Journal of International Economic Studies (2025), No.39, 237–257

Effect of Digital Data on Economic Growth: A Policy Analysis based on a Dynamic General Equilibrium Model incorporating Digital Data

Takashi Matsumura
Tokyo Denki University, School of Science and Technology for Future Life.

Abstract

This study analyzes the effects on macroeconomic growth of the transfer to firms digital data such as personal identification information and purchase history that are generated as a byproduct of household consumption. Specifically, we analyzed the effects of improvements in the quality of digital data, like advances in statistical analysis methods that create value from data and in the big data analysis used by firms in their marketing. We first expressed digital data generation as a by-product of consumption in a model, then constructed a dynamic general equilibrium (DGE) model incorporating digital data into firms' production function, and conducted a dynamic simulation analysis. The results showed that improving the quality of digital data positively affects output and leads to economic growth.

Keywords: dynamic general equilibrium model, digital data, production function, data economy, data economics
JEL Classification: E10, E22, E23, E60

1. Introduction

Digital data continues to accumulate owing to the increased information and communication technology (ICT) capital, and the development of an information society that utilizes this capital. Data have been called "the oil of the 21st century," and their utilization is expected to influence a country's economic development. Therefore, analyzing how information and digital data growth interacts to create new value, and affects the economy is becoming increasing essential.[1]

[1] Noguchi (2019) defines data capitalism as a society in which big data creates value.

Extant research has examined the effects of ICT capital, which is necessary for producing digital data, on the macroeconomy. In empirical studies, Japan Center for Economic Research (JCER; 2000), Shinozaki (2003a, b), Ministry of Internal Affairs and Communications (MIC; 2007, 2014), and Ministry of Internal Affairs and Communications, Information and Communications Bureau (MIC ICB; 2022) incorporate ICT capital into Cobb-Douglas-type macro production functions and estimate them using Japanese statistical data.

In a theoretical model study, Prettner (2019) analyzes the economic effects of artificial intelligence (AI)/robot capital on from a long-term economic growth perspective. Meanwhile, Lin and Weise (2019), Fueki and Maehashi (2019), and Matsumura (2021) analyze the effects from a short-term business cycle perspective using dynamic stochastic general equilibrium (DSGE) models. However, these studies focus on the effects of labor substitution by AI/robot capital on the economy, but not of digital data.

Next, studies have examined the effects of digital data on macroeconomy. First, as with ICT capital, many empirical studies have analyzed digital data's influence by adding digital data to macroeconomic production functions. MIC (2014) and Washio and Shinozaki (2021) use a similar approach. Specifically, they show that digital data, individually or together with ICT capital, has network externalities[2] with increasing returns to scale. Kono and Yoshimoto (2024) measure the value of digital data by evaluating the value of data, databases, and data analysis using a cost-accumulation method.

However, the cost-accumulation method results may significantly differ depending on how the scope of the data and activities related to data creation are considered. Additionally, the value of the digital data itself, which is automatically generated, and that which is analyzed instantly or automatically using a system or AI that has already been completed and used for corporate activities is zero.

Finally, theoretical model studies adopt three approaches depending on the modeling method.[3] The first approach is based on Jones and Tonetti (2020). They view data as ideas and formulate economic models in the form of being equivalent to productivity based on the role of idea in economic growth of Romer (1990). The second approach is that of Farboodi and Veldkamp (2022). They assume that using data can help reduce prediction errors, thereby improving the expected product quality and productivity. However, reducing prediction errors, and consequently, improving expected quality has limits. Hence, even if the data increase, the structure is such that

2 See MIC (2007) for a discussion of network externalities.

3 Studies on (digital) data economics are spreading to various fields. Here, we introduce some main approaches to digital data and macroeconomic fluctuations (growth), and explain this study's positioning in that context. Therefore, Veldkamp and Chung (2024) and Swallow and Haksar (2019) are useful references for how data economics is being treated in other fields.

indefinite economic growth is not possible. Both approaches treat data as a by-product of consumption and have a production structure that uses data to improve productivity. However, Jones and Tonetti (2020) consider data to be something that advance ideas and believe that data fully depreciate at the end of each period. Meanwhile, Farboodi and Veldkamp (2022) treat data as something that contributes to prediction, which makes it possible to accumulate data. Further, they introduce the concept of data depreciation, evaluating data's value as a stock. The third approach is that of Abis and Veldkamp (2024), who propose a method of separating the two aforementioned approaches based on the awareness that they equate data production and use. Specifically, regarding data use, the idea is that knowledge is produced by combining data and labor, and that this becomes productivity. Specifically, two types of knowledge are produced. The first type is created by AI analysts with machine learning skills using new technologies. This is formulated using a Cobb-Douglas production function based on AI analysts' labor input and structured data. The second type is produced by old technology analysts using conventional technology and similarly formulated. The sum of these two knowledge types is assumed to be productivity. Furthermore, the data produced are not raw but rather screened structured data that can be easily used for analysis. Structured data are modeled as being produced from data managers' labor input.

In summary, all three approaches provide significant findings. However, some points have not been fully considered. First, although studies discuss data as a by-product of consumption, they have not sufficiently modeled data production, or how the initial raw data are produced. Varian (2018) proposes a data pyramid, which divides value creation from data into four categories: data, information, knowledge, and action. Data are collected, organized, and analyzed to become information, and the knowledge gained from this is accumulated as knowledge and is ultimately supposed to foster action. Furthermore, owing to recent developments in ICT, AI, and machine learning, this process has been automated and accelerated for large amounts of data. Abis and Veldkamp (2024) separately analyze data production and use, and model data organization, analysis, and the subsequently generated information and knowledge, targeting structured data. However, they do not explicitly model the first part of data collection: the production of raw data.

Second, the data are modeled such that it is linked to productivity. For the structured data (not raw data) discussed by Abis and Veldkamp (2024), it is consistent with the model wherein data managers' labor input contributes to its production, and knowledge is produced by each data analyst using the structured data, resulting in productivity. However, directly linking raw data with productivity is difficult as data are a by-product of consumption. Households provide firms with data, like their own identification information and purchase history, generated because of their consumption behavior. However, they are often unaware about being compensated for this. That is, while a household originally owns such data, the rights to these data have been actually transferred to the firm. Ibarra et al. (2018) show that when the rights to data ownership are strongly held by firms that collect and process data, rather than by households, the aspect of "data as capital" becomes stronger rather than "data as labor." As such, a more desirable way to model raw data would be considering its aspect as capital.

