RIETI Discussion Paper Series 19-E-016

Use and sharing of big data, firm networks and their performance

KIM, YoungGak
Senshu University

MOTOHASHI, Kazuyuki
RIETI

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 19-E-016
March 2019

Use and sharing of big data, firm networks and their performance*

YoungGak KIM (Senshu University)
Kazuyuki MOTOHASHI (University of Tokyo, RIETI)

Abstract

RIETI conducted the Survey of Big Data Use and Innovation in Japanese Manufacturing Firms in 2015. This paper uses this survey data, linked with TSR data of inter-firm transactions, to examine the relationship between supplier and customer (business partner) network structures and the data sharing with these business partners. It is found that, in general, the number of suppliers is positively correlated with the likelihood of internal use of data and data sharing with suppliers, customers, and other third-party firms. On the contrary, the number of customers is negatively correlated with data use and sharing, especially with customers. The analysis results also show that long-term relationships with suppliers contribute negatively to data sharing, but contribute positively to data sharing with customers. Interestingly, the more customers a firm's suppliers have, or the more suppliers a firm's customers have in their transaction networks, the less likely it is that the firm shares big data with other third-party firms. We find that data sharing has a positive and significant impact on firm productivity. However, we find no positive contribution of data sharing to attracting new customers or suppliers. We do not find any significant effect of data sharing on the extensive margin of transactions.

Keywords: big data, data sharing, firm network, transaction network
JEL classification: D22, L14, L19, L25

* This study is conducted as a part of the Project "Empirical Analysis of Innovation Ecosystems in Advancement of the Internet of Things (IoT)" undertaken at Research Institute of Economy, Trade and Industry (RIETI). This study utilizes the micro data of the Survey of Big Data Use and Innovation in Japanese Manufacturing Firms conducted by Research Institute of Economy, Trade, and Industry (RIETI), and Tokyo Shoko Research (TSR) firm level data, which is provided by the Small and Medium Enterprise Agency. The authors are grateful for the generous provision of the data. The authors are also thankful for the helpful comments and suggestions by Makoto Yano (RIETI), Masayuki Morikawa (RIETI), Sadao Nagaoka (Tokyo Keizai University), and other Discussion Paper seminar participants at RIETI. Kim acknowledges the financial support of JSPS Kakenhi Grant Number 16K03659.

1. Introduction

Recent digital innovation enabled by networked computers, robot and smart equipment transforms organizations of economies, industry and firm much faster than in the past (Brynjolfsson, 2014). In the era of industrie 4.0, the big wave of digitalization in business process and transaction has significant impacts on Japanese manufacturers. In order to assess such impact, RIETI conducted the Survey of Big Data Use and Innovation in Japanese Manufacturing Firms (hereinafter referred to as the BDS) in 2015.

BDS reveals that big data are widely used in all business activities, including product development, mass production and after services. In addition, firms with a corporate level dedicated big data function are more likely to conduct big data activity by the company as a whole, i.e sharing the data across various departments inside a firm. Such firm is more likely to be involved with inter-firm data sharing, particularly with its business partner, such as suppliers and customers (Motohashi, 2017).

The whole structure of Industrie 4.0 hinges on the concept of cyber physical system, which allows efficient physical product allocation throughout a whole supply chain with digital data generation and communication among parts suppliers and assemblers. An idealistic goal of Industries 4.0 suggests flexible selection of suppliers, based on digital information generated in assembling process (VDMA, 2016). However, the real world is much more complex. The propriety information shared by specific supplier-customer generates higher values to both parties, so that both of them do not have any incentive to disclose such information to their competitors. This is particularly the case for automotive industry. Some empirical literature shows that the relationship specificity between supplier and OEM in Japanese automotive firm leads to its superior performance as compared to their wester competitors (Dyer, 1996; Sako, 1991). In addition, digitalization of manufacturing activities enables a firm to know more its suppliers and customers, so that interdependency between supply-chain partners may become stronger (Vendrell-Herrero et. al, 2017).

Therefore, it is not clear whether we will see the world described Industrie 4.0. However, it is certain that recent advancement of information technology drives fundamental changes in current structure of manufacturing supply chain. This paper shows quantitative evidences regarding the big data use (sharing with business partner), supply chain structure and productivity impact. There are some empirical studies investigating this subject (Gunasekaran et. al, 2017; Janssen et. al, 2017; Wamba et. al, 2017). However, all of them are based on a qualitative survey about big data management capability and performance perception. In contract, we have conducted a systematic evaluation based on quantitative data, such as the number of customers/suppliers and productivity, by constructing the linked data of RIETI's BDS and TSR (Tokyo Shoko Research).

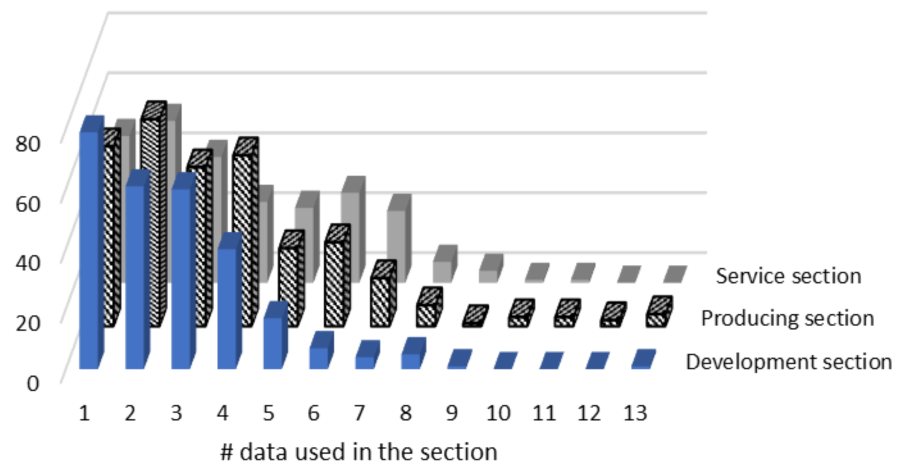
The organization of this paper is as follows. The next section shows the dataset and some descriptive statistics. Then the regression results on the determinants of data sharing with business partner and on the firm's productivity impact are supplied in section 3 and 4, respectively. Finally, this paper concludes with managerial and policy implications.

2. Data

Main sources of the data used in the analyses are (1) the micro data of BDS; and (2) firm level data of Tokyo Shoko Research (TSR) which includes information on domestic transactions as well as basic information on financial reports[1].

BDS is unique and valuable in that it is the first survey on the usage and sharing of big data in Japan. Among 498 respondents, 81% firms report to use big data in the corporate level, and 70% firms among 547 respondents report to use big data in one or more sections[2]. BDS also asks in which section among the sections in the firm, development section, production section, and service section each data is used. Data attributes are classified into 13 categories, such as production process data, POS data, operation data, CAD data, call-center data, and so on. Figure 1 shows how densely the big data is used in each section. Many firms uses only one or two kinds of big data in development section whereas more firms uses 2-6 kinds big data in producing or service sections.

Figure 1 How many kinds of big data are used in each section (BDS)



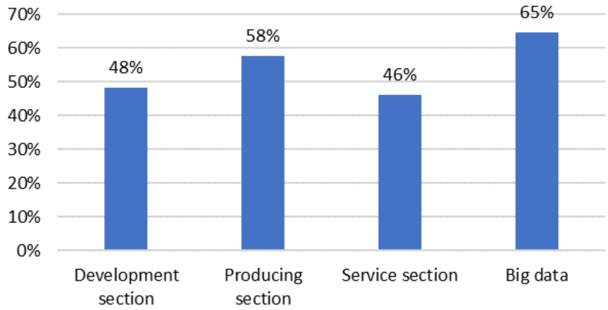
Source: Calculated by authors by using BDS (RIETI, 2015)
Note. Each bar reports how many firms use how many kinds of big data in each section. Each firm reports in which section they use each kind of big data such as production process data, POS data, operation data, CAD data, call-center data, and so on, total number of which is 13. For example, the left-bottom bar (79) means that 79 firms report to use only one kind of big data in development section. 2nd row and 2nd column bar shows that 69 firms use two kinds of big data in producing section.

---

[1] We are grateful to the Small and Medium Enterprise Agency and the Tokyo Shoko Research for the data use.
[2] Total number of the respondents is 592.

Figure 2 shows how many firms use at least one kind of big data in each section of the production stage. 58% of firms use big data one or more kinds of big data in producing section, whereas less firms (46%) use it in service section. 65% of firms use at least one kind of big data in their production stage.

