경제학 석사학위 논문

The Nonrivalry of Data and Economic Growth

– A Cross-Country Analysis –

데이터 비경합성과 경제성장

-실증 분석을 중심으로-

2023년 2월

서울대학교 대학원
경제학부 경제학 전공

정 영 주

The Nonrivalry of Data and Economic Growth

– A Cross-Country Analysis –

Advisor: Young Sik Kim

Submitting a master's thesis of Economics

December 2022

Graduate School of Economics
Seoul National University
Economics Major

Youngju Jeong

Confirming the master's thesis written by
Youngju Jeong
January 2023

Chair _____(Seal)
Vice Chair _____(Seal)
Examiner _____(Seal)

Abstract

We focus on the effect of data on the economic growth by testing implications from a model of Jones and Tonetti(2020). We focus not only on the quantity of data, but also the nonrivalry of data.

We use panel data on 35 countries over the period 2000−2020. Our empirical results show that data nonrivalry increases the economic growth rates in long term. Additionally, our paper gives implication about what proxy is proper to capture the nonrivalry of data. Our empirical results show that 'computer and software as a share of GDP' is appropriate. However, variables that reflect the extent of data openness is not reasonable as a proxy.

Consequently, the analysis offers new insights concerning the policy with regard to data ownership. The policy which derives the advantage of data nonrivalry is needed in today's data economy.

.

Keyword: Nonrivalry of Data, Data Economy, Economic Growth
Student Number: 2021−27125

Table of Contents

Chapter 1. Introduction

Data becomes more important nowadays. We so called it data economy. As the importance of data increases, the controversy over the ownership of data increases. First, is data ownership really meaningful? Second, who should have data ownership in terms of overall social welfare?

Observing the policy regarding data, i.e. EU's GDPR(General Data Protection Regulation), we can briefly answer above question. Nowadays, policymakers regard data is the thing that needs to be owned by some subjects. Data ownership to consumer becomes widely spread in overall countries.

We look into the policy enacted in the year of 2022, which is called "mydata" in Korea. Looking at the figure 1 below, We divide the group by whether the companies are in the group of mydata policy or not. The left of the figure 1 shows the ROA of companies in the group of mydata policy and the right of the figure 2 shows that of companies not in the group of mydata policy. Judging from the figure 1, we cannot precisely think that mydata policy is really effective for the growth of firm and the growth of country. Particularly since the start date of mydata policy is very recent, there is no enough data point to judge the effectiveness of policy.

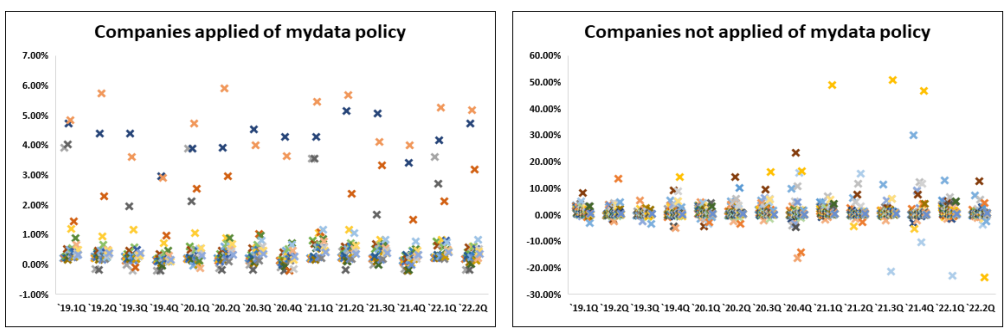


Figure 1. The ROA of companies

This is why we want to test some propositions with regard to the effect of data ownership to consumer to economic growth. As we will see in Jones and Tonetti(2020) in chapter 2, the nonrivalry of data is main feature when we consider the issue of data ownership. Therefore, our studies concentrates on the property of data, which is so called nonrivalriness. Since there exists enough data points with cross-country level data set, we testify some propositions with cross-country level data.

Chapter 2. Theoretical Basis

2.1. What is Data?

We classify data into 2 categories to well conceptualize our paper. First is the original data and second is the processed data. Original data refers to the data which is an input of production function, such as personal data. Processed data refers to the data which is an input and output at the same time. For example, if the consumer utilize the service of KAKAO, then the information about the pattern of consumers' using the service is processed, and this is what so called the processed data.

Jones and Tonetti(2020) focuses on the concept of processed data. However, in real world, it is difficult to accurately divide data into 2 categories, so we will not consider strictly between 2 concepts of data. Plus, From Jones and Tonetti(2020), "existing data can be used by any number of firms without being diminished", and this is so called the nonrivalry of data.

2.2. What is Data Economy?

We use the model of Jones and Tonetti(2020) to theoretically understand the data economy. In this part, we first look at the basic

environment introduced in Jones and Tonetti, and then look at the main result of Jones and Tonetti.

Jones and Tonetti(2020) suppose the economy consists of N varieties. Consumption of each variety produce an aggregate Y.

$$Y = \int_0^N Y_i^{\frac{\sigma-1}{\sigma}} di^{\frac{\sigma}{\sigma-1}} = N^{\frac{\sigma}{\sigma-1}} Y_i \quad (1)$$

Variety i is produced by combining an ides of quality A_i and labor L_i. Data D_i is used to increase the quality of idea A_i.

$$Y_i = A_i L_i \quad (2)$$

$$A_i = D_i^{\eta} \quad (3)$$

Putting these equations together,

$$Y_i = D_i^{\eta} L_i = D_i^{\eta} \frac{L}{N} = D_i^{\eta} v \quad (4)$$

where L is the total amount of labor, and v is the firm size measured by employment.

Data can be represented as below.

$$D_i = \alpha x Y_i + (1-\alpha) B = \alpha x Y_i + (1-\alpha) \tilde{x} N Y_i \quad (5)$$

$$Y_i = ([\alpha x + (1-\alpha)\tilde{x} N]^{\eta} v)^{\frac{1}{1-\eta}} \quad (6)$$

$$Y = N^{\frac{1}{\sigma-1}} ([\alpha x + (1-\alpha)\tilde{x} N]^{\eta} v)^{\frac{1}{1-\eta}} \quad (7)$$

where α measures the importance of specific firm's own data relative to the data bundle from other firms, x measures the fraction of specific firm's data which specific firm is allowed to use, and $\tilde{x}$ measures the fraction of other firm's data which specific firm is allowed to use.

It captures two features. First is the traditional expanding variety effect with the power of $\frac{1}{\sigma-1}$. Second is the nonrivalry of data which is raised to the term $\frac{\eta}{1-\eta}$.

The main result regarding economic growth is shown in equation (8) and (9).

$$Y_t^{alloc} = [v_{alloc}(1-\alpha)^{\eta}\tilde{x}_{alloc}^{\eta}]^{\frac{1}{1-\eta}}(\varphi_{alloc}L_t)^{1+\frac{1}{\sigma-1}+\frac{\eta}{1-\eta}} \quad for\ alloc \in \{sp, c, f\} \quad (8)$$

$$Y_t^{alloc} = [v_{OS}\alpha^{\eta}x_{os}^{\eta}]^{\frac{1}{1-\eta}}(\varphi_{os}L_t)^{1+\frac{1}{\sigma-1}} \quad for\ alloc \in \{sp, c, f\} \quad (9)$$

The implication of these equations is eventually same as we previously mentioned above, i.e. traditional expanding variety effect and the nonrivalry of data.