Drawing on studies that have analyzed the effects of digital data on the macroeconomy using theoretical models, we express the raw digital data generation process as a by-product of consumption using a model, which remains an understudied area. Additionally, we construct a model that allows raw digital data to accumulate, as it serves as capital rather than productivity. Furthermore, we reflect on prior research findings showing that digital data have increasing returns to scale and increase the overall economy's output. Accordingly, we construct a DGE model that considers these three aspects. It then analyzes the macroeconomic effects of improvements in the quality of digital data.

The remainder of this article proceeds as follows. Section 2 formulates the DGE model constructed, incorporating the process of generating raw data and digital data with increasing returns to scale. In Section 3, the parameters are calibrated for model analysis and a dynamic simulation analysis using the model is presented. Furthermore, a robustness check is performed based on multiple scenarios that change digital data-related parameters. Section 4 summarizes the conclusions of the study.

2. DGE model formulation

2.1. Model overview

The main study objective is analyzing the effect of improving the quality of digital data on the overall economy's output with increasing returns to scale by modeling the process of generating raw digital data as a by-product of consumption. Therefore, a simple DGE model is more suitable than a complex New Keynesian DSGE model because it is easier to determine the effect of this improvement in the quality of digital data. Here, we formulate a simple DGE model that incorporates digital data. The three economic agents in the model are households, firms, and the government, and all markets are assumed to be perfectly competitive.

First, we assume Ricardian households that exhibit optimal behavior at different time points. Households obtain positive (negative) utility from consumption (labor). They then lend capital to firms to receive rental costs and save money by purchasing government bonds. Digital data are automatically generated due to household consumption behavior, regardless of their intentions, in an economy with ICT capital. Ownership of this type of digital data has passed from households to firms. However, opinions differ regarding whether this is an economic transaction wherein households receive compensation from firms, a non-economic transaction wherein households are uncompensated, or whether they are not aware of it.[4] Because this consideration is not clear, we capture it as a non-economic transaction and model it.[5] Therefore, digital data are assumed to not affect household budget constraints.

The digital data we focus on are those data that firms directly obtain and use, like household identification information and purchase history, which emerge because of household consumption behavior. Firms emerge by borrowing traditional capital, ICT capital, and labor supplied by households as inputs, while paying households for each input. In our case, firms can use digital data automatically generated due to household consumption behavior without paying any costs to households.

The government receives tax revenue collected from households through a lumpsum tax and from the issuance of government bonds. The government intervenes in the economy by spending and paying interest on government bonds equal to its revenue.

Next, we describe each economic entity's formulation.

2.2. Households

Households behave as Ricardian households that optimize between different points in time. Ricardian households receive positive utility from consumption c_t and negative utility from labor supply n_t, and optimize consumption at different points in time by maximizing the discounted present value of their expected utility U_t as follows:

$$U_t = E_t \sum_{t=0}^{\infty} \beta^t \left[\frac{(c_t)^{1-\gamma}}{1-\gamma} - \frac{\chi (n_t)^{1+\sigma}}{1+\sigma} \right], \tag{1}$$

where t is the time point, β is the discount rate that discounts the expected utility to the present value, γ is the relative risk aversion coefficient, σ is the inverse of the elasticity of labor supply substitution, and χ is the parameter representing the ineffectiveness of labor. Households have labor n_t, traditional capital k_t, and ICT capital z_t, which they supply to firms and receive wages w_t, traditional capital rental costs r_t^k, and ICT capital rental costs r_t^z as compensation. They

---

[4] For example, point services offered by firms can be seen as a way of paying for data or strategy for acquiring customers.

[5] Therefore, the term "digital data capital" could be mistaken for the former economic transaction. Therefore, we use the term "digital data" rather than "capital," although it has the characteristics of accumulated capital. Here, ICT capital is considered as "electronic devices and computers that can be connected to information and communication networks" as defined by the MIC. Meanwhile, digital data are considered as the digitized data itself, which includes information on identification information and purchase history, among other digital information goods. Therefore, software, which refers to programs that run on computers, is generally classified as a digital information good. However, it is not data as described above and is considered here as a part of ICT capital and not digital data. For more information on the characteristics of information goods, see Shapiro and Varian (1998).

also invest in government bonds b_t. Each household's budget constraint equation for each period is as follows:

$$c_t + i_t^k + i_t^z + b_t = w_t n_t + r_t^k k_t + r_t^z z_t + R_{t-1}^b b_{t-1} - \tau_t. \tag{2}$$

Further, capital is formulated as follows:

$$k_{t+1} = (1 - \delta_k) k_t + i_t^k, \tag{3}$$

$$z_{t+1} = (1 - \delta_z) z_t + i_t^z, \tag{4}$$

where i_t^k is traditional capital investment, i_t^z is ICT capital investment, δ_k is depreciation rate of traditional capital, and δ_z is depreciation rate of ICT capital. The depreciation rate of ICT capital is higher than that of traditional capital because of the relative speed of performance improvements in ICT capital.[6]

By solving the household utility maximization problem and organizing the first-order conditions of optimization for consumption, labor, government bonds, traditional capital, and ICT capital, the following holds:

$$\chi {n_t}^{\sigma} = c_t^{-\gamma} w_t, \tag{5}$$

$$c_t^{-\gamma} = \beta R_t^b c_{t+1}^{-\gamma}, \tag{6}$$

$$R_t^b = r_{t+1}^k + 1 - \delta_k, \tag{7}$$

$$R_t^b = r_{t+1}^z + 1 - \delta_z. \tag{8}$$

2.3. Firms

Firms produce by borrowing traditional capital, ICT capital, and labor supplied by households as inputs in a perfectly competitive market, and pay households for each input. Furthermore, the quality-embedded digital data D_t are available at no cost. Therefore, we assume that production is based on the following Cobb–Douglas-type structure production function, where traditional capital, ICT capital, and labor are constant returns to scale, and quality-embedded digital data D_t are increasing returns to scale.