Figure 2 Use of big data in each production stages (BDS)



Source: Calculated by authors by using BDS (RIETI, 2015)

Table 1 shows how firms share the data with their transaction partners in BDS. A third of firms share the big data with their suppliers, and 40% with customers. Although two-way data sharing is more than half among the data-sharing firms, not a few firms share their data in one-way.

Table 1 How firms share their data

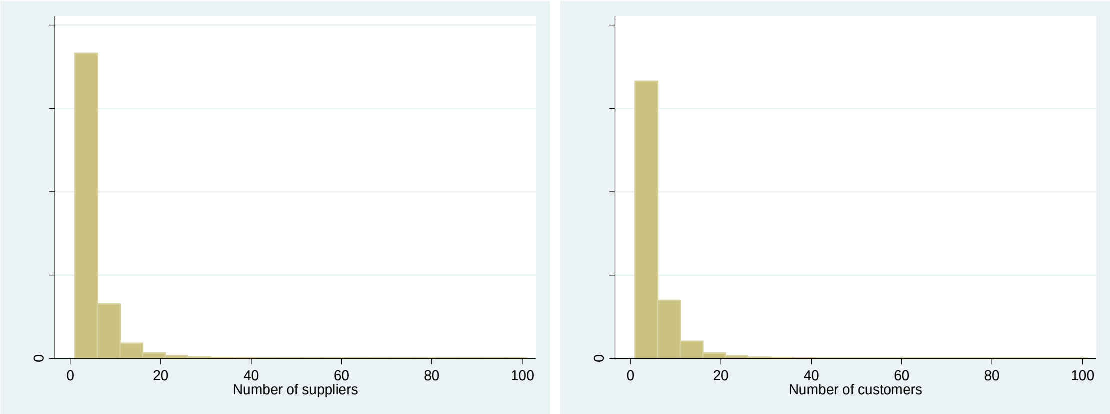
How to share data	with **supplier**	with **customer**
	Freq. (%)	Freq. (%)
not at all	374 (66.43)	343 (60.49)
provide only	31 (5.51)	19 (3.35)
use only	42 (7.46)	73 (12.87)
provide & use	116 (20.6)	132 (23.28)
Total	563 (100)	567 (100)

Source: Calculated by authors by using BDS (RIETI, 2015)

TSR data set is unique in the sense that it covers the transaction information and capital relationship over about 1 million firms every year between 2007 and 2016. Among them, about 700 thousand firms report their suppliers and almost the same number of firms reports their customers. Average number of suppliers and customers are 3.3 and 3.7. Maximum number of the transaction partner in the data set is 24, so that some large firms only report their major partners. Considering this limitation, we reversed the transaction matrix and utilized it, so that suppliers and customers of a firm are identified from the transaction partner sides as well as from the original transaction information.

The distributions of the identified suppliers and customers in 2015 are highly skewed as plotted in Figure 3. It is clear that most firms have only small number of transaction partners. Although average number of suppliers and customers are to 5.5 and 5.9 respectively in TSR data set, both of median number of suppliers and customers are 3. 285,445 firms of 940,163 have only one supplier in the year[3].

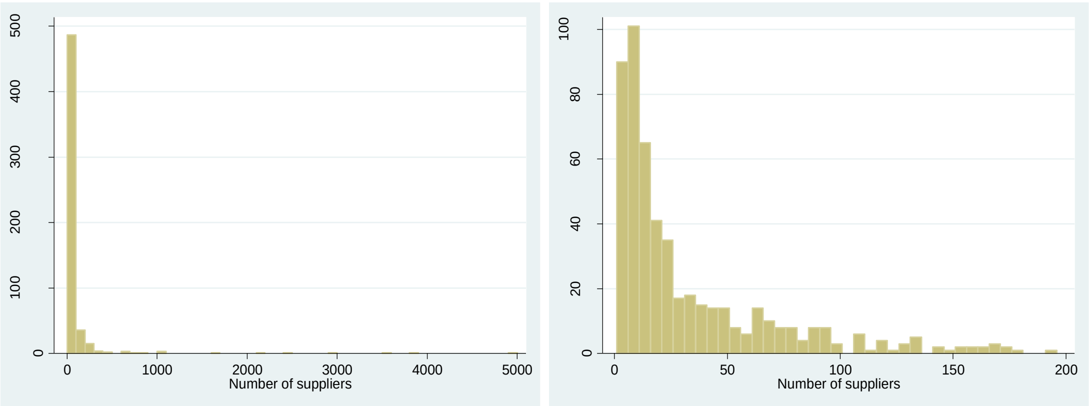
Figure 3 Distributions of the numbers of suppliers and customers (TSR, 2015)



Note: Calculated using TSR by authors. The number of firms with more than 100 suppliers is much less than 1%.

Firm level micro data of BDS are matched with the TSR data set for the analyses, resulting in the panel data set for 566 firms. The numbers of the transaction partner in the matched data set are also distributed in a highly skewed manner as plotted in Figures 4 and 5. Average number of suppliers and customers of the matched data are 92 and 74[4] which are much greater than those of original TSR data set. This is because many of the surveyed firms in the BDS are large firms. The number of firms with only one supplier in 2015 is 14 among 559 matched firms (2.5%) in the sample which is much lower than in the original TSR data set[5].

Figure 4 Distribution of the number of suppliers (TSR + BDS, 2015)

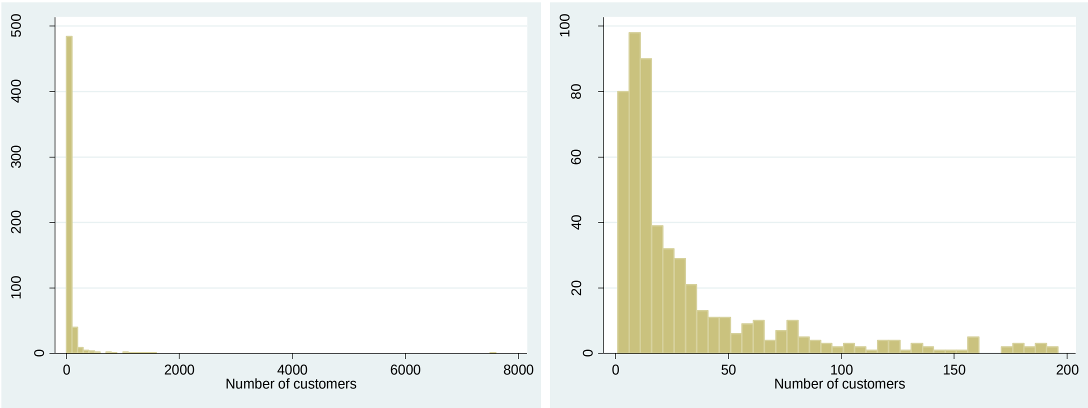


Note: Calculated by authors using TSR and BDS. Width of the bar in the left-hand graph is 100, and that of right-hand

[3] 251,324 firms of 885,541 have only one customer in 2015.
[4] Median number of suppliers and customers of matched data are 17 and 16 respectively.
[5] The number of firms with only one customer in 2015 is 10 among 555 firms.

graph is 5.

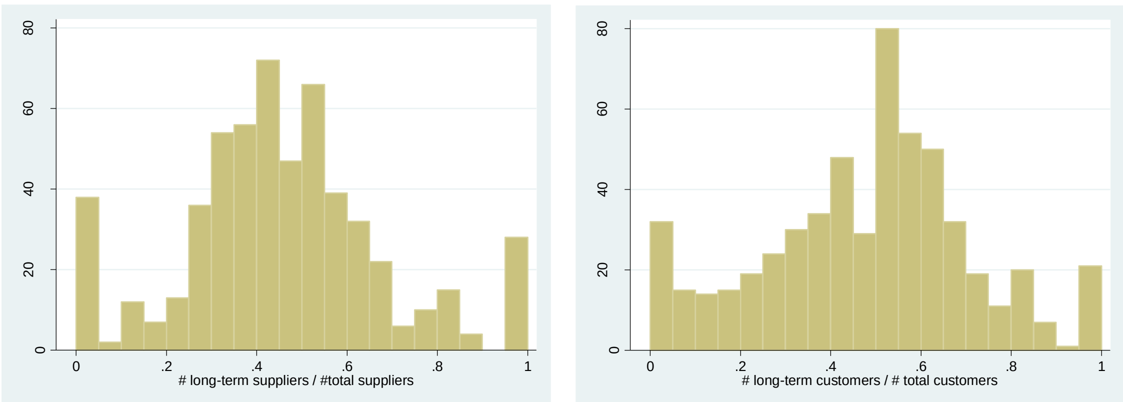
Figure 5 Distribution of the number of customers (TSR + BDS, 2015)



Note: Calculated by authors using TSR and BDS. Width of the bar in the left-hand graph is 100, and that of right-hand graph is 5.

