ARTICLE IN PRESS

Humanities and Social Sciences Communications

Article in Press

https://doi.org/10.1057/s41599-026-06747-6

Study on the impact of big data sharing on individuals' welfare—from the perspective of consumption and privacy

Received: 1 December 2024

Accepted: 11 February 2026

Hanghang Dong, Xiaoming Li, Yadi Liu & Chan Wang

Cite this article as: Dong, H., Li, X., Liu, Y. *et al.* Study on the impact of big data sharing on individuals' welfare—from the perspective of consumption and privacy. *Humanit Soc Sci Commun* (2026). https://doi.org/10.1057/s41599-026-06747-6

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Study on the Impact of Big Data Sharing on Individuals' Welfare

—from the Perspective of Consumption and Privacy

Abstract: This paper constructs a macro-level theoretical framework, grounded in the theory of creative destruction, to explain how big data sharing affects individuals' welfare from the perspectives of consumption and privacy. First, we treat data as a new type of production factor and endogenize it within the production function. We then propose an innovative view: individuals' welfare is influenced by both the privacy cost of big data sharing and their consumption levels. Consumption, in turn, is affected by the multiplier effect and the transformation patterns of R&D. Finally, we provide a theoretical analysis of the optimal level of big data sharing and its impact on the growth of individuals' welfare. Our results indicate that the optimal level of data sharing achieves the best balance between economic development and privacy, thereby maximizing individuals' welfare. In the short term, big data may inhibit welfare growth; however, in the long term, it promotes sustained improvements in individuals' welfare. Based on these findings, we propose new mechanisms through which big data affects individual welfare.

Keywords: Big Data; Individuals' Welfare; Multiplier Effect; Consumption Growth; Privacy Cost; Macro Model.

JEL Classification Numbers: O30, O33, O39

1. Introduction

In recent years, the importance of big data in the economy has become increasingly evident. As a strategic production factor, big data is driving a new technological revolution. It plays a transformative role in economic systems, prompting many countries to accelerate their capabilities in processing and utilizing big data. The proliferation of computing devices, network infrastructure, and advancements in machine learning have triggered an explosion in big data usage. In 2011, the world-renowned consulting firm Mckinsey released a report titled *"Big Data: The Next Frontier for Innovation, Competition, and Productivity"*, forecasting that the era of big data was imminent and that data would gradually become an essential production factor (Guillaume, 2011). Like roads, railways, ports, hydropower, and communication networks, big data is becoming part of the foundational infrastructure of modern society. However, unlike physical infrastructure, big data is non-rivalrous—it does not depreciate or lose value with use (Anagnostopoulos et al., 2016).

Big data, as a production factor, challenges long-standing assumptions about the value of experience, the nature of expertise, and the organization of production. Society is undergoing a technological revolution fueled by big data. Producers are now able—and will increasingly be able—to measure and understand far more about their operations. This increased knowledge can be directly translated into improved decision-making and enhanced productivity.

As big data becomes a crucial production factor and strategic asset, the need for data sharing and openness is becoming more urgent. However, existing economic theories do not yet adequately explain big data as a new production factor or offer theoretical guidance on its application and development. Drawing on modern theories of economic growth, this paper integrates big data into a macroeconomic model to explain how it influences individual welfare in the digital economy. We ask: As a driving force of technological change, through what paths and mechanisms does big data influence consumption levels and, consequently, welfare? What is the nature and magnitude of this impact on individuals' welfare?

Existing literature has outlined some critical characteristics of big data, which provide partial answers to these questions.

First, big data is non-rivalrous when shared. Goldfarb and Tucker (2019) argue that its low cost gives it characteristics similar to public goods. However, Oussous et al. (2018) point out that its replication, dissemination, and use are independent of time and space. These features make big data non-competitive (Jones and Tonetti, 2020, Yang et al., 2020) and thus different from ordinary public goods. For example, if an individual uses calculus or a special spreadsheet software, s/he cannot limit others' use of calculus or the same spreadsheet software. Anagnostopoulos et al. (2016), argue that big data is different from physical infrastructure because it will not depreciate with use. The authors also note that this lasting value was not a characteristic of previous general-purpose technologies that caused technological revolutions, such as steam engines and computers. Lynch (2008) points out that the low cost and non-competitive nature of big data make it easy to share, copy, and reorganize in open sources. According to Cohen et al. (2016) and Kim et al. (2014), the use of big data has improved consumer surplus and consumer welfare.

Second, the application of big data in production activities is increasing. The application of big data is becoming increasingly important for different sectors(Cong et al., 2021; Badshah et al.,2024). The key to the value created by big data lies in its application. With the rapid development of big data application technology, the use of big data in production activities has gradually increased. Chen et al. (2012) point out that, over the past two decades, business intelligence and analytics based on big data have become increasingly important in both the academic and business communities. Hughes-Cromwick and Coronado (2019) argue that digital platform companies can utilize government data in conjunction with their privately generated data to gain business insights and to make informed decisions. Oussous et al. (2018) showed that the breadth and depth of big data use continue to increase in many fields, such as economics, commerce, public management, national security, and scientific research (for specific examples, see Dong et al., 2017; Edelman, 2012; Lazer et al., 2014). Mikalef et al. (2018), indicate that big data increases the business value of production research and development (R&D) activities. In addition, Blazquez and Domenech (2018) claim that big data can explain the internal connections between economic indicators and the laws of economic development. Zhang et al. (2018) also point out that big data-driven machine learning has achieved great success in many areas, such as image analysis, speech recognition, and text recognition. McAfee et al. (2012) outline the resources needed for processing big data, including social networks, images, sensors, the web, and other unstructured data sources. The amount and variety of big data can generate a positive value proposition (Lawson-Body et al.,2024).

Third, big data has led to a transformation of R&D patterns. The significant growth of the digital economy is leading to some global changes. Yang et al. (2017) show that the processing technology and computing capabilities of big data are continually improving. Etzion and Aragon-Correa (2016) argues that big data offers researchers more attractive research directions and possibilities. Einav and Levin (2014) also point out that big data produces better indicators for researchers to explore industrial transformation. According to Calderaro (2015) traditional thinking models and technologies mainly analyze causality. Big data can overcome this limitation and help researchers make discoveries that go beyond causality and promote the transformation of existing production and R&D patterns(Chen and Zhang, 2014; Cong et al., 2022). Chen et al. (2012) point out that big data also helps enterprises better understand their business and market and make timely business decisions. However, similar to steam engines, electricity, and computers, the transformation caused by big data will require expensive reconstruction with high adjustment costs. Therefore, in the initial stage of its application, big data will perform poorly and necessitate a new set of intermediate goods, like with previous general-purpose technologies (Helpman and Trajtenberg, 1994).

Fourth, big data has a multiplier effect, which accelerates technological progress. According to the existing literature, big data can increase efficiency and drive technological innovation. For example, McAfee et al. (2012) point out that the use of big data by scientists can improve targeting in some decision-making areas, relying on data and rigor instead of intuition. The differences between big data analytics depend on the amount, speed, and diversity of the data. Big data can provide real-time information, which can make companies more competitive. Studies show that the vast potential and value contained in massive amounts of data can be used to mine factors that previously went unnoticed, which can lead to improved resource allocation and efficiency (Chen and Zhang, 2014; Glaeser et al., 2018; Etzion and Aragon-Correa, 2016). According to Sivarajah et al. (2017) and Begenau et al. (2018), big data can reduce companies' operating costs, as well as issues such as heterogeneous demand, networked supply, and decision errors. Furthermore, big data can drive technological progress by increasing the accuracy of artificial intelligence and machine learning, and by promoting the cross-fusion of technologies (Bajari et al., 2019; Zhang et al., 2018; Yang et al., 2020). Big data is a new engine that drives technological innovation, forming an ecosystem of accelerated technical iterations and industrial upgrading (Hilbert, 2016; Yaqoob et al., 2016; Farboodi and Veldkamp, 2021). Like many technological changes in history, the technological transformation caused by big data will make it easier to invent new products and new production processes, thereby bringing new secondary innovations and promoting long-term economic growth (Schaefer et al., 2014).