Jones and Tonetti classify the way of allocating data ownership into 3 categories. First category is ownership to consumer. Consumer can sell data to a data intermediary and choose how much data to sell. In this case, firms own zero data, and purchase data from the data intermediary instead. Second category is ownership to firm. Firms own data and decide whether to sell data and the quantity of data to sell. To strictly distinguish between first category and second category, Jones and Tonetti suppose that consumer cannot give exclusive right to use data to a specific firm. The final allocation is motivated by government policy, which limits the use of data.

Chapter 3. Data

In this paper, data covers 35 OECD countries for the period 2000–2020. Why we use country–level data is because data becomes important input in overall industries in each countries. From OECD(2020), "even in traditionally less data–intensive fields,

organizations are starting to leverage the large volumes of data." With cross-country data set, we can complement the analysis based on firm-level data in Korea what I briefly introduced in introduction. Why we use time-period 2000-2020 is because we consider period 2000-2020 is reasonable to analyze the concept of data economy. We consider the period before 2000 is not appropriate to analyze the effect of the property of data to economic growth.

The basic empirical framework is from Barro and Lee(1994). We add some variables to the model of Barro and Lee and exclude some variables from the model of Barro and Lee. That is, I relate growth rate of GDP to three kinds of variables: first, main variables, the stock of data which reflects the property of nonrivalriness; and second, initial levels of state variables, such as the stock of physical capital less the stock of data and the stock of human capital; and third, control or environmental variables, such as the ratio of government consumption to GDP, the life expectancy, and so on.

The basic information about key variables we consider is in Table 1.

Table 1
Key variables relevant for our study

Explanatory variable	Meaning	Source Max
log(GDP)	log of GDP	PWT.10.0, World Bank
HC	Human Capital Index	PWT.10.0
log(LIFE)	log of life expectancy	World Bank
DATA	Computer software and database / GDP (%)	OECD
ICT	Total ICT investment / GDP (%)	OECD
I/Y	Stock of other capital except the stock of data / GDP (%)	OECD, World Bank
G/Y	Government expenditure / GDP (%)	OECD, World Bank
Openness	data availability, data accessibility, government support of data usage	OECD Open, Useful and Re-usable data (OURdata) Index: 2019

3.1. The Stock of Data

Measuring the stock of data is a key challenge for our work. We choose 3 methods to measure this variables: first, the ratio of ICT investment to GDP; and second, the ratio of computer software and databases investment to GDP; and third, the extent of data accessibility, data availability, and government support of data use.

Before we concretely explain each methods, our key point is to comparing the ratio of ICT investment to GDP and the ratio of computer software and databases investment to GDP. The former captures the stock of data which does not reflect the nonrivalry of data, on the other hand the latter captures the stock of data which reflects the nonrivalry of data. Let us explain more clearly below.

First, we use the ICT investment as a % of GDP. This variable is needed to comparing with the proxy which reflects the nonrivalry of data.

Second, we use the computer software and databases as a % of GDP. There are 2 reasons why we regard this variable as an appropriate proxy. We have to clearly define the concept of the computer software and databases first. Following handbook of OECD(2010), computer software "consists of computer programs, program descriptions and supporting materials for both systems and applications software." Plus, database "consists of files of data organized in such a way as to permit resource effective access and use of the data."

The reason why we consider this variable as an appropriate proxy is following. As we see the equation (8), $[(1-\alpha)^{\eta}\tilde{x}_{alloc}^{\eta}]^{\frac{1}{1-\eta}}$ captures 2 properties. First, it is raised to the power η, because the data term is in production function; and second, it is also raised to the power $\frac{1}{1-\eta}$, because the data produced through production function

also affects economic growth which can be called as a feedback effect. Therefore, we use 2 terms which consist of databases and computer software. Since databases can capture the term with power η and computer software can capture the term with power $\frac{1}{1-\eta}$.

Third, we use the data accessibility, data availability, government support of data use as a proxy. We consider these as a proxy to capture the nonrivalry of data based on the meaning of data ownership in Jones and Tonetti(2020). In this paper, there are 2 systems of data ownership: first, data ownership to firms; and second, data ownership to consumers. The reason why social surplus of 2 systems is different is the nonrivalry of data. That is, when the data ownership is given to consumer, then all firms can utilize the data. On the other hand, when the data ownership is given to specific firm, then all firms cannot utilize the data, or even that specific firm can monopolize the right to utilize data. From this logic, the term of data openness can capture the nonrivalry of data. Therefore, we use the data accessibility, data availability, government support of data use to capture the nonrivalry of data.

Additionally, we once consider dummy variable to capture whether the policy of data ownership of each country at particular year is given to consumer or firm. However, we cannot appropriately classify each cases. Therefore, we do not include this dummy variable in our analysis, but we suggest that it is needed to be studied further.

3.2. Other Variables

The state variables are the ratio of other stock of capital(less of the stock of data) to GDP, denoted I/Y, and the human capital index, denoted HC. Instead of the data used in Barro and Lee, we use

human capital index to capture educational attainment. This is because we want to analyze the period 2000−2020, and human capital index from PWT.10.0 covers all these period. ①

The control variables are the countries dummies and time dummies. Instead of the black−market premium on foreign exchange used in Barro and Lee to capture market distortions, and the propensity to experience revolutions to capture country’ s specific situation, we use this dummies to control specific environment.

Other the control and environmental variables are the ratio of government consumption(less of spending on military and education) to GDP, denoted G/Y, the log of life expectancy, and the log of initial GDP.

Table 2
Summary statistics for panel data, 2000-2020

Explanatory variable	Mean	Min	Max
log(GDP)	12.9	9	16.9
HC	3.3	2.2	3.9
log(LIFE)	4.37	4.23	4.44
DATA	0	1.3	3.8
ICT	2.9	0.8	15.1
I/Y(except DATA)	22	11	36
I/Y(except ICT)	20	8	35
G/Y	14.7	5	27.7

3.3. Whether the Proxy is Appropriate

In this part, we want to check whether the proxy for the stock of

① Since pwt.10.0 covers period until 2020, I consider the human capital index of 2020 as same as that of 2019. Plus, for the country of BEL, HUN, data is missing even in period 2018,2019. So I consider the human capital index of 2018−2020 as same as that of 2017.