The data set also shows that about half of the transaction pairs are with long term transaction partners[6]. We defined long-term transaction partner as a firm with which the firm transact every period in the sample period. The average ratio of the number of the long-term suppliers over the total number of suppliers is 48%, and that of customers is 50%. Distributions of the number of the long-term transaction partners are plotted in Figure 6.

Figure 6 Distributions of the ratio of the long-term suppliers and customers (TSR + BDS, 2015)



Note: Calculated by authors using TSR and BDS. Width of the bar in the graphs is 0.05.

As described above, TSR data set is unique in the sense that it provides greater information on the transaction network. We calculated the average numbers of the transaction partners of the transaction partners of a firm. The customers of a firm's suppliers are usually her competitors or potential ones at least. The suppliers of a firm's customers may be competitors in the same manner. In

[6] Summary statistics are found in the appendix.

the decision of data sharing, these numbers play an important role[7]. As seen in Table A1 in the appendix, the average number of customers of a firm's suppliers is about 100 and the average number of suppliers of a firm's customers is 284. Interestingly enough, these numbers are negatively correlated with the decisions of data sharing with customers or suppliers (Table A2).

In the following sections, we analyze the determinants of the using and sharing big data, and their impact on the firm performance with the matched data between TSR and BDS described in this section. Although BDS is conducted in 2015, the time period on the usage and sharing of the big data is not specified in many questionnaires, so that matched TSR data between 2007 and 2016 are used in the analyses.

3. Determinants of using and sharing big data

This section investigates the factors affecting firms' decisions of data use and sharing of it. Although use of big data is a prerequisite for the data sharing in general, considering the fact that some firms are provided with big data and use it without providing others, we analyze the decision of big data use and that of data sharing as separate issues.

Use of big data

As described in the previous section, 80% of firms answers that they use big data internally, and 65% reports one or more sections in the production process use specific big data, if not all. Firms are expected to digitalize their production process to enhance the efficiency and productivity. Young and large firms, in general, are more likely to digitalize the production process. Firms producing and selling machinery or electric parts are also expected to have greater likelihood of digitalization.

We regress the data use on the firm characteristics and network information of the firm. Dependent variable is the dummy variable which takes value 1 if a firm reports they use one or more big data among 13 in any section of developing, producing, or service sections (extensive margin). As a first step, samples for the estimations are limited only to the observations in 2014, making the estimations cross-sectional. Since we do not know exactly when firms began to use big data, we restrict the observations for the regressions according to the survey year, so that the values of dependent and independent variables are for 2014.

Result in Model (1) of Table 2 means that larger firms are more likely to use big data. Since business type may affect firms' decision to use big data in the production, dummy variables for four kinds of business type[8]. Business type makes no significant difference. Controlling for the firm

---

[7] We do not include the number of suppliers of a firm's supplier or the number of customers of a firm's customer, because we cannot find any significant role of them in the decision of data sharing.
[8] Four categories of business forms are not exclusive to each other, so that we include all of them in

characteristics, the number of suppliers and customers has no overall impact on the decision of the use of big data.

We do same regressions for the pre-producing, producing, after-producing section in Models (2), (3), and (4). Dependent variable in Model (2) is the data use in developing section, that is, pre-production process. Models (3) and (4) are regressions for the data use in producing, and after-production process. Firms with more customers are less likely to use big data in producing sections with weak significance.

Table 2 Determinants of digitalization, extensive margin (2014)

	1 if using big data	in developing section	in producing section	in service section
	(1)	(2)	(3)	(4)
ln(# supplier$_t$+1)	0.0259	0.0661	0.0423	0.0254
	[0.0382]	[0.0418]	[0.0400]	[0.0421]
ln(# customer$_t$+1)	-0.0429	-0.0292	-0.0604*	-0.0175
	[0.0292]	[0.0309]	[0.0298]	[0.0312]
ln(# employee$_t$)	0.100***	0.0588*	0.119***	0.105***
	[0.0261]	[0.0284]	[0.0268]	[0.0286]
ln(firm age$_t$)	-0.0662	-0.0311	-0.0492	-0.0569
	[0.0367]	[0.0383]	[0.0382]	[0.0405]
1 if BtoB final goods	0.00127	0.0806	-0.00305	0.0317
	[0.0530]	[0.0550]	[0.0532]	[0.0560]
1 if BtoB parts	-0.000159	0.0327	0.0104	0.0203
	[0.0549]	[0.0551]	[0.0544]	[0.0570]
1 if BtoB material	-0.0782	-0.171*	-0.0546	-0.084
	[0.0643]	[0.0682]	[0.0671]	[0.0687]
1 if BtoC	0.126	-0.0306	-0.0163	0.116
	[0.0769]	[0.0825]	[0.0765]	[0.0795]
N	546	546	535	535
adj. R-sq	0.134	0.105	0.14	0.102

Note. Probit estimation. Sample period is 2014. Figures are marginal effect and those in brackets are robust standard error. * p<0.10, ** p<0.05, *** p<0.01.

As described above, we have no information on when firms exactly began to use the big data. Some data-using firms may be in their early stage of the big data use, whereas others in the matured stage. To deal with this shortage of the data, we try two sets of different estimations. First one is to run the same cross-sectional regressions but with longer time period of the independent variables. Note that dependent variable is the same, so that these regressions are cross-sectional. Another one is to revise the values of the dependent variables with limited information on the beginning point of the big data use. The survey questions how long firms have used big data for each kind of the big data. firms are supposed to answers with 6 categories, that is, less than one year, one to two, two to five, five to

each regression.

10, 10 to 15years, or longer than 16 years. We define the firms began to use big data in 2015, 2014, 2012, 2008, 2003, or before according to each corresponding answers. Longer period of independent variables are also used in the second set of the estimations.

Regression results of the first set of estimations are summarized in Models (1) – (4) in Table 3. Model (1) shows that Larger and younger firms with more suppliers and less customers are more likely to use big data. Estimations for the data use in each subsection give almost the same results.

Regressions with more updated dependent variables in Models (5) – (8) basically show the same results to those in (1) – (4).

Table 3 Determinants of digitalization, extensive margin (2007-2014)

	1 if using big data	in developing section	in producing section	in service section	1 if using big data	in developing section	in producing section	in service section
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln(# supplier$_t$+1)	0.0291*	0.0588***	0.0518***	0.0366*	0.0272*	0.0568***	0.0448**	0.0446**
	[0.0137]	[0.0155]	[0.0141]	[0.0147]	[0.0138]	[0.0155]	[0.0145]	[0.0147]
ln(# customer$_t$+1)	-0.0278**	-0.00445	-0.0403***	-0.0196	-0.0249*	-0.00375	-0.0372***	-0.0251*
	[0.0106]	[0.0114]	[0.0108]	[0.0111]	[0.0107]	[0.0114]	[0.0111]	[0.0111]
ln(# employee$_t$)	0.0890***	0.0502***	0.0941***	0.0979***	0.0928***	0.0520***	0.111***	0.100***
	[0.00965]	[0.0109]	[0.00989]	[0.0105]	[0.00976]	[0.0109]	[0.00994]	[0.0105]
ln(firm age$_t$)	-0.0644***	-0.0341*	-0.0641***	-0.0512***	-0.0664***	-0.0350*	-0.0779***	-0.0535***
	[0.0128]	[0.0136]	[0.0132]	[0.0141]	[0.0129]	[0.0136]	[0.0135]	[0.0141]
1 if BtoB final goods	-0.0146	0.0501*	-0.00205	0.0297	0.016	0.0518*	0.0299	0.0195
	[0.0199]	[0.0206]	[0.0198]	[0.0203]	[0.0201]	[0.0206]	[0.0204]	[0.0203]
1 if BtoB parts	-0.0438*	-0.0294	-0.023	-0.0181	-0.0126	-0.031	-0.00658	-0.0217
	[0.0208]	[0.0213]	[0.0206]	[0.0212]	[0.0209]	[0.0213]	[0.0211]	[0.0212]
1 if BtoB material	-0.0452	-0.0993***	-0.0162	-0.0739**	-0.0213	-0.0975***	-0.0465	-0.0630*
	[0.0244]	[0.0257]	[0.0252]	[0.0253]	[0.0245]	[0.0256]	[0.0257]	[0.0253]
1 if BtoC	0.133***	0.00862	-0.0291	0.0622*	0.151***	0.00927	0.00444	0.0708*
	[0.0316]	[0.0298]	[0.0291]	[0.0304]	[0.0313]	[0.0297]	[0.0298]	[0.0302]
N	4,174	4,160	4,174	4,174	4,174	4,160	4,119	4,174
adj. R-sq	0.168	0.157	0.192	0.158	0.164	0.157	0.175	0.161

Note. Probit estimation. Sample period is 2007-2014. Values of dependent variables in Models (5) – (8) are updated with information on categorical answers of the length of the data use of each firm. Dummy variables for year and 2 digit level SIC are included in each regression. Figures are marginal effects and those in brackets are robust standard error. * p<0.10, ** p<0.05, *** p<0.01.