Lastly, the sharing of big data will generate privacy costs. Chiou and Tucker (2017) show that the development brought by big data poses some privacy concerns. Acquisti et al. (2016) discuss the economics of privacy, explaining that consumers value the privacy of their data. In light of these concerns, scholars have studied the extent to which privacy issues can be manipulated by business and government interests (Acquisti et al., 2015; Cong et al., 2022). Goldfarb and Tucker (2011) note that advertisers use online customer data to create targeted marketing, which has caused consumer concerns about privacy and has led governments to adopt laws restricting data use and online tracking technologies for websites in order to protect consumer privacy. Ali and Abdelfettah (2019) studied the pricing and welfare implications of consumers' information disclosure on firms. Acquisti et al. (2016) analyzed the economic value and consequences of protecting and disclosing personal information, as well as consumers' understanding and decisions regarding the trade-offs associated with privacy and the sharing of personal data. However, Athey et al. (2017) showed that, although people care about privacy, they are willing to give up private data easily when they are motivated by interest.

To date, most literature about the influence of big data on consumption and welfare is descriptive. They tend to explain the use of big data for prediction, the characteristics of big data, its impact on production efficiency, its effect on economic development, and its impact on society from a descriptive perspective. There are fewer in-depth analyses that use theoretical models. Scholars have focused on specific industries by using data mining techniques, which mainly use big data as a sample for empirical tests (Varian, 2014; Cavallo and Rigobon, 2016). These studies treat big data as an information carrier and focus on the relationship between traditional factors of production, but ignore the essential characteristics of big data and how these can influence the welfare level of individuals.

Kshetri (2014) found the relation between characteristics of big data and privacy, security and consumer welfare issues are examined from the standpoints of data collection, storing, sharing and accessibility. This study argues that data elements have "non-competitive" and "low-cost" attributes, and privacy leaks occur when data is shared. Cong et al. (2021, 2022) also found that data in creative activities generates new knowledge and innovation, which can drive economic development. These studies share some similarities with this article, but also differ in some respects. They are similar in that they all examine the role of data in production and economic activities and consider the privacy and role of data. However, they differ in that some of these studies explain factors such as privacy through the logic of language and text, without constructing theoretical models involving privacy. Another part, such as Jones and Tonetti(2020) and Cong et al. (2021, 2022), construct theoretical models, but they mainly focus on the impact on economic growth and consider how to reduce privacy leakage through data intermediaries. They do not involve research on the welfare utility of residents in economic development. In addition, the theoretical frameworks used are different, and they do not conduct research from the perspective of improving the quality of intermediate products. So far, some studies treated big data as an endogenous factor, and few scholars have created economic models that endogenize big data. Existing research does not model the impact of big data on individuals' privacy and welfare based on its key characteristics.

Based on the theory of creative destruction, this article treats big data as a new type of production factor and endogenizes it into the production function. We construct a macro theoretical framework to study the influence of big data on individuals' welfare. We are particularly interested in how big data affects technology change in the economy, thus affecting consumption, privacy, and consumer welfare. We use the concept of multiplier effect to understand how big data upgrades the quality of intermediate goods and divides technical progress into production technology progress and intermediate goods quality improvement. We found that, although both the multiplier effect and the transformation of the R&D pattern affect the level of production technology, the multiplier effect also helps to improve intermediate goods quality. By constructing a model in which individuals' welfare is determined by consumption and privacy costs and in which consumption is influenced by production technology and intermediate goods quality, we show the path and mechanism through which big data can affect the level of welfare, from an atheoretical perspective. We also use numerical simulation to test our theoretical predictions. Our results show that development and privacy should be fully balanced when applying big data. Blindly increasing the degree of data sharing and using it for development has a negative impact on the level of individuals' welfare. Only through optimal sharing can individuals' welfare be maximized. In addition, big data has a restraining effect on individuals' welfare growth in the short term. However, in the long run, it has a significant growth effect on the utility.

Our study makes three contributions to the existing literature. First, by taking big data as an endogenous factor and incorporating it into the production function, we innovatively present the multiplier effect and the dynamic transformation of R&D patterns driven by big data. These processes capture the essential characteristics of big data's fluctuation effect on technological progress and economic growth. Second, unlike the study by Aghion and Howitt (1992), which assumes that technological progress comes only from the improvement of intermediate goods quality, we propose that big data affects production technology through both the multiplier effect and R&D transformation pattern. This more accurate macro model better reflects the realities of big data in modern economies. Third, by considering both consumption and privacy costs in the analysis of individuals' welfare—and by calculating the optimal degree of data sharing—we provide a macroeconomic model that offers theoretical guidance for determining the appropriate level of big data sharing.

The structure of this article is as follows: Section 2 presents the basic model based on the characteristics of big data. Section 3 describes the optimal sharing degree of big data and analyzes the growth of individuals' welfare under the influence of big data, from a theoretical perspective. Section 4 presents the parameter calibration and numerical simulation, and Section 5 the conclusion.

2. The Basic Model

The macro model developed in this paper is based on observations of the role of big data in social and economic activities. We endogenize big data as a production factor within the production function. The final product is determined by a combination of big data, production technology, labor, and intermediate goods. We assume that big data can replace some of the original intermediate inputs, and then change the production function of the final product. Big data has a multiplier effect on the promotion of the quality ladder of intermediate goods. Big data can also induce the transformation of R&D patterns of intermediate enterprises, which affects output and consumption. Furthermore, big data generates privacy costs during the data-sharing process. Together, consumption and privacy costs affect individuals' welfare levels.

2.1 Generating the Final Products

The production of final products requires big data D, labor L, intermediate goods x, and the production technology level A as inputs. Each final product is produced using a corresponding intermediate good. We assume that the labor force (L) is fixed and that the market for final products is perfectly competitive. Over a continuous time interval, each final product is manufactured using its respective intermediate input. The total output Y_t consists of N types of final products Y_{jt}. Assuming symmetry and following Yang et al. (2020), we construct the aggregate output Yt using a Cobb-Douglas production function of the following form:

$$Y_t = \sum_{j=0}^{N} D_{j,t}^{\alpha_2} L_j^{1-\alpha} A_{(\kappa_j,t)}^{1-\alpha_1} \tilde{x}_{j,t}^{\alpha_1} \tag{1}$$

j stands for the j-th final product. α_2 symbolizes the application level of big data in the production function, indicating its relative importance compared to other production factors. α, α_1, $\alpha_2 \in (0,1)$, and $\alpha_1+\alpha_2=\alpha$. $D_{j,t}$ represents the application scale of big data in period t. Considering the non-competitive nature of big data (Jones and Tonetti, 2020), we assume that each final product enterprise can use all the data available in the economy for production. (Similar to labor L and capital K, this paper takes big data as a production factor and uses the amount of big data D to endogenize it into the model. Whether big data belongs to the production side or the consumption side does not have a substantial impact on the amount of data represented in our model. Even data generated by consumers can be used for production decisions, so this paper does not consider the specific source of data, and big data in the production function is homogeneous.). The symmetrical production function means that the generation of the final product has the same labor demand. Let $L_j = L / N$ represents the labor employed by each final product enterprise, and $\tilde{x}_{j,t}$ represent the number of intermediate goods after quality adjustment(This paper is notation heavy and contains several equations stemming from different fields. To allow the reader to navigate better the analytical formulas,we summary construction a nomenclature table, where symbols are explained in detail. The nomenclature table can be found in Appendix A).

In the process of building the macro model, this paper assumes product homogeneity to simplify the analysis without affecting the core argument. The model focuses on overall economic performance rather than the specifics of individual markets. Its central concern is to examine the impact and dynamic effects of big data on the economic aggregate, rather than the relative prices of specific goods or micro-level allocative efficiency. Therefore, ignoring product heterogeneity allows for a more streamlined analysis and better highlights the key macroeconomic mechanisms.

The intermediate goods are developed and produced by the intermediate goods department, and the quality level moves along the quality ladder. We assume that the intermediate goods are non-durable and that, in each period, the final product manufacturer will choose the ones with the highest quality. Successful and innovative companies would monopolize the production and sales of the latest intermediate goods until the next innovation arrives. The quality-adjusted intermediate goods $\tilde{x}_{j,t}$ is determined by both its number x_{jt} and quality level $\bar{q}^{\kappa_j}$. Therefore, following Aghion and Howitt (1992), we construct the expression for quality-adjusted intermediate goods as follows:

$$\tilde{x}_{j,t} = \left(\bar{q}^{\kappa_j}\right)^{\frac{\alpha}{\alpha_1}} x_{j,t} = \left[\left(1+f(t)\right)q\right]^{\kappa_j \frac{\alpha}{\alpha_1}} x_{j,t} \tag{2}$$

$\bar{q}^{\kappa_j}$ is the quality adjustment coefficient of intermediate goods, $\bar{q}^{\kappa_j} > 1$. $\bar{q} = \left(1+f(t)\right)q$, and q is a constant. $f(t)$ stands for the "multiplier effect" of big data on the quality ladder promotion of intermediate goods, and the specific expression can be found in equation (8). κ_j stands for the highest quality level of the *j-th* intermediate goods, and each innovation would increase κ_j by 1.

