RIETI Discussion Paper Series 20-E-015

Data Sharing and Revenue Distribution Rule

OGURO, Kazumasa
RIETI

ISHIDA, Ryo
Policy Research Institute Ministry of Finance Japan

YASUOKA, Masaya
Kwansei Gakuin University

Research Institute of Economy, Trade & Industry, IAA
The Research Institute of Economy, Trade and Industry
https://www.rieti.go.jp/en/

RIETI Discussion Paper Series 20-E-015
March 2020

Data Sharing and Revenue Distribution Rule[1]

OGURO, Kazumasa
Hosei University / Research Institute of Economy, Trade and Industry

ISHIDA, Ryo
Policy Research Institute Ministry of Finance Japan

YASUOKA, Masaya
Kwansei Gakuin University

Abstract

The main purpose of this paper is to set a model in which there exist multiple firms producing data in a situation where each firm produces data and shares it voluntarily for new additional revenue. The model is used for theoretical examination of the revenue distribution rule and behaviors to maximize the social welfare. Consequently, the following three main results can be obtained. First, if the number of firms is sufficiently large and some conditions are assumed, the revenue distribution rule to maximize social welfare in a decentralized economy coincides with the elasticity of additional revenue with respect to the provided data. Second, if each firm maximizes profit in the decentralized economy, the firm can achieve allocations to maximize social welfare in a command optimum for any revenue distribution rule as long as the government provides the policy of lump-sum tax and subsidy appropriately. Third, if the subsidy for data sharing is financed by a flat rate tax for additional profit, each firm has an incentive to participate in the platform irrespective of the subsidy rate and revenue distribution rule.

Keywords: Data Sharing, Platform, Revenue Distribution Rule, Externality, Social Welfare
JEL classification: D21, D22, D83, H21, H23, L22, L50

[1] This study was supported by the Research Institute of Economy, Trade and Industry (RIETI). Any remaining errors are the authors' responsibility. The authors are grateful for helpful comments and suggestions by Discussion Paper seminar participants at RIETI.

1. Introduction

The main purpose of this paper is to set a model with multiple firms that produce data, where each firm produces data and shares it voluntarily for new additional revenue. The study also conducts theoretical examination of the revenue distribution rule and behaviors to maximize social welfare. As shown generally, as the Industrial Revolution progresses worldwide, new economies using combinations of data produced by firms and individuals are producing new value. This phenomenon is designated as data capitalism. Typical cases are "GAFA" in the U.S.A. comprises Google Inc., and Apple Corp. which is famous for the "iPhone" smartphone. Facebook is a major social network service. Amazon is a major marketing and distribution company. All exist today in an information market that includes big data. In addition, "BATH" in China comprises Baidu (百度), Alibaba (阿里巴巴集团), Tencent (騰訊), and Huawei (華為技術). All are gaining momentum in information markets.

As reported by the OECD (2015), data existing in the world in 2015 reached 8 trillion GB, but the amount is forecast to become more than forty times that amount by 2020. HM Treasury (2018) points out that data are "Non-rivalrous" and a "Positive externality," and that data have "Economies of scopes." Because data are non-rivalrous, a certain quantity of data can be used simultaneously in calculation and analysis. Combinations of data produce "Positive externalities" such as new discoveries, insights, and other phenomena. In addition, difficulties presented by asymmetric information exist, as pointed out by Akerlof (1970). By "Economies of scope," the efficiency of revenue and cost increases when we have and use data to the greatest extent possible rather than having each firm and individual possess and use it individually.

The following world can be created as the data industrial revolution progresses. First, "Brains" of artificial intelligence (AI) exist at the top of an internet of things (IoT), setting devices of high technology as an underlayer. This structure posits human beings as nodes, or "nerve cells." Naturally, these nerve cells contain the existing information network that spreads over in internet. Information of many kinds is produced there. Then, this information (big data) is pooled at specific places.

However, although Big data are meaningless without a brain, a third boom of artificial intelligence has started all over the world. Study of Deep learning has developed rapidly. The opening of Deep learning has started by the suggestion of an Auto Encoder by Hinton and Salakhutdinov (2006). Because of an increase in data that are distributed, by setting and controlling the purposes that humans are heading for, a stream of artificial intelligence spreads all over the world, analyzing and finding values of the data using Deep learning. Considering this meaning, Big data are "Food" that is necessary to advance artificial intelligence and which is economically regarded as an asset. Numerous values can be produced with combinations of many kinds of data.

As reported by the Ministry of Internal Affairs and Communications (2018), the "information bank" concept has started in Japan with attention to protection of individual information. The image of a bank is that a bank stewards our money safely. The money is used for lending to firms. Then, one can obtain interest from the deposit of money. For the information bank, the basic system is the same as that of an ordinary bank: the information bank collects numerous data all over the world and lends it to the firms which need the data. Even if a kind of data is a low value, the economic value can be extremely high because of the combination of many kinds of data.

The system produces profit as a result of using information. The profits are then distributed for the firms and individuals providing the data. However, even if it is clear that data sharing by firms and individuals have produced additional revenue, the economic criteria are not clear as to how the additional revenue produced by data sharing might be distributed.