On how heavily firms use the big data (intensive margin), we run the same regressions but with how many kinds of big data firms use as dependent variables. Table 4 summarizes the results of the estimation. Sample for the regressions are restricted to the observations of 2014, and big data using firms, so that the sample size is smaller. We could not find any factor affecting firms' decision of how densely to use the big data in the regressions.

Table 4 Determinants of digitalization, intensive margin (2014)

	ln(# big data used)	in developing section	in producing section	in service section
	(1)	(2)	(3)	(4)
ln(# supplier$_t$+1)	0.0186	0.0851	-0.207*	0.1
	[0.0705]	[0.0804]	[0.0837]	[0.0844]
ln(# customer$_t$+1)	0.012	0.0285	0.0771	-0.0313
	[0.0475]	[0.0631]	[0.0617]	[0.0616]
ln(# employee$_t$)	0.0617	-0.0639	0.108	-0.0375
	[0.0518]	[0.0580]	[0.0589]	[0.0674]
ln(firm age$_t$)	0.0038	-0.184**	0.0323	0.12
	[0.0550]	[0.0659]	[0.0782]	[0.0752]
1 if BtoB final goods	0.175	-0.165	0.0917	0.0948
	[0.102]	[0.132]	[0.109]	[0.122]
1 if BtoB parts	0.14	-0.152	0.136	0.0145
	[0.0999]	[0.125]	[0.111]	[0.120]
1 if BtoB material	-0.147	-0.396*	0.0702	-0.178
	[0.112]	[0.174]	[0.125]	[0.133]
1 if BtoC	-0.171	-0.0654	-0.391***	0.299
	[0.129]	[0.175]	[0.115]	[0.153]
N	357	268	318	257
adj. R-sq	0.076	0.037	0.022	0.037

Note. OLS. Sample period is 2014. Figures are coefficients and those in brackets are robust standard error. * p<0.10, ** p<0.05, *** p<0.01.

Table 5 summarizes the results of the estimations utilizing longer time period of data, that is, 2007-2014. Dependent variables in Models (5) – (8) of Table 5 are revised with the information on the length of big data use of each firm, as described above (Table 3). Model (6) implies that smaller (not larger) and young firms with more suppliers and more customers are more likely to use big data in their pre-producing, developing section. Interestingly, result in Model (7) is contrast to that in Model (6), meaning that larger and older firms with less suppliers are more likely to use big data in the producing section. The negative effect of the number of suppliers in Model (7) is also contrast to the positive effect of it on the decision of whether to use big data in Model (7) in Table 3.

Table 5 Determinants of digitalization, intensive margin (2007-2014)

	ln(# big data used)	in developing section	in producing section	in service section	ln(# big data used)	in developing section	in producing section	in service section
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln(# supplier$_t$+1)	0.0374	0.0775**	-0.157***	0.0802*	0.0487*	0.0768**	-0.159***	0.064
	[0.0254]	[0.0296]	[0.0290]	[0.0319]	[0.0248]	[0.0282]	[0.0308]	[0.0331]
ln(# customer$_t$+1)	-0.0139	0.0517*	0.0388	-0.0219	-0.00549	0.0655**	0.033	0.0316
	[0.0171]	[0.0226]	[0.0222]	[0.0232]	[0.0178]	[0.0221]	[0.0223]	[0.0241]
ln(# employee$_t$)	0.0462*	-0.0842***	0.0652**	-0.025	0.0518**	-0.0796***	0.0726***	-0.0426
	[0.0195]	[0.0233]	[0.0199]	[0.0256]	[0.0184]	[0.0227]	[0.0211]	[0.0258]
ln(firm age$_t$)	0.0106	-0.147***	0.0449	0.0827**	0.0221	-0.127***	0.0895**	0.0724**
	[0.0196]	[0.0253]	[0.0257]	[0.0284]	[0.0200]	[0.0249]	[0.0272]	[0.0276]
1 if BtoB final goods	0.230***	-0.180***	0.105*	0.141**	0.123**	-0.180***	0.0466	0.103*
	[0.0392]	[0.0500]	[0.0417]	[0.0471]	[0.0386]	[0.0490]	[0.0442]	[0.0455]
1 if BtoB parts	0.114**	-0.211***	0.121**	0.00302	0.0409	-0.180***	0.0889*	0.0422
	[0.0368]	[0.0474]	[0.0427]	[0.0471]	[0.0362]	[0.0465]	[0.0437]	[0.0440]
1 if BtoB material	-0.100*	-0.386***	0.0781	-0.130*	-0.194***	-0.379***	0.113*	-0.171**
	[0.0442]	[0.0654]	[0.0506]	[0.0555]	[0.0435]	[0.0645]	[0.0501]	[0.0522]
1 if BtoC	-0.180***	-0.0302	-0.403***	0.313***	-0.185***	-0.0714	-0.432***	0.311***
	[0.0509]	[0.0714]	[0.0519]	[0.0601]	[0.0509]	[0.0678]	[0.0518]	[0.0642]
N	2,756	2,091	2,474	1,986	2,709	2,086	2,289	1,929
adj. R-sq	0.119	0.137	0.104	0.079	0.128	0.134	0.087	0.114

Note. Probit estimation. Sample period is 2007-2014. Values of dependent variables in Models (5) – (8) are updated with information on categorical answers of the length of the data use of each firm. Dummy variables for year and 2 digit level SIC are included in each regression. Figures are coefficients and those in brackets are robust standard error. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Sharing of big data

On the decision of sharing their data generated from the design, production, and distribution processes, firms consider many factors. Sharing data may enhance the productivity of a firm through many channels, such as, better management of inventory, more accurate design and production with lower rate of defectives, rapid decision making, and so on. However, these merits of sharing the data may be cancelled out, since data sharing may weakens the competitiveness of the firm.

Both of merits and demerits are expected to increase as the number of the transaction partners gets greater. Which effect is greater may depend on the industry, firm size, firm age, and the product chain the firm locates, and should be answered by the empirical analyses. To answer to the question, we regress whether to share or not the data with other firms on firms' basic information (firm size and age) and transaction relationship (number of suppliers and customers).

Table 6 shows the estimation results. Sample period of the first four regressions are 2014, whereas last four are from 2007 to 2014. Model (5) in the table shows that firms with more suppliers and less customers are more likely to share the big data with other firms.

Model (6) implies that the number of suppliers is important on the decision of sharing data with suppliers, but the number of customers has no significant impact on it. The number of customers has negative and significant impact on sharing data with customers. Firms with more customers are less likely to share the data with customers. On sharing data with others, smaller firms with more transaction partner are more likely to share.

Besides, Firms whose business deals with BtoB final goods or BtoB parts but not BtoB materials are also more like to share the data. Firms with BtoC business are less likely to share the data with other third-party firms.