Then, combining equation (1) and (2), we obtain:

$$Y_t = \sum_{j=0}^{N} D_{j,t}^{\alpha_2} L_j^{1-\alpha} A_{(\kappa_j,t)}^{1-\alpha_1} \left(\bar{q}^{\kappa_j}\right)^{\alpha} x_{j,t}^{\alpha_1} \tag{3}$$

In Equation (3), Y_t is proportional to the application scale of big data $D_{j,t}$ in period t.

2.2 The Expression of Big Data

This paper simplifies the formation of big data as the by-product of production and consumption. Each product will produce one unit of data in the process of production or consumption, and data generated by one enterprise can be used by others after sharing. To simplify the expression, we assume that only the current data is valid for production and R&D, and that the previously generated data loses its corresponding effect due to timeliness. The non-competitive nature of big data is mainly reflected in its low propagation cost and rapid dissemination. For example, given a certain amount of big data used for machine algorithm learning of autonomous vehicles, it can be used for 1 car, 1,000 cars, or 1 million cars at the same time. Therefore, following Jones and Tonetti (2020), we construct the expression for big data as follows:

$$D_{j,t} = \left[m\mu_{j,t} + (1-m)\tilde{\mu}_{j,t}(N-1)\right]Y_{j,t} \tag{4}$$

where m is the importance of an enterprise’s own data compared with the data of other enterprises. $\mu_{j,t}$ is the proportion of the enterprise's data used for the production of its final products. And $\tilde{\mu}_{j,t}$ is the proportion of data generated outside product j but used for the production of type j (The parameters m,μ, are set to explain the impact of the relative importance of the degree of data sharing, data on production generated by one’s own enterprise and by other enterprises, consumption, R&D investment, and intermediate goods production. Therefore, we let parameters m,μ, be fixed exogenous variables.).

2.3 The Welfare of Individuals

We assume that all laborers in the economy are consumers, and that the total population remains unchanged. We assume that welfare is based on the size of the utility, by discounting future utility, we obtain the total current welfare:

$$U = \int_0^\infty e^{-\rho t} L_t u\left(c_t, \mu_{j,t}, \tilde{\mu}_{j,t}\right) dt \tag{5}$$

Where c_t is the consumption of individuals. ρ is the time discount rate. and the positive welfare comes from the consumption of final products. Data disclosure and sharing may lead to severe information leakage and even endanger personal and national security. Thus, we assume the public use of big data will generate privacy costs, which reduce welfare.

Following the construction of utility functions for nonmarketed goods proposed by Brown and Calsamiglia (2007), who provide necessary and sufficient conditions for efficient production of goods such as health, education, and environmental amenities, we address the issue of privacy in big data sharing. This issue is captured using a concave, monotone, continuous production function. We adopt a logarithmic loss function to represent privacy cost. Additionally, drawing on Jones and Tonetti (2020) and Cong et al. (2022), we construct the individuals’ welfare function as follows:

$$u\left(c_t, \mu_{j,t}, \tilde{\mu}_{j,t}\right) = \ln(c_t + 1) - \frac{\zeta}{2}\frac{1}{N^2}\sum_{j=0}^{N} \ln \mu_{j,t} - \frac{\tilde{\zeta}}{2}\frac{1}{N}\sum_{j=0}^{N} \ln \tilde{\mu}_{j,t} \tag{6}$$

where ζ and $\tilde{\zeta}$ indicate the privacy costs generated by an enterprise’s own data and by other enterprises’ data relative to consumption, respectively. The second term on the right side of equation (6) indicates the privacy cost of applying an enterprise’s own data, and the third term captures the privacy cost of using other enterprises’ data. Given that there are N types of products, we sum the privacy costs and compute the average to obtain the per-product average privacy cost. There is an additional 1/N scaling of the $\mu_{j,t}$ privacy cost. Since $\tilde{\mu}_{j,t}$ reflects the cost associated with data usage by all other firms in the economy, it is natural for there to be a factor of N difference between these costs.

The consumption of a single consumer is:

$$c_t = \sum_{j=0}^{N} \frac{C_{j,t}}{L} = \frac{C_t}{L} \tag{7}$$

2.4 The Multiplier Effect of Big Data

According to the famous s-shaped diffusion curve (Griliches, 1957), as with the gradual development of many general purpose technologies, the multiplier effect of big data on the quality ladder promotion of intermediate goods does not happen overnight. At the beginning stage, the multiplier effect of big data increases very slowly, as the efficient methods for achieving improve the quality ladder of intermediate goods have not yet been implemented. With the refinement and improvement of effective techniques, such as big data acquisition technology and big data mining etc., the growth rate of improving the quality ladder of intermediate goods gradually increases. In later stages, the multiplier effect is in full swing, and then converges onto a specific value. When the technology for utilizing and applying big data reaches a sufficiently high level, the multiplier effect reaches its peak value a, and $0<a<1$. That is, big data would have an a-fold effect on the quality ladder. We assume that the "multiplier effect" of big data in period t is $f(t)$, and that its growth rate is $df(t)/dt$, which is proportional to $f(t)$ and $(a-f(t))$ with a proportionality coefficient of k. Accordingly, we can construct the following differential equation:

$$\frac{df(t)}{dt}=\left[a-f(t)\right]kf(t)$$

We can know from the above formula the general solution to the multiplier effect of big data as(The detailed derivation process can be found in Appendix B)：

$$f(t)=\frac{ae^{lt+C_0}}{1+e^{lt+C_0}}=\frac{a}{1+Be^{-lt}} \tag{8}$$

where $B=1/e^{C_0}$, and $l=ak$, with l and B being positive constants. In the early days of application, the multiplier effect is minimal, which means that when $t=0$ and when B is sufficiently larger than a, $f(t)$ is close to zero. As time goes by, Be^{-lt} would gradually approach 0, and the multiplier effect of big data would reach the peak value $f(t)=a$.

2.5 The Transformation of R&D Patterns Driven by Big Data

In the initial stage, the fusion cost between big data and other production factors leads to the atransformation of R&D patterns. Specifically, investment in R&D for intermediate goods would become a complementary investment (Aghion et al., 2019), which would not generate economic benefits immediately. According to the model on major technological change through social learning elaborated by Aghion and Howitt (1998), intermediate manufacturing and R&D companies are in one of three states during the transformation of R&D patterns, State 0 indicates that the enterprise has not yet found a template yet, the template means to find a benchmark that can be used to transform of R&D patterns by big data. State 1 indicates that the enterprise has found a template but has not yet experienced the multiplier effect of big data. State 2 indicates that the enterprise has successfully achieved the transformation of R&D patterns under big data. Thus, State 1 is the transition phase from State 0 to State 2. We use n_0, n_1 and n_2 to indicate the proportion of intermediate manufacturing and R&D enterprises in each state and set the initial condition to $n_0=1$, $n_1=n_2=0$.

The intermediate manufacturing and R&D enterprise can upgrade from state 0 to State 1 if it finds a template independently, or imitates at least h similarly positioned companies that have achieved the transformation of R&D patterns (State 2). According to the difficulty of R&D, let the Poisson arrival rate that the enterprise finds a template independently be λ_0 ($\lambda_0<<1$), and let the Poisson arrival rate of imitating s ($s>h$) similarly positioned companies be 1. Therefore, by imitating similarly positioned companies, the probability of achieving R&D upgrade can be derived from the cumulative binomial:

$$\psi\left(s,h,n_2\right)=\sum_{b=h}^{s}\binom{s}{b}n_2^b\left(1-n_2\right)^{s-b}=\sum_{b=h}^{s}C_s^b n_2^b\left(1-n_2\right)^{s-b} \qquad (9)$$

The successful rate of transiting from State 0 to State 1 is the sum of finding a template and imitating the similarly-positioned firm, which is $\lambda_0+\psi(s,h,n_2)$. Therefore, the number of enterprises upgraded from State 0 to State 1 is equal to the product of n_0 (enterprises in State 0) and $\lambda_0+\psi(s,h,n_2)$.

