

Study on Data Supply Chain Pricing Decision Considering Network Externalities

Zhicheng Zhang^{1,*}, Yiming Zhang², Weibin Deng¹, Guangming Liu³

¹ School of Economics and Management, Chongqing University of Posts and Telecommunications, CO 400065, China

² 78111 troops, People's Liberation Army of China, CO 409626, China

³ Chongqing Branch of China Tower Corporation Limited, CO 401120, China

* Corresponding author: Zhang Zhi-Cheng (Email: 1079169190@qq.com)

Abstract: Nowadays, data has become an important productive factor in the world. Data as an asset to management and realizing its value are important factors for its flowing smoothly. However, how to price the data scientifically and reasonably has been a difficult and key problem for the data element market development. After analyzing the transaction process of data products from data resource acquisition to production and sales, a three-level data supply chain consisting of data providers, data trading platforms and users is constructed. In this chain, government subsidies for data trading platforms are considered. Based on that, a centralized and a decentralized differential game pricing model of data supply chain under the influence of network externalities are proposed respectively. The impact of network externalities and government subsidy elasticity coefficients, among others, on the two decision models is analyzed from a theoretical perspective.

Keywords: Data pricing, data supply chain, network externalities, government subsidies, differential game.

1. Introduction

With the widespread popularization and application of big data, the value of data resources is increasing. So far, dozens of data trading organizations have been established across the country, but compared with traditional resources, due to the intangible, variable, social and shared nature of data resources, resulting in a market with insufficient participants in data trading, low frequency of transactions, and small scale of transactions. Currently, research on data supply chain-based pricing strategies has made some progress in data supply chain operation, considering data quality and scale pricing. First, the research on the operation of data supply chain. Chen et al [1] optimized the profit of a two-level supply chain by constructing a Nash equilibrium bargaining model to achieve the global optimality of the supply chain. Barman et al [2] used the Stackelberg game model to evaluate the centralized and decentralized decision-making of the three-level green supply chain, and compared the optimal pricing of the supply chain with and without government subsidies. Pricing. Secondly, there are studies on data pricing considering data quality and scale. Batini et al [3] summarized data quality as accuracy, completeness, redundancy, readability, accessibility, consistency, usefulness, and trustworthiness. Taleb et al [4] proposed a new data quality framework that uses predefined quality metrics to accomplish more effective data quality assessment.

However, in actual data supply chain pricing is not solely influenced by data characteristics. From the market level, consumers' willingness to purchase data products is not only affected by their own needs and product factors, but also related to the size of the users who purchase the products, i.e., data products in the data supply chain have network externalities. Network externality is generated by the economies of scale of users and refers to the fact that the value of a product to its users increases with the increase in the number of consumers adopting the same product [5]. Tang et al [6] constructed a pricing model for the competition

between newcomer platforms and incumbent platforms by using the Hotelling model based on the consideration of cross-network externalities, and explored the future strategic choices of the two platforms in order to obtain more benefits. Zhu et al [7] explored the impact of the magnitude of network externalities on product quality, provider implementation mode and their profits.

For the node enterprises of data supply chain, how to improve the quality of data products and at the same time how to make reasonable data pricing decisions is a very challenging problem, most of the previous literature on data supply chain pricing focuses on the supply chain's own operation and pricing, and fails to consider the network externality and the government's subsidies for data trading platforms. Therefore, the innovations of this paper are reflected in the following points: (1) network externalities as a decision variable in the data pricing model, and further analyze its impact on the operation of the data supply chain, which is to make the model and the conclusion of the study more in line with the status quo of the development of the network economy. (2) In the existing research on data supply chain pricing, little consideration has been given to the impact of government subsidies on the pricing of data products, because in the actual operation of the data supply chain, government subsidies can play the role of incentives for data trading platforms to innovate, so as to enable them to develop higher-quality data products and promote the development of the data supply chain.