Table 6 Determinants of data sharing (overall)

	Year=2014				2007≤Year≤2014			
	1 if sharing big data	with suppliers	with customers	with other firms	1 if sharing big data	with suppliers	with customers	with other firms
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
ln(# supplier$_t$+1)	0.0973*	0.0604	0.0656	0.0251	0.0969***	0.0690***	0.0784***	0.0212*
	[0.0405]	[0.0393]	[0.0396]	[0.0247]	[0.0147]	[0.0144]	[0.0145]	[0.00930]
ln(# customer$_t$+1)	-0.0395	-0.000388	-0.0233	0.0189	-0.0448***	-0.0064	-0.0278*	0.0176*
	[0.0308]	[0.0301]	[0.0302]	[0.0204]	[0.0112]	[0.0110]	[0.0110]	[0.00757]
ln(# employee$_t$)	0.0117	0.00823	0.0224	-0.0379*	0.0171	0.00404	0.0152	-0.0355***
	[0.0281]	[0.0271]	[0.0275]	[0.0175]	[0.0103]	[0.0100]	[0.0102]	[0.00668]
ln(firm age$_t$)	0.0191	-0.0043	0.0149	0.0117	0.00895	-0.00519	0.00708	0.00728
	[0.0374]	[0.0366]	[0.0377]	[0.0235]	[0.0128]	[0.0125]	[0.0127]	[0.00806]
1 if BtoB final goods	0.0284	0.0205	0.00614	0.108**	0.0517*	0.0339	0.023	0.115***
	[0.0564]	[0.0567]	[0.0561]	[0.0375]	[0.0202]	[0.0204]	[0.0202]	[0.0135]
1 if BtoB parts	0.107	0.0609	0.143**	0.0516	0.122***	0.0725***	0.158***	0.0570***
	[0.0564]	[0.0572]	[0.0552]	[0.0378]	[0.0203]	[0.0206]	[0.0199]	[0.0136]
1 if BtoB material	-0.111	-0.0664	-0.0542	0.0328	-0.0873***	-0.0508*	-0.0347	0.0397*
	[0.0698]	[0.0719]	[0.0701]	[0.0438]	[0.0251]	[0.0258]	[0.0252]	[0.0155]
1 if BtoC	-0.0264	0.0297	0.00222	-0.0672	-0.0145	0.0277	0.00799	-0.0632**
	[0.0819]	[0.0812]	[0.0803]	[0.0642]	[0.0299]	[0.0295]	[0.0294]	[0.0235]
Observation	547	539	543	538	4,193	4,141	4,161	4,122
Pseudo R2	0.048	0.034	0.048	0.052	0.047	0.033	0.046	0.053
Industry F.E.	No	No	No	No	Yes	Yes	Yes	Yes
Year F.E.	No	No	No	No	Yes	Yes	Yes	Yes

Note. Probit estimation. Figures are marginal effects and those in brackets are robust standard error. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

It seems that determinants of data sharing differ according to whom the firms share the data with. In the following, we divide the analyses into three according to the counterpart with whom the big data is shared. Samples for the estimation also are restricted to the observation of the firm that has internal big data, which is to be clear on what decision making we are focusing.

Two sets of explanatory variables are added for the analyses to capture firms' information on the network firms are placed. First is the ratio of the long-term transaction partners. Long term relationship with transaction partners are expected to alleviate the negative effect of data sharing in the transaction. The ratio of the long-term suppliers is defined and measured as the ratio of the number of long-term suppliers[9] over the total number of suppliers every year. The ratio of long-term customers is defined in the same manner.

Another set of variables is the average number of the transaction partners of the transaction partners. Suppliers of a firm have their own customers and suppliers. Customers of the supplier of a firm may be competitors of the firm, so that its number may have significant impact on the decision of data sharing. In the same sense, the number of suppliers of the customer of a firm may be important when deciding whether to share the data[10] [11].

Data sharing with suppliers

Table 7 shows that smaller firms with more suppliers are more likely to share the data with suppliers. The effect of the number of customers is negative and significant (Models 1 and 3).

Long-term relationship with suppliers seems to play a negative and significant role on the decision of data sharing with suppliers. We do not find any other significant effect of the network variables on the decision of data sharing with suppliers. Controlling for variables for transaction and network characteristics, firm size has negative effect on the data sharing with suppliers.

Table 7 Determinants of sharing data with suppliers

---

[9] Long term supplier is defined as a supplier with which a firm transact in more than 90% of the sample period. Considering our data set covers 2007 – 2016, this criteria in fact means the firm transact with the long-term supplier every year.

[10] Although we get the information on the number of the suppliers of the suppliers and that of the customers of the customers, such information is not likely to have direct effect, so that we do not include them in the regressions.

[11] Although it is best to get the number of the customers of the supplier with which a firm shares the data, the survey does not ask with which suppliers or customers the firm shares the data, so that we take the average number of the customers of the suppliers of a firm for this variable. It is also applied to the customer information.

	1 if sharing data with suppliers			
	(1)	(2)	(3)	(4)
ln(# supplier$_t$+1)	0.104***	0.0997***	0.103***	0.0997***
	[0.0192]	[0.0205]	[0.0192]	[0.0205]
ln(# customer$_t$+1)	-0.0326*	-0.0309	-0.0315*	-0.0291
	[0.0149]	[0.0159]	[0.0153]	[0.0163]
# long-term supplier$_t$		-0.148**		-0.147**
/ # supplier$_t$		[0.0548]		[0.0546]
# long-term customer$_t$		0.0342		0.0364
/ # customer		[0.0517]		[0.0519]
ln(# customer of suppliers$_t$+1)			-0.00866	-0.00265
			[0.0102]	[0.0110]
ln(# supplier of customers$_t$+1)			0.0025	0.0038
			[0.00732]	[0.00787]
ln(# employee$_t$)	-0.0328*	-0.0338*	-0.0340*	-0.0348*
	[0.0136]	[0.0144]	[0.0137]	[0.0145]
ln(firm age$_t$)	0.00768	0.0159	0.00833	0.0163
	[0.0173]	[0.0188]	[0.0172]	[0.0188]
1 if BtoB final goods	0.165***	0.178***	0.164***	0.179***
	[0.0267]	[0.0284]	[0.0267]	[0.0284]
1 if BtoB parts	0.140***	0.144***	0.138***	0.144***
	[0.0278]	[0.0297]	[0.0278]	[0.0297]
1 if BtoB material	-0.00474	-0.00925	-0.00127	-0.00774
	[0.0349]	[0.0376]	[0.0352]	[0.0378]
1 if BtoC	-0.0149	-0.0209	-0.0163	-0.0213
	[0.0417]	[0.0449]	[0.0417]	[0.0448]
Observation	2,794	2,458	2,794	2,458
Pseudo R^2	0.081	0.086	0.081	0.086
Industry F.E.	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes

Note. Probit estimation. Figures are marginal effects and those in brackets are robust standard error. * p<0.10, ** p<0.05, *** p<0.01. Long term supplier (customer) is defined as a supplier (customer) with which a firm transacts in more than 90% of the sample period.

Data sharing with customers

The results in Table 8 indicate that smaller firms with more suppliers and less customers are more likely to share the data with customers. The negative impact of the number of customers on data sharing with customers is stronger than with suppliers. As described above, opening-up the information and sharing the data may weaken the competitiveness. This possible problem, however, may be alleviated by the long-term relationship with customers. Models (2) and (4) show that long-term relationship with customers does encourage the data sharing. Secondary network impact is not significant when sharing the data with customers.

Coefficients of the variables on product characteristics shows that firms dealing with BtoB parts and BtoB final goods are more likely to share the data.

Table 8 Determinants of sharing data with customers

	1 if sharing data with customers			
	(1)	(2)	(3)	(4)
ln(# supplier$_t$+1)	0.113***	0.112***	0.114***	0.113***
	[0.0190]	[0.0203]	[0.0191]	[0.0204]
ln(# customer$_t$+1)	-0.0498***	-0.0496**	-0.0478**	-0.0467**
	[0.0148]	[0.0157]	[0.0153]	[0.0163]
# long-term supplier$_t$		-0.163**		-0.165**
/ # supplier$_t$		[0.0551]		[0.0549]
# long-term customer$_t$		0.143**		0.146**
/ # customer		[0.0529]		[0.0530]
ln(# customer of suppliers$_t$+1)			0.00169	0.00409
			[0.0101]	[0.0110]
ln(# supplier of customers$_t$+1)			0.00437	0.0063
			[0.00733]	[0.00792]
ln(# employee$_t$)	-0.0315*	-0.0333*	-0.0323*	-0.0344*
	[0.0136]	[0.0145]	[0.0137]	[0.0146]
ln(firm age$_t$)	0.0185	0.0231	0.019	0.0237
	[0.0169]	[0.0186]	[0.0170]	[0.0186]
1 if BtoB final goods	0.0823**	0.0861**	0.0832**	0.0877**
	[0.0262]	[0.0282]	[0.0262]	[0.0282]
1 if BtoB parts	0.223***	0.226***	0.223***	0.225***
	[0.0263]	[0.0282]	[0.0263]	[0.0282]
1 if BtoB material	0.0484	0.041	0.0485	0.0408
	[0.0319]	[0.0343]	[0.0323]	[0.0348]
1 if BtoC	0.0119	0.00986	0.0115	0.00967
	[0.0390]	[0.0433]	[0.0390]	[0.0432]
Observation	2,769	2,442	2,769	2,442
Pseudo R^2	0.079	0.086	0.079	0.086
Industry F.E.	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes

Note. Probit estimation. Figures are marginal effects and those in brackets are robust standard error. * p<0.10, ** p<0.05, *** p<0.01. Long term supplier (customer) is defined as a supplier (customer) with which a firm transacts in more than 90% of the sample period.

