A Railway Ticket Data Pricing Model with Attenuation Coefficient Based on Stackelberg Game

Yuzhao Zhang, Shenyingze Gao, Weifeng Shi, Jianqiang Wang, Yanhua Wu

***Abstract*—After a long period of information development, the railway industry has accumulated rich railway data resources. To facilitate the comprehensive circulation of railway data factors, we propose a railway ticket data pricing model based on the Stackelberg game. The model covers three transaction participants, including the data owner, the middleman, and the data buyer. Based on railway data characteristics such as size, timeliness, privacy level, and price attenuation, we set up the expected profits of all transaction participants. The data purchase strategy and pricing strategy of transaction participants are derived from the Stackelberg game. Use Python programming language to realize the above process. The regularities of railway data pricing strategy and purchase strategy are found by adjusting the parameter values corresponding to different data characteristics. The conclusion proves that the timeliness of railway data is positively correlated with the income of all participants. The high data privacy level hurts the number of transactions, and the appropriate price attenuation strategy can mitigate the effect.**

***Index Terms*—Railway data, data characteristics, data pricing, Stackelberg game**

I. Introduction

DATA is gradually becoming a significant driving force to promote economic development. As the micro foundation of the digital economy, the data element has a strategic position and plays the role of an innovation engine [1]. According to the International Data Corporation (IDC), the global data size reached 64 zettabytes (1ZB=2^{70}B) in 2020, and by 2025, the number will be close to 180 zettabytes [2]. As a new factor of production, data has provided new momentum to economic development. Especially in recent years, with emerging technologies such as big data, artificial intelligence, Internet of Things, and cloud computing, the role of data factors has become more prominent. As one of the pillars of the national economy, the railway industry has achieved more development results of informatization and accumulated rich railway data resources. Among the numerous railway data, railway ticketing data, as an essential part of railway operation, has a profound impact on optimizing train operation and scheduling, predicting passenger travel demand, improving passenger service quality, and rationally allocating railway resources. Therefore, the realization of the circulation of railway ticketing data plays a vital role in breaking industry barriers, giving full play to the economic benefits of railway ticketing data, and stimulating the economic vitality of related industry markets.

At present, China has just started in the field of circulation of railway data elements, so it faces many difficulties: first of all, compared with physical products, the search, replication, transportation, tracking, and query costs of data products are significantly reduced [2]; especially the replication cost of data is close to zero so that the marginal cost is negligible [3]. As a result, the conventional pricing method that makes marginal cost equal to marginal benefit can't be used. Secondly, there is a lack of a complete system for the pricing and income distribution of railway data assets, which may hurt the willingness of all participants to join in the transaction. Moreover, the lack of unified and standard transaction channels and norms leads to low concentration and small transaction scale of data product transactions [4]. Therefore, a reasonably designed and standardized railway data pricing system is of great significance to facilitate the circulation of railway data. It can also accelerate the integration of railway data elements into production and life. The railway data pricing system will significantly promote the integration of digital economy and railway business, and help accelerate the transformation of digital railway.

This paper mainly introduces a railway ticket data pricing model based on the Stackelberg game. The main contributions of this paper can be classified as follows:

1) Combining the characteristics of data assets and the characteristics of the railway industry, find the characteristics of railway ticketing data in terms of pricing. The characteristics include large data scale, high data privacy level, and strong data timeliness. We design different transaction participants' profit functions based on

Manuscript received July 22, 2024; revised December 13, 2024.

This work was supported in part by the Scientific Research and Development Program of the China State Railway Group Co., Ltd (No. 2023F006-A(JB))

Yuzhao Zhang is a professor of the School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China (email: zhangyuzhao-81@mail.lzjtu.cn).

Shenyingze Gao is a postgraduate student of the School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China (corresponding author to provide e-mail: gaoshenyingze2023@163.com).

Weifeng Shi is a senior economist at the Institute of Computing Technology, China Academy Of Railway Sciences Corporation Limited, Beijing 100081, China (e-mail: shiweifeng@rails.cn).

Jianqiang Wang is an associate professor of the School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China (email: xinxiwjq@126.com).

Yanhua Wu is a researcher at the Institute of Computing Technology, China Academy Of Railway Sciences Corporation Limited, Beijing 100081, China (e-mail: wuyh2007@163.com).

these characteristics.

2) Construct a two-layer model from the original data owner to the middleman and then to the data buyer, in which the pricing strategy of the original data owner will affect the purchasing strategy of the middleman, and the pricing strategy of the middleman will affect the purchasing strategy of the data buyer. At the same time, the purchasing strategies of the data buyer and the middleman will also feedback and ultimately affect the processed data pricing and the original data pricing.

3) Derive the Stackelberg equilibrium points of three variables which include data purchase size, processed data price, and original data price. Design an iterative algorithm to obtain the optimal solution of the above variable under the given initial value. Use Python programming language to implement the algorithm.

4) Set different initial numerical conditions to analyze the general law of railway ticket data pricing and verify the performance of the pricing mechanism.

The rest of the article is structured as follows. Part II introduces the existing research results in the field of railway data pricing and general data pricing. Part III gives the overall model construction and algorithm design. In Part IV, the corresponding calculation examples are designed and analyzed. Part V offers some conclusions about the pricing of railway ticket data.

II. Research Status

The existing literature on the circulation of railway elements mainly constructs the basic framework of railway data valuation and pricing from the aspects of the internal and external ownership confirmation scheme of railway data [5], the value assessment of railway data assets [6, 7], and the improvement strategy of railway data service capability [8]. For example, Wang et al. [5] gave the subject of railway data rights, focused on the right to use railway data, and proposed the classification and management of railway data in multiple dimensions. Based on the characteristics of railway data, Ding [6] and Wu et al. [7] constructed multi-level indicators to evaluate the value of railway data assets. Li et al. [8] proposed a value-driven data collaborative service model by analyzing data value and service. However, the existing results tend to be qualitative evaluation and analysis and lack specific methods for railway data pricing. As the source of railway data transactions, valuating and pricing of railway data have a significant impact on the subsequent transaction process.

In the solution to the data pricing, the profit distribution model based on game theory is widely used. It fully considers the needs of all participants and pays attention to the process of price discovery. It also makes use of different characteristics of the data trading market to unify the price between data owners and data buyers [9]. In the axiomatic formulation for solutions in game theory, consistency is an important property. Consistency declares the independence of an outcome with respect to fixing several agents with its allotted payoffs [10]. The research on data transaction pricing based on game theory can be divided into three categories: pricing model based on the non-cooperative game, pricing model based on the Stackelberg game, and pricing model based on the bargaining game.

The Nash equilibrium of the non-cooperative game requires sellers to know each other's strategy and announce their strategy simultaneously, so its application in real life will be limited to some extent. Luong N C et al. [11] provided a state-of-the-art literature review on economic analysis and pricing models for data collection and wireless communication in the Internet of Things (IoT). The bargaining game requires both the supply and demand sides to reach an agreement through negotiation, which mainly guarantees the privacy and fairness of both sides of the transaction. It takes less consideration in profit distribution, which consumes more time and resources. Jung K et al. [12] proposed a fair negotiation method, which used the Rubinstein bargaining model to determine the price of data and the value of privacy loss to ensure fair trade.

Stackelberg game considers information asymmetry between buyers and sellers, and describes the decision scheme with priority between leaders and followers. Moreover, it gives more consideration to the problem of maximizing the benefits of all participants in a transaction. Therefore, the Stackelberg game is widely used in market competition and enterprise strategy. Li et al. [13] discussed the Stackelberg game model between the data owner and the middleman for pure bundle pricing and separate pricing. Then they studied the pricing mode of the data owner under the condition that both participants maximize their benefits. Liu [14] and Xu et al. [15] built a two-stage Stackelberg game model to solve the price and purchase problems of data demanders, considering the cases of single seller and multi-seller, respectively. Hong [16] uses fuzzy set theory to examine the optimal decision of each member of a two-stage supply chain, which includes a manufacturer and a retailer.

To sum up, the Stackelberg game can be used to build a three-party game model between the owner of railway ticketing data, the middleman, and the buyer. The game consists of two stages, carried out between the data owner and the middleman, the middleman and the data buyer respectively. The game shows each party's buying strategy, selling strategy, and income distribution.