---

6 See Lin and Weise (2019).

$$y_t = a_t k_t^{\alpha} z_t^{1-\gamma_n-\alpha} n_t^{\gamma_n} D_t^{\nu}, \qquad \alpha + \gamma_n = 1, \nu > 0, \tag{9}$$

where y_t is output, a_t is productivity, α is traditional capital share, γ_n is labor share, $1-\gamma_n-\alpha$ is ICT capital share, and ν is quality-embedded digital data effect ($\nu > 0$).[7] Firms perform their activities according to a Cobb–Douglas-type production function.

Quality digital data are formulated as follows:

$$D_t = \xi_t d_t, \tag{10}$$

where d_t is digital data and ξ_t is the quality of digital data. Digital data d_t are raw data like personal identification information and purchase history, which are generated as a by-product of household consumption. Specifically, these are sales record data linked to household information accumulated on the corporate side through e-commerce use, point-card use for fee payments, point-of-sales (POS) data, identification POS data, and other means. Such digital data are a by-product of household consumption, but households do not necessarily intentionally provide digital data to firms. Here, firms can accumulate digital data without paying households.[8]

Furthermore, utilizing such digital data, either directly or indirectly, in firms' production activities is difficult. Digital data must be processed to a sufficiently high quality to be used in firm production and business, or value must be created from it. This is done via data analysis, like statistical analyses. Specifically, by analyzing data, raw digital data can be converted into high-quality digital data that can be used in a firm's production. Such data analysis can be conducted by the firm itself or outsourced to a specialist company. Furthermore, as ICT and AI advance further, this type of ICT capital with advanced functions may continue to accumulate and evolve, and eventually be automatically processed by free AI and other systems. Therefore, we define digital data as an exogenous change.

Therefore, we formulate and digital data investment as follows:

$$d_{t+1} = (1-\delta_d) d_t + i_t^d, \tag{11}$$

$$i_t^d = c_t^{\mu_c} z_t^{\mu_z}. \tag{12}$$

---

7 The formulation is based on empirical studies, MIC (2007, 2014), and Washio and Shinozaki (2021).

8 Consequently, firms do offer personalized advertisements to households based on individual preferences, discount services that differ for each household, loyalty points, and so on. However, personalized advertising, for example, has both positive and negative effects on utility: while it can increase household convenience, it can also cause an invasion of privacy. This point needs further examination via research on households' consumption behavior. Therefore, we constructed the model assuming that firms do not affect households' consumption behavior regarding the generation of data capital, and that firms have access to data capital without payment of any costs.

where δ_d is the depreciation rate of digital data and i_t^d is the digital data investment. Digital data investment is formulated by assuming that it is generated based on a Cobb-Douglas function from household consumption and ICT capital. As noted before, we consider digital data to be a by-product of consumption that is directly used by firms. Thus, household consumption behavior provides the foundation, while digital data generation depends on household consumption. The generation of digital data as digital information goods is expected to depend not only on household consumption but also on the quantity and quality of ICT capital present in the economy as a whole. This is because digital data cannot be created in a world in which ICT capital does not exist. Therefore, we consider both household consumption and ICT capital to be necessary for generating digital data and formulate these two components in a Cobb–Douglas-type structure. The returns to scale of the input factors can be diminishing, constant, or increasing. ICT development has made it easier to generate large amounts of digital data. Furthermore, social factors, like the social tolerance of firms holding personal information, and systems and customs, have eased. Thus, there is a possibility of increasing returns to scale. However, if the personal identification information and purchase history that can be obtained has limitations, decreasing returns to scale may occur even if the input elements increase. Originally, this should be determined through an empirical analysis using statistical data; unfortunately, such public statistics do not yet exist. Therefore, we use simulations to clarify how the effect on the overall economic output changes in each case.

Next, using the production function, we consider the firms' profit-maximization problem. Note that quality-embedded digital data D_t are available at no cost. That is, they are not included in the costs. The firm's profit function π_t is:

$$\pi_t = y_t - r_t^k k_t - r_t^z z_t - w_t n_t. \tag{13}$$

By solving the firm's profit-maximization problem and organizing the first-order conditions of optimization for traditional capital, ICT capital, and labor, the following equations hold:

$$r_t^k = \alpha \frac{y_t}{k_t}, \tag{14}$$

$$r_t^z = (1 - \gamma_n - \alpha) \frac{y_t}{z_t}, \tag{15}$$

$$w_t = \gamma_n \frac{y_t}{n_t}. \tag{16}$$

2.4. Government

First, we consider government revenue. The government issues total government bonds b_t and charges taxes from households as a lump-sum tax τ_t. Second, we consider government spending. The government makes interest payments on government bonds, $R^b_{t-1}b_t$, and spending g_t. In summary, the government follows the following budgetary constraints:

$$b_t + \tau_t = R^b_{t-1}b_{t-1} + g_t. \tag{17}$$

Additionally, government expenditure is treated as an exogenous variable.

Furthermore, as a policy rule to stabilize government bonds, the constraint that the lumpsum tax should be within the range of government spending in a log-linear approximation around the steady state is imposed:

$$\tau_t = b_{t-1}{}^{\frac{b}{\tau}\phi}. \tag{18}$$

2.5. Resource constraint and exogenous variable

The resource constraint condition is:

$$y_t = c_t + i^k_t + i^z_t + g_t. \tag{19}$$

Finally, we consider three exogenous variables: productivity, quality of digital data, and government expenditure. Each variable is assumed to follow the dynamic equation of a first-order autoregressive process with log-linear approximations around the steady state. Here, $\rho_a, \rho_\xi, and\ \rho_g$ are the parameters representing the persistence of the change, and $e^a_t, e^\xi_t, and\ e^g_t$ are the exogenous change whose steady state is 0, respectively:

$$a_{t+1} = a^{1-\rho_a} a_t^{\rho_a} exp(e^a_t), \tag{20}$$

$$\xi_{t+1} = \xi^{1-\rho_\xi} \xi_t^{\rho_\xi} exp\left(e^\xi_t\right), \tag{21}$$

$$g_{t+1} = g^{1-\rho_g} g_t^{\rho_g} exp\left(e^g_t\right). \tag{22}$$

3. Simulation analysis of the DGE model

Section 3 considers an equation system comprising first-order conditional equations for the optimization problem formulated in Section 2. These equations were log-linearized around the steady state and calibrated. Specifically, the benchmark case parameters are set based on empirical analysis, actual economic statistical data, previous studies, and assumed scenarios. A dynamic simulation analysis is conducted to determine how a persistent increase in the quality of digital data affects output. Next, robustness checks are performed using dynamic simulation analysis based on multiple scenarios with varying parameters related to digital data.