Companies in State 1 would allocate resources to the new production and R&D mode so, in this stage, the companies have to produce with the original technology. As a result, companies in Sate 1 have no contribution to technology level A until they upgrade to State 2. We assume the Poisson arrival rate that the intermediate manufacturing and R&D enterprise upgrades from State 1 to State 2 is λ_1. Thus, the number of companies upgraded from State 1 to State 2 is the product of n_1(companies in State 1) and λ_1.

In sum, the differential equation for n_1 and n_2 are:

$$\dot{n}_1=\left[\lambda_0+1\times\psi\left(s,h,n_2\right)\right]\left(1-n_1-n_2\right)-\lambda_1 n_1 \qquad (10)$$

$$\dot{n}_2=\lambda_1 n_1 \qquad (11)$$

In equations(10) and (11), under the initial conditions, $n_1(0)=0$, $n_2(0)=0$, and n_0, n_1 and n_2 have the following identity relationship:

$$n_0=1-\left(n_1+n_2\right) \qquad (12)$$

Plugging equations (9) and (12) into equation (10), and then combining equations (10) and (11) we obtain the number of intermediate manufacturing and R&D enterprises in the three states (The detailed derivation process can be found in Appendix C).

$$n_0 = 1 - \frac{\left(t + \lambda_1 t^2\right) \sum_{b=h}^{s} C_s^b n_2^b \left(1 - n_2\right)^{s-b}}{1 + \lambda_1 t + \left(t + t^2 \lambda_1\right) \sum_{b=h}^{s} C_s^b n_2^b \left(1 - n_2\right)^{s-b}} \tag{13}$$

$$n_1 = \frac{t \sum_{b=h}^{s} C_s^b n_2^b \left(1 - n_2\right)^{s-b}}{1 + \lambda_1 t + \left(t + t^2 \lambda_1\right) \sum_{b=h}^{s} C_s^b n_2^b \left(1 - n_2\right)^{s-b}} \tag{14}$$

$$n_2 = \frac{\lambda_1 t^2 \sum_{b=h}^{s} C_s^b n_2^b \left(1 - n_2\right)^{s-b}}{1 + \lambda_1 t + \left(t + t^2 \lambda_1\right) \sum_{b=h}^{s} C_s^b n_2^b \left(1 - n_2\right)^{s-b}} \tag{15}$$

2.6 The Impact of Big Data on Production Technology Level

The non-competitive nature of big data means that it not only serves as a factor of production influencing the generation of final products, but also exerts a multiplier effect on the promotion of intermediate goods along the quality ladder. This, in turn, affects the level of production technology and the growth of individuals' welfare. Under the original R&D pattern, intermediate manufacturing and R&D enterprises can promote the intermediate goods quality from level κ_j to level $\kappa_j+1$with a success rate p_κ in each period, increasing by q-1 times. When the transformation of R&D patterns is finished, every successful innovation would be affected by the multiplier effect with the quality ladder being promoted by $(q-1)(1+f(t))$ times. Therefore, given the technical level in period t, the technical level in period $t+1$ is expressed by the difference equation:

Finally, the technical level of production, A_t, evolves according to the difference equation

$$\dot{A}_{(\kappa_j,t)} = n_0 p_{\kappa_j} \left(q - 1\right) + n_2 p_{\kappa_j} \left(1 + f(t)\right)\left(q - 1\right) \tag{16}$$

where p_κ is the success rate of R&D. The level of production technology is affected by the state of intermediate goods manufacturing and R&D enterprises. Enterprises in State 1 are undergoing a transformation of R&D patterns, which cannot improve the production technology.

3. Optimal Sharing of Big Data and Growth of Individuals' Welfare

3.1 Model equilibrium

We normalize the price of the final product to 1, let the price of the intermediate goods be $p_{j,t}$. The R&D and production of each intermediate goods require one final product. Enterprises that have successfully innovated can get the monopoly price, which equals the marginal value of the corresponding final product. That is:

$$P_{j,t} = \frac{\partial Y_{j,t}}{\partial x_{j,t}} = \alpha_1 D_{j,t}^{\alpha_2} L_j^{1-\alpha} A_{(\kappa_j,t)}^{1-\alpha_1} \bar{q}^{\kappa_j \alpha} x_{j,t}^{\alpha_1 - 1} \tag{17}$$

Equation (17) is also the inverse demand function of the *j*-th intermediate goods. It can be seen that the demand function for the intermediate tilts to the lower right, indicating the incumbent enterprises can get monopoly profits through monopoly production. By solving the formula outlined above. The demand function for the intermediate goods can be written as:

$$x_{j,t}=\left(\frac{\alpha_1 D_{j,t}^{\alpha_2} L_j^{1-\alpha} A_{(\kappa_j,t)}^{1-\alpha_1} \bar{q}^{\kappa_j \alpha}}{P_{j,t}}\right)^{\frac{1}{1-\alpha_1}} \tag{18}$$

The monopoly profit owned by enterprises that have successfully innovated and with the highest quality of intermediate goods is:

$$\pi_j=\left(P_{j,t}-1\right)\left(\frac{\alpha_1 D_{j,t}^{\alpha_2} L_j^{1-\alpha} A_{(\kappa_j,t)}^{1-\alpha_1} \bar{q}^{\kappa_j \alpha}}{P_{j,t}}\right)^{\frac{1}{1-\alpha_1}} \tag{19}$$

According to the first-order condition of price P_{jt} in equation (19), and the principle of price monopoly and maximization of sales profit, the monopoly price for the intermediate goods is: $P_{j,t}=1/\alpha_1$. Substituting $P_{j,t}=1/\alpha_1$ into equation (18) leads to the following inverse demand curve of the intermediate goods:

$$x_{j,t}=\alpha_1^{\frac{2}{1-\alpha_1}} D_{j,t}^{\frac{\alpha_2}{1-\alpha_1}} L_j^{\frac{1-\alpha}{1-\alpha_1}} A_{(\kappa_j,t)} \bar{q}^{\kappa_j \frac{\alpha}{1-\alpha_1}} \tag{20}$$

Based on the formula outlined above, we know that the demand for intermediate goods is positively related to the level of intermediate goods quality, production technology, and the application-level of big data. Equation (20) together with equation (3) give the production function of the final product Y_j:

$$Y_{j,t}=\alpha_1^{\frac{2\alpha_1}{1-\alpha_1}} D_{j,t}^{\frac{\alpha_2}{1-\alpha_1}} L_j^{\frac{1-\alpha}{1-\alpha_1}} A_{(\kappa_j,t)} \bar{q}^{\kappa_j \frac{\alpha}{1-\alpha_1}} \tag{21}$$

Combining equations (4) and (21), we obtain:

$$Y_{j,t}=\alpha_1^{\frac{2\alpha_1}{1-\alpha}} L A_{(\kappa_j,t)}^{\frac{1-\alpha_1}{1-\alpha}} \bar{q}^{\kappa_j \frac{\alpha_1}{1-\alpha}} \left[m\mu_{j,t}+(1-m)\tilde{\mu}_{j,t}(N-1)\right]^{\frac{\alpha_2}{1-\alpha}} \tag{22}$$

As more products are produced and consumed, more data is generated, and these data are collected and utilized subsequently, which increases productivity and promotes output and consumption (The detailed derivation process can be found in Appendix D). Equation (4) together with equation (20) give the demand for intermediate goods:

$$x_{j,t}=\alpha_1^{\frac{2}{1-\alpha_1}} L_j^{\frac{1-\alpha}{1-\alpha_1}} A_{(\kappa_j,t)} \bar{q}^{\kappa_j \frac{\alpha_1}{1-\alpha_1}} \left\{\left[m\mu_{j,t}+(1-m)\tilde{\mu}_{j,t}(N-1)\right] Y_{j,t}\right\}^{\frac{\alpha_2}{1-\alpha_1}}$$

Substituting equation (22) into the equation outlined above, the expression for the demand of intermediate goods can be written as:

$$x_{j,t}=\alpha_1^{\frac{2}{1-\alpha_1}+\frac{2\alpha_1}{1-\alpha}\frac{\alpha_2}{1-\alpha_1}} L_j A_{(\kappa_j,t)}^{\frac{1-\alpha_1}{1-\alpha}} \bar{q}^{\kappa_j \frac{\alpha_1}{1-\alpha}} \left[m\mu_{j,t}+(1-m)\tilde{\mu}_{j,t}(N-1)\right]^{\frac{\alpha_2}{1-\alpha}} \tag{23}$$

The detailed derivation process can be found in Appendix E.