III. Railway Ticket Data Pricing Model

A. *Characteristics of Railway Ticket Data*

As an emerging production factor, railway data elements not only contain the general regularities of data elements, but are also affected by the characteristics of the railway industry. Therefore, both should be considered comprehensively when designing the pricing method of railway ticket data. Specifically, railway ticket data mainly includes the following characteristics:

1) Large data scale. As one of the main means of transportation, the railway processes a large amount of ticket data daily. These data include ticket sales information, passenger booking information, train schedule changes, station information, and so on. With the continuous expansion of the railway network and the growth of the

number of passengers, the amount of ticketing data is also increasing, showing the characteristics of a large data scale. According to rough statistics, at present, the total amount of railway data assets is more than 10PB, and the daily growth amount is more than 1TB [7], among which ticketing data occupies a considerable proportion. For example, during the Spring Festival in 2024 alone, the railway department sold 91.828 million train tickets in total, and the corresponding stored data may reach hundreds of GB. At the same time, the current railway data collection methods are more advanced, and the collection density increases constantly, so that more data can be stored in digital ways, which reduces the data storage cost and increases the data storage capacity.

2) Strong data timeliness. The normal data asset has the characteristics of strong timeliness [17], and the railway ticketing data also needs to be updated in time to reflect the latest ticketing information and train operation status. Passengers' travel plans may change due to various reasons, so the ticket system needs to process different requests and update the corresponding data in time. What's more, the railway dispatching system also needs to obtain real-time ticketing data, and adjust the train operation according to the needs of passengers. Only by timely mastering ticket data, we can accurately predict the passenger flow, and rationally arrange the capacity resources. Then we will ensure the safety, efficiency, and convenience of railway transportation. These highly dynamic and random data lead to a faster updating speed of railway data.

3) High data privacy level. Railway ticket data contains much personal information of passengers, such as ID number, name, and contact information. The information refers to personal privacy, which needs to be strictly protected. It may bring serious threats to the privacy and property security of users once leaked. In addition, ticket data is often associated with the user's payment information, such as bank card number and credit card number, which is directly related to the user's property security. Railway authorities need to comply with relevant laws and privacy protection policies when processing and using ticket data to ensure the security and privacy of passenger information.

B. Expected Profit for All Participants

The rights of relevant data can be divided into three parts [18]: the right to hold, the right to process, and the right to operate products. In this problem, we choose to build a two-layer Stackelberg game pricing model, in which the three participants involved in the game are the owner of railway ticketing data (the major railway bureaus), the middleman (the trusted intermediary platform) and the buyer of railway ticketing data (the enterprise with demand for railway ticketing data). The data owner holds the ownership of the original data and is responsible for selling the right to process the original data to the middleman. The middleman will fulfill the right of processing and process the original data to some extent. Then the middleman classifies and packages the data according to the needs of the data buyer. The data buyer purchases the data from the middleman, who exercises the right to operate the data product. The whole game process is divided into three stages: determining the income function of each party, deducing the maximum point of the income function, and iterating to obtain the optimal Stackelberg equilibrium point.

1) Expected profit of the data owner

The expected profit of the data owner is mainly determined by the selling price of the original data, transaction and packaging costs, the size of the purchase by the middleman. p_r represents the unit price of raw data. e represents the transaction fee and the unit cost of packaged data, and x represents the number of packets sold of a particular type. Only when $p_r > e$, the data owner chooses to sell the data. The specific formula is shown in equation (1):

$$B_1 = \left(p_r - e\right)x \tag{1}$$

2) Expected profit of the middleman

After purchasing the raw data from the data owner, the middleman needs to process the raw data according to users' needs and make data products that can be sold. According to the above introduction, the essence of data trading is to sell the right to use data rather than the data ownership. This

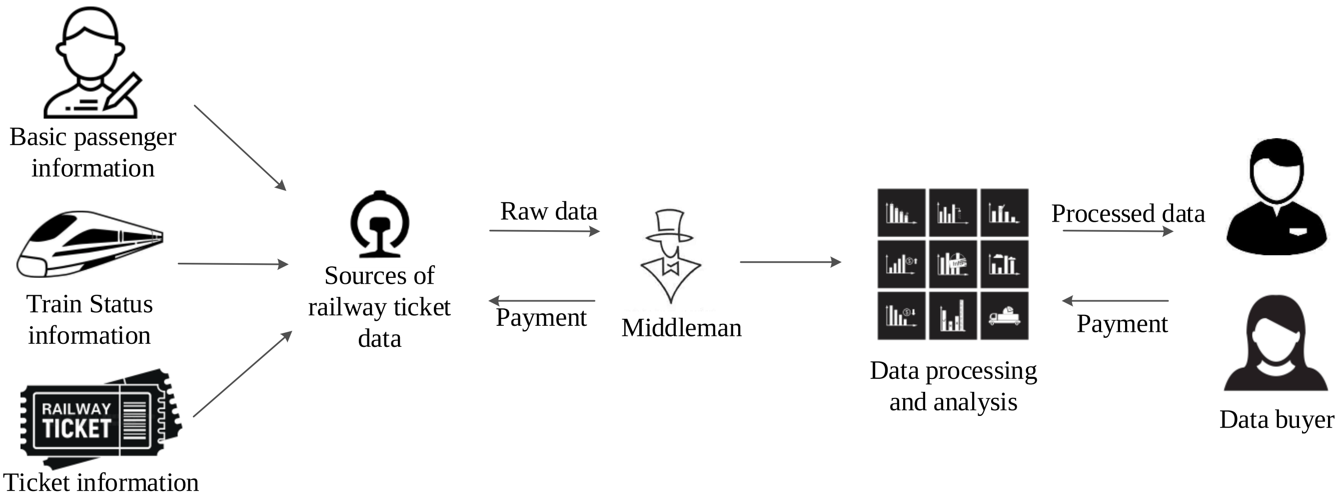

Fig. 1. The whole course of data transaction

leads to an essential difference between data trading and general labor product trading: the same data product can be sold repeatedly. It can be seen that the processing of data products only occurs at the beginning of the entire transaction process. Therefore, we need to redefine the cost of data processing and spread it across all transactions, so that each transaction bears a corresponding cost. As the number of transactions increases, the cost per transaction will decrease. At the same time, compared with general labor products, the transportation and storage costs of data products are almost negligible, so the calculation of processing costs becomes more important.

The apportioned data processing cost p_i is expressed by formula (2), where C is the cost required to process this type of data packet once. p_d is the unit price of data after processing. The expected return of the middleman B_2 is expressed as formula (3).

$$p_i = \frac{C}{x} \tag{2}$$

$$B_2 = (p_d - p_r)x - p_i \tag{3}$$

Due to the strong timeliness of a considerable part of data products, which means the benefits brought by data products will decline as time goes by. Balazinska *et al.* point out that the easy replication of data makes the variable cost of data assets deficient [19], which also exacerbates this process. To sum up, it is possible to increase data sales by gradually reducing data pricing to maintain the seller's profit as much as possible. Based on this, formula (3) can be modified into formula (4), where the value of λ can be taken as needed between (0, 1).

$$B_2 = \sum_{i=1}^{x}[1-(i-1)\lambda]p_d - p_r x - p_i \tag{4}$$

3) Expected profit of the buyer

In the third part of the model, the expected benefit of the buyer buying the processed data from the middleman has a strong correlation with user satisfaction, and user satisfaction depends on the data quality. Data quality is related to data scale, data timeliness and data privacy level. Data quality under the influence of three factors is explained:

According to the characteristics of large-scale railway ticketing data, the part related to data quality in user satisfaction can be defined as equation (5), where and are the curve fitting parameters obtained by the actual experiment [20]:

$$N(x) = \alpha_1 + \alpha_2 \ln(1+x) \tag{5}$$

Railway ticket data has strong timeliness. Timeliness t_i indicates the old and new degree of the data in the data packet. The update time of the data in the data packet is arranged from small to large, and its value is (0, 1). represents the median of all update times. t_m represents the minimum value of all update times, and t_e represents the maximum value of all update times:

$$t_i = \frac{t_m - t_e}{t_l - t_e} \tag{6}$$

According to the privacy level of the data, the data quality can be defined as formula (7), where λ_1, λ_2, λ_3 are the curve fitting parameter obtained through practical experiments[21]:

$$Q_j = \lambda_1 - \lambda_2 e^{\lambda_3 r_j} \tag{7}$$

The buyer of railway ticket data can determine the data size, timeliness, and privacy level according to their own needs, so the expected benefit of the buyer can be expressed as formula (8):

$$B_3 = Q_j N(x) t_i - p_d x \tag{8}$$

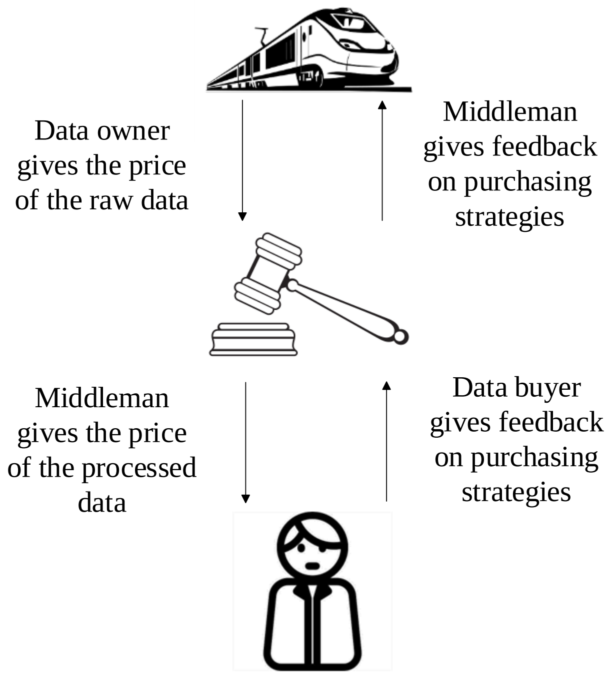

Fig. 2. Relationship map of the participants to the game

C. *Maximum Points of Participants' Expected Profits*

The model is divided into three parts, which include data owners, intermediaries and data buyers. The game is divided into two stages. The first stage includes middlemen and data buyers, and the second stage includes data owners and middlemen. The game's purpose is to maximize the benefits of all participants under certain conditions, which means finding the equilibrium point of the game. Therefore, based on the goal of maximizing all participants' income functions, maximum value points of decision variables in three parts can be obtained by using the inverse method. Decision variables contain the scale of the user's purchase of the data x^*, the pricing of the middleman for the processed data p_d^*, and the pricing of the data owner for the original data p_r^*.

Since the game in this process has a sequence, which means the followers can observe the leader's behavior and give a response plan according to the leader's decision, the game process here is a dynamic game. In the dynamic game, the leader will consider followers' interests when making decisions. Only the followers in the last stage can make decisions without any restrictions. When the followers' decisions are made, the restricted leader's decision will become direct and accessible. Therefore, we can use backward induction for dynamic games, which means constantly deducing from the last stage to find the game equilibrium points.

1) The first stage game

$$\max B_3 \tag{9}$$

$$s.t.\, x \geq 0 \tag{10}$$

Let the price of the processed data p_d be fixed, and Take the partial of B_3 with respect to x , then the formula (11) can be obtained:

$$\frac{\partial B_3}{\partial x} = \frac{\alpha_2 t_i Q_j}{1+x} - p_d \tag{11}$$

And set its partial derivative to 0, so that the corresponding solution x^* can be obtained:

$$x^* = \frac{\alpha_2 t_i Q_j}{p_d} - 1 \tag{12}$$

Take the derivative of x again with equation (11), and get the second partial derivative of B_3 with respect to x:

$$\frac{\partial^2 B_3}{\partial x^2} = -\frac{\alpha_2 t_i Q_j}{(1+x)^2} \tag{13}$$

It is easy to know the formula (13) is less than 0, so x^* is the maximum point. After the user gives his own purchasing strategy, the middleman can improve his pricing strategy to maximize his utility in the following ways:

$$\max B_2 \tag{14}$$

$$s.t.\, x \geq 0 \tag{15}$$

Since the user has given the optimal procurement strategy at this time, it is regarded as a fixed value. Let $A = \alpha_2 t_i Q_j$. When x^* is substituted into B_2, the result is as follows:

$$\begin{aligned} B_2 = p_d[\frac{A}{p_d} - 1 - \frac{1}{2}\lambda(\frac{A}{p_d} - 1)(\frac{A}{p_d} - 2)] \\ - p_r(\frac{A}{p_d} - 1) - p_i \end{aligned} \tag{16}$$

Also we can get the partial derivative of B_2 with respect to p_d , in order to find the equilibrium point of the Stackelberg game. The results obtained are as follows:

$$\frac{\partial B_2}{\partial p_d} = \frac{\lambda A^2 + 2Ap_r}{2(p_d)^2} - \lambda - 1 \tag{17}$$

And set the partial derivative equal to 0 to p_d^* get as shown in equation (18):

$$p_d^* = \sqrt{\frac{\lambda A^2 + 2p_r A}{2(\lambda+1)}} \tag{18}$$

The second-order partial derivative can also be obtained as equation (19):

$$\frac{\partial^2 B_2}{\partial {p_d}^2} = -\frac{\lambda A^2 + p_r A}{(p_d)^3} \tag{19}$$

It is easy to know that formula (19) is less than 0, so p_d^* is the maximum point. It can be concluded that the maximum values of B_2 and B_3 are reached at this time, so x^* and p_d^* are the balance points of the game at this stage.

2) The second stage game

After receiving the pricing strategy feedback from the middleman, the data owner can adjust the original data pricing strategy based on the feedback to maximize its utility function:

$$\max B_1 \tag{20}$$

$$s.t.\, p_r > e \tag{21}$$

At this step, the data buyer and the middleman in the first stage adopt the optimal buying strategy and the optimal pricing strategy respectively. Then x^* and p_d^* can be regarded as fixed values, and they should be substituted into B_1 to obtain the equation (22):

$$B_1 = (p_r - e)(\frac{A}{\sqrt{\frac{\lambda A^2 + 2p_r A}{2(\lambda+1)}}} - 1) \tag{22}$$

According to the research in the previous section, the Stackelberg equilibrium point p_r^* can be expressed as:

$$p_r^* = \arg\max B_1 \tag{23}$$

$$s.t.\, (x^*, p_d^*) = \arg\max B_2 \tag{24}$$

$$x^* = \arg\max B_3 \tag{25}$$

Take the second partial derivative of B_1 with respect to p_r . The equation (26) can be obtained:

$$\begin{aligned} \frac{\partial^2 B_1}{\partial {p_r}^2} = \sqrt{2A(\lambda+1)} \\ \cdot[3(p_r - e)(\lambda A + 2p_r)^{-\frac{5}{2}} - 2(\lambda A + 2p_r)^{-\frac{3}{2}}] \end{aligned} \tag{26}$$

After mathematical derivation, it can be known that $\frac{\partial^2 B_1}{\partial {p_r}^2} < 0$, so p_r^* is the maximum point.

D. Iterative Algorithm

The above game process has derived the balance point after one game. To obtain the optimal solution, the above process needs to be carried out for several times. We can use the relationship between the obtained variables to get the optimal solution that meets the accuracy requirements. The specific algorithm is as follows:

1) Assign initial values to each variable. Let $x = x^{(0)}$, $p_d = p_d^{(0)}$, $p_r = p_r^{(0)}$, $k = 0$. $C = \left|x^{k+1} - x^k\right|$ represents the precision control variable, which initial value is set to 1; and $\varepsilon = 10^{-3}$ is the precision requirement.

2) If $C > \varepsilon$, then order $x^{k+1} = \frac{A}{p_d^k} - 1$. It indicates that the data buyer makes the corresponding purchase decision based on the pricing of the middleman;

3) Let $p_d^{k+1} = \sqrt{\frac{\lambda A^2 + 2p_r^k A}{2(\lambda+1)}}$, which means that the middleman makes corresponding price decisions based on the pricing of the data owner;

4) Let $p_r^{k+1} = \arg\max B_1(x^{k+1}, p_d^{k+1}, p_r^k)$, which means the pricing decision made by the data owner under the premise of satisfying the interests of the middleman and the data buyer;

5) Order $k = k+1$ and test whether $C = \left|x^{k+1} - x^k\right| < \varepsilon$ is true. If it is true, stop the iteration and enter the step (6); Conversely, return to (2) to continue the iteration;

6) Output the qualified result about x^k, p_d^k, p_r^k.

IV. Numerical Results

In this section, we will analyze the numerical results obtained. For the above algorithm, use Python to write a program and consider the influence of different factors on the three variables. Set the transaction cost and the unit cost of the packaged data e=0.1, the attenuation coefficient λ=0.95, and the remaining experience coefficients α_1=0.5, α_2=1, λ_1=3, λ_2=0.1, λ_3=2. The initial values of data purchase size x, processed data price p_d, and raw data price p_r are 0.50, 1.00, and 0.50, respectively.