3.1. Setting production function parameters

The firms' production function incorporates quality-embedded digital data in a form with increasing returns to scale. However, to our knowledge, no study has conducted an empirical analysis that explicitly states the digital data directly used by firms, as assumed here, while separating traditional and ICT capital. Without any prior work to rely on, we set the estimated production function obtained from the following empirical analysis as the parameter.[9] The data for output, traditional capital, ICT capital, and labor (number of workers *x* working hours) used to estimate the production function are from the "Annual Report on the National Accounts" published by the Economic and Social Research Institute, Cabinet Office, Government of Japan, and the accompanying data from the "Survey on Economic Analysis of ICT, " compiled and published by the MIC based on these data. For the capacity utilization rate of traditional capital, we use data from the "Indices of Industrial Production" of the Ministry of Economy, Trade and Industry. The estimation period was 1980-2021 (some periods are linked and combined).

For estimation, the production function constructed in the DGE model must be somewhat consistent because no public statistics corresponding to digital data are available. Hence, we assume constant returns to scale for the production factors of traditional capital, ICT capital, and labor, and estimated the following equation by taking the logarithm of both sides. We employ the generalized method of moments (GMM) estimate, not the ordinary least squares (OLS) estimate, because of the possible presence of endogeneity in the explanatory variables. Table 1 presents the results.

$$ln\left(\frac{y_t}{n_t}\right) = \alpha_0 + \alpha_1\, ln\left(\frac{k_t \cdot RCU_t}{n_t}\right) + \alpha_2\, ln\left(\frac{z_t}{n_t}\right) + \alpha_3 trend \tag{23}$$

9 MIC (2014) and InfoCom Research (2015), discussed below, do not separately estimate distribution rates for traditional and ICT capital, although they explicitly include data capital that is directly used by firms, as assumed here.

Table 1 GMM Estimation of Production Function

Parameter	Estimate	P-Value
α_0	-2.0517	0.0000
α_1	0.2892	0.0337
α_2	0.0896	0.0262
α_3	0.0081	0.0000

Instrument specification: α_0, ln(k(-1)*rcu(-1)/n(-1)), ln(z(-1)/n(-1)), trend(-1), trend(-2), trend(-3)
Endogenous variables to treat as exogenous: ln(k*rcu/n) ln(z/n)
J-stats 6.192149, P-Value 0.0452
Source: Created by the author

where RCU_t is the capacity utilization rate of traditional capital. The instrumental variables were constant terms, $ln(k_t \cdot RCU_t/n_t)$, and $ln(z_t/n_t)$ variables with lags up to one periods as well as trend variable with lags up to three periods. All coefficients are statistically significant at the 5% level. The endogeneity test results show that the p-value is sufficiently small at 0.0452, and the null hypothesis that $ln(k_t \cdot RCU_t/n_t)$ and $ln(z_t/n_t)$ are exogenous variables is rejected at the 5% significance level.[10] Further, the traditional capital share $\alpha = 0.29$, ICT capital share $1 - \gamma_n - \alpha = 0.090$, and labor share $\gamma_n = 0.62$.

Next, the magnitude of this coefficient is compared with the value calculated from the statistical data to verify its consistency. First, the labor share is calculated as the sum of "compensation of employer" and "taxes on production and imports" divided by "gross domestic product" from the "Annual Report on the National Accounts" data published by the Economic and Social Research Institute. The average value from 1980 to 2021 is 0.59, with a minimum of 0.55 and maximum of 0.62. The most recent trend is upward from 2015 to 2021. Meanwhile, capital share, which is the sum of traditional and ICT capital, is calculated as one minus the above labor share. The average, maximum, and minimum values are 0.42, 0.45, and 0.38, respectively. The most recent trend is downward from 2015 to 2021. The GMM results fall within the historical range of values produced from statistical data and are generally consistent with reality.

Digital data distribution rates are taken from estimates by MIC (2014) and InfoCom Research (2015), who conducted a large-scale questionnaire survey to understand the actual big data utilization by firms and others, and estimated the data traffic volume by industry and media. Then, the relationship between data traffic volume and economic growth is specified by a Cobb Douglas-type production function, and estimated using generalized least squares (GLS) with a pooled dataset. In the media, "customer data" and "sales record data in e-commerce" are the data which are similar to the digital data assumed here. Therefore, the result of these estimates,

10 This implies, namely, that OLS estimation causes problems.

(together) 0.06, is used as the digital data distribution rate.

Overall, the parameters used in the DGE model are the traditional capital share $\alpha = 0.29$, ICT capital share $1 - \gamma_n - \alpha = 0.09$, labor share $\gamma_n = 0.62$, and quality-embedded digital data effect $\nu = 0.06$.

3.2. Setting other parameters

The other parameter values are mainly obtained from Eguchi (2011), whose estimates were made using data on the Japanese economy. For the capital depletion rate, the ICT capital depreciation rate is calculated in reverse, such that the value after service life is 10%. This is because ICT capital generally has a shorter service life than traditional capital. According to the MIC ICB (2022), the service life of information capital is four to six years. Furthermore, intangible assets generally depreciate faster than tangible assets. The digital data assumed here are the identification information and purchase histories of households, which emerge as a by-product of household consumption behavior. Consequently, the type of business is expected to be a firm in the B-to-C market, like convenience stores and supermarkets where consumers make frequent purchases, and home appliance retailers which handle seasonal products. Firms generally tend not to use long-term historical data; however, marketing data analyses may involve year-on-year comparisons. Therefore, we assume that data is depreciated over a period of two years, which enables year-on-year comparisons from the perspective of data use, and set the benchmark case as a depreciation rate of 0.435, which is less than 1% of the residual value after two years on a quarterly basis.