Data sharing with other third-party firms

Estimation results of data sharing with other third-party firms in Table 9 are different from those with suppliers or customers. On data sharing with third-party firms, the numbers of suppliers and customers have no significant impact. It is interesting that the estimated coefficients of the number of customers of the suppliers, the number of suppliers of the customers, and the ratio of the long-term suppliers are all negative and significant, which means that in the case where a firm's suppliers have more customers (supposedly potential competitors), its customers have more suppliers (supposedly potential competitors), or the firm has more long-term suppliers, the firm is less likely to share the data with other firms. Saying in other way, if a firm is in the intensive firm network, they may be less likely to join the open network. It is a common result that smaller and BtoB business firms are more likely to share the data.

Table 9 Determinants of sharing data with other third-party firms

	1 if sharing data with third-party firms			
	(1)	(2)	(3)	(4)
ln(# supplier$_t$+1)	0.0142	0.0129	0.0077	0.00702
	[0.0132]	[0.0140]	[0.0128]	[0.0137]
ln(# customer$_t$+1)	0.0232*	0.0217	0.0135	0.0114
	[0.0112]	[0.0118]	[0.0108]	[0.0115]
# long-term supplier$_t$		-0.135***		-0.116**
/ # supplier$_t$		[0.0391]		[0.0382]
# long-term customer$_t$		0.0247		0.0131
/ # customer		[0.0375]		[0.0357]
ln(# customer of suppliers$_t$+1)			-0.0362***	-0.0346***
			[0.00689]	[0.00744]
ln(# supplier of customers$_t$+1)			-0.0205***	-0.0210***
			[0.00501]	[0.00531]
ln(# employee$_t$)	-0.0388***	-0.0379***	-0.0363***	-0.0351***
	[0.00966]	[0.0103]	[0.00965]	[0.0103]
ln(firm age$_t$)	-0.00127	0.00333	-0.000752	0.00413
	[0.0108]	[0.0116]	[0.0109]	[0.0118]
1 if BtoB final goods	0.172***	0.183***	0.168***	0.177***
	[0.0184]	[0.0198]	[0.0184]	[0.0199]
1 if BtoB parts	0.0923***	0.0993***	0.100***	0.105***
	[0.0183]	[0.0196]	[0.0185]	[0.0199]
1 if BtoB material	0.0609**	0.0623*	0.0848***	0.0829***
	[0.0224]	[0.0246]	[0.0223]	[0.0243]
1 if BtoC	-0.0252	-0.0268	-0.0371	-0.0407
	[0.0330]	[0.0356]	[0.0317]	[0.0342]
Observation	2,575	2,265	2,575	2,265
Pseudo R^2	0.106	0.116	0.127	0.136
Industry F.E.	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes

Note. Probit estimation. Figures are coefficients and those in brackets are robust standard error. * p<0.10, ** p<0.05, *** p<0.01. Long term supplier (customer) is defined as a supplier (customer) with which a firm transacts in more than 90% of the sample period.

Sharing data does not always mean two-way. Some firms only provide the data for the partner, and others only use the data, although more firms do both way or no way. We expect that decision making of data-sharing may be different from each other. BDS asks about the data sharing whether they only provide, only use, provide and use, or do nothing.

As for sharing data with suppliers, the number of suppliers is important when the firm only provide or provide-and-use the data, but not when the firm only use the data. The assemblers may enjoy greater productivity gain when they provide their data to their suppliers to enhance the production efficiency, for example, by lowering the defective rate. But such productivity gain is not expected when a firm only use the date of the supplier (Model 2 of Table 10).

The estimated coefficient of the number of suppliers of customers (=ln(number of suppliers of customers+1)) is significantly negative in Model (4) indicating that firms which supply to customer firms with many suppliers are less likely to provide the data, because such behavior supposedly lowers the firms' competitiveness. Contrastingly, its sign is positive when the firm only use the big data of the customers, because the more suppliers the customers have, the more likely the firm get more information on the market they compete. Alleviation effect of the long-term relationship with customer is only important in the cases of one-way data sharing.

Table 10 Determinants of data sharing, provide only, use only, or both?[12]

Dep. Var. : 1 if sharing data	with Suppliers			with Customers		
	provide only	use only	provide & use	provide only	use only	provide & use
	(1)	(2)	(3)	(4)	(5)	(6)
ln(# supplier$_t$+1)	0.388*	0.297	0.493***	0.895***	0.602***	0.389***
	[0.169]	[0.187]	[0.102]	[0.183]	[0.154]	[0.0990]
# long-term supplier$_t$	-0.4	0.633	-0.896***			
/ # supplier$_t$	[0.485]	[0.373]	[0.265]			
ln(# customers of suppliers$_t$+1)	-0.00165	0.0425	-0.08			
	[0.0933]	[0.0781]	[0.0591]			
ln(# customer$_t$+1)	0.196	-0.0389	-0.172*	-0.791***	-0.115	-0.142
	[0.109]	[0.134]	[0.0815]	[0.206]	[0.117]	[0.0772]
# long-term customer$_t$				1.293*	1.445***	0.117
/ # customer				[0.554]	[0.340]	[0.235]
ln(# suppliers of customers$_t$+1)				-0.253**	0.158**	-0.039
				[0.0783]	[0.0556]	[0.0381]
ln(# employee$_t$)	-0.363**	-0.0157	-0.182**	-0.325*	-0.136	-0.0784
	[0.116]	[0.124]	[0.0701]	[0.128]	[0.109]	[0.0689]
ln(firm age$_t$)	0.0205	-0.0187	0.176	0.281	0.0898	0.0512
	[0.164]	[0.133]	[0.0969]	[0.229]	[0.135]	[0.0863]
1 if BtoB final goods	-0.137	-0.124	0.803***	0.552	0.0755	0.393**
	[0.282]	[0.214]	[0.141]	[0.319]	[0.159]	[0.144]
1 if BtoB parts	-0.449	0.312	0.825***	0.870**	1.040***	1.013***
	[0.257]	[0.225]	[0.149]	[0.311]	[0.159]	[0.147]
1 if BtoB material	-34.53***	0.694**	-0.185	1.572***	-1.408***	0.00802
	[0.323]	[0.255]	[0.173]	[0.264]	[0.317]	[0.161]
1 if BtoC	0.046	-0.0934	-0.442*	-31.88***	0.226	-0.36
	[0.259]	[0.361]	[0.210]	[0.474]	[0.228]	[0.207]
Observation		2,500			2,500	
Pseudo R^2		0.081			0.085	
Industry F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes

Note. Multinomial logit estimation. Figures are coefficients and those in brackets are robust standard error. * p<0.10, ** p<0.05, *** p<0.01. Long term supplier (customer) is defined as a supplier (customer) with which a firm transact in more than 90% of the sample period.

[12] Multinomial logit estimation. It is known that multinomial probit (MNP) is better than multinomil logit (MNL) because MNL makes the often-erroneous independence of irrelevant alternatives (IIA) assumption. MNP is free from the IIA assumption. However, Dow and Endersby (2004) shows that MNL is often preferable to MNP, even when the IIA assumption is violated. We adopt MNL estimation here for this reason.

4. Effects of data sharing on firm performance

Productivity

This section investigates the effect of data sharing on firms' performances. Although decision of data sharing is never exogeneous as discussed in the previous section, we do not address the endogeneity issue any more here, mainly because the sample size is small and key variables (whether and with whom to share data) are not time variant. OLS and simple IV estimation (2 stage least squares) are applied to the analyses.

We regress simplest and most intuitive performance variable, productivity, on the data sharing variables, as a first step. Log value of the sales per employee is regressed on four types of data sharing variables (data sharing, data sharing with suppliers, customers, and other third-party firms) as well as the firm size (log value of the number of employee) and capital labor ratio (log value of the ratio of the value of the nominal tangible fixed asset over the number of employees)[13].

Table 11 shows the results of OLS and IV regression. 2-5 period lagged values of the number of employees, the numbers of suppliers and customers, rates of long-term suppliers and customers, the number of customers of suppliers, and the number of suppliers of the customers, as well as the lagged value of capital labor ratio and the number of employees are used as IV in Models (6) – (9). Regression results show that data sharing significantly enhances firms' productivity.