Set the range of privacy levels to be $r_j = \left[0.2, 0.4, 0.6, 0.8, 1.0\right]$. According to the privacy level, the railway ticket data can be divided into the following categories:

1) Non-sensitive data. Non-sensitive data includes train frequency information, train type and facilities, fare information, station facilities information, etc. The privacy level of this type of data can be set as $r_j = 0.2$.

2) Low sensitive data. Low sensitive data includes desensitized user registration information, order statistics, ticket sales reports, etc. You can set the privacy level for this type of data as $r_j = [0.4, 0.6]$.

3) Medium sensitive data. Medium sensitive data includes user ticket purchase records, user payment information, refund and change records, etc. The privacy level of this type of data can be set as $r_j = [0.8, 1.0]$.

4) Highly sensitive data. Highly sensitive data includes the user's complete identity information, the user's bank card information, the user's biometric information. This kind of information is highly sensitive and may cause serious data security accidents once leaked, so it should be strictly controlled and not transacted.

Set the range of timeliness to be $t_i = \left[0.1, 0.3, 0.5, 0.7, 0.9\right]$. According to the timeliness of the data, the railway ticket data can be divided into the following categories:

1) Real-time data. The real-time data includes the current ticketing status, train real-time location and running time, station passenger flow statistics, station real-time information and so on. The timeliness of this kind of data can be set as $t_i = [0.7, 0.9]$.

2) Periodic data. The periodic data includes ticket sales reports, train schedule adjustment notices, passenger satisfaction surveys, etc. The timeliness of this type of data can be set as $t_i = [0.3, 0.5]$.

3) Historical data. The historical data includes ticketing history data, passenger travel history and train operation history data. The timeliness of this type of data can be set as $t_i = 0.1$.

A. Convergence Judgment of Results

If the data buyer chooses the data with the privacy level of 0.6 and the timeliness of 0.9, A=2.40 can be obtained, and λ=0.95 is substituted into the program for calculation. Get the result is x=0.35, p_d=1.78, p_r=1.43. As the number of iterations increases, the changes in data purchase size x, processed data price p_d, and original data price p_r are shown in Figure 3.

Fig. 3 shows the change in the purchase size x of data buyers with the number of iterations. It can be seen that under the initial conditions, buyers tend to purchase more data to obtain greater benefits. When the middleman and the data provider change their pricing strategies, the purchase quantity of the buyer rapidly converges until the Stackelberg equilibrium point is reached. The processed data price p_d and the original data price p_r also tend to be fixed after 2 and 1 iterations respectively, which means they reach their respective Stackelberg equilibrium points. The convergence order of the decision variables is $p_r > p_d > x$, which is also consistent with the algorithm design, and verifies the accuracy of the algorithm.

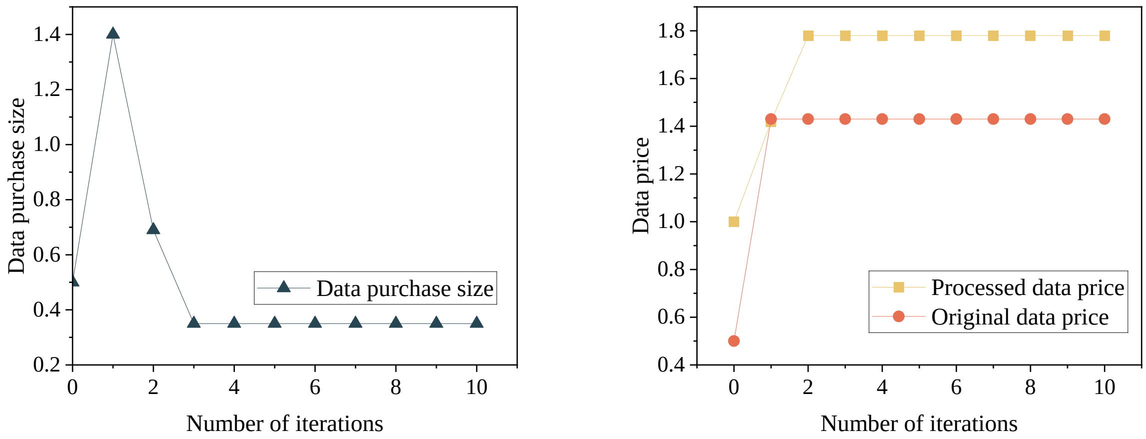

Fig. 3. Data purchase size and data pricing at different stages vary with the number of iterations

B. *Sensitivity Analysis*

1) Single factor analysis

a. Timeliness

Select data with r_j =0.8, and make the initial value of data purchase size x =0.20, processed data price p_d =0.20, and original data price p_r =0.15. Then substitute them into the program to get the result:

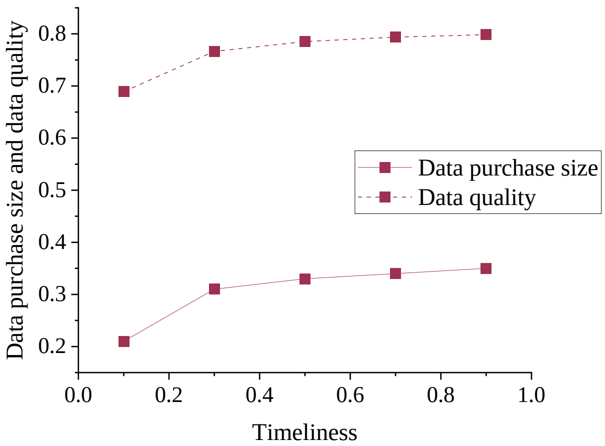

Fig. 4. Impacts of timeliness on data purchase size and data quality

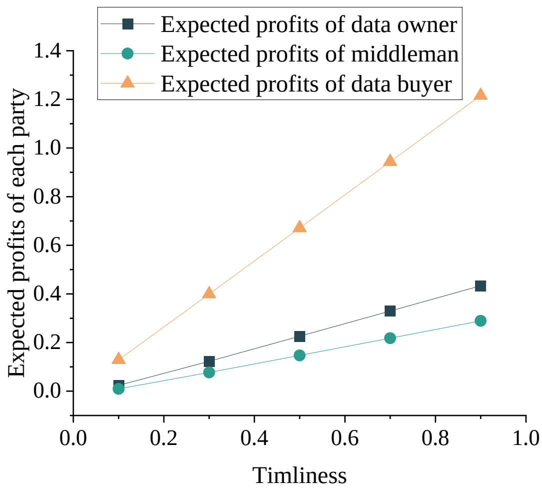

Fig. 5. Impacts of timeliness on expected profit of each party

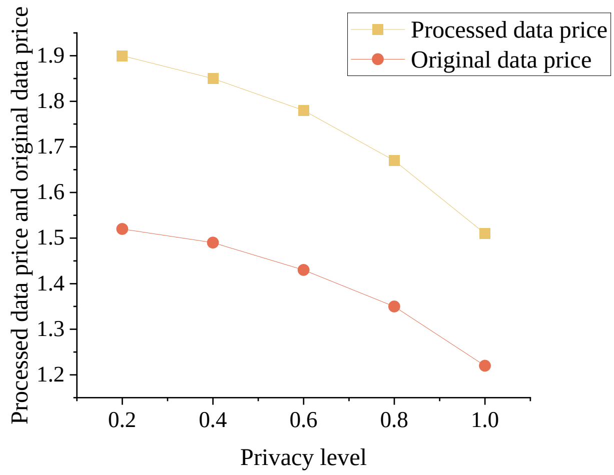

Fig. 6. Impacts of timeliness on processed data price, original data price

Fig. 4 shows the relationship between data purchase size, data quality and timeliness. It can be seen that with the improvement of data timeliness, the amount of data purchased by users and the data quality determined by the purchase size are also increasing. As can be seen from Fig. 5, with the improvement of data timeliness, the benefits of all participants have increased to varying degrees, among which the benefits of data buyers have increased most significantly. As shown in Fig. 6, the relationship between data pricing after data processing and raw data pricing and timeliness shows that both data pricing and timeliness are positively correlated. Therefore, it can be concluded that the higher the validity of the transaction data, the higher the data turnover and the benefits of the trading participants, showing a win-win situation in general.

b. Privacy level

Select data with t_i =0.9 and make the initial value of data purchase size x =0.20, processed data price p_d =0.20 and original data price p_r =0.15. Then substitute them into the program to get the result:

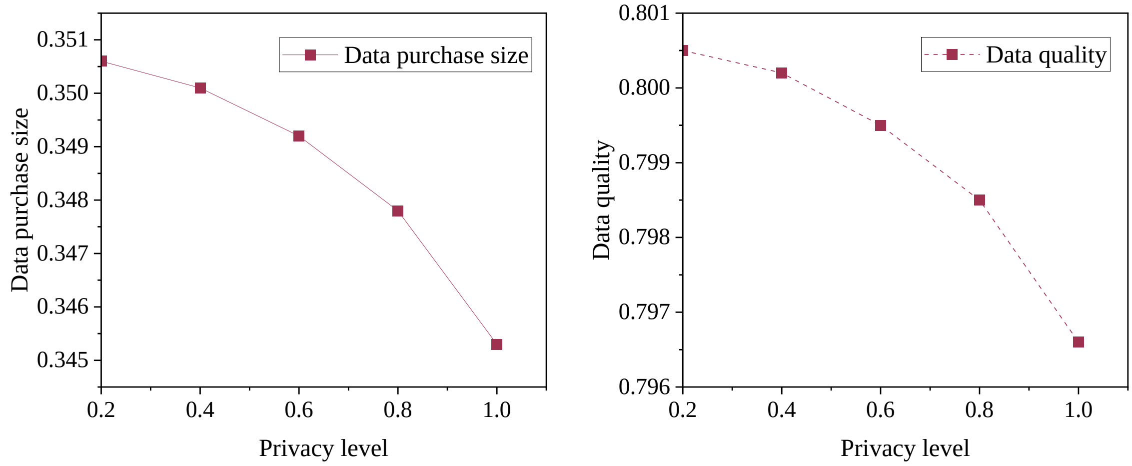

Fig. 7. Impacts of privacy level on data purchase size and data quality

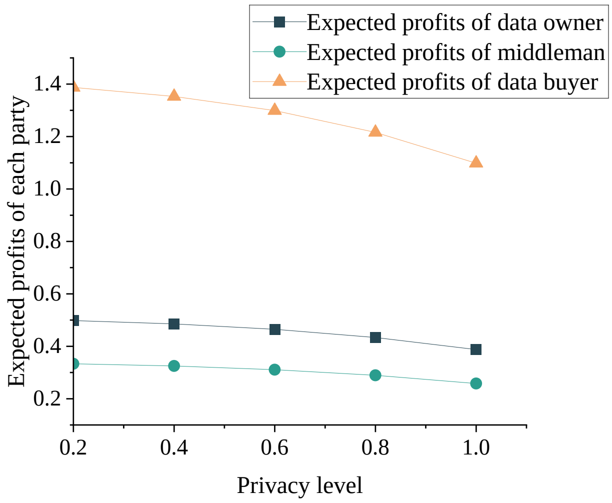

Fig. 8. Impacts of privacy level on expected profits of each party

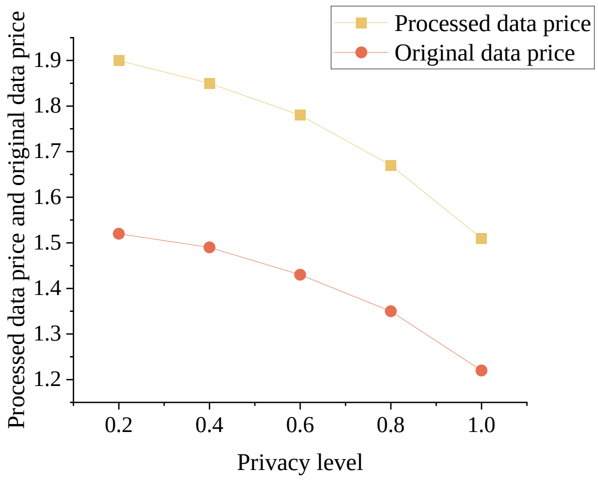

Fig. 9. Impacts of privacy level on processed/original data price

According to Fig. 7, the relationship among data purchase size, data quality and privacy level indicates that with the increase of data privacy level, data purchase size and data quality all show the downward trend. From Fig. 8 and Fig. 9, we can presume that the increase of privacy level hurts data price and each party's expected profits. Combined with formula (7), the reason for this situation is that the data with high privacy level needs more confidential processing, so that the scope of application is smaller. These factors lead to the decline of data quality, and thus affect the subsequent transaction process.

c. Attenuation coefficient

Select the data with r_j =0.8, t_i =0.9, and set the attenuation coefficients λ as 0.95, 0.85, 0.75, 0.65, 0.55 respectively. Make the initial value of data purchase size x =0.20, processed data price p_d =0.20 and original data price p_r =0.15. Then substitute them into the program to get the result:

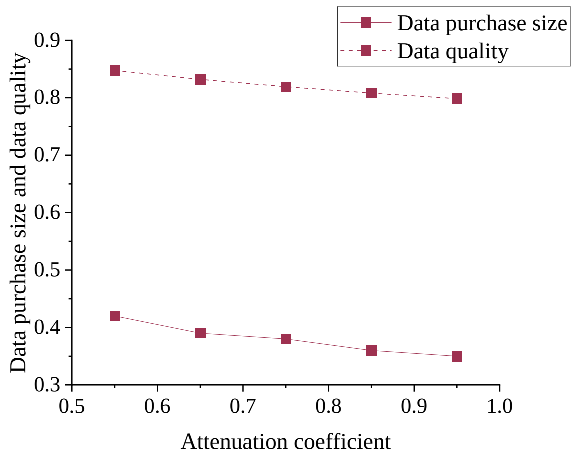

Fig. 10. Impacts of attenuation coefficient on data purchase size and data quality

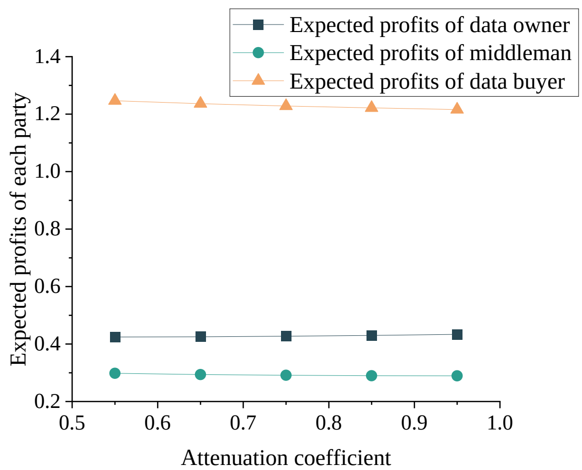

Fig. 11. Impacts of attenuation coefficient on expected profits

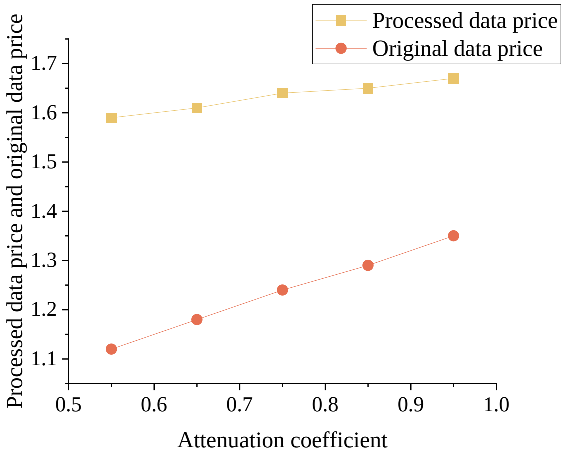

Fig. 12. Impacts of attenuation coefficient on data price

The attenuation coefficient describes the price changes of data products in the trading process. It can be seen that the price attenuation of data products will have a positive impact on its trading size from Fig. 10. However, from the impacts of attenuation coefficient on data price in Fig. 12, the price attenuation of data products will lead to the reduction of processed data price and original data price. Under the joint action of two factors with opposite effects, the macro impact on the income of all participants is basically unchanged, as shown in Fig. 11.