The coefficients μ_c and μ_z of the Cobb-Douglas-type digital data investment function, which uses household consumption and ICT capital as input factors, have not been empirically analyzed in studies. Further, confirming them using data is difficult. Accordingly, this study assumes constant returns to scale for the benchmark case, where household consumption and ICT capital have the same effect on digital data investment, and sets this at 0.5.[11] For steady-state values, the steady-state government expenditure to output ratio and government bonds to output ratio are set from statistical data. Furthermore, an increase in the quality of digital data is not expected to be temporary, but is characterized by structural changes. Therefore, to simulate the model's dynamic properties, we set parameter ρ_{id} representing persistence of values as close to 1 as possible.[12] Thus, these changes are represented as continuing over a long period and, after the onset of a change, approximately express the transition from one steady state to another.

The main parameter and steady-state values for the benchmark case are listed in Table 2.

---

[11] However, the constant returns to scale assumption need not necessarily hold. Accordingly, we consider the cases of decreasing and increasing returns to scale in the scenario analysis described below.
[12] That is, it has a similar shape to a unit root shock.

Table 2 Parameters and Steady State Values (Benchmark case)

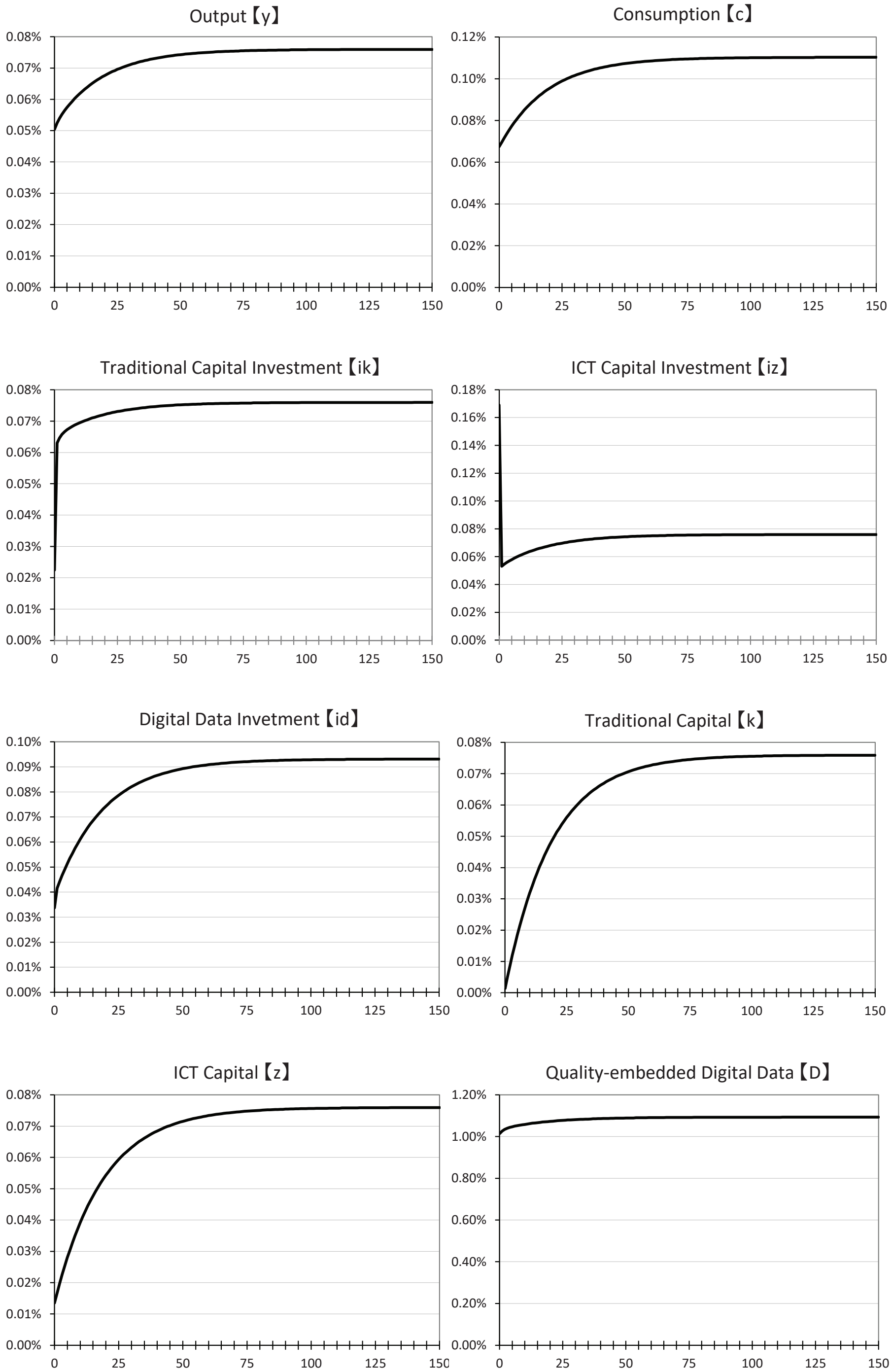
Parameter and Steady State	Value
Discount rete	$\beta = 0.996$
Inverse of the elasticity of substitution between different points in time	$\gamma = 1.5$
Inverse of the elasticity of labor supply	$\sigma = 2.0$
Disutility of labor	$\chi = 1.0$
Depreciation rate of traditional capital	$\delta_k = 0.06$
Depreciation rate of ICT capital	$\delta_z = 0.08$
Depreciation rate of digital data	$\delta_d = 0.435$
Traditional capital share	$\alpha = 0.29$
Labore share	$\gamma_n = 0.62$
Quality-embedded digital data effect	$\nu = 0.06$
Consumption share of digital data investment	$\mu_c = 0.5$
ICT capital share of digital data investment	$\mu_z = 0.5$
Persistence of productivity increases	$\rho_a = 0.9$
Persistence of government expenditure increases	$\rho_g = 0.9$
Persistence of increases of quality of digital data	$\rho_\xi = 0.999999$
Steady state of government expenditure to output ratio	$g/y = 0.2$
Steady state of government bonds to output ratio	$b/y = 2.5$

Source: Created by the author

3.3. Dynamic simulation analysis

The constructed DGE model is log-linearized around the steady-state. A dynamic simulation analysis of output, consumption, traditional capital investment, ICT capital investment, digital data investment, traditional capital, ICT capital, and digital data is conducted for a 1% increase in the quality of digital data. Since these changes are expected to be structural rather than temporary changes, they are assumed to continue over a long period and represent an approximation of a steady-state transition situation. Figure 1 presents the dynamic simulation analysis results using the constructed DGE model.

