A Model of Pricing Data and Their Constituent Variables Traded in Two-Sided Markets with Resale: A Subject Experiment

Toshihiko Nanba[a], Kazuhito Ogawa[b], Naoki Watanabe[c], Teruaki Hayashi[d], Hiroki Sakaji[d]

[a]KUAS (econ), [b]Kansai Univ. (socio), [c]Keio Univ. (business), [d]U. of Tokyo (eng)

August 13, 2023
Economic Theory Workshop in Morioka

Part of this work was published in Proceedings of 2022 IEEE Int'l Conference on Big Data from IEEE Xplore in January 2023.

These slides are available at my website.

- First, search for my website using naoki watanabe and keio as keywords. Then, find the url of (presentation slide) at [2] in <Publications Written in English>.

  `http://labs.kbs.keio.ac.jp/naoki50lab/index.html`

- Alternatively,

  `http://labs.kbs.keio.ac.jp/naoki50lab/DataExPresentation3.pdf`

Imagine the following data service,
as a simple familiar example.

Creating a new market for used digital communication devices

- various factors (variables) that determine the price of a used device
  - in-store or online retail prices and specifications of new products
  - transaction information in online auctions of second-hand products, including damage, network restrictions, and so on
  - others
- a set of “processed and summarized” variables (data) of a used device specified for each user

1. Introduction: Research Question

- Each instance of data consists of several variables.
  - variable: address, gender, # of households, ... (for real-estate markets)
  - data: a set these variables constitute that makes sense when used for a given purpose
- Data and variables are easily replicable. $\Rightarrow$ not scarce $\Rightarrow$ difficult to put a price on variables and data

How can we price the data and its constituent variables?

1. Introduction: Major Features

a new model from the perspective of cooperative games

1. Muto and Nakayama (1992): resale of information (without production)
2. Watanabe and Muto (2008): disadvantage arising due to not owning information (under patent protection)

The prices of variables are exogenously set at the initial round, and they are updated for the next round based on the outcomes of trades made in the current round.

- Conjecture: backward induction $\Rightarrow$ prices never move in any rounds that proceed after the initial round.
- There is no budget constraint on subjects' decision-making. (We conducted additional sessions under budget constraints recently.)

1. Introduction: Main Results

- The prices of data and the constituent variables fluctuated over rounds (probably due to no budget constraint), but
  the average prices of those variables were not far from the initial values we set theoretically.
- No significant difference in amounts of producer surplus, regardless of whether subjects were informed of values of gross profits trades generated (whether the information was complete or incomplete).
  - Real trades are conducted with incomplete information.
  - Theoretical inference of this work was derived with complete information.

2. Model

- 2 instances of data, each consisting of 3 items of variables, traded in 12 rounds ($r = 0, 1, 2, \ldots, 12$) by 4 traders.
- At each round, the prices of the variables are given to traders and updated for the next round immediately at the end of the round, based on the outcomes of trades.
  - $p_i(r)$: price of variable i in round r
  - replicability $\Rightarrow$ Traders purchase one unit of the variable.
  - The production cost of data is the sum of the expenditure for the constituent variables.
  - $t_j(r)$: price of data j in round r
  - replicability $\Rightarrow$ Users of data j purchase one unit of the data.
  - no budget constraints

- the gross profits for an individual user and those for a non-user of the data satisfy the following relationship:

$$\begin{aligned} W_j(1) \geq W_j(2) \geq \cdots \geq W_j(n) > L_j(0) \\ \geq L_j(1) \geq L_j(2) \geq \cdots \geq L_j(n-1). \end{aligned} \quad (1)$$

  When there are s owners of data j, a user obtains $W_j(s)$ while a non-user obtains $L_j(s)$. (disadvantages for non-users)
- The data can be replicated and resold freely.
  - each owner of the data proposes the price and the number of instances of the data, and
  - traders who do not own the data may purchase the data at the proposed price.
- $t_i(r)$ is the price at which the initial owner sells. The resale process stops for all data $\Rightarrow$ transactions proceed to the next round.
- Any data expire for the use within each round in order to avoid dynamic competition among instances of data at the markets.