Therefore, the number of intermediate goods used to produce final products is：

$$X_t = \alpha_1^{\frac{2}{1-\alpha_1}+\frac{2\alpha_1}{1-\alpha}\frac{\alpha_2}{1-\alpha_1}} A_{(\kappa,t)}^{\frac{1-\alpha_1}{1-\alpha}} L \sum_{j=0}^{N}\left[m\mu_{j,t}+(1-m)\tilde{\mu}_{j,t}(N-1)\right]^{\frac{\alpha_2}{1-\alpha}} \bar{q}^{\kappa_j\frac{\alpha}{1-\alpha}} \tag{24}$$

We obtain the sum the final products using equation(22), and the total output function can be written as:

$$Y_t = \alpha_1^{\frac{2\alpha_1}{1-\alpha}} A_{(\kappa_j,t)}^{\frac{1-\alpha_1}{1-\alpha}} L \sum_{j=0}^{N}\left[m\mu_{j,t}+(1-m)\tilde{\mu}_{j,t}(N-1)\right]^{\frac{\alpha_2}{1-\alpha}} \bar{q}^{\kappa_j\frac{\alpha}{1-\alpha}} \tag{25}$$

Finally, substituting equations (24) and (25) into $Y_t = C_t + X_t$, we obtain the expression for total consumption:

$$C_t = \left(\alpha_1^{\frac{2\alpha_1}{1-\alpha}} - \alpha_1^{\frac{2}{1-\alpha_1}+\frac{2\alpha_1}{1-\alpha}\frac{\alpha_2}{1-\alpha_1}}\right) L A_{(\kappa,t)}^{\frac{1-\alpha_1}{1-\alpha}} \sum_{j=0}^{N}\left[m\mu_{j,t}+(1-m)\tilde{\mu}_{j,t}(N-1)\right]^{\frac{\alpha_2}{1-\alpha}} \bar{q}^{\kappa_j\frac{\alpha}{1-\alpha}} \tag{26}$$

3.2 The degree of data sharing

Corollary 1: In the process of applying big data for production, individuals' welfare is affected by consumption and privacy costs. The blind pursuit of data sharing produces negative impacts on individuals' welfare. Only through the optimal sharing of big data can individuals' welfare be maximized.

From equation (31), it can be seen that the consumption of the final product has a linear relationship with the sharing parameters of big data μ, $\tilde{\mu}$. As the degree of sharing increases, the consumption in the economy rises accordingly, thereby increasing the welfare of individuals. Regardless of other factors, when big data is fully shared and utilized (that is, when $\mu = 1$, $\tilde{\mu} = 1$), consumption in each period will reach its maximum. However, the role of big data needs to be supported by data involving many individuals, and excessive sharing of data may lead to privacy leakage, which reduces individual welfare and total social welfare. Although it is necessary to promote the development of big data by encouraging data sharing, the protection of privacy must not be ignored.

The socially optimal allocation is a dynamic path of consumption and privacy costs. We maximize the intertemporal welfare of the representative consumer(5), subject to (6), (7), and (26):

$$\max_{\{L,\mu_{j,t},\tilde{\mu}_{j,t}\}} \int_0^\infty e^{-\rho t} L_t u\left(c_t, \mu_{j,t}, \tilde{\mu}_{j,t}\right) dt \tag{27}$$

$$\text{s.t. } c_t = \frac{C_t}{L}$$

$$C_t = \left(\alpha_1^{\frac{2\alpha_1}{1-\alpha}} - \alpha_1^{\frac{2}{1-\alpha_1}+\frac{2\alpha_1}{1-\alpha}\frac{\alpha_2}{1-\alpha_1}}\right) L A_{(\kappa,t)}^{\frac{1-\alpha_1}{1-\alpha}} \sum_{j=0}^{N}\left[m\mu_{j,t}+(1-m)\tilde{\mu}_{j,t}(N-1)\right]^{\frac{\alpha_2}{1-\alpha}} \bar{q}^{\kappa_j\frac{\alpha}{1-\alpha}}$$

$$u\left(c_t, \mu_{jt}, \tilde{\mu}_{j,t}\right) = \ln(c_t+1) - \frac{\zeta}{2}\frac{1}{N^2}\sum_{j=0}^{N}\ln\left(\mu_{j,t}+1\right) - \frac{\tilde{\zeta}}{2}\frac{1}{N}\sum_{j=0}^{N}\ln\left(\tilde{\mu}_{j,t}+1\right)$$

We further solve the optimization conditions, and can obtain the optimal proportion of the enterprise's own data used for the production of its final products, μ_{jt}, can be written as (The detailed derivation process can be found in Appendix F):

$$\mu_{j,t}=\frac{2\frac{\alpha_2}{1-\alpha}m-m\tilde{\zeta}+(1-m)\zeta}{m\tilde{\zeta}-2\frac{\alpha_2}{1-\alpha}m} \tag{28}$$

And the optimal proportion of data generated by other enterprises but used for the production of type j, $\tilde{\mu}_{jt}$, can be expressed as:

$$\tilde{\mu}_{jt}=\frac{m\tilde{\zeta}}{(1-m)\zeta}\left(\frac{(1-m)\zeta}{m\tilde{\zeta}-2\frac{\alpha_2}{1-\alpha}m}\right)-1 \tag{29}$$

From equations (28) and(29), we know that the optimal sharing degree for big data is directly proportional to both the importance of big data α_2 and to the capital factor allocation ratio α. Decision makers not only need to increase the level of welfare by maximizing consumption but also need to consider the privacy costs brought by the excessive sharing of big data. On the one hand, data sharing can fully benefit from the non-competitive advantages of big data, improve the efficiency of production and R&D, and thereby increase individuals' welfare level. On the other hand, data sharing also faces problems such as privacy leakage, which creates privacy costs and reduces individuals' welfare. Therefore, by optimizing the degree of data sharing, it is necessary to obtain the optimal degree of sharing for big data in order to maximize individuals' welfare.

3.3 The growth of individuals' welfare

Corollary 2: In the long-term, big data enhances individuals' welfare, while in the short term, it inhibits their welfare.

Assuming that the degree of data sharing required for the production of various final products is homogeneous, and plugging equations (26), (28), and (29) into equation (6), we obtain the optimal welfare level:

$$u\left(c_t,\mu_{jt},\tilde{\mu}_{j,t}\right)=\ln(c_t+1)-\frac{\zeta}{2}\frac{1}{N}\ln\left(\frac{\tilde{\zeta}m-\zeta(1-m)}{m\tilde{\zeta}+2m\frac{\alpha_2}{(1-\alpha)}}+1\right)-\frac{\tilde{\zeta}}{2}\ln\frac{m\tilde{\zeta}}{(1-m)\zeta}\left(\frac{\tilde{\zeta}m-\zeta(1-m)}{m\tilde{\zeta}+2m\frac{\alpha_2}{1-\alpha}}+1\right) \tag{30}$$

Where $c_t=\left(\alpha_1^{\frac{2\alpha_1}{1-\alpha}}-\alpha_1^{\frac{2}{1-\alpha_1}+\frac{2\alpha_1}{1-\alpha}\frac{\alpha_2}{1-\alpha_1}}\right)A_{(\kappa,t)}^{\frac{1-\alpha_1}{1-\alpha}}\sum_{j=0}^{N}\left[m\mu_{j,t}+(1-m)\tilde{\mu}_{j,t}(N-1)\right]^{\frac{\alpha_2}{1-\alpha}}\bar{q}^{\kappa_j\frac{\alpha}{1-\alpha}}$.