Figure 1 Dynamic simulation of increases of quality of digital data



The figure illustrates dynamic simulation when quality of digital data increases by 1% in benchmark case.
Source: Created by the author

First, we consider a case wherein the quality of digital data increases. First, quality-embedded digital data directly increases by 1% and its effect is assumed to be long-lasting. Therefore, immediately after a change in the quality of digital data occurs, quality-embedded digital data increases by 1.01%. An increase in quality-embedded digital data affects output through the production function, which indicates a firm's production technology. The effect on the output depends on the magnitude of the quality-embedded digital data effect parameter $\nu = 0.06$. Initially, the output will be under pressure to increase by 0.05%. Furthermore, as digital society expands and increasing returns to scale in digital data strengthen, the parameter ν is expected to become larger. Therefore, increases in output are expected to become even larger. An increase in output affects consumption, traditional capital investment, and ICT capital investment through the equilibrium conditions in the goods market.

In the parameter settings for this benchmark case, the quality of digital data increases. Therefore, initially, traditional capital investment and ICT capital investment react strongly. However, in the long term, everything, including consumption, increases to a constant value, and eventually the steady-state value increases by approximately 0.11% for consumption, and 0.076% for both traditional and ICT capital investments. Because of the increase in each investment, the accumulation of traditional and ICT capital progresses. Eventually, the steady-state value of each capital increases by approximately 0.076%.

Capital accumulates according to the transition equation; therefore, the time required to complete the transition to steady-state values is longer than that for investment. Increased consumption and ICT capital resulting from an increase in ICT capital investment increase the pressure on digital data investment. Digital data investment initially increases by 0.03% in response to the increasing change in the quality of digital data. Then, the increasing pressure persists because of the spillover effects described above, eventually increasing to a steady-state value of 0.093%. The initial change of the increase in the quality of digital data persists for a long time. Furthermore, the subsequent spillover effects sustain. Ultimately, the quality-embedded digital data itself shifts to a steady-state value that increases by 1.09%. Thereafter, the output increases under pressure and the steady-state value of the output shifts, resulting in an increase of 0.076% than that before the change occurred. This suggests that when the quality of digital data in society increases, economic output also increases, contributing to economic growth.

Next, we take a look at the current situation in Japan. According to the MIC (2020), the digital data utilization rate of Japanese firms is not as high as it was in 2020. The firm survey results on digital data analysis methods show that "data browsing" and "aggregation" have a high overall utilization rate of about 70%, but "statistical analysis (correlation analysis, variance analysis, etc.)" and "prediction using AI such as machine learning and deep learning" are utilized by only slightly less than 50% and more than 10% of firms, respectively. Clearly, the data analysis system within the firm is inadequate and there is a lack of specialized human resources.

The results suggest that promoting policies that increase the value of digital data can increase the economy's total output, thus contributing to economic growth. These policies are essentially policies to increase the number of human resources specializing in data analysis in society and create an environment and develop technologies that facilitate data analysis.

3.4. Robustness checks for parameters in the dynamic simulation analysis

Next, we perform a robustness check for the parameters related to digital data in the dynamic simulation analysis. Specifically, we check how much the simulation results change compared to the benchmark case when the parameters are changed in the four scenarios (eight cases), as shown in Table 3.

Table 3 Robustness check scenarios for dynamic simulation analysis

No	Scenarios	Case	δ_d	ν	μ_c	μ_z
1	Change in depreciation rates for digital data (δ_d)	Benchmark	0.435	0.06	0.5	0.5
		Case1	0.99	0.06	0.5	0.5
		Case2	0.1	0.06	0.5	0.5
2	Change in the quality-embedded digital data effect (ν)	Benchmark	0.435	0.06	0.5	0.5
		Case3	0.435	0.1	0.5	0.5
		Case4	0.435	0.01	0.5	0.5
3	Change in consumption share of digital data investment function under the constant returns to scale	Benchmark	0.435	0.06	0.5	0.5
		Case5	0.435	0.06	0.7	0.3
		Case6	0.435	0.06	0.3	0.7
4	Changes in returns to scale of digital data investment function	Benchmark	0.435	0.06	0.5	0.5
		Case7	0.435	0.06	1.0	1.0
		Case8	0.435	0.06	0.3	0.3

Source: Created by the author

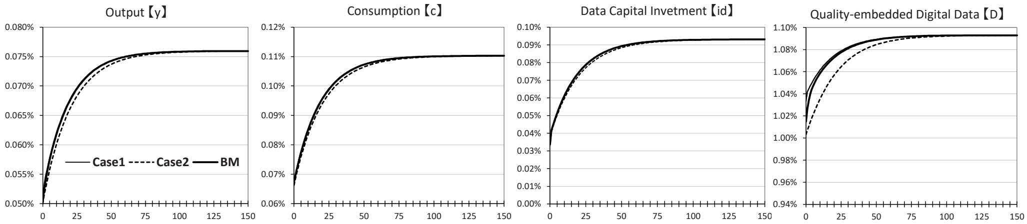
The first check is the change in the digital data attrition rate δ_d. The rapid obsolescence of digital data means that the value of the long-term data used by firms to make decisions decreases. This means that household consumption behavior is becoming less influenced by past consumption behavior and household consumption decisions are becoming increasingly short-term. In Case 1, the obsolescence of digital data is set to $\delta_d = 0.99$, which is faster than benchmark case and the residual value is set to 1% or less in one quarter. Meanwhile, in Case 2, we set the balance of the traditional capital depreciation rate $\delta_k = 0.06$ and ICT capital depreciation rate $\delta_z = 0.08$, as it takes longer for digital data to become obsolete. $\delta_d = 0.1$ is a value that shows that the residual value falls to 10% or less after 6 years and 2 quarters. In Scenario 1 of Figure 2, the dynamic simulation results showed that the faster the digital data became obsolete (i.e., the greater the depreciation rate), the shorter the period required for many variables to shift to new steady-state values. Furthermore, they did not contribute much economic growth. However, the benchmark case and Case 1 are almost superimposed on the graph and visually confirming the differences between them is difficult. That is, the digital data in the benchmark case becomes obsolete relatively quickly and effect of the digital data becoming obsolete more quickly than in the benchmark case is not very significant.