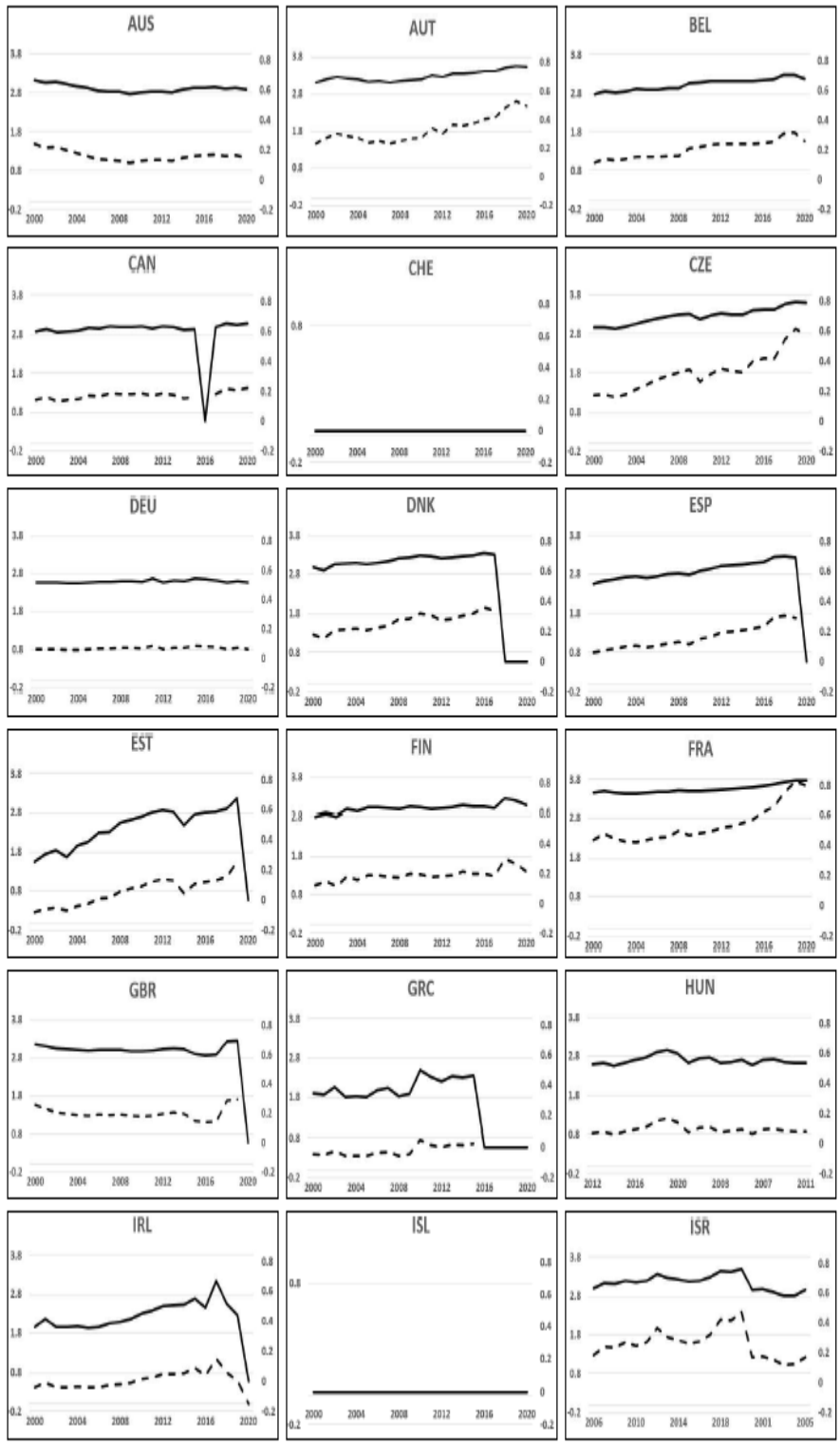
data is proper. I compare the ratio of computer software and databases investment to GDP with η which is represented in Jones and Tonetti(2020). η is numerically derived from the result of optimization in Jones and Tonetti(2020). The equation below is one of the result in paper.

$$\left(\frac{N_t(p_{at}D_{at}+p_{bt}D_{bt})}{Y_t}\right)^c = \frac{\eta}{1-\eta}\frac{\sigma-1}{\sigma} \tag{10}$$

$$\left(\frac{N_t(p_{at}Y_{it}+p_{bt}D_{bt})}{Y_t}\right)^f = \frac{\eta}{1-\eta\frac{\epsilon-1}{\epsilon}}\frac{\sigma-1}{\sigma} \tag{11}$$

In the case of the data ownership to firm, since Jones and Tonetti assume $\epsilon = 50$, the equation is almost same as in the case of data ownership to consumer. So I do not consider whether the data ownership is to firm or consumer in specific country in this part.

Looking at the figure 2 below, I compare the ratio of computer software and databases investment to GDP with η. With the logic we explained in previous part, we consider this proxy appropriate to capture the nonrivalry of data.


AUS
AUT
BEL
CAN
CHE
CZE
DEU
DNK
ESP
EST
FIN
FRA
GBR
GRC
HUN
IRL
ISL
ISR

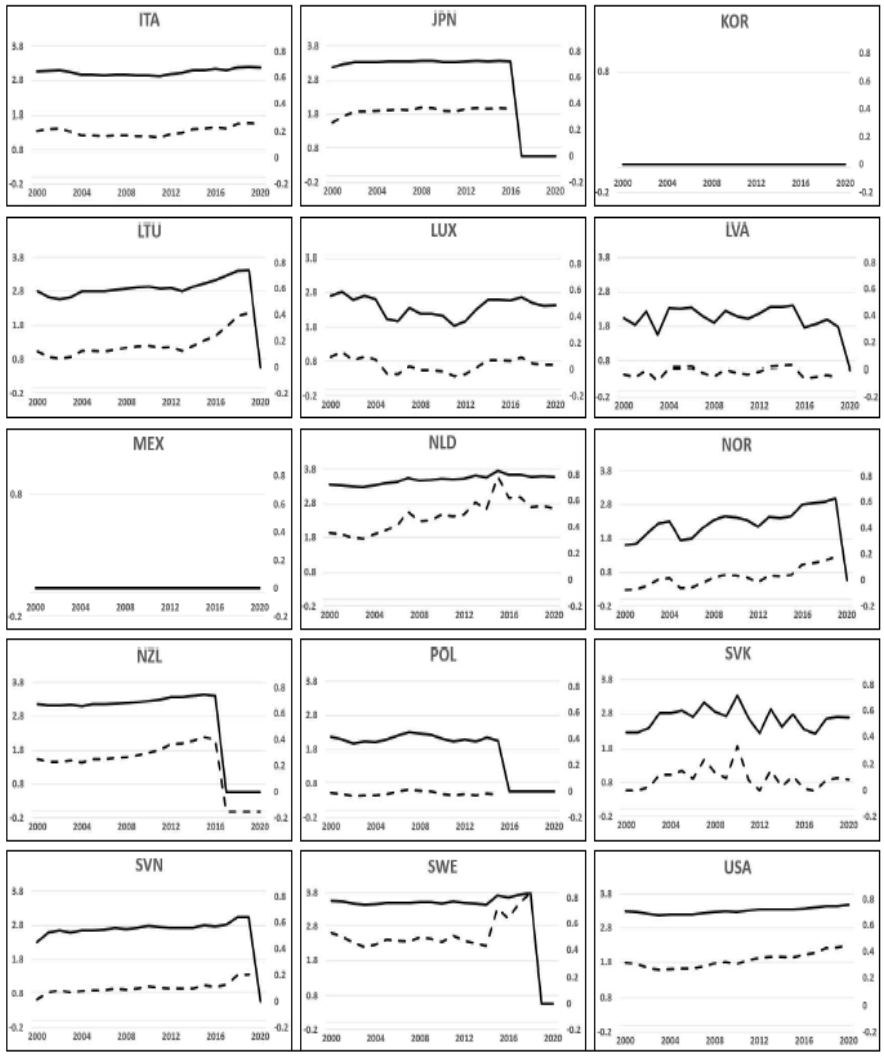


Figure 2. η (solid line) and 'computer software and database/GDP' (dotted line)②

Chapter 4. Empirical Analysis

We use the cross-country data set and relevant econometric techniques to analyze the basic implications stemming from Jones and Tonetti(2020). The basic implication from Jones and Tonetti model is "Giving data ownership to consumer is more socially

② We excluded the countries whose data of computer software and databases investment is not available.

efficient than giving data ownership to firm." To empirically test this implication, we translate the this implication as follow, "Not only the quantity of data, but the nonrivalry of data increases the economy growth."