Table 11 Data sharing and productivity

---

[13] In principle, dependent variable in this section should be the value-added instead of sales. However, TSR used in this research has little information on the intermediate input, so that calculating value-added with the limited information gives us too small number of samples, which makes regressions less-runnable. It is also pointed out that possible correlation with data-sharing and intermediate input may gives seemingly positive relationship between sales and data sharing. We are thankful to the attendants of Discussion paper review seminar for these points, and leave them for the future works.

Dep. Var.: ln(Sales/emp)$_t$	OLS					IV			
Year=2015~2016	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ln(K/L)$_{t-1}$	0.068*	0.069*	0.068*	0.075*	0.062*	0.083*	0.083*	0.088**	0.078*
	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)	(0.043)	(0.043)	(0.044)	(0.042)
ln(EMP)$_{t-1}$	0.233***	0.226***	0.224***	0.230***	0.234***	0.158***	0.169***	0.171***	0.201***
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.025)	(0.025)	(0.023)	(0.020)
1 (data sharing)		0.104**				0.699**			
		(0.044)				(0.271)			
1 (data sharing with supplier)			0.191***				0.726**		
			(0.043)				(0.307)		
1 (data sharing with customer)				0.106**				0.615**	
				(0.043)				(0.255)	
1 (data sharing with others)					0.222***				1.345***
					(0.063)				(0.491)
Observation	826	826	815	818	810	776	768	768	760
Adjusted R^2	0.503	0.506	0.504	0.511	0.505	0.406	0.434	0.451	0.34
Industry F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note. Instrument variables used in Models (6) - (9) are 2-5 period lagged values of the number of employees, the numbers of suppliers and customers, rates of long-term suppliers and customers, the number of customers of suppliers, and the number of suppliers of the customers, as well as the lagged value of capital labor ratio and the number of employees. Figures in brackets are robust standard error. * p<0.10, ** p<0.05, *** p<0.01.

New customers

Does data sharing contribute to or hinder the new innovation? Data sharing may speed up the flow of information between the firms resulting in the new products or process innovation. However, stronger and more stable relationship with suppliers and customers may hinder the flow of new product, methods, and materials, and new idea. If a firm innovates a new product, it may search and find a new customer, in which case the number of new customers indicates the degree of the product innovation of the firm[14]. In the same vein, getting new suppliers mean the process innovation.

We regress whether a firm gets new customer or not on the data sharing variables with other control variables. Table 12-a show the results of probit estimation. In the models, data sharing has weak effect on pioneering new customers. Model (3) shows that data sharing with customers positively contributes to pioneering new customers.

Estimation with the number of new customers as a dependent variable in Table 12-b is of little difference from those of Table 12-a.

Table 12-a Data sharing and new customers

[14] The net number of new customers (the number of new customers – the number of customers with which the transaction are ceased) may be another candidate for the index of product innovation. However, the estimation results are same, so that we do not report them.

Dep. Var.: 1 if# NEW customers$_t$>0	Probit				
Year=2015~2016	(1)	(2)	(3)	(4)	(5)
ln(# customer$_{t-1}$+1)	0.735***	0.741***	0.743***	0.725***	0.751***
	(0.074)	(0.074)	(0.075)	(0.073)	(0.076)
1 (data sharing)	0.241**				
	(0.098)				
1 (data sharing with supplier)		0.148			0.026
		(0.105)			(0.132)
1 (data sharing with customer)			0.182*		0.151
			(0.100)		(0.126)
1 (data sharing with others)				0.172	0.114
				(0.175)	(0.179)
ln(EMP)$_{t-1}$	0.147***	0.146***	0.149***	0.160***	0.151***
	(0.052)	(0.051)	(0.052)	(0.051)	(0.053)
ln(Age)$_t$	0.02	0.013	0.011	0.033	0.015
	(0.097)	(0.098)	(0.098)	(0.098)	(0.099)
Observation	1073	1058	1065	1055	1042
Pseudo R^2	0.303	0.302	0.302	0.299	0.305
Industry F.E.	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes

Note. Figures in brackets are robust standard error. * p<0.10, ** p<0.05, *** p<0.01.

Table 12-b Data sharing and the number of new customers

Dep. Var.: ln(# NEW customers$_t$+1)	OLS				
Year=2015~2016	(1)	(2)	(3)	(4)	(5)
ln(# customer$_{t-1}$+1)	0.597***	0.601***	0.599***	0.592***	0.597***
	(0.026)	(0.026)	(0.026)	(0.026)	(0.025)
1 (data sharing)	0.065*				
	(0.038)				
1 (data sharing with supplier)		0.02			-0.057
		(0.040)			(0.052)
1 (data sharing with customer)			0.076**		0.096*
			(0.038)		(0.049)
1 (data sharing with others)				0.109	0.093
				(0.071)	(0.072)
ln(EMP)$_{t-1}$	0.077***	0.077***	0.077***	0.084***	0.081***
	(0.020)	(0.020)	(0.020)	(0.019)	(0.020)
ln(Age)$_t$	-0.178***	-0.180***	-0.179***	-0.174***	-0.178***
	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)
Observation	1104	1089	1096	1086	1073
Adjusted R^2	0.667	0.667	0.667	0.668	0.67
Industry F.E.	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes

Note. Figures in brackets are robust standard error. * p<0.10, ** p<0.05, *** p<0.01.

New suppliers

Similarly, we regress whether a firm finds new suppliers or not on the data sharing variables, to get the results in Tables 13-a and b. We find no evidence that, as a whole, data sharing with other firms contribute to searching and finding new suppliers.

Table 13-a Data sharing and new suppliers

Dep. Var.: 1 if# NEW suppliers$_t$>0	Probit				
Year=2015~2016	(1)	(2)	(3)	(4)	(5)
ln(# supplier$_{t-1}$+1)	0.460***	0.472***	0.460***	0.447***	0.453***
	(0.087)	(0.088)	(0.087)	(0.089)	(0.091)
1 (data sharing)	-0.016				
	(0.102)				
1 (data sharing with supplier)		0.127			0.214
		(0.113)			(0.145)
1 (data sharing with customer)			-0.02		-0.156
			(0.106)		(0.133)
1 (data sharing with others)				0.017	0.002
				(0.170)	(0.178)
ln(EMP)$_{t-1}$	0.390***	0.393***	0.387***	0.405***	0.413***
	(0.070)	(0.071)	(0.070)	(0.071)	(0.073)
ln(Age)$_t$	0.146	0.122	0.148	0.137	0.125
	(0.103)	(0.103)	(0.102)	(0.103)	(0.104)
Observation	1074	1059	1066	1048	1035
Pseudo R^2	0.338	0.342	0.336	0.337	0.341
Industry F.E.	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes

Note. Figures in brackets are robust standard error. * p<0.10, ** p<0.05, *** p<0.01.

Table 13-b Data sharing and the number of new suppliers

Dep. Var.: ln(# NEW suppliers$_t$+1)	OLS				
Year=2015~2016	(1)	(2)	(3)	(4)	(5)
ln(# supplier$_{t-1}$+1)	0.592***	0.598***	0.596***	0.589***	0.596***
	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)
1 (data sharing)	0.046				
	(0.036)				
1 (data sharing with supplier)		0.056			0.03
		(0.039)			(0.051)
1 (data sharing with customer)			0.044		0.018
			(0.037)		(0.048)
1 (data sharing with others)				0.071	0.043
				(0.061)	(0.064)
ln(EMP)$_{t-1}$	0.150***	0.148***	0.149***	0.156***	0.152***
	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
ln(Age)$_t$	-0.126***	-0.132***	-0.129***	-0.129***	-0.136***
	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)
Observation	1109	1094	1101	1091	1078
Adjusted R^2	0.749	0.75	0.749	0.748	0.75
Industry F.E.	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes

Note. Figures in brackets are robust standard error. * p<0.10, ** p<0.05, *** p<0.01.

5. Conclusion

This paper uses the data of BDS, linked with TSR data of inter-firm transactions, to investigate the relationship between the structure of supplier and customer (business partner) networks and the big data use and sharing with these business partners. It is found that, in general, the number of suppliers is positively correlated with the likelihood of internal use of data and its sharing with suppliers, customers, and other third-party firms. On the contrary, the number of customers is negatively correlated with the data use and sharing. This negative effect of the number of customers is more prominent in the data sharing with customers. It may be because firms with more customers usually produce more general product. which may not be controlled for fully with industry dummy variables and product characteristics variables. The estimation results also show that long-term relationship with suppliers contribute negatively to the data sharing, but that with customers positively contributes to the data sharing with customers. Interestingly, the more customers a firm's suppliers have, or the more suppliers a firm's customers have in her transaction network, the less likely the firm shares big data with other third-party firms.