- $\#$ of users of data j ($j = 1, 2$) $\Rightarrow$ demand for variable i
- $q_i(r)$: threshold for variable i in round r. This threshold value is used for updating its price $p_i(r)$ that is set at the market in the next round.
- The initial values of $q_i(0)$ and $p_i(0)$ are exogenously given.
- Demand for variable i in round t exceeds $q_i(t) \Rightarrow$ the price in round $t+1$ is updated to $p_i(t+1) = p_i(t) + a_i$. The threshold in round $t+1$ is also updated to $q_i(t+1) = q_i(t) + b_i$.
- Demand falls below $q_i(t)$. $\Rightarrow$ $p_i(t+1) = \max(0, p_i(t) - a_i)$ and $q_i(t+1) = \max(0, q_i(t) - b_i)$.

  $\cdots$ a_i and b_i are positive constants exogenously given.

3. Experimental Design: Gross Profits

Data 1 and Data 2 are traded.

The gross profits are specified as follows.

Data 1: $W_1(1) = 200$, $W_1(2) = 150$, $W_1(3) = 70$, $W_1(4) = 50$,
$L_1(0) = 40$, $L_1(1) = 30$, $L_1(2) = 20$, $L_1(3) = 10$.

Data 2: $W_2(1) = 200$, $W_2(2) = 120$, $W_2(3) = 70$, $W_2(4) = 50$,
$L_2(0) = 40$, $L_2(1) = 30$, $L_2(2) = 20$, $L_2(3) = 10$.

- complete information environment: the values of gross profits are shown to the subjects.
- incomplete information environment: subjects are informed of the orders of those gross profits. In practice, traders have to estimate those values.

3. Experimental Design: Demands for Variables

$\#$ of users of data j ($j = 1, 2$) $\Rightarrow$ demand for variable i

First, find the largest integer e such that $2 \leq e \leq 3$ and

$$W_j(e) \geq W_j(4) + (4 - e)(W_j(4) - L_j(e)).$$

- When $e = 3$, $W_j(3) = 70$ and $W_j(4) + (4 - 3)(W_j(4) - L_j(3)) = 90$. Namely, when there are 3 data holders, each one obtains 70 but he can obtain 90 by reselling the data, but the amount is at most 90.
- When $e = 2$, $W_1(2) = 150$ and $W_2(2) = 120$, but $W_j(4) + (4 - 2)(W_j(4) - L_j(2)) = 110$ for $j = 1, 2$. Namely, when there are 2 data holders, each data holder does not have an incentive to resell the data.

The solution of the above problem is thus $e = 2$.

Second, consider a situation in which there are 2 data holders.

- If one of the data holders resells to a trader, then he can receive $W_j(3) - L_j(2) = 50$ for Data $j = 1, 2$ from the trader.
- In the case of $e = 3$, however, another resale is expected to occur, as shown previously. Thus, the data holders can eventually obtain at most $W_j(3) - L_j(2) + W_j(4) + (W_j(4) - L_j(3)) = 140$ for Data $j = 1, 2$.
- In the case of Data 1, any resale should not occur when $e = 2$, because $W_1(2) = 150$.
- Even in the case of Data 2, any resale should not occur when $e = 2$; Truly, $W_2(2) = 120$, but when $e = 3$, resale should be competitive among 3 resellers and thus the resale price would decrease to the smallest possible monetary unit; $W_2(3) - L_2(2) + W(4) = 100$.

Resale does not occur when there are 2 data holders.

Finally, compare the case of $e = 2$ and $e = 1$.

- If the initial holder of Data 1 does not sell the data, then he obtains $W_1(1) = 200$, but he can obtain
  $W_1(2) + (2-1)(W_1(2) - L_1(1)) = 270$
  through monetary transfer by selling the data to a trader.
- Similarly, consider the case of Data 2. $W_2(1) = 200$, but he can obtain $W_2(2) + (2-1)(W_2(2) - L_2(1)) = 210$
  through monetary transfer by selling the data to a trader.

According to the backward induction, 2 data holders should not resell the data to any traders.

3. Experimental Design: Note

Muto and Nakayama (1992) showed that these outcomes are in a variant of Bargaining Set recurrently defined in a dynamic context, when

$$\begin{aligned} W_j(1) \geq W_j(2) \geq \cdots \geq W_j(n) > L_j(0) \\ = L_j(1) = L_j(2) = \cdots = L_j(n-1) = 0, \end{aligned} \tag{2}$$

and one instance of data is traded.