As we can see, welfare is determined by consumption and by the cost of privacy. When the degree of sharing is optimal, the cost of privacy will no longer change, and changes in consumption will determine the variation of welfare. Therefore, individuals' welfare will grow at the same rate as consumption. The growth rate of consumption is determined by the increase in intermediate goods quality and the increase in production technology. Based on the leading edge of the quality ladder κ, the improvement of the intermediate goods quality level is determined by the R&D success rate p_κ and the quality ladder promotion of intermediate goods. According to equation (24), the quality level of intermediate goods after one phase of R&D is:

$$\tilde{q}_{t+1} = p_{\kappa}\bar{q}^{\frac{(\kappa_j+1)\alpha}{1-\alpha}} + (1-p_{\kappa})\bar{q}^{\frac{\kappa_j\alpha}{1-\alpha}} \tag{31}$$

Therefore, the change in the quality level of the intermediate goods is:

$$\dot{\tilde{q}}(t) = \left((1+f(t))^{\frac{\alpha}{1-\alpha}} q^{\frac{\alpha}{1-\alpha}} - 1 \right) p_{\kappa} \tag{32}$$

According to formula (13), the shift in production technology level *A* is:

$$\dot{A}(t) = n_0 p_{\kappa}(q-1) + n_2 p_{\kappa}(1+f(t))(q-1) \tag{33}$$

From equations (26), (32), and (33), we can obtain the growth rate of individuals' welfare under the influence of big data. When deriving the growth rate of various economic indicators, we assume that the relative importance of the data between enterprises and the sharing parameters m, μ, are fixed values, which means they do not affect the growth of individuals' welfare. This setting only omits changes in the sharing parameter and affects the size of D, but does not have a substantial effect on the trend of technological progress and on the growth of individuals' welfare as influenced by big data.

$$\gamma_u = \gamma_C = \frac{1-\alpha_1}{1-\alpha}\dot{A}(t) + \dot{\tilde{q}}(t) = \frac{1-\alpha_1}{1-\alpha}\left[n_0 p_{\kappa}(q-1) + n_2 p_{\kappa}(1+f(t))(q-1)\right] + \left((1+f(t))^{\frac{\alpha}{1-\alpha}} q^{\frac{\alpha}{1-\alpha}} - 1 \right) p_{\kappa} \tag{34}$$

In the equation above, we can see that the growth rate of individuals' welfare depends on the production technology progress and on the intermediate goods quality improvement brought by successful R&D. Before the application of big data, that is, when $n_0=1$, $n_1=n_2=0$, $\alpha_2=0$, and $f(t)=0$ in equation (34), the growth rate of individuals' welfare can be written as:

$$\gamma_{u0} = \left[p_{\kappa}(q-1)\right] + \left(q^{\frac{\alpha}{1-\alpha}} - 1 \right) p_{\kappa} \tag{35}$$

When big data is fully applied and the transformation of R&D patterns is completed, that is when $n_0=0$, $n_1=0$, $n_2=1$, and $f(t)=a$, the growth rate of welfare γ_{u1} is:

$$\gamma_{u1} = \frac{1-\alpha_1}{1-\alpha}\left[p_{\kappa}(1+a)(q-1)+1\right] + \left(\left[(1+a)q\right]^{\frac{\alpha}{1-\alpha}} - 1 \right) p_{\kappa} \tag{36}$$

Comparing formula (35) and formula (36), where $0<\alpha_1<\alpha<1$, then $(1-\alpha_1)/(1-\alpha)>1$, $\alpha/(1-\alpha)>0$ and $q>1$, $a>0$. Taking these conditions into account, the growth rate of individuals' welfare after big data is used is higher than that before big data is used, that is, $\gamma_{u1}>\gamma_{u0}$. Therefore, big datapromotes the growth of individuals' welfare in the long run.

By comparing equations (34) and (35), we can see that, in the beginning, when intermediate enterprises are in the stage of model transformation, n_0 decreases rapidly, but n_2 increases slowly. That is, when n_1 is large, and $f(t)$ is small, the growth rate of individuals' welfare will be less than that before the application of big data, indicating there exists $\gamma_{u1}>\gamma_{u0}$. Therefore, big data has an inhibitory effect on the growth rate of individuals' welfare in the short term.

To maintain the growth effect of big data while weakening its short-term negative impact, it is crucial to achieve a sustainable and stable improvement in residents' welfare. The key lies in accelerating the "R&D mode transformation" of R&D enterprises—specifically, increasing the probability that R&D firms independently discover innovation templates, speeding up the transition from intermediate state 1 to state

2, and reducing n_1. Simultaneously, it is important to fully leverage the "multiplier effect" of big data by increasing the proportional coefficient l in $f(t)$, enhancing the multiplier effect during economic slowdowns, and offsetting the impact of slowed technological progress. Moreover, it is essential to mitigate the privacy costs associated with big data, since data sharing can significantly affect residents' utility in the short term due to privacy breaches. Therefore, actively promoting the breadth and depth of big data applications and fully unlocking its inherent potential will accelerate the transformation speed of R&D enterprises and enhance the multiplier role of big data. This approach can better stimulate consumption, reduce the negative impact of privacy concerns, and ultimately improve residents' welfare more effectively.

4. Parameter Calibration and Numerical Simulation

In this section, we use the numerical method to simulate the model of Part 3. We show the dynamic process of the transformation of R&D patterns and changes in the growth rate of individuals' welfare caused by big data, thus verifying the rationality of Corollary 2.

4.1 Parameter Calibration

The parameters are set based on the characteristics and existing research. The primary goal is to make the simulation results close to social reality. Currently, there are only a few studies on the impact of big data on the growth of individuals' welfare. Therefore, so some parameters are difficult to obtain from the existing research and can only be selected based on the nature of the model and the empirical facts. For these parameters, we carry out the sensitivity analysis to test the robustness of the conclusions.

Because previous research has not addressed the multiplier effect of big data and the transformation of R&D patterns directly, we draw on studies that have analyzed these phenomena for general poupous technology in . Based on the study by Schaefer et al. (2014), we set the extreme value for the multiplier effect a as 0.03, which means $1+a=1.03$. We set the coefficient B in $f(t)$ as 200 and the proportionality coefficient l as 0.2. Jones and Tonetti（2020）assume that the share of big data in production (α_2) falls between the interval of [0.03, 0.12] after it has been fully applied. Based on their settings, we set α_2 as 0.06 and use $\alpha_2=0.06$ and $\alpha_2=0.12$ for sensitivity analysis.

With reference to the parameter settings in the study by Aghion and Howitt (1998), we assume the Poisson arrival rate where intermediate manufacturing and R&D enterprises can separately find a "template" and transit from State 0 to State 1 to be $\lambda_0=0.01$. We set the Poisson arrival rate that the intermediate manufacturing and R&D enterprise upgrades from State 1 to State 2 as $\lambda_1=0.4$. The number of similarly positioned companies in the industry is $s=10$, and the least number for imitating companies that have successfully transformed is $h=3$.

According to the setting of Schaefer et al. (2014), the intermediate goods quality improvement caused by successful R&D is $q=1.06$ and the total labor supply $L=1$. It is relatively common to set the return on capital to one-third of total production, that is, $\alpha=1/3$. Therefore, we first set $\alpha=1/3$ to test the generality of the inferences and then set $\alpha=0.4$ and $\alpha=0.3$ for sensitivity analysis. Let the benchmark consumption growth rate without the influence of big data be $\gamma_{0Y}=0.03$, which is the world average in recent years, and we can solve the success rate of R&D as $p_{0\kappa}=0.333$.

4.2 Numerical Simulation Results and Evaluation

(1)Benchmark Model

Figure 1 shows the time trajectories of intermediate manufacturing and R&D enterprises being in states n_0, n_1, and n_2 during the transformation of R&D patterns. n_2 (enterprises that have completed the transformation) increases gradually at the beginning stage and then accelerates. The maximum growth rate for n_2 appeared at a certain point in the middle of the region. After that, the growth rate gradually decreased, and eventually, n_2 converged to 1. The largest value for n_1 (enterpreises in the transitions stage) appeared in the 16th year, and the starting and ending point of n_1 are both 0. As for n_0, it has a downward trend in the entire interval, indicating that the number of enterprises with the original R&D pattern decreases gradually until the transformation has been completed. If the arrival rate λ_0 for finding the template is minimal, then both n_1 and n_2 will rise slowly and the values will be small for a long time.