The additional revenue distribution rule is important because of “positive externalities” of data. “Positive externalities” arise in discussions of network externalities (Jackson, 2008; Ioannides, 2012) The discussion starts at pioneer studies and progresses theoretically. One can find the same situation in our paper: if firms that participate in the data sharing scheme increase, the additional revenue that each firm can gain increases. However, for instance, in a case without an additional revenue distribution rule, no firm considers positive externalities for other firms. Consequently, the supply of data sharing can be less than the socially optimal level.

One means of resolving this difficulty is to set an appropriate additional revenue distribution rule. This problem has the same mechanism of the discussion by which a voluntary supply of public goods is generally less than the optimal level (Yamashige, 2006; Slavov, 2014). Villanacci and Zenginobuz (2006) examine how a government should intervene for the voluntary supply of public goods. For instance, Boadway, Pestieau and Wildasin (1989) set a model of voluntary supply of public goods and demonstrate the possibility that the optimal level can be derived through subsidy for the supply in cases of a non-cooperative supply of public goods.

In the case of data sharing, one can consider the subsidy policy for data sharing to internalize externalities. Nevertheless, no report of the relevant literature describes such an analysis.

Now, this paper sets a simple model with multiple firms that produce data. Using the model, one can analyze the revenue distribution rule and behaviors theoretically to maximize social welfare. Concretely, the structure of this paper is described as follows. Section 2 derives the revenue distribution rule to maximize social welfare in a situation of data sharing by firms in a decentralized economy. Moreover, we examine how the revenue distribution rule changes in command optimum and the policy to obtain social welfare in the case of a command optimum. Then, we derive some propositions. Section 3 expands the model setting given by section 2 and examines the case of $N > 2$. Concretely, based on some assumptions, we derive the proposition that the revenue distribution rule to maximize social welfare in a decentralized economy is expected to be equivalent to the elasticity of additional revenue with respect to provided data. Finally, section 4 concludes our paper and describes avenues for future work.

2. Model Setting and Brief Analysis

First, a model is set with the profit and cost of a firm that produces data. This model economy has N firms, of which firm j (j=1,2,3,⋯,N) produces new data D_j. The revenue produced using the data (D_j) and the cost to produce the data (D_j) are defined respectively as $\theta_j D_j$ and $\frac{c_j}{2} D_j^2$. Also, d_j is defined as the share data that firm j has. We assume $d_j \leq D_j$. The cost to share the data is defined as $\frac{m_j}{2} d_j^2$.

Second, we set the model for additional revenue produced by the combination of share data. d_{i_k} is defined as the data that are mutually shared, but which firm i_k (k=1,2,3,⋯,m) possesses. Then, the additional revenue produced by the combination of data sharing can be shown as $\gamma \prod_{k=1}^{m} d_{i_k}{}^{\rho_{i_k}}$. Each $\rho_{i_k} \geq 0$ holds the condition of $\sum_{k=1}^{m} \rho_{i_k} = 1$ and is changed by the combination of shared data.

The additional revenue produced by the combination of share data ($\gamma \prod_{k=1}^{m} d_{i_k}{}^{\rho_{i_k}}$) is distributed to each firm i_k by the distribution rate ζ_{i_k}; $\sum_{k=1}^{m} \zeta_{i_k} = 1$ is assumed.

<Decision Making in a Decentralized Economy ($N = 2$)>

If the number of firms $N = 2$, then the profit of firm j π_j ($j = 1, 2$) can be expressed as the following equations according to the revenue distribution rule of data sharing (ζ_1, ζ_2) because of the model setting explained above.

$$\pi_1 = \theta_1 D_1 + \zeta_1 \gamma d_1{}^{\rho_1} d_2{}^{\rho_2} - \frac{c_1}{2} D_1{}^2 - \frac{m_1}{2} d_1{}^2 \quad (1)$$
$$\pi_2 = \theta_2 D_2 + \zeta_2 \gamma d_1{}^{\rho_1} d_2{}^{\rho_2} - \frac{c_2}{2} D_2{}^2 - \frac{m_2}{2} d_2{}^2$$

Each firm decides to produce of data D_j. Then, the profit maximization condition can be shown as presented below.

$$D_1 = \theta_1 / c_1 \quad (2)$$
$$D_2 = \theta_2 / c_2$$

By substituting (2) into (1), the profit of each firm can be shown as follows.

$$\pi_1 = \frac{\theta_1{}^2}{2c_1} + \zeta_1 \gamma d_1{}^{\rho_1} d_2{}^{\rho_2} - \frac{m_1}{2} d_1{}^2 \quad (3)$$

$$\pi_2 = \frac{\theta_2{}^2}{2c_2} + \zeta_2 \gamma d_1{}^{\rho_1} d_2{}^{\rho_2} - \frac{m_2}{2} d_2{}^2$$

Under the revenue distribution rule of data sharing (ζ_1, ζ_2), each firm decides the data sharing d_j （$j = 1, 2$） at the constraint $d_j \leq D_j$ and $\rho_1 + \rho_2 = 1$. The profit maximization condition in the decentralized economy is shown as

$$\zeta_1 \gamma \rho_1 \left(\frac{d_2}{d_1}\right)^{\rho_2} = m_1 d_1 \quad \text{and} \quad \zeta_2 \gamma \rho_2 \left(\frac{d_1}{d_2}\right)^{\rho_1} = m_2 d_2 \; .$$