2. Problem Description and Assumptions

In this paper, we consider a data supply chain with monopoly position consisting of a data provider and a data trading platform. Among them, the data provider obtains raw data of quality q_1 and size n through a reward mechanism, removes personal attributes through technical processing and then sells them to the data trading platform at price P_1 . The

data trading platform produces data products by processing the raw data obtained through its own technical resources, and then sells them to users at a certain price P_2 . The government, in order to promote the technological innovation of the data trading platform, will subsidize the platform according to the quality of the data products it produces at a certain price P_3 . In the operation of this data supply chain, both data providers and data trading platforms incur a certain amount of data processing costs, and they need to consider their own revenues and costs in order to maximize profits. In order to illustrate more clearly the study of profit maximization decision-making of data supply chain under the consideration of network externality, the relevant symbols involved in the model construction are first defined, as shown in Table 1.

Table 1. Description of relevant symbols

Symbol	Definition
P_1	Price of unit quantity of data purchased from data providers by data trading platforms
P_2	Price of data products developed by the data trading platform
P_3	Cost of government subsidies to encourage technological innovation by data trading
q_1	Quality of raw data acquired by data providers
q_2	Quality of data products developed by data trading platforms
n	Size of raw data acquired by data providers
α	Cost coefficient of raw data acquired by the data provider, $\alpha \in (0, 1)$
k_1	Cost coefficient of raw data processing by the data provider, $k_1 \in (0, 1)$
k_2	Coefficient of data product development cost of the data trading platform, $k_2 \in (0, 1)$
u	Network externality, $u \in (0, 10)$
a	User market size
θ	Elasticity coefficient of the selling price of data products, $\theta \in (0, 1)$
w	Elasticity coefficient of the cost of subsidies given to the platform by the government, $w \in$
Q	Market demand for the data product
π_1	Profit function of the data provider
π_2	Profit function of the data trading platform
π_c	Total profit function of the data supply chain

Assumption 1 Data providers need to pay a certain cost when acquiring raw data, the more cost they pay, the larger the scale and quality of raw data they acquire, so this paper assumes that the cost of raw data acquisition for data providers is $\alpha q_1^2 n$, where α denotes the cost coefficient of raw data acquisition for data providers.

Assumption 2 The improvement of raw data quality and scale helps data trading platforms to develop higher quality data products, so this paper assumes that the relationship function of q_2 , q_1 and n is:

$$q_2 = q_1 n \quad (1)$$

Assumption 3 The subsidy cost of the government to encourage technological innovation of the data trading platform and the quality of data products is a linear relationship between the higher the quality of data products obtained by the user, the more government subsidies, but the subsidy is not an infinite increase, and the relationship function is:

$$P_3 = w q_2 = w q_1 n \quad (2)$$

Assumption 4 Assume that end users will only purchase data products on this one data trading platform, and the product production of the data trading platform is the market demand. In order to consider the impact of network externalities on the data supply chain, the demand function of the market is defined as:

$$Q = (1 + u)a - \theta P_2 \quad (3)$$

where u is the network externality, a is the end-user size, and θ is the elasticity coefficient of the sales price of data products.

Assumption 5 The profits of the data provider and the data trading platform are π_1 and π_2 , respectively, and the total profit of the data supply chain is π_c . When the data scale increases, the data processing cost and the product development cost also increase, and the growth rate increases gradually, so this paper assumes that the data processing cost of the provider and the product development cost of the platform are $k_1 n^2$ and $k_2 n^2$, respectively, where k_1 denotes the original data processing cost coefficient of the data provider, and k_2 is the data product development cost coefficient of the data trading platform.

The profit function of the data provider is:

$$\pi_1 = P_1 n - \alpha q_1^2 n - k_1 n^2 \quad (4)$$

The data trading platform receives government subsidies by producing high-quality data products and also earns profits by selling data products. The profit function of data trading platform is:

$$\pi_2 = P_2 Q - P_1 n - k_2 n^2 + w q_1 n \quad (5)$$

The total profit function of the data supply chain is:

$$\pi_c = \pi_1 + \pi_2 \quad (6)$$

When the size of raw data is too large, data providers and data trading platforms will give up the processing of raw data and the development of data products because the cost is higher than the benefit.