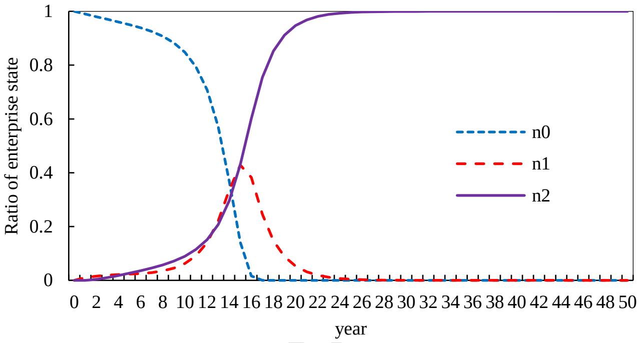

Figure 1 The transformation of R&D patterns influenced by big data

The change process of the multiplier effect of big data is presented in Figure 2. In the first ten years, the application technology and methods of big data are not yet mature. Thus, the multiplier effect of big data is relatively small. With the accumulation of experience, the multiplier effect continues to rise from the 10th to the 30th year, and it gradually converges to 0.03 in the later period. The simulation shows that the multiplier effect of big data presents an upward trend similar to the logistics curve. It fits well with the multiplier effect of big data that promotes the quality ladder of intermediate goods and is consistent with the analysis of the previous theoretical model.

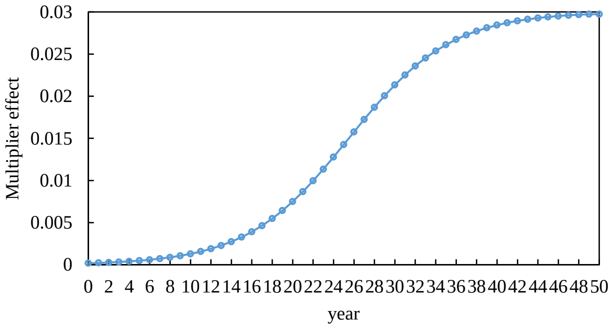

Figure 2 The multiplier effect of big data

Figure 3 shows the impact of big data on the growth rate of individuals' welfare. As predicted by the theoretical framework, in the first 11 years, the impact of big data on the growth rate of individuals' welfare is minimal. During this period, the growth rate of individuals' welfare is relatively stable, with a slight decline of 0.1% compared to the base period. The growth of welfare begins to experience a period of recession when the proportion of enterprises in State 1 surges and the growth rate falls by more than 0.8%

compared with the base period. Therefore, the numerical simulation shows that, in the short term, because intermediate R&D enterprises need to make complementary investments and adapt to the new pattern, a significant downturn of welfare growth is generated, which is consistent with the theoretical expectation of Corollary 2.

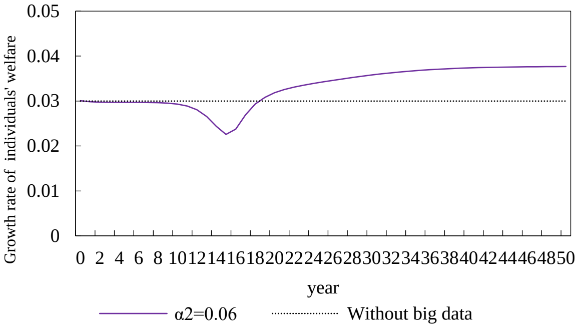

Figure 3 The impact of big data on the growth of individuals' welfare

Starting in the 11th year, the growth of individuals' welfare began to decline significantly, and the growth rate went down in the 15th year, reaching about 2.25%. The growth of individuals' welfare gradually increased in the 16th year, reaching a 3.35% increase in the 23rd year. During this period, the multiplier effect of big data increased rapidly, and enterprises completed the transformation, so the growth of individuals' welfare also multiplied. During the 20 years from the 23rd to the 42nd, there was an increase in the growth of individuals' welfare to 3.75% in the 42nd year. After the 42nd year, the growth rate gradually converged to γ=3.77% and formed a stable state, which exceeded 0.77% of the base period. We can see from numerical simulations that, after big data is fully infiltrated in intermediate R&D enterprises, it has a long-term growth effect on the individuals' welfare, which is in line with the theoretical expectation of Corollary 2.

(2)Sensitivity Analysis

In this section, we perform the sensitivity analysis on some of the parameters. That is, we change the value of a parameter while keeping the others unchanged to see whether the impact of big data on the growth of individuals' welfare varies substantially.

First, we focus on the arrival rate for the intermediate manufacturing and R&D enterprises to discover a template, independently, during the transformation of R&D patterns. The arrival rate λ_0 is previously set to 0.01, and now we change it to 0.007 and 0.013 to represent a lower and higher arrival rate, respectively. The results of the numerical simulations are presented in Figure 4. We can see that changes in λ_0 vary the rate to which enterprises transit from State 0 to States 1 and 2, and then affect the period and scale of the individuals' welfare growth downturn. Also, the dynamic evolution paths of the model differ with the arrival rate $\lambda_{0,}$ as larger values in λ_0 make the transition process faster. Thus, changes in λ_0 only have a slight influence on the size and time of the recession, but they do not change the growth trend in individuals' welfare nor the steady-state growth rate.

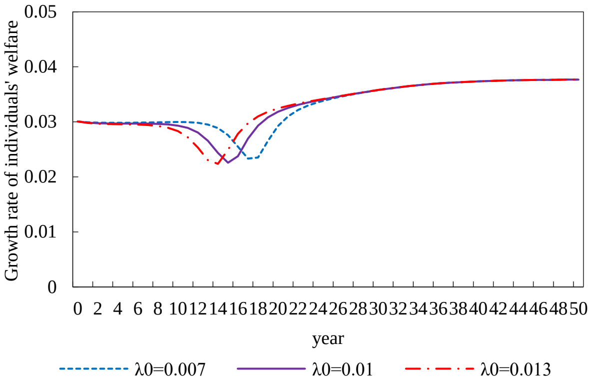

Figure 4 The impact of different arrival rates on the growth of individuals' welfare

Second, we pay attention to changes in the maximum application-level of big data α_2. Based on the benchmark model, we reset α_2 to 0.03 and 0.12 as control groups, and the result are shown in Figure 5. They show that the larger α_2 brings a higher steady-state growth in individuals' welfare. As this article focuses on the direction of the impact of big data on the growth of individuals' welfare, the change in the application-level of big data does not affect the direction and trend of individuals' welfare growth.

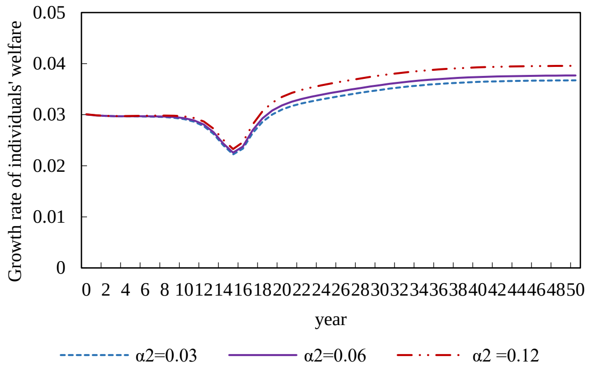

Figure 5 The impact of the different application-levels of big data on the growth in individuals' welfare

In sum, the impact of big data on individuals' welfare is not sensitive to changes in the arrival rate to new templates or to changes in the extreme value of the multiplier effect. This indicates that our results are robust.

5. Conclusions and Recommendations

The development of big data has become key concern given that it is a strategic production factor spearheading a new round of scientific, technological, and industrial revolution. Nevertheless, the impact of big data on individuals' welfare has remained undertheorized. Based on the theory of creative destruction, we treated big data as an endogenous factor and included it into the production function. We explained how big data affects consumption and generates privacy costs, thereby influencing individuals' welfare. We also constructed an innovative macro model explaining how the multiplier effect and the transformation of R&D patterns driven by big data lead to technological progress and consumption growth in the long term. Thus, we analyzed the impact of big data on individuals' welfare from the perspective of consumption and privacy, and we described these effects theoretically.

Our results show that economic development and privacy protection should be fully balanced when applying big data. Blindly increasing the degree of data sharing will harm individuals’ welfare, and only be establishing the optimal degree of sharing can individuals’ welfare be maximized. In the short term, big data has an inhibiting effect on the growth of individuals' welfare, while in the long run, it has a significantly increases individuals’ welfare. To observe the role of big data on welfare growth, it is necessary to not only fully considerthe multiplier effect, but also to guide the smooth transformation of R&D production models and to reduce the scale and duration of the downturn brought by the transformation. If the privacy issues caused by data sharing cannot be solved, enterprises will be reluctant to publish data. As a result, there will be a reduction in the sharing of data resources, and then it will be difficult to effectively mine and utilize big data. In this case, big data will fail to play its essential role in promoting production and consumption.