3. Experimental Design: Social Welfare

Compute the maximal amounts of producer surplus. (The payments are cancelled between payers and recipients, when summing up their profits.)

- Data 1: the total sum of profits (producer surplus) is maximized at $e = 2$; $2W_1(2) + 2L_1(2) = 2 * 150 + 2 * 20 = 340$.
- Data 2: it is maximized at $e = 1$;
  $W_2(1) + 3L_2(3) = 200 + 3 * 30 = 290$.
  $(2W_2(2) + 2L_2(2) = 2 * 120 + 2 * 20 = 280)$

  Therefore, in terms of "social welfare" of traders, the initial holder of Data 2 should not sell Data 2 to any users.

3. Experimental Design: Theoretical Predictions

- payoff maximization: The initial holders of Data 1 and Data 2 should sell their data to a trader, respectively. Resale should not occur.
- social welfare: It is maximized when there are 2 data holders in the case of Data 1, while it is maximized when the initial holder of Data 2 does not sell the data to any traders.

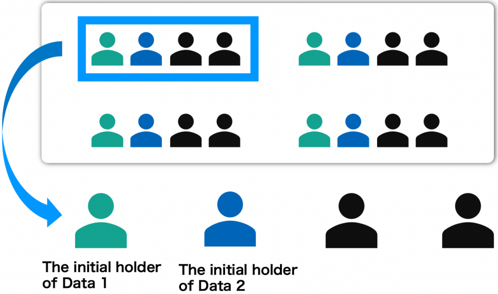
3. Experimental Design: Session Details

We conducted 4 experimental sessions, each consisting of 12 rounds.

	comp xy	comp xz	incomp xy	incomp xz
Data 1	x , y , z	x , y , z	x , y , z	x , y , z
Data 2	x , y	x , z	x , y	x , z
information	complete	complete	incomplete	incomplete

- The initial prices of variables x, y and z : $p_1(0) = 10$, $p_2(0) = 20$, $p_3(0) = 30$.
- $q_1(0) = q_2(0) = q_3(0) = 2$ for Data 1, and $q_1(0) = q_2(0) = 2$ or $q_1(0) = q_3(0) = 2$ for Data 2 (by the previous backward induction)
- The constants for updating those prices and thresholds are $a_1 = a_2 = 10$ and $b_1 = b_2 = 10$, respectively.

In every session, 16 subjects divided into 4 groups of 4 subjects at the beginning of the session.



The initial holder of each data is determined randomly and does not change in all rounds.

- Data 1 $\Rightarrow$ Data 2: The transactions of Data 2 proceed in a manner similar to those for Data 1.
- After the transactions for Data 1 and Data 2, the price of the variables are updated.
- Then, the transactions for the next round start in a similar manner.

Resale can be made at most twice.

The payoff for the initial holder of Data j = (gross profit obtained from Data j − prices of Data j's constituent variables + selling price of Data j + total resale price of Data j) + (gross profit obtained from Data i - purchase price of Data i + total resale price of Data i),

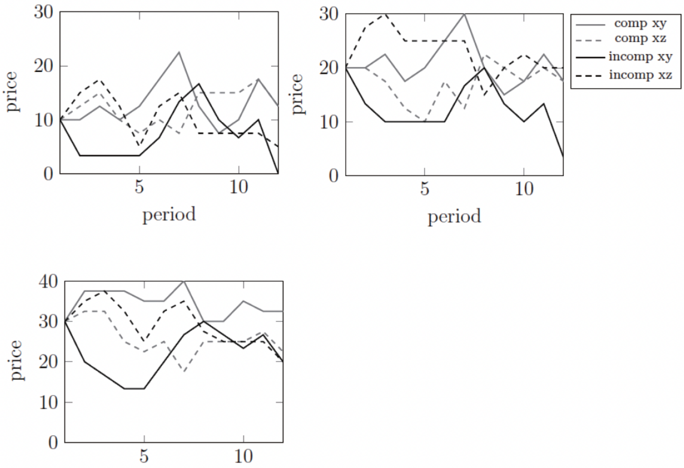
where $i = 1, 2$ and $i \neq j$.