We use two proxies to empirically test the above proposition: first, the ratio of ICT investment to GDP, denoted as ICT; and second, the ratio of computer software and databases investment to GDP, denoted as DATA.

The following models are used to conduct the empirical analysis:

$$g_{it} = \beta_0 \log(GDP)_{i,t-1} + \beta_1 \mathrm{HC}_{i,t-1} + \beta_2 \log(LIFE)_{i,t-1} + \beta_3 \mathrm{ICT}_{i,t-1} + \beta_4 \frac{I}{Y}_{i,t-1} + \beta_5 \frac{G}{Y}_{i,t-1} + u_{it} \quad (12)$$

$$g_{it} = \beta_0 \log(GDP)_{i,t-1} + \beta_1 \mathrm{HC}_{i,t-1} + \beta_2 \log(LIFE)_{i,t-1} + \beta_3 \mathrm{DATA}_{i,t-1} + \beta_4 \frac{I}{Y}_{i,t-1} + \beta_5 \frac{G}{Y}_{i,t-1} + u_{it} \quad (13)$$

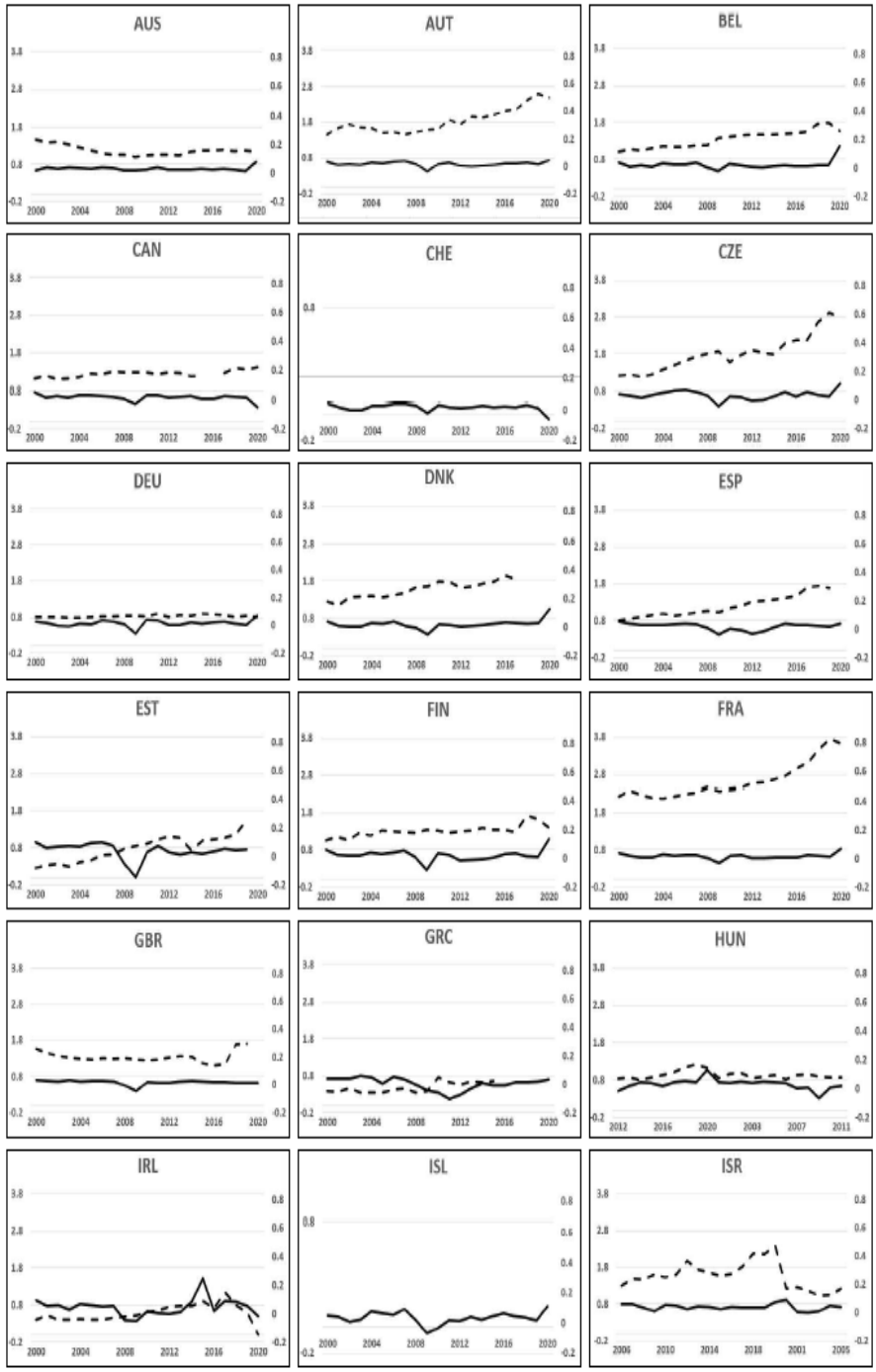
where g is the 1 year growth rate of GDP in time t, $\log(GDP)_{i,t-1}$ is the log of GDP in time t−1, $\mathrm{HC}_{i,t-1}$ is the human capital index in time t−1, $\log(LIFE)_{i,t-1}$ is the log of life expectancy in time t−1, $\mathrm{ICT}_{i,t-1}$ is 'ICT investment / GDP' in time t−1, $\mathrm{DATA}_{i,t-1}$ is 'computer and software / GDP' in time t−1, $\frac{I}{Y}_{it}$ is the ' gross domestic investment / GDP (less than the computer and software investment)' in time t−1 if the term $\mathrm{DATA}_{i,t-1}$ is used, and $\frac{I}{Y}_{i,t-1}$ is the ' gross domestic investment / GDP (less than ICT investment)' in time t−1 if the term ICT is used, $\frac{G}{Y}_{i,t-1}$ is ' government consumption / GDP (less than the spending on military and noncapital expenditures on education)'.

$$g_{it} = \beta_0 \log(GDP)_{i,t-5} + \beta_1 \mathrm{HC}_{i,t-5} + \beta_2 \log(LIFE)_{i,t-5} + \beta_3 \mathrm{DATA}_{i,t-5} + \beta_4 \frac{I}{Y}_{i,t-5} + \beta_5 \frac{G}{Y}_{i,t-5} + u_{it} \quad (14)$$

where g is the 5 years growth rate of GDP in time t, and the other term is same as above (14) except it captures the period t−5 not t−1.