To make better use of big data and to increase individuals’ welfare, we make the following policy recommendations. First, it is necessary to support the construction of new infrastructure for big data mining and big data circulation platforms. This will optimize the big data ecosystem, activate the multiplier effect, and thereby reduce the privacy costs that would initially decline individuals' welfare. Second, it is necessary to speed up legislation on big data sharing and gradually implement the classified sharing principle. Specifically, we must promote the “one class, one policy” rule to control different types of data with varying levels of sharing. Finally, sharing compensation mechanisms must be formulated for various sources and types of data. We must implement a transparent collection of privacy-related data to reduce privacy costs caused by data leakage, and to inactive data sharing in these circumstances.

References

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. Science, 347(6221), 509-514.

Acquisti, A., Taylor, C., & Wagman, L. (2016). The economics of privacy. Journal of economic Literature, 54(2), 442-492.

Aghion, P., & Howitt, P. (1992). A Model of Growth Through Creative Destruction. Econometrica, 60, 323-351.

Aghion, P., Bergeaud, A., Boppart, T., Klenow, P. J., & Li, H. (2019). Missing growth from creative destruction. American Economic Review, 109(8), 2795-2822.

Aghion, P., Howitt, P. (1998).Endogenous Growth Theory. Cambridge MA, MIT Press.

Ali, A., & Abdelfettah, B. (2019). Financial disclosure information, board of directors, and firm characteristics among French CAC 40 listed firms. Journal of the Knowledge Economy, 10(3), 941-957.

Anagnostopoulos, I., Zeadally, S., & Exposito, E. (2016). Handling big data: research challenges and future directions. The Journal of Supercomputing, 72, 1494-1516.

Athey, S., Catalini, C., & Tucker, C. (2017). The digital privacy paradox: Small money, small costs, small talk (No. w23488). National Bureau of Economic Research..

Badshah, A., Daud, A., Alharbey, R., Banjar, A., Bukhari, A., & Alshemaimri, B. (2024). Big data applications: Overview, challenges and future. Artificial Intelligence Review, 57(11), 290.

Bajari, P., Chernozhukov, V., Hortaçsu, A., & Suzuki, J. (2019). The Impact of Big Data on Firm

Performance: An Empirical Investigation. Aea Papers and Proceedings, 109, 33-37.

Begenau, J., Farboodi, M., & Veldkamp, L. (2018). Big data in finance and the growth of large firms. Journal of Monetary Economics, 97, 71-87.

Blazquez, D., & Domenech, J. (2018). Big Data sources and methods for social and economic analyses. Technological Forecasting and Social Change, 130, 99-113.

Brown, D. J., & Calsamiglia, C. (2007). The nonparametric approach to applied welfare analysis. Economic Theory, 31(1), 183-188.

Calderaro, A. (2015). Big Data: A Revolution That Will Transform How We Live, Work, and Think. Media Culture & Society, 37, 1113-1115.

Cavallo, A., & Rigobon, R. (2016). The billion prices project: Using online prices for measurement and research. Journal of Economic Perspectives, 30(2), 151-178.

Chen, C. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information sciences, 275, 314-347.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. MIS quarterly, 1165-1188.

Chiou, L., & Tucker, C. (2017). Search engines and data retention: Implications for privacy and antitrust (No. w23815). National Bureau of Economic Research.

Cohen, P., Hahn, R., Hall, J., Levitt, S., & Metcalfe, R. (2016). Using big data to estimate consumer surplus: The case of uber (No. w22627). National Bureau of Economic Research.

Cong L W, Xie D X, Wei W S, Zhang L T. (2022). Endogenous growth under multiple uses of data. Journal of Economic Dynamics and Control, 141, 104395.

Cong L W, Xie D X, Zhang L T. (2021). Knowledge accumulation, privacy, and growth in a data economy. Management science, 67(10), 6480-6492.

Dong, L., Chen, S., Cheng, Y., Wu, Z., Li, C., & Wu, H. (2017). Measuring economic activity in China with mobile big data. EPJ Data Science, 6, 1-17.

Edelman, B. (2012). Using internet data for economic research. Journal of Economic Perspectives, 26(2), 189-206.

Einav, L., & Levin, J. (2014). Economics in the age of big data. Science, 346(6210), 1243089.

Etzion, D., & Aragon-Correa, J. A. (2016). Big data, management, and sustainability: Strategic opportunities ahead. Organization & Environment, 29(2), 147-155.

Farboodi M., Veldkamp L. . (2021). A model of the data economy (No. w28427). Cambridge, MA, USA: National Bureau of Economic Research.

Glaeser, E. L., Kominers, S. D., Luca, M., & Naik, N. (2018). Big data and big cities: The promises and limitations of improved measures of urban life. Economic Inquiry, 56(1), 114-137.

Goldfarb, A., & Tucker, C. (2019). Digital economics. Journal of economic literature, 57(1), 3-43.

Goldfarb, A., & Tucker, C. E. (2011). Privacy regulation and online advertising. Management science, 57(1), 57-71.

Griliches, Z. (1957). Hybrid Corn: An Exploration in the Economics of Technological Change, *Econometrica*, 25, 501-522.

Guillaume Dupuy d'Angeac, 2011. Big data, The Next Frontier for Innovation, Competition, and Productivity. New York, Mckinsey and Company.

Helpman, E., & Trajtenberg, M. (1994). A Time to Sow and a Time to Reap: Growth Based on General Purpose Technologies. *NBER Working Papers*.

Hilbert, M. (2016). Big data for development: A review of promises and challenges. Development Policy Review, 34(1), 135-174.

Hughes-Cromwick, E., & Coronado, J. (2019). The value of US government data to US business decisions. Journal of Economic Perspectives, 33(1), 131-146.

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

Kim, G. H., Trimi, S., & Chung, J. H. (2014). Big-data applications in the government sector. Communications of the ACM, 57(3), 78-85.

Kshetri, N. (2014). Big data′ s impact on privacy, security and consumer welfare. Telecommunications Policy, 38(11), 1134-1145.

Lawson-Body, A., Illia, A., Lawson-Body, L., Rouibah, K., Akalin, G., & Tamandja, E. M. (2024). Big Data Analytics and Culture: Newly Validated Measurement Instruments for Developing Countries' Value Proposition. Journal of Organizational and End User Computing (JOEUC), 36(1), 1-30.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: traps in big data analysis. science, 343(6176), 1203-1205.

Lynch, C. (2008). How do your data grow?. Nature, 455(7209), 28-29.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. Harvard business review, 90(10), 60-68.

Mikalef, P., Pappas, I. O., Krogstie, J., & Giannakos, M. (2018). Big data analytics capabilities: a systematic literature review and research agenda. Information systems and e-business management, 16, 547-578.

Oussous, A., Benjelloun, F. Z., Ait Lahcen, A., & Belfkih, S. (2018). Big Data technologies: A survey. Journal of King Saud University-Computer and Information Sciences, 30(4), 431-448.

Schaefer, A., Schiess, D., & Wehrli, R. (2014). Long-term growth driven by a sequence of general purpose technologies. Economic Modelling, 37, 23-31.

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. Journal of business research, 70, 263-286.

Varian, H. R. (2014). Big data: New tricks for econometrics. Journal of economic perspectives, 28(2), 3-28.

Yang, C., Huang, Q., Li, Z., Liu, K., & Hu, F. (2017). Big Data and cloud computing: innovation opportunities and challenges. International Journal of Digital Earth, 10(1), 13-53.

Yang, J., Li, X M., Huang, S J. (2020). Impacts on environmental quality and required environmental regulation adjustments: A perspective of directed technical change driven by big data. Journal of Cleaner Production, 275, 124126.

Yaqoob, I., Hashem, I. A. T., Gani, A., Mokhtar, S., Ahmed, E., Anuar, N. B., & Vasilakos, A. V. (2016). Big data: From beginning to future. International Journal of Information Management, 36(6), 1231-1247.

Zhang, Q., Yang, L. T., Chen, Z., & Li, P. (2018). A survey on deep learning for big data. Information Fusion, 42, 146-157.

Ethical statements

This article does not contain any studies with human participants performed by any of the authors.

Data availability

The datasets analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare no competing interests.