The second check is the change in the quality-embedded digital data effect ν. An increase in this value indicates that economic growth drivers shift from tangible to intangible capital and society becomes one with more characteristics of an information economy. This means that the importance of digital data increases in economic production activities. In Scenario 2 of Figure 2, the benchmark case is set to ν=0.06. However, in Case 3, we set ν=0.1, which is more effective than the benchmark case. Finally, in Case 4, we set ν=0.01, which is less effective than the benchmark case. The dynamic simulation results show that in Case 3, besides the steady-state value of quality-embedded digital data increasing from 1.09% to 1.17% of the benchmark case, the steady-state values of consumption and digital data also increase. The steady-state value of output increased from approximately 0.076% to approximately 0.135%. However, Case 4 shows the opposite result to Case 3. Thus, as the characteristics of the information economy strengthen and importance of digital data increases, improvements in the quality of digital data contribute more to economic growth.

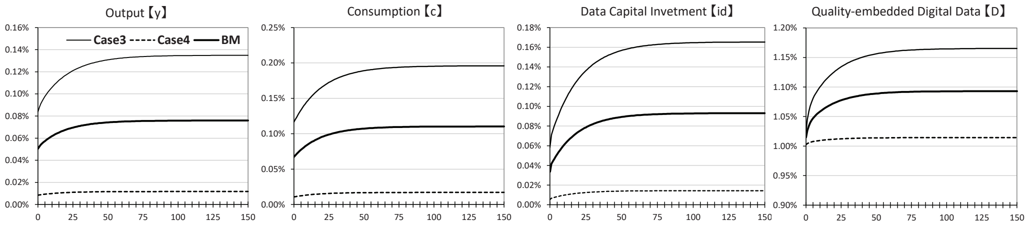
The third check is the asymmetry between the consumption share μ_c and ICT capital share μ_z which is important when producing digital data by aggregating consumption and ICT capital using Cobb-Douglus functions. Here, we assume a constant returns to scale. Asymmetry $\mu_c > \mu_z$ means that there is sufficient ICT capital and performance to generate digital data, and that consumption is relatively important for generating digital data. That is, consumption is more scarce in this economy. From Scenario 3 in Figure 2, the benchmark case ($\mu_c = \mu_z = 0.5$) is compared with Case 5 ($\mu_c = 0.3, \mu_z = 0.7$), which has a low consumption share, and Case 6 ($\mu_c = 0.7,\ \mu_z = 0.3$), which has a high consumption share. As in the benchmark case, we assume constant returns to scale, enabling us to verify only the effect of asymmetric shares. The direction of change in the overall dynamic simulation is not significantly different from that in the benchmark case. However, higher consumption shares tended to increase the steady-state value of the output. Thus, as societies become more characterized by an information economy with sufficient amounts and performance of ICT capital to generate digital data, improvements in the quality of digital data contributes more to economic growth.

Figure 2 Dynamic simulation of increases of quality of digital data by robustness checks for parameters by scenarios

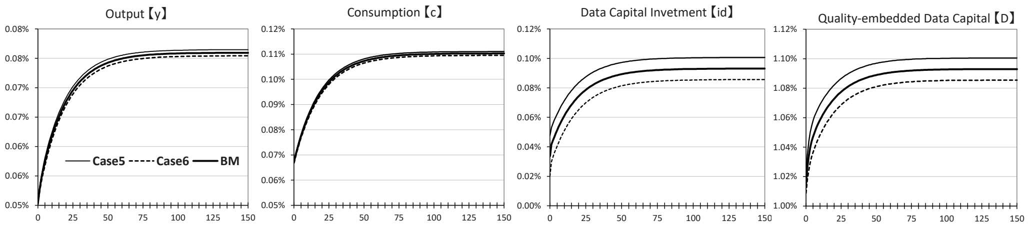
Scenario 1



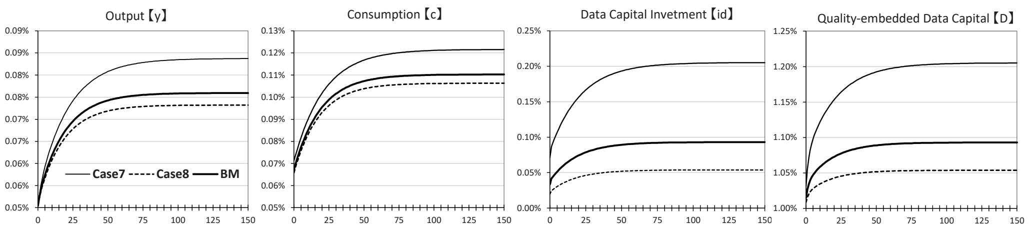
Scenario 2



Scenario 3



Scenario 4



The figure illustrates dynamic simulation by Table3's scenario when quality of digital data increases by 1%.
Source: Created by the author