2) Multi-factor analysis

a. Timeliness and privacy level

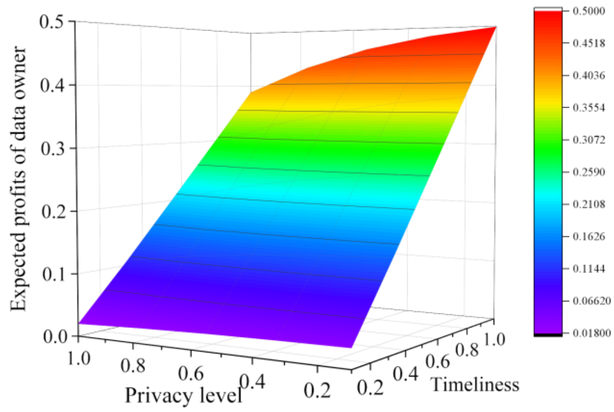

Fig. 13. Combined impacts of privacy level and timeliness on expected profits of data owner

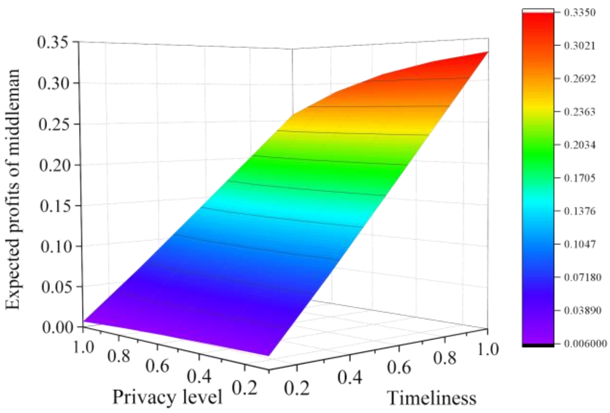

Fig. 14. Combined impacts of privacy level and timeliness on expected profits of middleman

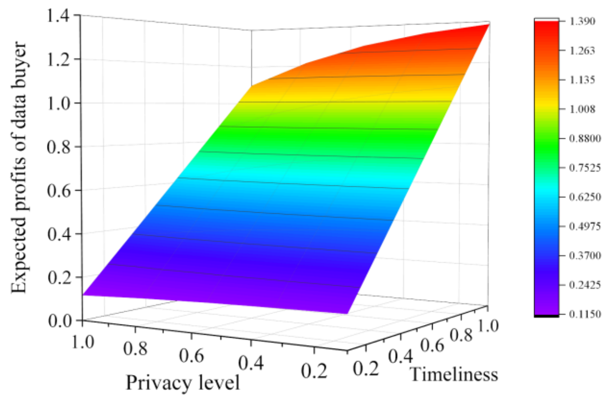

Fig. 15. Combined impacts of privacy level and timeliness on expected profits of data buyer

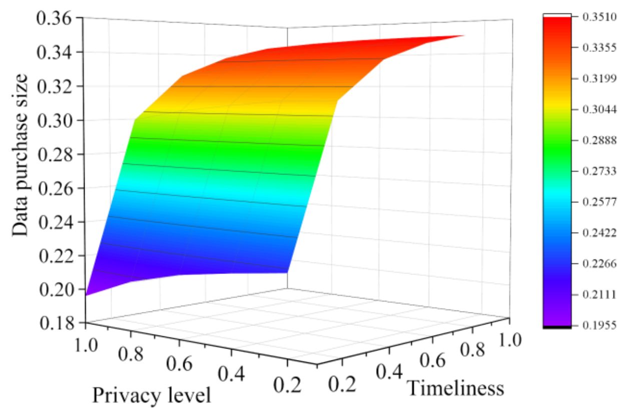

Fig. 16. Combined impacts of privacy level and timeliness on data purchase size

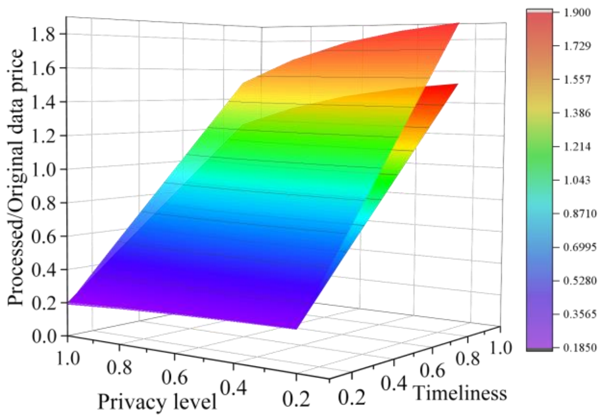

Fig. 17. Combined impacts of privacy level and timeliness on processed/original data price

From Fig. 13 to Fig. 15, it can be seen that when timeliness t_i =0.9 and privacy level r_j =0.2, the benefits of each party can reach the maximum. At the same time, the privacy level is negatively correlated with the profit function. In contrast, the timeliness is positively correlated with the profit function, which is the same as the single factor analysis above. As the timeliness of data increases, the impact of data privacy level on expected profits becomes more and more obvious. In addition, according to the changes of participants' profits in the curved surface diagram, timeliness is the main factor of the change of each party's income. It reflects the importance of data timeliness in the process of data transaction. When the timeliness of data reaches a high level, the privacy level of data will have a more prominent impact.

From Fig. 16, it can be concluded that the privacy level of data is negatively correlated with the volume of data purchased. In contrast, the timeliness of data is positively correlated with the volume of data purchased. As the data timeliness reaches higher levels, the growth in data purchase size starts to slow. It can also be seen that the impact of data privacy level on the volume of data purchased is less than that of data timeliness. Therefore, under the condition of a certain decay coefficient, the timeliness of data is the main factor determining the volume of data purchased.

From Fig. 17, it can be concluded that the privacy level of data is negatively correlated with the price of data transactions. In contrast, the timeliness of data is positively correlated with the price of data transactions. It can also be seen that the impact of data privacy level on the processed/original data price is less than data timeliness. As the timeliness increases, data privacy level's impact on processed/original data price becomes more obvious. The price of processed data is always higher than the pricing of raw data, which is consistent with the data product processing process.

b. Timeliness and attenuation coefficient

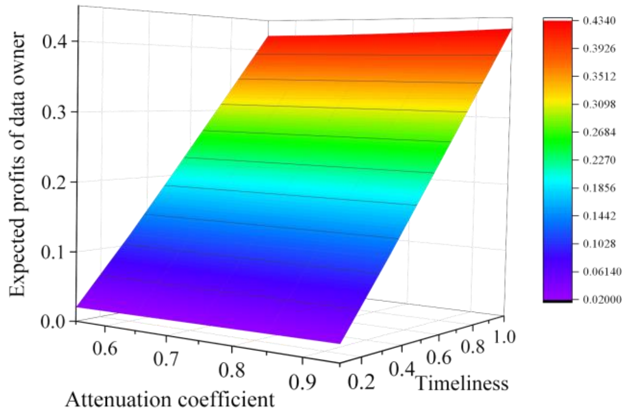

Fig. 18. Combined impacts of attenuation coefficient and timeliness on expected profits of data owner

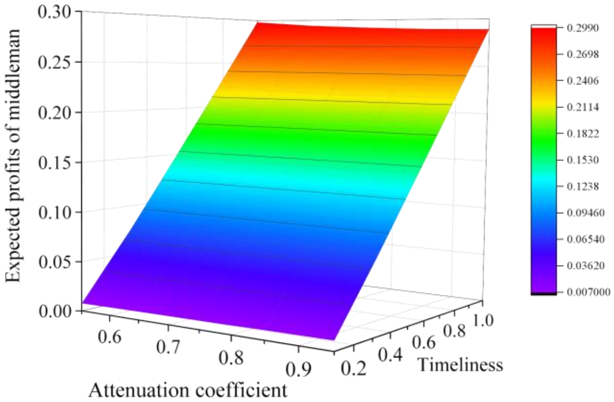

Fig. 19. Combined impacts of attenuation coefficient and timeliness on expected profits of middleman

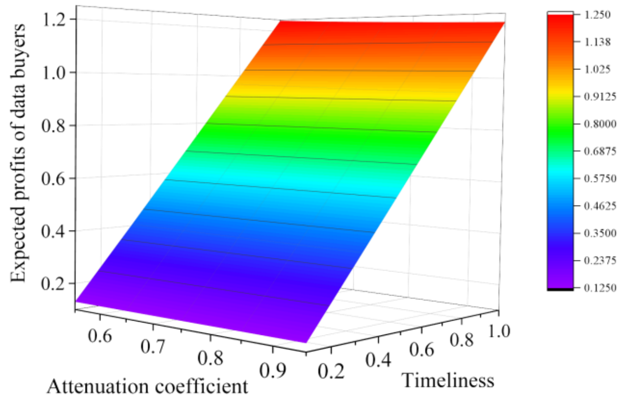

Fig. 20. Combined impacts of attenuation coefficient and timeliness on expected profits of data buyers