We also investigate the impact of big data use and firm's performance in terms of intensive and extensive margin. We find that the data sharing has positive and significant impact on firm's productivity. However, we partially find a weak contribution of data sharing with customers to pioneering new customers. We do not find any other effect of data sharing on the extensive margin of transaction.

These results imply that the expected negative impact of data sharing should be taken into account when making a rule for the data sharing in the economy. Even though data sharing contributes to the productivity positively in general, firms with complicated network may be reluctant to share the big data. Rules that contribute to both sides of supplier and customer should be carefully designed to deal with such complex interests between firms.

There are large numbers of studies addressing firm level information network use and firm's performance (Aral et. al, 2007: Motohashi, 2007; Shin, 2000). In contrast to most of existing literature ignoring the contents communicated though the network, this paper takes into account the information whether the data contents are shared with its supply chain partners, to see its impact on firm performance. However, there are some possible venues in terms of further investigation of this issue.

First, the productivity increase can be achieved either by cost reduction or by output increase. A typical management impact of digitalization of supply chain is inventory reduction by timely delivery (acceptance) of materials and parts. However, information of manufacturing process can be used for new product developments (Nguyen et. al, 2018). The productivity increase comes from cost reduction in the former case, while the latter case is an example of output increase. These two factors behind productivity increase can be disentangled by using the detail data from BDS.

Second, TSR customer-supplier information can be further broken down. For example, only numbers of customers and suppliers are used at current version, but bargaining power in supply chain relationship depends on relative size of firms. In addition, it is possible to look more detail in the type of industry of suppliers and customers to see the organization of supply chain structure clearer. Finally, the relationship between the dynamics of business partners and data sharing should be further investigated, since the key question regarding Industrie 4.0 is whether digitalization of manufacturing process leads to flexibility in supplier selection or close coordination between business partners.

References

Aral, E. Brynjolfsson, M. van Alstyne (2007). Productivity Effects of Information Diffusion in Networks, D-space@MIT, MIT, Cambridge MA

Brynjolfsson, E. (2014), The Second Machine Age: Work, Progress and Prosperity in Time of Brilliant Technologies, W. W. Norton & Company

Dow, J. K., & Endersby, J. W. (2004). Multinomial probit and multinomial logit: a comparison of choice models for voting research. *Electoral studies*, 23(1), 107-122.

Dyer, J. H. (1996), Specialized supplier networks as a source of competitive advantage: evidence from the auto industry, *Strategic Management Journal* 17(1996); 271-291

Gunasekaran, A. T. Papadopoulos, R. Dubey, S. F. Wamba, S. J. Childe, B. Hazen, and S. Akter (2017), Big data and predictive analytics for supply chain and organizational performance, Journal of Business Research 70(2017), 308-317

Janssen, M., H. can deer Voort and A. Wahyundi (2017), Factors influencing big data decision making quality, Journal of Business Research 70 (2017), 338-345

Motohashi, K (2017), Survey of Big Data Use and Innovation in Japanese Manufacturing Firms, RIETI Policy Discussion Paper, 17-P-027, August 2017.

Motohashi, K (2007), Firm-level analysis of information network use and productivity in Japan, Journal of The Japanese and International Economies 21(1) :121-137

Nguyen, T., L. Zhou, V. Spiegler, P. Ieromonachou and Y. Lin (2018), Big data analytics in supply chain management: A state of the art literature review, Computers and Operations Research 98 (2018), 254-264

Sako, M. (1991), The role of “trust” in Japanese buyer-supplier relationships, *Ricerche Economiche*, XLV: 449-474

Shin, I. (2001). Use of Information Network And Organizational Productivity: Firm-Level Evidence In Korea, Economics of Innovation and New Technology, 9(5): 447-646

Vendrell-Herrero, F., Bustinza O., Parry, G. and N. Georgantzis (2017), Servitazation, digitalization and supply chain interdependency, International Marketing Management, 60(2011) 69-81

VDMA. (2016). Guideline Industrie 4.0：Guideline Principles for the implementation of Industrie 4.0 in small and medium sized businesses. VDMA Verlag. ISBN 978-3-8163-0687-0

Wamba, S.F, A. Gunasekaran, S. Akter, S. J. Ren, R. Dubey and S. J. Childe (2017), Big data analytics and firm performance: effects of dynamic capabilities, Journal of Business Research 70 (2017), 356-365

Appendix

Table A1. Summary statistics

	Variable	N	Mean	SD	Med.	Min.	Max.
【1】	Data sharing	5,614	0.49	0.5	0	0	1
【2】	Data sharing with supplier	5,535	0.34	0.47	0	0	1
【3】	Data sharing with customer	5,574	0.4	0.49	0	0	1
【4】	Data sharing with third party firms	5,524	0.1	0.31	0	0	1
【5】	ln(Sales/# employee)	5,411	7.88	0.88	7.86	2.3	13.13
【6】	# suppliers	5,414	91.95	389.69	17	1	5276
【7】	# customers	5,384	74.35	324.73	16	1	7997
【8】	ln(# suppliers+1)	5,414	3.12	1.37	2.89	0.69	8.57
【9】	ln(# customers+1)	5,384	3.09	1.31	2.83	0.69	8.99
【10】	# longterm suppliers/total suppliers	4,903	0.48	0.22	0.47	0	1
【11】	# longterm customers/total customers	4,877	0.49	0.23	0.5	0	1
【12】	ln(average # customer of the suppliers)	5,414	99.97	175.35	48.32	1	2620.6
【13】	ln(average # supplier of the customers)	5,384	284.72	448.99	123.49	1	6080.67
【14】	ln(# employee)	5,439	4.75	1.65	4.7	0	10.62
【15】	ln(K/L)	3,937	9.09	1.28	9.25	-0.43	12.98
【16】	ln(firm age)	5,456	3.68	0.72	3.89	0	4.84

Table A2. Correlation table

	Variable	【1】	【2】	【3】	【4】	【5】	【6】	【7】	【8】	【9】	【10】	【11】	【12】	【13】	【14】	【15】	【16】
【1】	Data sharing	1															
【2】	Data sharing with supplier	0.7256	1														
【3】	Data sharing with customer	0.8188	0.5977	1													
【4】	Data sharing with third party firms	0.349	0.1757	0.1515	1												
【5】	ln(Sales/# employee)	0.1511	0.2077	0.1669	0.0103	1											
【6】	# suppliers	0.1667	0.1201	0.1372	-0.0038	0.1949	1										
【7】	# customers	0.1296	0.0684	0.1511	0.0433	0.2011	0.5661	1									
【8】	ln(# suppliers+1)	0.2447	0.2249	0.244	0.0365	0.5361	0.6004	0.4471	1								
【9】	ln(# customers+1)	0.1466	0.1673	0.164	0.0414	0.5572	0.4963	0.5242	0.8385	1							
【10】	# longterm suppliers/total suppliers	-0.0151	-0.0256	-0.0218	0.0121	-0.0725	-0.0044	-0.0568	-0.0595	-0.0367	1						
【11】	# longterm customers/total customers	0.0605	0.0751	0.0831	-0.0089	0.1017	0.0144	0.0111	0.106	0.0941	0.4222	1					
【12】	ln(average # customer of the suppliers)	-0.1108	-0.1222	-0.1209	0.0282	-0.1347	-0.1112	-0.0837	-0.2683	-0.1938	0.0181	-0.0314	1				
【13】	ln(average # supplier of the customers)	-0.0366	-0.0515	-0.0038	-0.0619	-0.2111	-0.1152	-0.1245	-0.2747	-0.3553	0.0226	-0.0507	0.0335	1			
【14】	ln(# employee)	0.2234	0.2032	0.2377	-0.0119	0.4116	0.5398	0.3847	0.8962	0.7434	-0.0103	0.1357	-0.2725	-0.222	1		
【15】	ln(K/L)	0.074	0.0801	0.1016	-0.0137	0.3872	0.1123	0.099	0.3895	0.3451	-0.0099	0.1131	-0.0571	-0.164	0.3433	1	
【16】	ln(firm age)	0.1155	0.1212	0.1277	-0.0119	0.3108	0.16	0.0268	0.5278	0.4508	0.0391	0.1986	-0.1881	-0.211	0.5291	0.4472	1