3. Model Construction and Analysis

3.1. Pricing model analysis under centralized decision-making

Under the centralized decision-making model, this paper is based on the hypothetical three-stage data supply chain, in which data providers charge data trading platforms the subscription price of raw data P_1 , data trading platforms receive government subsidies for the production of data products, and ultimately data trading platforms charge end users the price of data products P_2 . This study establishes the transaction process between data providers and data trading platforms as a two-stage game. In this study, the transaction process between the data provider and the data trading platform is modeled as a two-stage game, where the first stage is for the data provider to decide the price of the original data P_1 , and the second stage is for the data trading platform to decide the price of the data product P_2 . The model can be expressed as follows.

$$\pi_c = \pi_1 + \pi_2 = P_2 Q - \alpha q_1^2 n - (k_1 + k_2)n^2 + wq_1 n = (1+u)aP_2 - \theta P_2^2 - \alpha q_1^2 n - (k_1 + k_2)n^2 + wq_1 n \quad (7)$$

Again taking the second order partial derivative with respect to P_2 yields:

$$\frac{\partial^2 \pi_c}{\partial P_2^2} = -2\theta < 0 \quad (8)$$

Since its second order partial derivative is less than zero, there exists an extreme value of π_c . The optimal pricing is achieved when the first order partial derivative is equal to zero, at this time:

$$P_2^* = \frac{(1+u)a}{2\theta} \quad (9)$$

Substituting equation (9) into π_c can be obtained:

$$\pi_c = \frac{(1+u)^2 a^2}{4\theta} - \alpha q_1^2 n - (k_1 + k_2)n^2 + wq_1 n \quad (10)$$

Then the second order partial derivatives can be obtained for q_1 and n respectively:

$$\frac{\partial^2 \pi_c}{\partial q_1^2} = -2\alpha n < 0, \quad \frac{\partial^2 \pi_c}{\partial n^2} = -2(k_1 + k_2) < 0 \quad (11)$$

One can find q_1 and n respectively:

$$q_1 = \frac{w}{2\alpha}, \quad n = \frac{w^2}{8\alpha(k_1+k_2)} \quad (12)$$

The maximum profit of the data supply chain is:

$$\begin{aligned} \max \pi_c &= \frac{a^2(1+u)^2}{4\theta} + \frac{w^4}{64\alpha^2(k_1+k_2)} \\ &= \frac{16a^2(1+u)^2\alpha^2(k_1+k_2)+w^4\theta}{64\theta\alpha^2(k_1+k_2)} \end{aligned} \quad (13)$$

Proposition 1 Raw data quality q_1 is inversely proportional to the data acquisition cost coefficient α and directly proportional to the government subsidy elasticity coefficient w . Raw data size n is inversely proportional to the coefficient α as well as to the data processing cost coefficients k_1 and k_2 , and directly proportional to the government subsidy elasticity fee coefficient w .

Proof: As can be seen from equation (12), when the coefficient w increases, the original data quality q_1 and the original data size n both increase; when the data acquisition cost coefficient α increases, the original data quality q_1 and the original data size n both decrease; when the data processing cost coefficients k_1 and k_2 both increase, size n decreases.

Proposition 2 Both π_c and the price of the data product P_2 are positively affected by network externalities.

Proof: From equation (9) and (13), and when the network externality is increased, the total profit of the data supply chain and the price of the data product P_2 are both increased.

Proposition 3 The total profit of the data supply chain is inversely proportional to the data acquisition cost coefficient α and the data processing cost coefficients k_1 , k_2 , and directly proportional to the elasticity coefficient of government subsidy w .

Proof: From equation (13), when the coefficient α and the cost coefficients k_1 , k_2 increase, the maximum total profit of the data supply chain decreases; when the government subsidy elasticity coefficient w increases, the maximum total profit of the data supply chain increases, so that proposition 3 is proved.