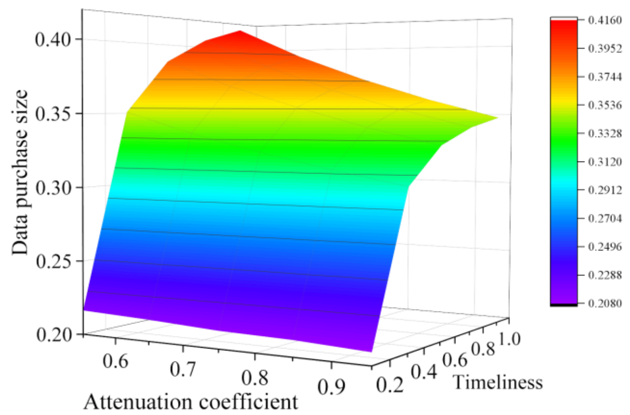

Fig. 21. Combined impacts of privacy level and timeliness on data purchase size

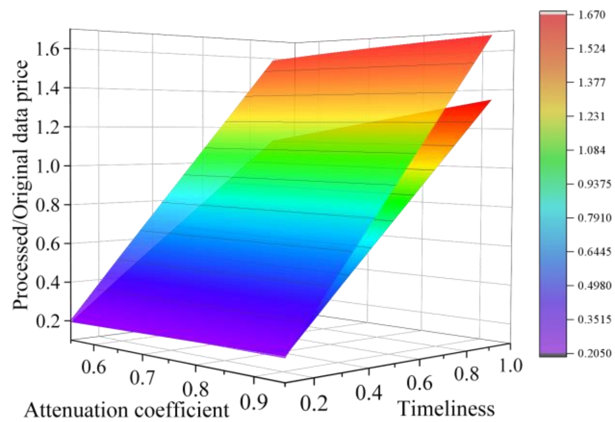

Fig. 22. Combined impacts of privacy level and timeliness on processed/original data price

From figure 18 to figure 20, it can be seen that when attenuation coefficient λ =0.95 and timeliness t_i =1.0, data owner's return B1 is the largest; When attenuation coefficient λ =0.55 and timeliness t_i =1.0, the profit of middleman B2 and the profit of data buyer B3 are the largest. The attenuation coefficient is positively correlated with the profits of the data owner and negatively correlated with the profits of the other two participants, which verifies the results of the single factor analysis above. This result shows that when the price of data products declines slowly, the income of data owners can be guaranteed. However, this may affect the purchase willingness of middlemen and data buyers, resulting in a decline in the income of the other two participants.

From Fig. 21, it can be concluded that the price decay coefficient is inversely proportional to the volume of data purchased, which means that appropriately reducing the price can promote the willingness to trade for all parties. In addition, this phenomenon is particularly obvious when the data timeliness is at a high level. The data timeliness is positively proportional to the volume of data purchased.

From Fig. 22, it can be concluded that the transaction price and data timeliness are positively proportional to the decay coefficient. As the data timeliness increases, the impact of the price decay coefficient on the processed/original data price becomes more significant. What's more, the impact of data freshness is more significant.

c. Privacy level and attenuation coefficient

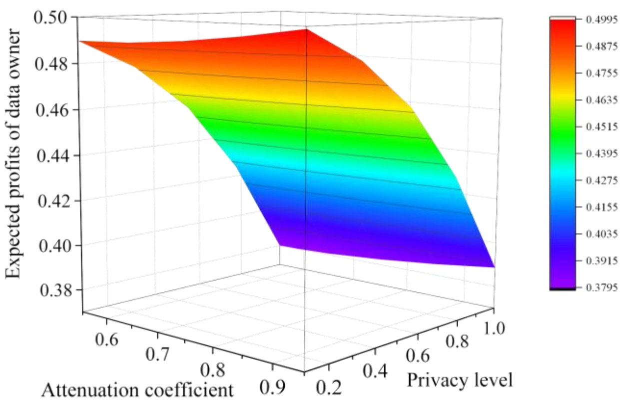

Fig. 23. Combined impacts of privacy level and attenuation coefficient on expected profits of data owner

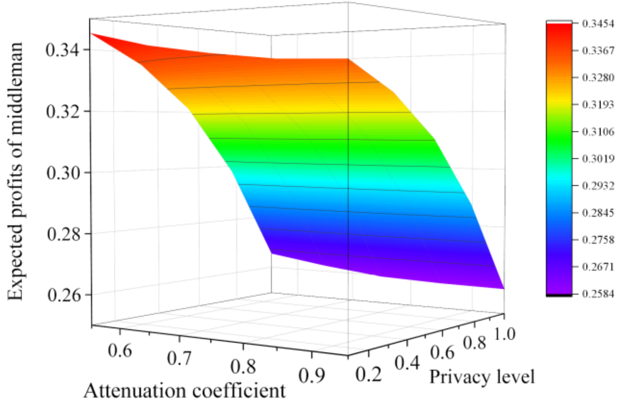

Fig. 24. Combined impacts of privacy level and attenuation coefficient on expected profits of middleman

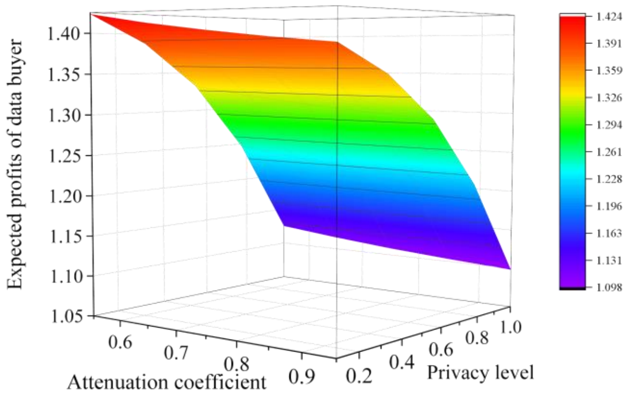

Fig. 25. Combined impacts of privacy level and attenuation coefficient on expected profits of data buyer

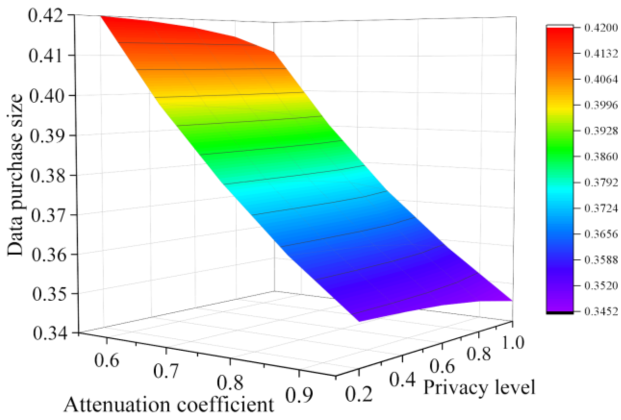

Fig. 26. Combined impacts of privacy level and timeliness on data purchase size

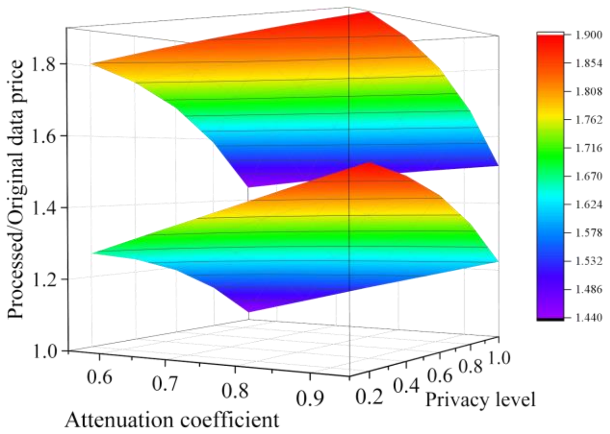

Fig. 27. Combined impacts of privacy level and timeliness on processed/original data price

Through the analysis of Fig. 23 to Fig. 25, it can be seen that the data privacy level r_j has a more critical impact on the profit of each party compared with the attenuation coefficient λ. When the attenuation coefficient λ=0.95 and the privacy level r_j=0.2, the expected profits of data owner reach the maximum. However, when the attenuation coefficient λ=0.55 and the privacy level r_j=0.2, the expected profits of other two participants reach the maximum. As a result, the higher data price will increase the data owner's profits, but it will also decrease other participants' profits. With the increase of data privacy level, the profit of all participants presents a decreasing trend, which is also consistent with the separate analysis of privacy level above.

From Fig. 26, it can be concluded that, under the constant data timeliness, both the price decay coefficient and data privacy level are inversely proportional to the data purchase volume. It shows that the higher privacy level and price will have a negative impact on users' purchase intentions. However, the main influencing factor is the price decay coefficient. From Fig. 27, it can be seen that the data transaction price is positively proportional to the price decay coefficient, but inversely proportional to the data privacy level. The processed data price is always higher than the original data price. Based on the research results above, the high privacy level will impact users' purchase intentions and the transaction price ultimately.