These equations can derive the following equations.

$$d_1 = \overline{d_1} \equiv \sqrt{\frac{\zeta_1 \rho_1}{m_1} \tau {\zeta_1}^{\rho_1} {\zeta_2}^{\rho_2}} \quad \text{and} \quad d_2 = \overline{d_2} \equiv \sqrt{\frac{\zeta_2 \rho_2}{m_2} \tau {\zeta_1}^{\rho_1} {\zeta_2}^{\rho_2}} \tag{4}$$

$$\pi_1 = \frac{{\theta_1}^2}{2c_1} + \frac{1}{2}(2\zeta_1 - \zeta_1 \rho_1)\tau {\zeta_1}^{\rho_1} {\zeta_2}^{\rho_2} \tag{5}$$

$$\pi_2 = \frac{{\theta_2}^2}{2c_2} + \frac{1}{2}(2\zeta_2 - \zeta_2 \rho_2)\tau {\zeta_1}^{\rho_1} {\zeta_2}^{\rho_2}$$

In those equations, $\sqrt{\tau} = \gamma \sqrt{\frac{{\rho_1}^{\rho_1} {\rho_2}^{\rho_2}}{{\mathrm{m}_1}^{\rho_1} {\mathrm{m}_2}^{\rho_2}}}$.

This model has no household sector. Therefore, we derive the revenue distribution rule (ζ_1, ζ_2) to maximize total profit shown by (5) $(\Pi = \pi_1 + \pi_2)$ as the condition to maximize social welfare. First, considering $\zeta_1 = \zeta$ and $\zeta_2 = 1 - \zeta$, the total profit shown by (5) can be expressed as follows.

$$\Pi = \pi_1 + \pi_2$$

$$= \frac{{\theta_1}^2}{2c_1} + \frac{{\theta_2}^2}{2c_2} + \frac{1}{2}(2 - (\zeta \rho_1 + (1-\zeta)\rho_2))\tau \zeta^{\rho_1} (1-\zeta)^{\rho_2} \tag{6}$$

Considering $\rho_1 = \rho$ and $\rho_2 = 1 - \rho$, one can obtain the following total profit maximizing condition and Proposition 1.

$$\frac{\rho}{\zeta} - \frac{1-\rho}{1-\zeta} - \frac{2\rho - 1}{2 - (\zeta\rho + (1-\zeta)(1-\rho))} = 0$$

Proposition 1 If the number of firms is $N = 2$ and each firm considers optimization of profit in a decentralized economy, then the revenue distribution rule (ζ_1, ζ_2) can be given as $\zeta_1 = \zeta$ and $\zeta_2 = 1 - \zeta$, where ζ is shown as presented below.

$$\zeta = \frac{\left(1 + \rho + \frac{1+\rho}{2\rho - 1}\right) - \sqrt{\left(1 + \rho + \frac{1+\rho}{2\rho - 1}\right)^2 - 8\rho \frac{1+\rho}{2\rho - 1}}}{4}$$

<Decision Making at the Command Optimum ($N = 2$)>

In this section, we derive the condition to maximize social welfare in the case of decision-making at the command optimum. Therefore social welfare W coincides with the total profit of each firm. One can obtain the following equation.

$$W = \pi_1 + \pi_2$$

$$= \frac{{\theta_1}^2}{2c_1} + \frac{{\theta_2}^2}{2c_2} + \gamma {d_1}^{\rho_1} {d_2}^{\rho_2} - \frac{m_1}{2} {\mathrm{d}_1}^2 - \frac{m_2}{2} {\mathrm{d}_2}^2$$

The condition to maximize social welfare W can be shown as presented below.

$$\gamma \rho_1 \left(\frac{d_2}{d_1}\right)^{\rho_2} = m_1 d_1 \quad \text{and} \quad \gamma \rho_2 \left(\frac{d_1}{d_2}\right)^{\rho_1} = m_2 d_2$$

Then, from these equations, one can obtain the following.

$$d_1^* = \sqrt{\frac{\rho_1}{m_1}\tau} \quad \text{and} \quad d_2^* = \sqrt{\frac{\rho_2}{m_2}\tau} \tag{7}$$

$$W = \frac{\theta_1{}^2}{2c_1} + \frac{\theta_2{}^2}{2c_2} + \frac{1}{2}\tau \qquad \left(\sqrt{\tau} = \gamma\sqrt{\frac{\rho_1{}^{\rho_1}\rho_2{}^{\rho_2}}{m_1{}^{\rho_1}m_2{}^{\rho_2}}} \right) \tag{8}$$

Comparing (7) with (4) shows that, because of $\overline{d_j} \le d_j^* (j = 1, 2)$, the data sharing in the case of decentralized economy is less than in the case of the command optimum. In addition, comparing (8) with (6) because of $\Pi < W$, (6) in the case of a decentralized economy is less than social welfare (8) in the case of the command optimum.