3.2. Pricing model analysis under decentralized decision-making

The pricing model under decentralized decision-making is based on the three-level data supply chain composed of data providers, data trading platforms and users, and a Stackelberg game model between data providers and data trading platforms is constructed by taking into account the original data transaction process between data providers and data trading platforms, and the data product transaction process between data trading platforms and users. The game sequence of the model is: the data provider first chooses the original data price P_1 to maximize its own revenue π_1 , and then the data trading platform chooses the data product sales price P_2 according to the P_1 set by the data provider to maximize its revenue π_2 . The model can be expressed as:

$$\max \pi_1 = P_1 n - \alpha q_1^2 n - k_1 n^2 \quad (14)$$

$$n(P_1) = \arg \max_{P_2, n} \{P_2 Q - P_1 n - k_2 n^2 + wq_1 n\} \quad (15)$$

This Stackelberg game model takes the data provider as the leader and the data trading platform as the follower, reflecting the important position of the owner of the data in the data supply chain. At the same time, the model correlates the size and quality of raw data with the cost of data providers and data trading platforms, and also considers government subsidies to data trading platforms, reflecting the impact of data characteristics on data product design and the government's role in promoting the data market in real life.

The profit function of the data trading platform is:

$$\pi_2 = P_2 Q - P_1 n - k_2 n^2 + wq_1 n \quad (16)$$

Taking the second order partial derivatives of π_2 for the data product subscription price P_2 and data size n , respectively, yields:

$$\frac{\partial^2 \pi_2}{\partial P_2^2} = -2\theta < 0, \quad \frac{\partial^2 \pi_2}{\partial n^2} = -2k_2 < 0 \quad (17)$$

It can be seen that the second order partial derivatives of both decision variables are less than zero, and there is a unique global optimal solution for π_2 . Let the first order partial derivatives be zero, it can be obtained:

$$P_1^* = wq_1 - 2k_2n, \quad P_2^* = \frac{(1+u)a}{2\theta} \quad (18)$$

Bringing equation (18) into (14) we can get π_1 is:

$$\pi_1 = (wq_1 - 2k_2n)n - \alpha q_1^2 n - k_1 n^2 \quad (19)$$

Then the second order partial derivatives of π_1 for the original data quality q_1 and the original data size n respectively can be obtained:

$$\frac{\partial^2 \pi_1}{\partial q_1^2} = -2\alpha n < 0, \quad \frac{\partial^2 \pi_1}{\partial n^2} = -4k_2 - 2k_1 < 0 \quad (20)$$

It can be seen that the second order partial derivatives of both decision variables are less than zero. Therefore, there is a unique global optimal solution for $\max \pi_1$. According to the fact that the first order partial derivatives are equal to zero, it can be obtained:

$$q_1 = \frac{w}{2\alpha}, \quad n = \frac{w^2}{8\alpha(k_1+2k_2)} \quad (21)$$

Therefore the original data price P_1^* is:

$$P_1^* = \frac{w^2}{2\alpha} \left[1 - \frac{1}{2\left(\frac{k_1}{k_2}+2\right)} \right] \quad (22)$$

At this point the profit function π_1 of the data provider is:

$$\pi_1 = wq_1n - \alpha q_1^2 n - (k_1+2k_2)n^2 = \frac{w^4}{64\alpha^2(k_1+2k_2)} \quad (23)$$

The profit function π_2 of the data trading platform is:

$$\pi_2 = (1+u)aP_2 - \theta P_2^2 - P_1n - k_2n^2 + wq_1n \\ = \frac{16a^2(1+u)^2\alpha^2(k_1+k_2)+w^4\theta}{64\theta\alpha^2(k_1+k_2)} \quad (24)$$

Proposition 4 The quality of raw data q_1 is inversely proportional to the cost coefficient of data acquisition α and directly proportional to the elasticity coefficient of government subsidy w . The size of raw data n is inversely proportional to the coefficient α as well as to the cost coefficients of k_1 and k_2 , and directly proportional to the elasticity coefficient of government subsidy w .

Proof: From equation (21), it can be seen that when the coefficient w increases, the original data quality q_1 and the original data size n both increase; when the data acquisition cost coefficient α increases, the original data quality q_1 and the original data size n both decrease; when the data processing cost coefficient k_1 and coefficient k_2 both increase, the original data size n decreases.

Proposition 5 Both the profit of data trading platform and the price of data products P_2 are positively affected by network externalities.