The payoff for a trader who is not the initial holder = gross profit obtained from Data 1 − buying price of Data 1 + total resale price of Data 1 + gross profit obtained from Data 2 − buying price of Data 2 + total resale price of Data 2.

4. Result: Prices of Variables

Result 1

Without budget constraints, the prices of variables fluctuated, but for any $i = x, y, z$, the averages are not far from its initial price $p_i(0)$.



session	variable x	variable y	variable z
comp xy	12.917 (10.097)	20.625 (12.949)	34.375 (7.118)
comp xz	12.292 (11.530)	17.292 (8.688)	25.833 (13.657)
incomp xy	10.417 (8.742)	16.875 (10.750)	25.000 (11.109)
incomp xz	10.208 (7.852)	22.917 (8.495)	29.167 (9.187)

Table: Average prices of variables. No budget constraints. Values in parentheses represent standard deviations

Note: The initial prices of variables are set as $p_1(0) = 10$, $p_2(0) = 20$, and $p_3(0) = 30$.

session	variable x	variable y	variable z
comp xy	8.889 (9.428)	13.472 (9.952)	23.472 (11.279)
comp xz	9.500 (8.719)	22.167 (7.831)	25.000 (12.003)
incomp xy	13.646 (10.966)	21.458 (12.813)	29.271 (11.262)
incomp xz	12.708 (9.394)	22.292 (6.270)	29.583 (12.831)

Table: Average prices of variables. Under budget constraints. Values in parentheses represent standard deviations.

Note: The initial prices of variables are set as $p_1(0) = 10$, $p_2(0) = 20$, and $p_3(0) = 30$.

4. Result: Prices of Data

Result 2

Without budget constraints, for both Data 1 and Data 2, there are some rounds in which soaring prices of data were observed in a session.

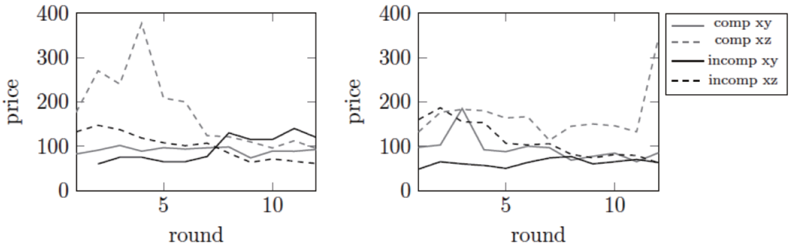


Fig. 2: Time series of prices of data determined by the initial data holder.

Note: The graphs on the left-hand (right-hand) side represent the time series of prices of Data 1 (Data 2).

session	Data 1	Data 2	p-value
comp xy	90.644 (39.448)	91.097 (44.582)	0.632
comp xz	144.293 (69.744)	150.119 (66.961)	0.7762
incomp xy	91.176 (32.908)	62.581 (20.036)	$<$**0.001**
incomp xz	100.404 (57.896)	110.220 (56.987)	0.214

Table: Average prices of data offered by the initial data holders. No budget constraints

Note: Transactions were not made at extraordinarily high prices. Those prices are excluded in computing the average listed above. The p-values for the Brunner-Munzel test are also listed, where the null hypothesis is that the prices are on average the same between Data 1 and Data 2.

session	Data 1	Data 2	p-value
comp xy	95.758 (38.793)	93.111 (51.899)	0.395
comp xz	92.889 (39.910)	83.618 (21.500)	0.694
incomp xy	101.961 (37.197)	92.978 (40.622)	0.089
incomp xz	78.450 (26.330)	79.568 (34.506)	0.574

Table: Average prices of data offered by the initial data holders. Under budget constraints

Note: The p-values for the Brunner-Munzel test are also listed.

4. Result: Social Welfare 1

Result 3

Without budget constraints, the efficiency rates observed in the incomplete information environment were not significantly different from those observed in the complete information environment.

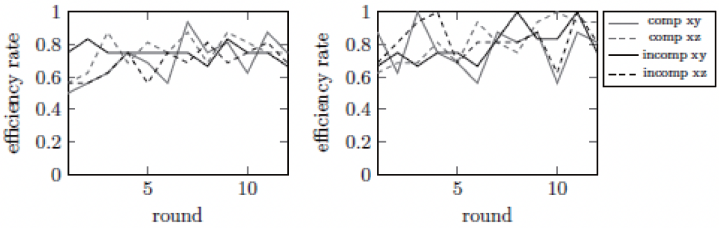


Fig. 3: Time series of average efficiency rates.
Note: The graph on the left-hand (right) side stands for the efficiency rates generated in trades of Data 1 (Data 2).