4.1. The Nonrivalry of Data and Economic Growth

We briefly visualize the relationship between 'growth rate of GDP' and 'computer software and database/GDP' and looking at the Figure 3. In the case of CZE, FRA, NLD, SWE, USA, although the ratio of computer software and database investment to GDP is obviously increasing, corresponding growth rate of GDP does not show any change obviously. This occurs with 2 scenarios. First, the nonrivalry of data does not affect the economic growth. Second, the nonrivalry of data effect the economic growth in particular case, such as with sufficient time lags or with more longer term. To accurately analyze the causality of the nonrivalry of data to economic growth, we use several specification methods.


AUS
AUT
BEL
CAN
CHE
CZE
DEU
DNK
ESP
EST
FIN
FRA
GBR
GRC
HUN
IRL
ISL
ISR

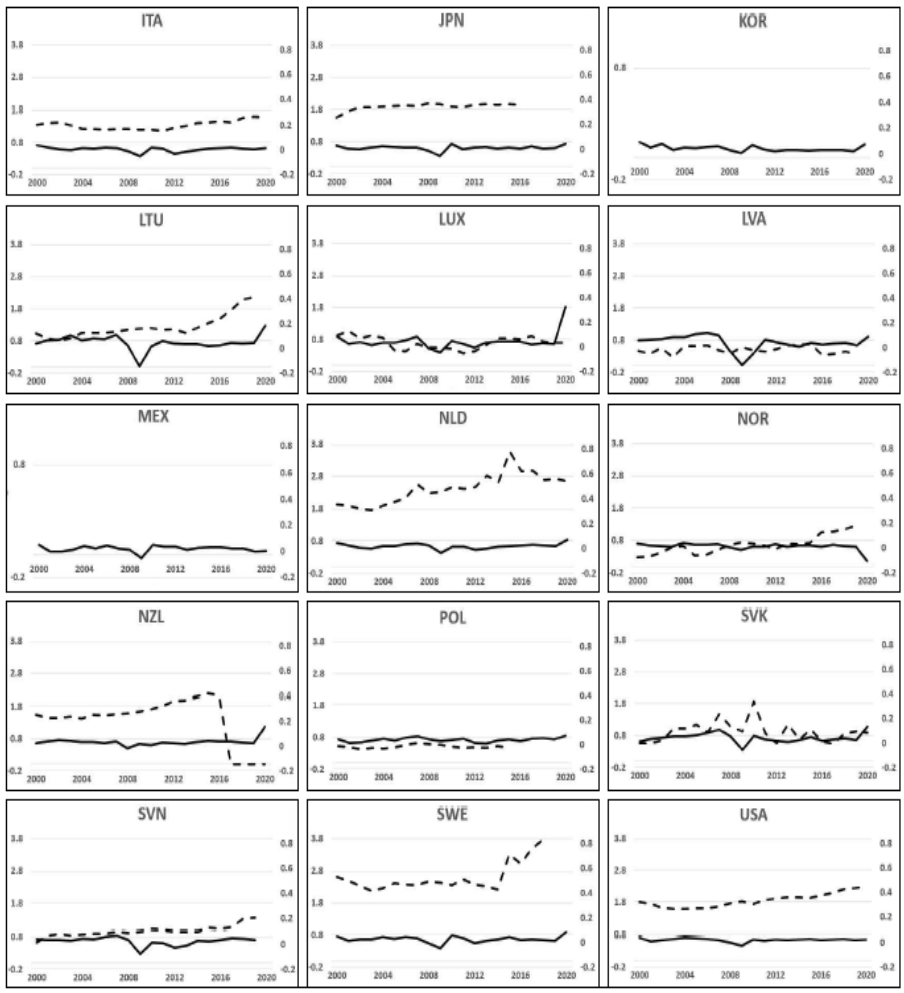


Figure 3. Growth rate of GDP(solid line) and ‘computer software and database/GDP’ (dotted line)③

4.2. Basic Regression on Short-Term Growth Rate

The regression (1) do not include time dummies and country dummies. The regression (2) only includes country dummies, and the regression (3) only includes time dummies. The regression (4) include both dummies. Looking at the all the result of regression,

③ We excluded the countries whose data of computer software and databases investment is not available.

Table 3-6, time dummies is not proper to control one's specific environment.

Looking at the Table 3 and Table 4, ICT is used as an explanatory variable as a proxy of the stock of data. For the Table 5 and Table 6, DATA is used as an explanatory variable. For the Table 3 and Table 5, 1 year growth rate of GDP is dependent variable, and for the Table 4 and Table 6, 5 year growth rate of GDP is dependent variable which follows Barro and Lee(1994).

The variable log(GDP) is an observation for initial state of GDP. For the regression (2) in Table 4 and (2) in Table 6, the estimated coefficient is -2.5636 and -2.28 respectively. As similar with the result in Barro and Lee(1994), this shows the tendency for conditional convergence. That is, the lower the starting GDP, the higher the growth rate of GDP in that year.

The variable HC is to looking for the effect of human capital on the growth of GDP. For the regression (2) in Table 4 and (2) in Table 6, it turns out to be negative.

The variable log(LIFE) is to look for the effect of basic environmental factor on the growth of GDP. It shows negative in regression (1) in Table 3, (1) and (2) in Table 5, which is opposite with the result in Barro and Lee(1994). This is because we analyze the period after 2000. That is, the improvement of productivity which is enhanced by investment, i.e. R&D, is more important than the basic factor to keep the quality of life. On the other hand, it shows positive in regression (2) in Table 4 and (2) in Table 6, whose dependent variable is the 5 years growth rate of GDP. We interpret this results as the factor that supports the basic quality of life influences the growth rate not instantaneously, but slowly.

The variable G/Y is the government expenditure to GDP. In all regressions in Table 3 and Table 5, the estimated coefficients

which is −0.001754, −0.003404, −0.001832, −0.003349 respectively are significant.

The variable I/Y is the stock of capital to GDP except the stock of data. In all regressions in Table 3 and Table 5, the estimated coefficients are strongly significant which is 0.002294, 0.002832, 0.002415, 0.003022 respectively. Comparing with the result of log(LIFE), investment is more important for increasing economic growth than the basic environment factor in the period 2000−2020. That is the input that influences the growth of GDP through production function is more important than the input which is basically needed to keep the quality of life.