Proof: From equation (18) and (24), when the network externality is increased, the maximal profit of the data trading platform and the price of the data product P_2 are both increased.

Proposition 6 The profit of the data trading platform is

inversely proportional to the data acquisition cost coefficient α and the data processing cost coefficients k_1 and k_2 , and directly proportional to the elasticity coefficient of government subsidy w .

Proof: From equation (24), it can be seen that when the data acquisition cost coefficient α , and the data processing cost coefficients k_1 and k_2 are increased, the maximum profit of the data trading platform will be decreased, and when the coefficient w is increased, the maximum profit of the data trading platform will be increased, so proposition 6 is proved.

Proposition 7 The effects of the incentive coefficient α , the data processing cost coefficients k_1 , k_2 and the government subsidy elasticity coefficient w on the pricing of unit raw data P_1 are as follows.

(1) When the data acquisition cost coefficient α increases, P_1 decreases;

(2) P_1 decreases when the data processing cost coefficients k_1 and k_2 increase;

(3) The elasticity coefficient of the government subsidy w paid by users to the data trading platform increases, and P_1 increases.

Proof: From equation (22), when the coefficient α and the coefficients k_1 and k_2 increase, the original data pricing will decrease; when the government subsidy elasticity coefficient w increases, the original data pricing will increase, so proposition 7 is proved.

4. Conclusion

In this paper, in the data supply chain with data providers and data trading platforms as the main body, parameters such as network externality, government subsidies, data quality, data scale, data acquisition and processing costs are introduced to analyze the optimal decision-making of data providers and data trading platforms in the data supply chain under two decision-making modes respectively, we compare the two decision-making models, and discuss the impacts of network externality and the subsidies given to data trading platforms by the government on the supply chain. The impact of network externalities and government subsidies given to data trading platforms on supply chain pricing is discussed. The study finds that:

(1) The optimal product prices under centralized and decentralized decision-making are the same, and the profit of data trading platform, the total profit of data supply chain and the price of data products are positively affected by network externality, while the profit of data provider under decentralized decision-making remains unchanged with the increase of network externality. (2) The quality of raw data and data size under both decision-making modes are proportional to the elasticity coefficient of government subsidy, and the trend of government subsidy is more obvious under decentralized decision-making compared with centralized decision-making. (3) With the increase of network externality and government subsidy elasticity coefficient in both centralized and decentralized decision-making, the total profit of the supply chain shows a rising trend, and the overall profit of the supply chain under centralized decision-making is larger than that under decentralized decision-making.

Acknowledgment

This work is supported by the Chongqing Social Science Planning Project (2022ZDZK30) and the Chongqing

Technology Innovation and Application Development Project (CSTB2022TIAD-GPX0015).

References

- [1] Chen J X, Chen J. Supply Chain Profit Game and Equilibrium Pricing[J]. Chinese Journal of Management Science, 2022, 30(9): 128-139.
- [2] Barman A, De P K, Chakraborty A K, et al. Optimal Pricing Policy in a Three-layer dual-channel Supply Chain under Government Subsidy in Green Manufacturing[J]. Mathematics and Computers in Simulation, 2023, 204: 401-429.
- [3] Batini C, Scannapieco M. Data and Information Quality: Dimensions, Principles and Techniques[M]. Springer International Publishing, 2018.
- [4] Taleb I, Serhani M A, Bouhaddioui C, et al. Big Data Quality Framework: a Holistic Approach to Continuous Quality Management[J]. Journal of Big Data, 2021, 8(1): 1-41.
- [5] Li Q X, Zhang J, Huang Y M, Qi E S. Research on the Marketing Strategy of Community Group Buying Supply Chain Based on Network Externality[J/OL]. Chinese Journal of Management Science:1-12[2022-10-04].
- [6] Tang F C, Guo H. Pricing Strategy of Platform Competition Considering Crossnetwork Externality [J/OL]. Chinese Journal of Management Science: 1-20 [2022-10-18].
- [7] Zhu B X, Ma Z Q, Leon Williams. Research on Pricing and Quality Decision of Information Goods Based on Free Strategy under Piracy and Network Externalities[J]. Management Review,2021,33(9):143-154.