The fourth check compares the returns to scale of both factors, which are important when generating digital data, by aggregating consumption and ICT capital using the Cobb-Douglus functions. Here, the symmetry between the consumption share μ_c and ICT capital share μ_z are maintained. An economy in which digital data generation has increasing returns to scale is thought to be achieved through the further deepening of the digital society. In this case, from Scenario 4 in Figure 2, the benchmark case with the assumption of constant returns to scale is set, and Case 7 ($\mu_c = 1.0,\ \mu_z = 1.0$), which has increasing returns to scale, and Case 8 ($\mu_c = 0.3,\ \mu_z = 0.3$), which has decreasing returns to scale, are set and checked. As with the benchmark case, we assume the symmetry of the shares. Hence, the effect of returns to scale can be checked. The overall trend of the dynamic simulation results is not significantly different from that of the benchmark case. However, Case 7 showed a tendency to increase the steady-state values overall compared to the benchmark case. Meanwhile, Case 8 showed a tendency to decrease the steady-state values overall compared to the benchmark case. The steady-state values of the output are 0.084% for Case 7, 0.076% for the benchmark case, and 0.073% for Case 8. Thus, as the degree of increasing returns to scale in digital data increases, the improvement in the quality of digital data is more likely to contribute to economic growth. For example, if the legal system for firms to hold personal identification information is further expanded, stricter information management systems are enacted, and social acceptance increases, the digital economy will deepen further and the quality of digital data will be more likely to contribute to economic growth.

Finally, in all four robustness checks for parameters, no cases showed a response that significantly differed from the dynamic propagation path of the benchmark case. This demonstrates the robustness of the parameters of the benchmark case.

4. Conclusion

This study analyzes the effects of the transfer of digital data, like personal identification information and purchase history that are generated as a by-product of household consumption, to firms on macroeconomic growth. We first expressed digital data generation as a by-product of consumption, and then constructed a DGE model where a firm's production factors comprise four elements: traditional capital, ICT capital, labor, and digital data. Subsequently, we incorporated a production function wherein digital data has increasing returns to scale and conducted a dynamic simulation analysis.

The results showed that an increase in the quality of digital data positively affects output and leads to economic growth. Thus, promoting policies which increase the value of digital data can increase output, thereby contributing to economic growth. For example, policies should increase individuals specializing in data analysis, and promote the development of an environment and technology that facilitates the statistical analysis of data. Finally, our robustness checks of the digital data-related parameters also examined the cases that contribute more to economic growth.

References

Abis, S. and Veldkamp, L. (2024), The Changing Economics of Knowledge Production, *The Review of Financial Studies*, 37 (2024), pp.89–118.

Eguchi, M. (2011), *Dougakuteki ippankinko moderu niyoru zaiseiseisaku no bunnseki (Analysis of Fiscal Policy in a Dynamic General Equilibrium Model)*, Mitsubishi keizai kenkyujyo, Japan.

Farboodi, M. and Veldkamp, L. (2022), A Model of the Data Economy, *NBER Working Paper*, 28427.

Fueki, T., and Maehashi, K. (2019), Inflation Dynamics in the Age of Robots: Evidence and Some Theory, *Bank of Japan Working Paper Series*, No.19-E-9, June 2019.

Ibarra, I. A., Goff, L., Hernández, D. J., Lanier, J., and Weyl, E. G. (2018), Should We Treat Data as Labor? Moving beyond "Free", *AEA Papers and Proceedings*, Vol.108, pp.38–42.

InfoCom Research (2015), Bigdeta no ryutsuryo no suikei oyobi bigdata no katsuyojittai ni kansuru chosakenkyu hokokusho (Report of Research Study on Estimation of Big Data Circulation Volume and Actual Utilization of Big Data).

JCER (2000), Nihonkeizai no saisyuppatsu II IT kakushin no syougeki to sono hyouka (Japan's Economy Restart II Impact of IT innovation and its evaluation), Japan Center for Economic Research, Japan.

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

Lin, T-t. T., and Weise, C. (2019), A New Keynesian Model with Robots: Implications for Business Cycles and Monetary Policy, *International Atlantic Economic Society*, Vol. 47, pp. 81-101.

Matsumura, T. (2021), Analysis of Basic Income Policy Based on AI/Robot Capital Income Taxation Using a Dynamic Stochastic General Equilibrium Model, *The Hosei University Economic Review*, Vol.89, No.1, pp. 85-104.

MIC (2007), *Information and Communications in Japan*, Ministry of Internal Affairs and Communications, Japan.

MIC (2014), *Information and Communications in Japan*, Ministry of Internal Affairs and Communications, Japan.

MIC (2020), *Information and Communications in Japan*, Ministry of Internal Affairs and Communications, Japan.

MIC ICB (2022), *ICT no Keizai bunseki ni kansuru chosa* (*Survey on Economic Analysis of ICT*), Information and Communications Bureau, Ministry of Internal Affairs and Communications, Japan.

Noguchi, Y. (2019), *De-ta Shihonsyugi (Data Capitalism)*, Nihonkeizai Sinbun Syuppansya, Japan.

Prettner, K. (2019), A Note on the Implications of Automation for Economic Growth and the Labor share, *Macroeconomic Dynamics*, 23, 2019, pp.1294-1301.

Romer, Ｐ．Ｍ. (1990), Endogenous Technological Change, *Journal of Political Economy*, Vol.98, No.5, Part.2, pp.71-102.

Shapiro, C., and Varian, H, A. (1998) *Information Rules: A Strategic Guide to the Network Economy*, Harvard Business Review Press.

Shinozaki, A (2003a), *Jouhou gijyutu kakushin no Keizai kouka (The Economic Effects of Information Technology Innovation)*, Nihonhyouronsya, Japan.

Shinozaki, A (2003b), Capital Investment in Telecommunication Industry and It's Impact on the Economy: Production Stimulus, Job Creation, and Network Externality, InfoCom Research, No31(8), pp.36-45.

Swallow, Y. C. and Haksar, V. (2019), The Economics and Implications of Data an Integrated Perspective, *IMF Departmental Paper Series*, No19/16.

Varian, H. R. (2018), Artificial Intelligence, Economics, and Industrial Organization, *NBER Working Paper*, 24839.

Veldkamp, L. and Chung, C. (2024), Data and the Aggregate Economy, *Journal of Economic Literature*, 62(2), pp.458–484.

Washio, T., and Shinozaki, A. (2021), Seisan kansu moderu niyoru de-ta katsuyou no keizaikouka bunseki (The Economic Effects Analysis of Data Utilization by Production Function Model), SLRC Discussion Paper Series, Vol.17, No.1, July.2021.