<Lump-sum Taxation・Subsidy Policy>

We consider subsidy β_j for a constant rate for data sharing of firm *j* (*j*=1,2), which is financed by lump-sum taxation Φ for each firm to solve the difficulty described above. Then, (3) can be changed as shown below.

$$\pi_1 = \frac{\theta_1{}^2}{2c_1} + \zeta_1\gamma d_1{}^{\rho_1} d_2{}^{\rho_2} - \frac{m_1}{2}(1-\beta_1)d_1{}^2 - \Phi \tag{9}$$

$$\pi_2 = \frac{\theta_2{}^2}{2c_2} + \zeta_2\gamma d_1{}^{\rho_1} d_2{}^{\rho_2} - \frac{m_2}{2}(1-\beta_2)d_2{}^2 - \Phi$$

Each firm decides data sharing d_j $(j = 1, 2)$ under the revenue distribution rule of data sharing (ζ_1, ζ_2). Then, the condition to maximize the profit can be presented as shown below.

$$\zeta_1\gamma\rho_1\left(\frac{d_2}{d_1}\right)^{\rho_2} = (1-\beta_1)m_1 d_1 \quad \text{and} \quad \zeta_2\gamma\rho_2\left(\frac{d_1}{d_2}\right)^{\rho_1} = (1-\beta_2)m_2 d_2$$

The following condition shows that the above condition is equivalent to (7).

$$\zeta_1 = 1-\beta_1 \quad \text{and} \quad \zeta_2 = 1-\beta_2 \tag{10}$$

Then, the following proposition can be established.

Proposition 2 If the number of firms is $N = 2$ and each firm maximizes profit in the case of a decentralized economy, then one can obtain social welfare that is maximized in the case of the command optimum as long as the lump-sum and subsidy policy are provided appropriately for any revenue distribution rule (ζ_1, ζ_2).

Proof If the subsidy policy by which (10) holds is provided, then one can obtain the data sharing (d_1^*, d_2^*) as shown by (7), and total profit (9) can be expressed as shown below.

$$\pi_1 + \pi_2 = \frac{\theta_1{}^2}{2c_1} + \frac{\theta_2{}^2}{2c_2} + \gamma d_1^{*\rho_1} d_2^{*\rho_2} - \frac{m_1}{2}(1-\beta_1)d_1^{*2} - \frac{m_2}{2}(1-\beta_2)d_2^{*2} - 2\Phi$$

$$=\frac{\theta_1{}^2}{2c_1}+\frac{\theta_2{}^2}{2c_2}+\frac{1}{2}\tau+\frac{m_1}{2}\beta_1 d_1^{*2}+\frac{m_2}{2}\beta_2 d_2^{*2}-2\Phi$$

Because of the government budget constraint ($\frac{m_1}{2}\beta_1 d_1^{*2}+\frac{m_2}{2}\beta_2 d_2^{*2}=2\Phi$), it is apparent that that the total profit calculated above (8) is equivalent to (8). (Q.E.D)

In addition, because of Proposition 2, the revenue distribution rule (ζ_1,ζ_2) can be shown as $\zeta_1=\zeta_2=1-\beta$ if the subsidy is set by the constant rate ($\beta\equiv\beta_1$=β_2). However, because $\zeta_1+\zeta_2=1$ must hold, one obtains $\beta=1/2$ and $\zeta_1=\zeta_2=1/2$.

3. Expansion of the Model Setting ($N>2$)

This section presents expansion of the above-described model to N firms model. Then, the profit of firm j (j=1,2,3,⋯,N)can be shown as follows under the revenue distribution rule of data sharing $(\zeta_1,\zeta_2,\dots,\zeta_N)$ and $\sum_{k=1}^{N}\rho_k=1$,

$$\pi_j=\theta_j D_j+\zeta_j\gamma\prod_{k=1}^{N}d_k{}^{\rho_k}-\frac{c_j}{2}D_j{}^2-\frac{m_j}{2}d_j{}^2 \tag{11}$$

The second term $\gamma\prod_{k=1}^{N}d_k{}^{\rho_k}$ shows the additional revenue produced by data sharing of N firms. Therefore, the produced data (D_j) to maximize profit (11) is $D_j=\theta_j/c_j$. One can obtain the following equation because of $D_j=\theta_j/c_j$ and (11).

$$\pi_j=\frac{\theta_j{}^2}{2c_j}+\zeta_j\gamma\prod_{k=1}^{N}d_k{}^{\rho_k}-\frac{m_j}{2}d_j{}^2 \tag{12}$$

Firm j decides data sharing (d_j) under the revenue distribution rule of data sharing. Therefore $d_j\leq D_j$, the condition to maximize the profit in the decentralized economy, can be shown as

$$\zeta_j\gamma\rho_j\frac{\prod_{k=1}^{N}d_k{}^{\rho_k}}{d_j}=m_j d_j\ .$$

Then, one can obtain the following equations.