The main focus of our paper is the coefficient of ICT and DATA. If the coefficient of DATA is positive, we can interpret in two ways. First, the quantity of data itself is meaningful to the growth of GDP as an original input of the firm's production function. Second, the nonrivalry of data is also an important property when utilizing data as an input. In the case of ICT, it turns out to be positive in the regression (2) in Table 3. In the case of DATA, it turns out to be positive which is weakly significant in the regression (1) and (2) in Table 5. We can interpret as a following ways. First, ICT investment increases the economic growth instantly, and DATA also influences the economic growth instantly although its terms is not strongly significant. That is, the nonrivalry of data is not significantly important for economic growth, rather the quantity of data itself have the source of economic growth. Second, comparing the size of coefficient, we even conclude that the nonrivalry of data is not strongly important as a main factor to determine the social welfare of overall economic system. For the regression (2), the coefficient of ICT in Table 3 is 0.017016, and the coefficient of DATA in Table 5 is 0.013025. Since we consider ICT representing the stock of data not reflecting the nonrivalry of data and DATA representing the stock of data reflecting the nonrivalry of data, it shows that the latter does not have enormous power for economic

growth. Third, comparing the size of coefficient of Table 5 and Table 6, the coefficient of Table 6 is larger than that of Table 5. It shows that the nonrivalry of data influences the economic growth with some time lags.

Table 3
Regressions for Growth Rate of GDP(1yr) with ICT investment

Explanatory variable	(1)	(2)	(3)	(4)
constant	0.574639**	0.733039	0.733039	0.574639**
log(GDP)	-0.001043	0.011319	0.011319	-0.001043
HC	0.005314	0.013924	0.013924	0.005314
log(LIFE)	-0.132392**	-0.276132	-0.276132	-0.132392**
ICT	0.001124	0.017016***	0.017016***	0.001124
I/Y	0.002294***	0.002832***	0.002832***	0.002294***
G/Y	-0.001754***	-0.003404***	-0.003404***	-0.001754***
Time dummy	No	No	Yes	Yes
Country dummy	No	Yes	Yes	No
R squared	0.132	0.291	0.291	0.132
Adjusted R	0.172	0.239	0.239	0.172

Note:
' denotes significance at 10% levels.
* denotes significance at 5% levels.
** denotes significance at 1% levels.
*** denotes significance at 0.1% levels.

Table 4
Regressions for Growth Rate of GDP(5yr) with ICT investment

Explanatory variable	(1)	(2)	(3)	(4)
constant	-1.42590	-111.6303***	-111.6303***	-1.42590
log(GDP)	-0.04503	-2.5636**	-2.5636**	-0.04503
HC	0.15255	-7.5021***	-7.5021***	0.15255
log(LIFE)	0.39871	38.6819***	38.6819***	0.39871
ICT	-0.01179	0.1218	0.1218	-0.01179
I/Y	0.00525	0.0635**	0.0635**	0.00525
G/Y	-0.00939	-0.0102	-0.0102	-0.00939
Time dummy	No	No	Yes	Yes
Country dummy	No	Yes	Yes	No
R squared	0.00916	0.421	0.421	0.00916
Adjusted R	-0.00768	0.358	0.358	-0.00768

Note:
' denotes significance at 10% levels.
* denotes significance at 5% levels.
** denotes significance at 1% levels.
*** denotes significance at 0.1% levels.

Table 5
Regressions for Growth Rate of GDP(1yr) with DATA proxy

Explanatory variable	(1)	(2)	(3)	(4)
constant	0.524412**	1.858314**	1.858314**	0.524412**
log(GDP)	-0.001912'	-0.000558	-0.000558	-0.001912'
HC	0.006773	0.017498	0.017498	0.006773
log(LIFE)	-0.121539**	-0.445148*	-0.445148*	-0.121539**
DATA	0.005153'	0.013025'	0.013025'	0.005153'
I/Y	0.002415***	0.003022***	0.003022***	0.002415***
G/Y	-0.001832***	-0.003349***	-0.003349***	-0.001832***
Time dummy	No	No	Yes	Yes
Country dummy	No	Yes	Yes	No
R squared	0.176	0.265	0.265	0.176
Adjusted R	0.166	0.21	0.21	0.166

Note:
' denotes significance at 10% levels.
* denotes significance at 5% levels.
** denotes significance at 1% levels.
*** denotes significance at 0.1% levels.

Table 6
Regressions for Growth Rate of GDP(5yr) with DATA proxy

Explanatory variable	(1)	(2)	(3)	(4)
constant	-0.95015	-74.6*	-74.6*	-0.95015
log(GDP)	-0.04793	-2.28*	-2.28*	-0.04793
HC	0.12982	-6.18***	-6.18***	0.12982
log(LIFE)	0.31970	28.4***	28.4***	0.31970
DATA	0.04816	0.0271	0.0271	0.04816
I/Y	0.00399	0.0567*	0.0567*	0.00399
G/Y	-0.01527	-0.0000637	-0.0000637	-0.01527
Time dummy	No	No	Yes	Yes
Country dummy	No	Yes	Yes	No
R squared	0.00875	0.407	0.407	0.00875
Adjusted R	-0.00943	0.341	0.341	-0.00943

Note:
' denotes significance at 10% levels.
* denotes significance at 5% levels.
** denotes significance at 1% levels.
*** denotes significance at 0.1% levels.

4.3. Basic Regression on Long-Term Growth Rate

For interpreting relationship between the nonrivalry of data and the economic growth as a causality not just correlation, we consider

new dependent variables which is represented below.

$$g_{it} = \beta_0 \log(GDP)_{i,t-5} + \beta_1 \mathrm{HC}_{i,t-5} + \beta_2 \log(LIFE)_{i,t-5} + \beta_3 \mathrm{DATA}_{i,t-5} + \beta_4 \frac{I}{Y_{i,t-5}} + \beta_5 \frac{G}{Y_{i,t-5}} + u_{it} \quad (15)$$

where g is the moving average of 5 years growth rate of GDP in time t, and the other term is same as above.

$$g_{it} = \beta_0 \log(GDP)_{i,t-5} + \beta_1 \mathrm{HC}_{i,t-5} + \beta_2 \log(LIFE)_{i,t-5} + \beta_3 \mathrm{DATA}_{i,t-5} + \beta_4 \frac{I}{Y_{i,t-5}} + \beta_5 \frac{G}{Y_{i,t-5}} + u_{it} \quad (16)$$

where g is the moving average of 11 years growth rate of GDP in time t, and the other term is same as above.

The regression (1) do not include time dummies and country dummies. The regression (2) only includes country dummies, and the regression (3) only includes time dummies. The regression (4) include both dummies. Looking at the all the result of regression, Table 7−10, time dummies is not proper to control one's specific environment.

Looking at the Table 7 and Table 8, ICT is used as an explanatory variable as a proxy of the stock of data. For the Table 9 and Table 10, DATA is used as an explanatory variable. For the Table 7 and 9, MA 5 years growth rate of GDP is dependent variable, and for the Table 8 and 10, MA 11 years growth rate of GDP is dependent variable.

Above all, comparing with the results of 4.2 which consider the short term of growth rate, the significance of each coefficients in part 4.3 is more stronger than that of each coefficients in part 4.2.

For the coefficient of HC, it turns out to be positive in regression (1) in Table 7, 8, 9 and 10, and negative in regression (2) in Table

8. Human capital positively influences the long term growth rate of GDP. It is apposite to the result shown in 2.4 that human capital negatively influences the short term growth rate of GDP.

For the variable log(GDP), log(LIFE), and G/Y, it shows consistent result with 2.4 and Barro and Lee(1994).

