)\left(1-\frac{n_t+\delta_0}{\bar{b}}\right)-(1+\theta)u_t\right]\left(1-\frac{n_t+\delta_0}{\bar{b}}-\frac{1}{1-\theta}u_t\right)}, \tag{D.31}$$

$$\mathcal{E}^*=-\frac{\frac{\gamma}{1-\theta}}{1-\frac{n_t+\delta_0}{\bar{b}}-\frac{1}{1-\theta}u_t}+\frac{\theta\left(1-\frac{n_t+\delta_0}{\bar{b}}\right)}{\left[(1-\theta)\left(1-\frac{n_t+\delta_0}{\bar{b}}\right)-(1+\theta)u_t\right]\left(1-\frac{n_t+\delta_0}{\bar{b}}-\frac{1}{1-\theta}u_t\right)}, \tag{D.32}$$

$$\mathcal{F}^*=\bar{b}\left[\varepsilon\left(1-\frac{n_t+\delta_0}{\bar{b}}-\frac{1+\theta}{1-\theta}u_t\right)+\frac{\theta}{1-\theta}u_t\right]-\sigma_P(\gamma-1)\frac{\dot{h}_t}{h_t}-[\varepsilon(\gamma-1)-1]n_t-\bar{\rho}-\delta_0. \tag{D.33}$$