$$d_j=\bar{d}_j\equiv\sqrt{\frac{\zeta_j\rho_j}{m_j}\sigma\prod_{k=1}^{N}\zeta_k{}^{\rho_k}}\qquad\left(\sqrt{\sigma}=\gamma\prod_{k=1}^{N}\sqrt{\frac{\rho_k}{m_k}}^{\rho_k}\right) \tag{13}$$

$$\pi_j=\frac{\theta_j{}^2}{2c_j}+\frac{\gamma}{2}\left(2\zeta_j\gamma\prod_{k=1}^{N}\left(\frac{\zeta_k\rho_k}{m_k}\right)^{\rho_k}-\zeta_j\rho_j\gamma\prod_{k=1}^{N}\left(\frac{\zeta_k\rho_k}{m_k}\right)^{\rho_k}\right)$$

Therefore, the social welfare ($W=\sum_{j=1}^{N}\pi_j$) can be shown as follows.

$$W=\sum_{j=1}^{N}\frac{\theta_j{}^2}{2c_j}+\sum_{j=1}^{N}\frac{\gamma}{2}\left(2\zeta_j\gamma\prod_{k=1}^{N}\left(\frac{\zeta_k\rho_k}{m_k}\right)^{\rho_k}-\zeta_j\rho_j\gamma\prod_{k=1}^{N}\left(\frac{\zeta_k\rho_k}{m_k}\right)^{\rho_k}\right)$$

$$=\sum_{j=1}^{N}\frac{\theta_j{}^2}{2c_j}+\frac{\gamma^2}{2}\prod_{k=1}^{N}\left(\frac{\zeta_k\rho_k}{m_k}\right)^{\rho_k}\sum_{j=1}^{N}(2\zeta_j-\zeta_j\rho_j)$$

$$=\sum_{j=1}^{N}\frac{\theta_j{}^2}{2c_j}+\frac{\gamma^2}{2}\prod_{k=1}^{N}\left(\frac{\rho_k}{m_k}\right)^{\rho_k}\prod_{k=1}^{N}\zeta_k{}^{\rho_k}\left(2-\sum_{j=1}^{N}\zeta_j\rho_j\right) \tag{14}$$

The condition to maximize social welfare W is equivalent to $\max\big[\log(\prod_{k=1}^{N}\zeta_k{}^{\rho_k}\,(2-$

$\sum_{j=1}^{N}\zeta_j\rho_j)) + \lambda(1 - \sum_{k=1}^{N}\zeta_k)]$. The condition can be shown as follows. λ denotes the Lagrangian multiplier of $\sum_{k=1}^{N}\zeta_k = 1$.

$$\frac{\rho_j}{\zeta_j} = \frac{\rho_j}{2-\sum_{j=1}^{N}\zeta_j\rho_j} + \lambda \quad \Leftrightarrow \quad \rho_j(2 - \sum_{j=1}^{N}\zeta_j\rho_j) = \zeta_j\rho_j + \lambda\zeta_j(2 - \sum_{j=1}^{N}\zeta_j\rho_j) \qquad (15)$$

Then, because of this equation and constraint ($\sum_{k=1}^{N}\zeta_k = 1$ and $\sum_{k=1}^{N}\rho_k = 1$), one can obtain λ as shown below.

$$\lambda = \frac{2(1-\sum_{j=1}^{N}\zeta_j\rho_j)}{2-\sum_{j=1}^{N}\zeta_j\rho_j}$$

If we substitute this equation into (15) and define $\mathrm{Z} \equiv \sum_{j=1}^{N}\zeta_j\rho_j$, one can obtain $\rho_j(2 - Z) = \zeta_j\rho_j + 2\zeta_j(1 - Z)$ and the following.

$$\zeta_j = \frac{\rho_j(2-Z)}{\rho_j+2(1-Z)} \qquad (16)$$

Substituting this equation into $\mathrm{Z} = \sum_{j=1}^{N}\zeta_j\rho_j$, one can obtain $\mathrm{Z} = \sum_{j=1}^{N}(2 - Z){\rho_j}^2/[\rho_j + 2(1 - Z)]$. Then, if one calculates Z analytically, one can obtain the revenue distribution rule to maximize social welfare (14) $(\zeta_1, \zeta_2, \dots, \zeta_N)$ by considering (16). Then, the following proposition can be established.

Proposition 3 If the number of firms *N* is large and if $\max(\rho_j) < s/N^{\varepsilon}$ $(\varepsilon > 0)$ holds, then (16) is $\zeta_j \approx \rho_j$.

Proof Because of $\max(\rho_j) < s/N^{\varepsilon}$, $\mathrm{Z} = \sum_{j=1}^{N}\zeta_j\rho_j < \frac{s}{N^{\varepsilon}}\sum_{j=1}^{N}\zeta_j = \frac{s}{N^{\varepsilon}}$ holds. Then, $\lim_{N\to\infty} Z = 0$ because $\mathrm{Z} \geq 0$. Therefore, one can obtain $\zeta_j \approx 2\rho_j/(\rho_j + 2)$ and $2/(\rho_j + 2) \approx 1$. That is, $\zeta_j \approx \rho_j$ because of $\rho_j = o(1/N^{\varepsilon})$. (Q.E.D)

Actually, ρ_j denotes the parameter showing the elasticity of additional revenue with respect to provided data brought about by the combination of data sharing ($\gamma\prod_{k=1}^{N}{d_k}^{\rho_k}$). If the assumption to derive Proposition 3 holds, then Proposition 3 shows that the revenue distribution rule to maximize social welfare ζ_j under a decentralized economy can be expected to be consistent with the elasticity of additional revenue with respect to provided data ($\gamma\prod_{k=1}^{N}{d_k}^{\rho_k}$).