In all regressions in Table 7, Table 8, Table 9 and Table 10, the estimated coefficients of I/Y are strongly significant and turns out to be positive. Comparing the result of Table 7 and Table 8, the size of coefficient of I/Y with dependent variable of MA5 is bigger than that of MA11, which is 0.001536 in Table 7 and 0.000589 in Table 8 respectively.

The main focus of our paper is the coefficient of ICT and DATA, and it shows some interesting results. Looking at the result of the coefficient of ICT, it shows significant result only with the dependent variable of MA5, not with the dependent variable of MA11. Unlike the result of the coefficient of ICT, DATA significantly influences the economic growth in both terms of MA5 and MA11, although its significance is relatively weak. From this, ICT investment which contributes to the quantity of data itself positively increases the economic growth in long term. However depending just on ICT investment is not a good policy, since it does not influence the economic growth in more long term which means it does not contribute to the fundamental power of economic growth. On the other hand, computer software and databases investment, denoted as DATA, can be interpreted as the source of fundamental economic growth. Since the coefficient of DATA in Table 9 is 0.003769 in regression (1) and 0.009243 in regression (2), and the coefficient of DATA in Table 10 is 0.002873 in regression (1) and 0.004572 in regression (2). Of course, for some reasons, such as government's policy, the connectivity of DATA and economic growth is relatively weaker than that of ICT investment. Plus, comparing the coefficient in Table 9 and Table 10, the effect of

DATA is more powerful in MA5 than in MA11 which is same result with the ICT.

Table 7
Regressions for Growth Rate of GDP(MA5) with ICT investment

Explanatory variable	(1)	(2)	(3)	(4)
constant	0.625837***	0.306681*	0.306681*	0.625837***
log(GDP)	−0.001546**	−0.023737*	−0.023737*	−0.001546**
HC	0.005647*	0.013876	0.013876	0.005647*
log(LIFE)	−0.140731***	−0.150530	−0.150530	−0.140731***
ICT	0.000687'	0.009252***	0.009252***	0.000687'
I/Y	0.001536***	0.002087***	0.002087***	0.001536***
G/Y	−0.001297***	−0.002153***	−0.002153***	−0.001297***
Time dummy	No	No	Yes	Yes
Country dummy	No	Yes	Yes	No
R squared	0.326	0.524	0.524	0.326
Adjusted R	0.318	0.489	0.489	0.318

Note:
' denotes significance at 10% levels.
* denotes significance at 5% levels.
** denotes significance at 1% levels.
*** denotes significance at 0.1% levels.

Table 8
Regressions for Growth Rate of GDP(MA11) with ICT investment

Explanatory variable	(1)	(2)	(3)	(4)
constant	0.867059***	1.75***	1.75***	0.867059***
log(GDP)	−0.001512***	−0.0275***	−0.0275***	−0.001512***
HC	0.003871*	−0.0164*	−0.0164*	0.003871*
log(LIFE)	−0.191877***	−0.300***	−0.300***	−0.191877***
ICT	0.000386	0.00101	0.00101	0.000386
I/Y	0.000589***	0.000284	0.000284	0.000589***
G/Y	−0.000825***	−0.300***	−0.300***	−0.000825***
Time dummy	No	No	Yes	Yes
Country dummy	No	Yes	Yes	No
R squared	0.459	0.829	0.829	0.459
Adjusted R	0.452	0.813	0.813	0.452

Note:
' denotes significance at 10% levels.
* denotes significance at 5% levels.
** denotes significance at 1% levels.
*** denotes significance at 0.1% levels.

Table 9
Regressions for Growth Rate of GDP(MA5) with DATA proxy

Explanatory variable	(1)	(2)	(3)	(4)
constant	0.609703***	1.609670***	1.609670***	0.609703***
log(GDP)	−0.002220***	−0.034181**	−0.034181**	−0.002220***
HC	0.007296**	0.021210	0.021210	0.007296**
log(LIFE)	−0.137836***	−0.274552**	−0.274552**	−0.137836***
DATA	0.003769*	0.009243*	0.009243*	0.003769*
I/Y	0.001587***	0.002232***	0.002232***	0.001587***
G/Y	−0.001309***	−0.002138***	−0.002138***	−0.001309***
Time dummy	No	No	Yes	Yes
Country dummy	No	Yes	Yes	No
R squared	0.326	0.517	0.517	0.326
Adjusted R	0.317	0.481	0.481	0.317

Note:
' denotes significance at 10% levels.
* denotes significance at 5% levels.
** denotes significance at 1% levels.
*** denotes significance at 0.1% levels.

Table 10
Regressions for Growth Rate of GDP(MA11) with DATA proxy

Explanatory variable	(1)	(2)	(3)	(4)
constant	0.813132***	1.963842***	1.963842***	0.813132***
log(GDP)	−0.002281***	−0.029211***	−0.029211***	−0.002281***
HC	0.006276***	−0.013469'	−0.013469'	0.006276***
log(LIFE)	−0.179936***	−0.345726***	−0.345726***	−0.179936***
DATA	0.002873*	0.004572*	0.004572*	0.002873*
I/Y	0.000614***	0.000356'	0.000356'	0.000614***
G/Y	−0.000833***	−0.001089***	−0.001089***	−0.000833***
Time dummy	No	No	Yes	Yes
Country dummy	No	Yes	Yes	No
R squared	0.438	0.83	0.83	0.438
Adjusted R	0.43	0.815	0.815	0.43

Note:
' denotes significance at 10% levels.
* denotes significance at 5% levels.
** denotes significance at 1% levels.
*** denotes significance at 0.1% levels.

Summarizing the result of 4.1 and 4.2, we can derive some implications. The nonrivalry of data do not effectively increase the economic growth in short term comparing with the quantity of data itself. However, the nonrivalry of data needs enough time to

effectively influence the economic growth. Therefore, at the aspects of economic growth in long term, the nonrivalry of data is the main source of fundamental economic growth.

4.4. Further Study

We use another empirical specification method to analyze following: first, to analyze the implication of Jones and Tonetti(2020) we mentioned above, and second, to find the proper proxy to measure the nonrivalry of data.

$$\Delta g_i = \beta_0 + \beta_1 \Delta \log(GDP)_i + \beta_2 \Delta \mathrm{HC}_i + \beta_3 \Delta \log(LIFE)_i + \beta_4 \Delta \mathrm{DATA}_i + \beta_5 \Delta \frac{I}{Y_i} + \beta_6 \Delta \frac{G}{Y_i} + \beta_7 \Delta \mathrm{Openness}_i + u_i \quad (17)$$

where, Δ indicates the first difference between the year of 2017 and 2019④, $\mathrm{Openness}_i$ is the extent of openness of data which can be captured by 3 variables consisting of data accessibility, data availability, and the government support of data usage.

The regression (1) do not include openness variable, and other regression include openness variable. The regression (2) use data availability, (3) use data accessibility, and (4) use the government support of data use as a proxy for openness.

④ The period is chosen by availability of data of openness.