Through a longitudinal comparison of these three factors, it can be found that the timeliness of data has the most significant impact on the profit function of all participants, followed by the data privacy level. In comparison, the price attenuation coefficient has the most negligible effect on the profit function of all participants. Therefore, in actual transactions, data with high timeliness should be used as much as possible.

What’s more, the data privacy level is negatively correlated with the data purchase size and the profits of participants, which means the excessive data privacy level will hurt the transaction process. It leads to carrying out more detailed discussion in the actual production. Other compensation mechanisms also should be adopted to compensate the losses of middlemen and data providers.

Besides, we found that to promote the trade of data products, price attenuation can be appropriately adopted to keep the trading market active. In actual production and life, we should consider all kinds of factors comprehensively and design a trading scheme that can take into account all participants’ interests, so that data elements will play a more significant role in the future.

V. CONCLUSIONS

With the rapid growth of railway data scale, railway data elements play a significant role in production. As an essential part of railway data elements, the complete railway ticket data pricing and profit distribution system is conducive to promoting the active participation of all participants in railway data transactions. It will also effectively release the value of railway data elements, and lay a foundation for future related research. Based on the Stackelberg game principle and the characteristics of railway data, the paper designs a railway ticket data pricing model with three participants in the transaction. Through quantitative calculation, we find the optimal pricing strategy and purchase strategy corresponding to different transaction scenarios. Furthermore, this paper shows the regularities of the railway data transaction process, and provides a reference for the actual transaction process.

Through the analysis of case studies with given parameters, we verify the results’ convergence and the algorithm’s effectiveness. As the number of iterations increased, the data purchase size, processed data price, and original data price tend to stabilize. Then, we analyze the indicators of timeliness, privacy level, and price attenuation coefficient from single-factor and dual-factor levels. We study the impact of different factors on the data purchase size, processed data price, and original data price. Ultimately, we find that data timeliness has a positive impact on the entire transaction process. It can promote the growth of data purchase size and all participants’ profits. An increase in privacy level will hurt the transaction, making data circulation difficult. It also causes a decrease in data purchase size and all participants’ profits. An appropriate price attenuation coefficient will increase users' willingness to purchase. What’s more, it improves the benefits of middlemen and data buyers, but reduces the benefits of data owners. Among the three factors, the impact of data timeliness on the results is the greatest, followed by the privacy level, while the impact of the price attenuation coefficient on all participants is the smallest.

In the future, the transaction details of railway data can be added, considering more influential factors and transaction scenarios. We should also continue to optimize the railway data pricing model, and constantly promote the in-depth study of railway data pricing. In that way we can do more theoretical research for the value release of data elements in the railway field.

The highlights of this study lie are as follows:

1) Aiming at the lack of quantitative research in the field of circulation of railway data elements, combined with the characteristics of railway ticketing data in reality, we present a quantitative pricing model of railway ticketing data. The expected profits functions of participants, data pricing strategies and data purchase strategies are described in detail.

2) Build a model based on Stackelberg game theory, which can fully consider the benefit maximization of each participant, and find the optimal variable values satisfying the constraint conditions.

3) Considering the characteristics that data products are easy to depreciate in the transaction process, a suitable attenuation coefficient is designed for the model, so that we can depict the price changes of railway data products in the transaction process. It is believed that will provide reference for actual production and life.

REFERENCES

[1] Q. Q. Xiong, and K. Tang, “Research progress on the right delimitation, exchange and pricing of data,” *Economic Perspectives*, vol. 2, pp. 143-158, 2021.

[2] A. Goldfarb, and C. Tucker, “Digital economics,” *Journal of economic literature*, vol. 57, no. 01, pp. 3-43, 2019.

[3] H. Chen, Q. C. Li, C. Z. Zhai, “Pricing and Circulation of Data Elements and Its Security Governance,” *Academic Exchange*, vol. 04, pp. 107-124, 2022.

[4] H. B. Peng, Y. J. Zhou, “Data Pricing Mechanism Status and Development Trends,” *Journal of Beijing University of Posts and Telecommunications*, vol. 01, pp. 120-125, 2019.

[5] Y. Wang, L. Huang, “A Study on Internal and External Rights Confirmation Scheme of Railway Data,” *Railway Transport and Economy*, vol. 10, pp. 92-97, 2020.

[6] C. C. Ding, “An Initial Exploration of Methods for Value Evaluation of Railway Data,” *Railway Economic Research*, vol. 04, pp. 16-23, 2023.

[7] J. Wu, X. N. Ma, D. Zou, “Value Evaluation Method for Railway Data Assets Based on AHP-FCE,” *Railway Transport and Economy*, vol. 12, pp. 80-86, 2021.

[8] W. W. Li, J. Q. Gou, W. X. Mu, “Research on Value-Driven Improvement Strategy of Railway Data Service Capability,” *Railway Transport and Economy*, vol. 9, pp. 112-118, 2023.

[9] X. W. Zhang, D. Jiang, “A Survey of Game Theory and Auction-based Data Pricing,” *Big Data Research*, vol. 7, no. 04, pp. 61-79, 2021.

[10] Yu-Hsien Liao, Chia-Hung Li, Yen-Chin Chen, Li-Yang Tsai, Yu-Chen Hsu, and Chih-Kuan Chen, “Agents, Activity Levels and Utility Distributing Mechanism: Game-theoretical Viewpoint,” IAENG International Journal of Applied Mathematics, vol. 51, no. 4, pp867-873, 2021.

[11] N. C. Luong, D. T. Hoang, P. Wang, D. Niyato, D. I. Kim, Z. Han, “Data Collection and Wireless Communication in Internet of Things (IoT) using economic analysis and pricing models: A survey,” *IEEE Communications Surveys & Tutorials*, vol. 18, no. 04, pp. 2546-2590, 2016.

[12] K. Jung, S. Park, “Privacy Bargaining with Fairness: Privacy-price Negotiation System for Applying Differential Privacy in Data Market Environments,” *IEEE International Conference on Big Data (Big Data)*, pp. 1389-1394, 2019.

[13] J. M. Li, W. Li, K. Nie, “Pricing Decision Analysis for Information Services of the Internet of Things Based on Stackelberg Game,” *LISS 2012: Proceedings of 2nd International Conference on Logistics, Informatics and Service Science*, Springer Berlin Heidelberg, pp. 1097-1104, 2013.

[14] K. Li, X. Y. Qi, W. H. Chen, X. Chen, Z. Zheng, “Optimal Pricing Mechanism for Data Market in blockchain-enhanced Internet of Things,” *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 9748-9761, 2019.

[15] C. Z. Xu, K. Zhu, C. Y. Yi, R. Wang, “Data Pricing for Blockchain-based Car Sharing: A Stackelberg Game Approach,” *GLOBECOM 2020-2020 IEEE Global Communications Conference*, pp. 1-5, 2020.

[16] Mei-xiang Hong, Lei Meng, “A Stackelberg Game for Retailer-Led Supply Chains Considering the Selling Effort in a Fuzzy Environment,” IAENG International Journal of Applied Mathematics, vol. 50, no. 4, pp817-827, 2020.

[17] Y. J. Tang, C.H. Tang, “Pricing Decision Analysis for Information Services of the Internet of Things Based on Stackelberg Game,” *Changbai Journal*, vol. 01, pp. 98-106, 2022.

[18] C. H. Wang, X.D. Fang, “The Core Points of Establishing Data Property Rights System,” *Journal of Nanjing University of Posts and Telecommunications (Social Science)*, vol. 25, no. 01, pp. 19-32, 2023.

[19] M. Balazinska, B. Howe, P. Koutris, D. Suciu, P. Upadhyaya, “A discussion on pricing relational data,” *In Search of Elegance in the Theory and Practice of Computation: Essays Dedicated to Peter Buneman*, pp. 167-173, 2013.

[20] Y. T. Jiao, P. Wang, S. H. Feng, D. Niyato, “Profit maximization mechanism and data management for data analytics services,” *IEEE Internet of Things Journal*, vol. 5, no. 03, pp. 2001-2014, 2018.

[21] M. A. Alsheikh, D. Niyato, D. Leong, P. Wang, Z. Han, “Privacy management and optimal pricing in people-centric sensing,” *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 04, pp. 906-920, 2017.