Therefore, because the profit maximization condition is $\gamma\rho_j\rho_j\prod_{k=1}^{N}{d_k}^{\rho_k} = m_j{\mathrm{d}_j}^2$, (12) can be changed as expressed below.

$$\pi_j \approx \frac{{\theta_j}^2}{2c_j} + (1 - \frac{1}{2}\rho_j)\rho_j\gamma\prod_{k=1}^{N}{d_k}^{\rho_k}$$

Without data sharing, the profit is $\frac{{\theta_j}^2}{2c_j}$. Then, because of $2 - \rho_j > 0$, firm *j* has an incentive to engage in data sharing.

We present the condition to maximize social welfare for firms $N > 2$ under the command optimum. Because of (12), social welfare ($W = \sum_{j=1}^{N} \pi_j$) can be expressed as

$$W = \sum_{j=1}^{N} \frac{\theta_j^{\,2}}{2c_j} + \gamma \prod_{k=1}^{N} d_k^{\,\rho_k} - \sum_{j=1}^{N} \frac{m_j}{2} \mathrm{d}_j^{\,2} \quad . \tag{17}$$

The condition to maximize (17) can be shown as

$$\gamma \rho_j \frac{\prod_{k=1}^{N} d_k^{\,\rho_k}}{d_j} = m_j d_j \quad . \tag{18}$$

Then, the following equation can be derived.

$$d_j^* = \sqrt{\frac{\rho_j}{m_j}\sigma} \qquad \left(\sqrt{\sigma} = \gamma \prod_{k=1}^{N} \sqrt{\frac{\rho_k}{m_k}}^{\rho_k}\right) \tag{19}$$

$$W^* = \sum_{j=1}^{N} \frac{\theta_j^{\,2}}{2c_j} + \frac{1}{2}\sigma \tag{20}$$

Comparison of (19) with (13) reveals that the data share in a decentralized economy is less than that in the command optimum case because of $\overline{d_j} \leq d_j^*$ $(j = 1, 2, \dots N)$. In addition, comparison of (20) with (14), reveals that (14) in the decentralized economy is less than social welfare (20) in the command optimum case because $W < W^*$. To resolve the difficulty described above, even if one considers a decentralized economy, the allocations in the decentralized economy coincide with the allocations to maximize the social welfare in the command optimum case as long as the lump-sum and subsidy policy are provided appropriately, as shown by Proposition 2: one can establish the following Proposition 4.

Proposition 4 If the number of firms $N > 2$ and if each firm maximizes profit in a decentralized economy, each firm can achieve allocations to maximize social welfare in the command optimum case as long as lump-sum taxation and subsidy policy are provided appropriately for any revenue distribution rule $(\zeta_1, \zeta_2, \dots, \zeta_N)$.

Proof We consider a policy by which a subsidy is provided for data sharing costs of firm j at the constant rate β_j. It is financed by lump-sum taxation Φ for each firm. Then, (12) can be changed as

$$\pi_j = \frac{\theta_j^{\,2}}{2c_j} + \zeta_j \gamma \prod_{k=1}^{N} d_k^{\,\rho_k} - \frac{m_j}{2}(1 - \beta_j)\mathrm{d}_j^{\,2} - \Phi \quad . \tag{21}$$

If each firm decides data sharing d_j under the revenue distribution rule of data sharing $(\zeta_1, \zeta_2, \dots, \zeta_N)$, then the profit maximization condition in a decentralized economy is

$$\zeta_j \gamma \rho_j \frac{\prod_{k=1}^{N} d_k^{\,\rho_k}}{d_j} = (1 - \beta_j) m_j d_j \quad . \tag{22}$$

The condition under which this equation is equivalent to (18) is $\beta_j = 1 - \zeta_j$ $(j = 1, 2, \dots N)$. If the lump-sum taxation Φ is set to hold the government budget constraint $\left(\sum_{j=1}^{N} \beta_j \frac{m_j}{2} d_j^{*2} = N\Phi\right)$, then total profit (21) coincides with social welfare (20). (Q.E.D)

Considering Proposition 4, if the subsidy is provided at constant rate $\beta_j = \beta$ $(j = 1, 2, \ldots N)$ as Proposition 2, then the revenue distribution rule can be derived as $\zeta_j = 1 - \beta (j = 1, 2, \ldots N)$. However, because $\sum_{j=1}^{N} \zeta_j = 1$ must hold, one can obtain $\beta = 1 - 1/N$ and $\zeta_j = 1/N$ $(j = 1, 2, \ldots N)$. If the number of firms $N = 10$, then one can obtain $\zeta_j = 0.1$ and $\beta_j = 0.9$. Proposition 4 considers the lump-sum taxation Φ policy to finance subsidy β_j. However, one can consider another policy by which the flat rate taxation at the rate μ is levied on the additional profit of data sharing. Then, (21) can be changed as presented below.