Table 11
First Difference Regressions for Growth Rate of GDP with Openness

Explanatory variable	(1)	(2)	(3)	(4)
constant	0.0258	−0.0174	−0.0184	−0.0090
log(GDP)	−0.0029*	0.0003	0.0004	−0.0005
HC	0.0113'	0.1267	0.1112	0.1263
log(LIFE)	−0.0077	0.0036	0.0033	0.0033
DATA	0.0050	0.0386	0.0002	0.0014
I/Y	0.0020***	0.0003	0.0002	0.0002
G/Y	−0.0016***	0.0009	0.0008	0.0007
Openness	·	0.0386	0.0447	0.0231
Openness proxy	NO	data availability	data accessibility	government support of data use
R squared	0.1882	0.2373	0.2298	0.2199
Adjusted R	0.1634	−0.0296	−0.03977	−.0532

Note:
' denotes significance at 10% levels.
* denotes significance at 5% levels.
** denotes significance at 1% levels.
*** denotes significance at 0.1% levels.

To firstly say the result of empirical analysis, there not exists any significant coefficients when we add the variable of $\Delta Openness_i$. Although the result is not significant, we can derive some implications from these result. This is because the result when we do not add the variable $\Delta Openness_i$ shows significant result. That is, the proxy we use to capture openness of data, further the nonrivalry of data is not so appropriate.

Chapter 5. Conclusion

Most literature on endogenous growth consider capital, such as physical capital and human capital, as an important factor for economic growth. On the other hand, Jones and Tonetti(2020) considers data as an important factor for economic growth. In this paper, we first focus on the effect of policy with regard to data ownership at a firm level. Since the data is scant to precisely analyze the effect of the policy, we identified some implications on Jones and Tonetti(2020) with a cross country level data.

Our empirical results give implication in terms of two aspects. First implication is whether data affects economic growth. We investigate that the quantity of data itself increases the economic growth. Plus, the nonrivalry of data increases the economic growth in long term. That is, not only the quantity of data is important, but also the nonrivalry of data is important in economic growth. Therefore data affects economic growth not only through production function, but also through feedback effects. Second implication is what proxy is appropriate to measure the nonrivalry of data. Our empirical analysis shows that 'computer software and database / GDP' is appropriate, but 'openness of data' is not appropriate as a proxy for the nonrivalry of data. Two implications are the key differences with previous literature regarding data economy.

This study offers important policy implications in connection with the data ownership over time. Indeed, the worldwide trend toward giving data ownership to consumer will increase the economic growth in these data economy. This is because consumer distributes more chances to more firms a right to using data, and eventually, the nonrivalry of data becomes more powerful. To enhance economic growth in long term, data ownership should be given to consumer.

A potential deficiency of our empirical analysis is the method of estimation we used. Even Barro and Lee(1994) which we benchmarked is criticized by many literatures. To solve this problem, we tried to add instrument variables such as an earlier value of log(GDP), a lagged value of I/Y, and a lagged value of G/Y. However, it faces multicollinearity problem, so we cannot use these as an instrument variables. In addition to this issue, there exists data issues to capture the nonrivalry of data. Therefore more studies with refined econometric methods and proper proxy for the nonrivalry of data is needed further.

References

[1] Bajgar, M., C. Criscuolo and J. Timmis. (2021), "Intangibles and industry concentration: Supersize me", *OECD Science, Technology and Industry Working Papers*, No. 2021/12, OECD Publishing, Paris.

[2] Barro, Robert J. and Lee, Jong-Wha, (1994), "Sources of economic growth", *Carnegie-Rochester Conference Series on Public Policy, Elsevier*, vol. 40(1), pages 1-46, June.

[3] Chirwa, Themba G. and Themba G and Odhiambo, Nicholas M. (2018), "Exogenous and Endogenous Growth Models: a Critical Review", *Comparative Economic Research. Central and Eastern Europe*, ISSN 2082-6737, De Gruyter, Warsaw, Vol.21, Iss.4, pp. 63-84.

[4] Colecchia, A. and Schreyer, P.,(2002), "ICT Investment and Economic Growth in the 1990s: is the Unites States a unique case? : a comparative study of nine OECD countries", *Rev.Econ.Dynam.5 (2)*, 408-422.

[5] Howitt, Peter. (2000), "Endogenous Growth and Cross-Country Income Differences", *American Economic Review*, 90 (4): 829-846.

[6] J. B. Taylor and Harald Uhlig (ed.), (1999), "Handbook of Macroeconomics", *Handbook of Macroeconomics, Elsevier,* volume 1A, p. 670-737.

[7] Jones, Charles I., and Christopher Tonetti. (2020), "Nonrivalry and the Economics of Data", *American Economic Review*, 110 (9): 2819-58.

[8] Ker, D. and E. Mazzini (2020), "Perspectives on the value of data and data flows", *OECD Digital Economy Papers*, No. 299, OECD Publishing, Paris.

[9] Mitchell, J., D. Ker and M. Lesher (2021), "Measuring the economic value of data", *Going Digital Toolkit Note*, No. 20.

[10] Nestor Duch-Brown, Bertin Martens and Frank Mueller-Langer, (2017), "The economics of ownership, access and trade in digital data; Digital Economy Working Paper 2016-10", *JRC Technical Reports.*

[11] OECD (2019), "A measurement roadmap for the future", in Measuring the Digital Transformation: A Roadmap for the Future, OECD Publishing, Paris.

[12] OECD (2020), "Mapping Approaches to Data and Data Flows", OECD Publishing,

[13] OECD (2009), "Handbook on Deriving Capital Measures of Intellectual Property Products", OECD Publishing, Paris.

[14] OECD(2020), "OECD Policy Paper on Public Governance No.1".

[15] Veldkamp, L. and Chung C. (2019), "Data and the aggregate economy", *J.Econ.Liter.*, forthcoming, in press.

[16] Yuan, Bocong, and Jiannan Li. (2019), "The Policy Effect of the General Data Protection Regulation (GDPR) on the Digital Public Health Sector in the European Union: An Empirical Investigation", *International Journal of Environmental Research and Public Health* 16, no. 6: 1070.

초록

이 논문에서는 Jones and Tonetti(2020)의 명제를 검증하는 실증 분석을 통해 데이터가 경제성장에 미치는 영향을 제시한다. 본 논문은 데이터의 양 자체뿐 아니라 데이터의 비경합적 성질에 주목하고 있다.

2000-2020년 기간 동안 OECD 35개국의 패널데이터가 주요 분석 대상이다. 실증 분석 결과 데이터 비경합성은 장기적으로 경제성장률을 증가시킨다. 또한 본 논문은 데이터 비경합성을 측정할 수 있는 대리변수가 무엇인지에 대한 함의도 제시한다. “GDP 대비 컴퓨터 소프트웨어와 데이터베이스”는 대리변수로 적합하다고 판단되나, “데이터의 개방성”은 대리변수로 적합하지 않다고 판단된다.

결론적으로, 본 논문의 분석은 데이터 소유권과 관련된 정책에 시사점을 준다. 현 데이터 경제에서는 데이터 비경합성의 이점을 극대화 할 수 있는 방향의 정책이 요구된다.

키워드: 데이터 비경합성, 데이터 경제, 경제성장

학 번: 2021-27125