	Data 1	Data 2	n
comp xy	0.813 (0.174)	0.915 (0.122)	48
comp xz	0.849 (0.145)	0.940 (0.105)	48
incomp xy	0.820 (0.131)	0.911 (0.118)	48
incomp xz	0.797 (0.159)	0.937 (0.109)	48

Table: Average efficiency rates without budget constraints

	Data 1	Data 2	n
comp xy	0.827 (0.120)	0.924 (0.114)	72
comp xz	0.829 (0.172)	0.940 (0.107)	60
incomp xy	0.824 (0.152)	0.886 (0.122)	96
incomp xz	0.860 (0.160)	0.881 (0.128)	48

Table: Average efficiency rates under budget constraints

There would be a significant difference in efficiency rates for Data 2.

4. Result: Social Welfare 2

Result 4

There was no significant difference in efficiency rates between Data 1 and Data 2, except in a session. (No budget constraints)

Table III shows the results of a two-way ANOVA.

sessions pooled	complete info	Data 1	rounds
comp xy and incomp xy	0.949	**0.043**	
	0.948	**0.039**	0.089
comp xz and incomp xz	0.502	**0.002**	
	0.496	**0.002**	0.130

TABLE III: p-values for a two-way ANOVA.

Note: The dependent variable is arcsin(sqrt(efficiency rate)). Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

negative test results: Normality of samples in some sessions and homoscedasticity of samples in Data 2. $\Rightarrow$ Brunner-Munzel test

	Data 1	Data 2	p-value	n
comp xy	1.102	1.211	0.187	48
incomp xy	1.091	1.214	0.109	36
p-value	0.637	0.992		
comp xz	1.145	1.260	0.241	48
incomp xz	1.058	1.276	**0.003**	48
p-value	0.147	0.808		

TABLE IV: Average transformed efficiency rates and p-values for the Brunner-Munzel test.

Result 5

Social welfare declines as the frequency of resale increases. (No budget constraints)

	resales	Data 1	Data 2
comp xy	0	1.272 (n=35)	1.336 (n=39)
	1	0.681 (n=10)	0.711 (n=7)
	2	0.524 (n=3)	0.524 (n=2)
p-value		$p <$**0.001**	$p <$**0.001**
incomp xy	0	1.129 (n=32)	1.309 (n=30)
	1	0.785 (n=4)	0.785 (n=5)
	2	——- (n=0)	0.524 (n=1)
p-value		**0.003**	**0.003**
comp xz	0	1.268 (n=38)	1.346 (n=42)
	1	0.750 (n=7)	0.720 (n=4)
	2	0.524 (n=3)	0.524 (n=2)
p-value		$p <$**0.001**	$p <$**0.001**
incompxz	0	1.167 (n=37)	1.371 (n=42)
	1	0.707 (n=10)	0.698 (n=3)
	2	0.524 (n=1)	0.524 (n=3)
p-value		$p <$**0.001**	$p <$**0.001**

TABLE V: Averages of transformed efficiency rates and p-values of the Kruskal-Wallis test.

5. Final Remarks: For Future Research

- Conducting additional treatment.
  - Result 4 may imply that the order of trading Data 1 and Data 2 affected the difference in efficiency rates.
  - more realistic budget constraints
- Extension to a more general model.
  - We need to introduce asymmetric gross profits in to our model.

Are the outcomes derived by the backward Induction in the Bargaining Set defined in the spirit of Muto and Nakayama (1992)?

Ads: 2023 IEEE Int'l Conference on Big Data

- Date: 15-18 December, 2023, Venue: Sorrento, Italy
- http://bigdataieee.org/BigData2023/
- Int'l Conference (acceptance rate = 18%),
  Special Sessions (acceptance rates = 30-40%),
  Workshops (acceptance rates = 40-80(?)%)

WS: Large-scale Data Utilization in Economics of Information and Management Sciences: Theory, Computation, and Experiment

http://labs.kbs.keio.ac.jp/naoki50lab/Workshop_IEEE_BigData2023.pdf