$$\pi_j = \frac{\theta_j^2}{2c_j} + (1-\mu)\left[\zeta_j \gamma \prod_{k=1}^{N} d_k^{\rho_k} - \frac{m_j}{2}(1-\beta_j) d_j^2\right] \tag{23}$$

In this case, the profit maximization condition of firm *j* in a decentralized economy is shown as (22). If the flat rate tax μ is set to hold the constraint $\beta_j = 1 - \zeta_j$ and if the government budget constraint as shown as follows, then it is apparent that that total profit (23) coincides with social welfare (20).

$$\sum_{j=1}^{N} \beta_j \frac{m_j}{2} d_j^2 = \sum_{j=1}^{N} \mu \left[\zeta_j \gamma \prod_{k=1}^{N} d_k^{\rho_k} - \frac{m_j}{2}(1-\beta_j) d_j^2\right]$$

$$\Leftrightarrow \quad \mu = \frac{\sum_{j=1}^{N} \beta_j \frac{1}{2(1-\beta_j)} \zeta_j \rho_j}{\sum_{j=1}^{N} (\zeta_j - \frac{1}{2}\zeta_j \rho_j)} \quad \Leftrightarrow \quad \mu = \frac{\sum_{j=1}^{N} (1-\zeta_j)\rho_j}{\sum_{j=1}^{N} (2-\rho_j)\zeta_j} \tag{24}$$

Additionally, we designate the "place" (field) that each firm shares as a data "platform." It depends on the decision of the firm whether each firm participates in the platform or not. Moreover, if the government does not provide a policy of lump-sum taxation and subsidy and if each firm voluntarily constructs the platform by which each firm shares data, one can consider a lump-sum tax Φ and a flat rate tax μ as participation fees for the data sharing platform.

Next we consider the case in which there exist $(N-1)$ firms; we construct the platform. Then, if new firm *j* does not participate in the platform, the profit is $\pi_j = \theta_j^2/2c_j$. If new firm *j* participates in the platform, then the profit can be shown as (21) or (23). Therefore, with the following inequality, firm *j* has an incentive to participate in the platform.

In the case of lump-sum tax Φ : $\quad \zeta_j \gamma \prod_{k=1}^{N} d_k^{\rho_k} - \frac{m_j}{2}(1-\beta_j) d_j^2 - \Phi > 0 \qquad (25)$

In the case of flat rate tax μ : $\quad (1-\mu)\left[\zeta_j \gamma \prod_{k=1}^{N} d_k^{\rho_k} - \frac{m_j}{2}(1-\beta_j) d_j^2\right] > 0 \qquad (26)$

The condition under which this inequality holds depends on subsidy rate (β_j),revenue distribution rule (ζ_j), and the elasticity of additional revenue with respect to provided data (ρ_j). For instance, if the subsidy rate is $\beta_j = 1 - 1/N$ and the revenue distribution rule is $\zeta_j = 1/N$, then (22) can be changed by $\frac{m_j}{2} \mathrm{d}_j{}^2 = \frac{1}{2}\rho_j\Omega$ as $\Omega \equiv \gamma \prod_{k=1}^{N} d_k{}^{\rho_k}$. In the case of a lump-sum tax Φ, substituting this equation into the budget constraint of the platform ($\sum_{j=1}^{N} \beta_j \frac{m_j}{2} d_j^{*2} = N\Phi$), one can obtain $\Phi = \frac{1}{2N}\Omega(1 - \frac{1}{N})$. By substituting these equations into (25), one can obtain $\rho_j < 1 + \frac{1}{N}$, which ρ_j holds this condition. If the subsidy rate is $\beta_j = 1 - 1/N$ and the revenue distribution rule is $\zeta_j = 1/N$, then firm *j* has the incentive to participate in the platform. Whether $(N-1)$ firms allow new firm *j* to participate in the platform or not depends on how the profit of each firm is changed by the firm *j* participation. If the participation of firm *j* raises the profit of $(N-1)$ firms, then firm *j* can always participate in the platform. Otherwise, participation depends on the management rule of the platform.

In the case of a flat rate tax μ, by substituting (22) into (26), one can obtain the condition of $\rho_j < 2$ with a brief calculation irrespective of the subsidy rate (β_j) and revenue distribution rule (ζ_j). Here, ρ_j holds in this condition; the following proposition can be established.

Proposition 5 If the subsidy for data sharing is financed by a flat rate tax for additional profit, then each firm has an incentive to participate in the platform irrespective of the subsidy rate (β_j) and revenue distribution rule (ζ_j).

4. Conclusions and Future Research

The main purpose of this paper was to present consideration of a situation in which each firm produces or shares data voluntarily to gain new additional revenue. We set a simple model with multiple firms producing the data. Theoretically, we used it to analyze the revenue distribution rule and behaviors to maximize social welfare. Our analysis yielded the following results.

First, if the number of firms (N) is sufficiently large and some settings are assumed, the revenue distribution rule to maximize social welfare (ζ_j) in a decentralized economy coincides approximately to the elasticity of additional revenue with respect to provided data (ρ_j).

Second, if each firm maximizes the profit in a decentralized economy, the firm can achieve allocations to maximize social welfare in the command optimum for any revenue distribution rule as long as the government appropriately provides a policy of lump-sum tax and subsidy.

Third, if the subsidy for data sharing is financed by a flat rate tax for additional profit, each firm has an incentive to participate in the platform irrespective of subsidy rate (β_j) and revenue distribution rule (ζ_j).

However, the following three works should be examined for future study in this area.

The first is empirical research exploring the elasticity of additional revenue with respect to provided data (ρ_j). In our paper, we examine a model that assumes additional revenue with the combination of data sharing, shown by $\gamma \prod_{k=1}^{m} d_{i_k}{}^{\rho_{i_k}}$. However, checking the robustness of Proposition 2 is important to examine empirically whether the additional revenue can be shown in the above-described form, or not. Empirical confirmation must be done of the stability of elasticity of additional revenue with respect to provided data (ρ_j) and other matters.

The second is related with cooperative games. In the real world, each firm considers the interests of others and sets the platform cooperatively. The platform of data sharing is the same; the notion of the core of cooperative games is important (Peleg & Sudhölter, 2008). The condition to hold the allocations that are included in the core is designated as cooperative rationality. We can consider a case in which multiple firms partially cooperate, except for total cooperation. The additional revenue is distributed for firms. In this case, not a platform, but some platforms can exist. Moreover, although it is important to examine the notion of Shapley value reported by Shapley (1953) and the Humanity "Jin (仁)" described by Schmeidler (1969) discussed in cooperative games, our paper does not address such analysis. Nevertheless, this analysis is important.

The third is about protection of individual information and information security policies. Our model does not consider the negative externalities of data. For instance, even if each does not select data sharing of information, other persons can find "privacy" by which each does not want to announce if they combine and analyze data that exists on the internet. This effect reduces the reservation utility of each actor compared with the optimal level. It produces losses of social welfare. Choi et al. (2019) report a business-stealing effect pointed by Mankiw and Whinston (1986). Reportedly, it works and presents the possibility that websites collect excessive data including the customer information if brokers buy and sell data. To eliminate this difficulty, rules of protection for individual information are discussed in economically developed countries. The additional revenue distribution rule and other matters related to the model should be analyzed in light of the discussion above and considering costs of the policy for information security.

References

· Akerlof, G. (1970) "The market for lemons: quality uncertainty and the market mechanism", *Quarterly Journal of Economics*, 84(3), pp.488-500.

• Boadway, R., Pestieau P. and Wildasin, D.E. (1989) "Non-cooperative Behavior and Efficient Provision of Public Goods," *Public Finance*, 44(1), pp.1-7.

• Choi, J.P., Jeon, D-S., and Kim, B-C. (2019) "Privacy and personal data collection with information externalities," Journal of Public Economics, 173, pp.113-124.

• Hinton, G.E., and Salakhutdinov, R.R. (2006) "Reducing the dimensionality of data with neural networks," science 313 (5786), pp.504-507.

• HM Treasury (2018) "The economic value of data: discussion paper"
https://bit.ly/2Q3j1BU

• Ioannides, Y.M. (2012) "Complexity and organizational architecture," Mathematical Social Sciences, 64 (2), pp.193–202.

• Jackson, M.O. (2008), Social and Economic Networks, Princeton University Press.

• MacCarthy, M. (2011) "New Directions in Privacy: Disclosure, Unfairness and Externalities," I/S: A Journal of Law and Policy for the Information Society, 6 (3), pp.425–512.

• Mankiw, N.G. and Whinston, M.D. (1986) "Free Entry and Social Inefficiency," Rand Journal of Economics, 17, pp.48-58.

• Ministry of Internal Affairs and Communications (2018) "Activity for Social Implementation for "information Bank" (Joho Ginko no Shakai Jisso ni Muketa Torikumi (in Japanese))" https://bit.ly/2MxTsHV

• OECD (2015) "Data-driven Innovation: Big Data for Growth and Well-being" OECD Publishing

• Peleg, B. and Sudhölter, P. (2008) *Introduction to the Theory of Cooperative Games*, second ed., Springer-Verlag.

• Shapley, L. S. (1953) "A value for n-person games," *Contributions to the Theory of Games II*, H. Kuhn and A. W. Tucker (eds.), Princeton University Press, pp. 307– 317.

• Schmeidler, D. (1969) "The nucleolus of a characteristic function game," *SIAM Journal on Applied Mathematics*, 17, pp. 1163–1170.

• Slavov, S.N. (2014) "Public Versus Private Provision of Public Goods," *Journal of Public Economic Theory*, 16(2), pp.222-258.

• Villanacci, A., and Zenginobuz, Ü. (2006) "Pareto Improving Interventions in A General Equilibrium Model with Private Provision of Public Goods," *Review of Economic Design*, 10(3), pp.249-271.

• Yamashige S. (2013) "Economic Analysis of Family and Society Change of Japan Society and Policy Response (Kazoku to Shakai no Keizai Bunseki Nihonteki Shakai no Henyou to Seisakuteki Taiou in Japanese)" University of Tokyo Press