Statistical Theory and Related Fields

ISSN: 2475-4269 (Print) 2475-4277 (Online) Journal homepage: www.tandfonline.com/journals/tstf20

Optimal pricing approaches for data markets in market-operated data exchanges

Yangming Lyu, Linyi Qian, Zhixin Yang, Jing Yao & Xiaochen Zuo

To cite this article: Yangming Lyu, Linyi Qian, Zhixin Yang, Jing Yao & Xiaochen Zuo (27 Nov 2025): Optimal pricing approaches for data markets in market-operated data exchanges, Statistical Theory and Related Fields, DOI: 10.1080/24754269.2025.2588857

To link to this article: https://doi.org/10.1080/24754269.2025.2588857

Published online: 27 Nov 2025.

STATISTICAL THEORY AND RELATED FIELDS
https://doi.org/10.1080/24754269.2025.2588857

Optimal pricing approaches for data markets in market-operated data exchanges

Yangming Lyu[a], Linyi Qian[b,c], Zhixin Yang[d], Jing Yao[e] and Xiaochen Zuo[f]

[a]School of Economics and Management, East China Normal University, Shanghai, People's Republic of China; [b]Key Laboratory of Advanced Theory and Application in Statistics and Data Science-MOE, School of Statistics, East China Normal University, Shanghai, People's Republic of China; [c]China Inclusive Ageing Finance Research Center, East China Normal University, Shanghai, People's Republic of China; [d]Department of Mathematical Sciences, Ball State University, Muncie, IN, USA; [e]Financial Engineering Research Center, Soochow University, Suzhou, Jiangsu, People's Republic of China; [f]Shanghai Data Exchange, Shanghai, People's Republic of China

ABSTRACT

This work contributes to the theoretical foundation for pricing in data markets and offers practical insights for managing digital data exchanges in the era of big data. We propose a structured pricing model for data exchanges transitioning from quasi-public to market-oriented operations. To address the complex dynamics among data exchanges, suppliers, and consumers, the authors develop a three-stage Stackelberg game framework. In this model, the data exchange acts as a leader setting transaction commission rates, suppliers are intermediate leaders determining unit prices, and consumers are followers making purchasing decisions. Two pricing strategies are examined: the Independent Pricing Approach (IPA) and the novel Perfectly Competitive Pricing Approach (PCPA), which accounts for competition among data providers. Using backward induction, the study derives subgame-perfect equilibria and proves the existence and uniqueness of Stackelberg equilibria under both approaches. Extensive numerical simulations are carried out in the model, demonstrating that PCPA enhances data demander utility, encourages supplier competition, increases transaction volume, and improves the overall profitability and sustainability of data exchanges. Social welfare analysis further confirms PCPA's superiority in promoting efficient and fair data markets.

ARTICLE HISTORY

Received 18 May 2025
Accepted 9 November 2025

KEYWORDS

Data exchange; data market; digital economy; perfectly competitive pricing approach; Stackelberg game

1. Introduction

In recent years, the digital economy has expanded rapidly (Ma & Zhu, 2022), and the data produced from diverse production and management activities has multiplied. Massive datasets have become a desirable commodity as the demand for training data for machine learning and external data for management decision-making increases (Munappy et al., 2022). The effective use of acquired datasets is a critical issue, which has given rise to an emerging market, the data market (Farboodi & Veldkamp, 2023). The primary idea behind the data market is to establish an online platform for data producers and users to list, purchase, and trade data (Nguyen et al., 2021). Data exchanges have emerged to promote global

CONTACT Linyi Qian lyqian@stat.ecnu.edu.cn Key Laboratory of Advanced Theory and Application in Statistics and Data Science-MOE, School of Statistics, East China Normal University, Shanghai 200062, People's Republic of China China Inclusive Ageing Finance Research Center, East China Normal University, Shanghai 200062, People's Republic of China

data trade and progress the data capitalization process by providing services such as data trading infrastructure services, data product registration, and data product trading (Große-Bley & Kostka, 2021; Todaro, 2023).

Data is seen as a virtual commodity with replicability, large amounts, and low marginal cost, so the traditional market price mechanism must be modified for an effective data market (Aaltonen et al., 2021). Many data exchanges have been established to lead and support the development of digital businesses, like the Shanghai Data Exchange and BDEX (Y. N. Li et al., 2020; Tang et al., 2020). However, the quasi-public service organization-type operation model of early data exchanges is not conducive to the sustainable development of the data market, as there is a lack of a clear profit model, and it is difficult to rely solely on policy support to maintain the long-term operation of data exchanges, which is why data exchange profitability must be improved.

The purpose of this work is to develop a structured pricing approach for data markets that is centered on data exchanges. For an effective centralized data market, the following difficulties must be overcome.

(1) Developing a profitability model for data exchanges with market-based operations to accomplish high-level objectives such as increasing operational independence and achieving sustainable development. This is because market-oriented transformation is an unavoidable solution for data exchanges to overcome the survival challenge.
(2) Develop effective negotiating and pricing approaches for data exchange-centralized data markets. Approaches to pricing for data products must be carefully designed to provide participant advantages, fairness, and incentives.

To overcome the first difficulty, we apply the Stackelberg structure for data markets with data exchanges (Li & Sethi, 2017). The Stackelberg structure is a leader-follower structure, where there is a difference in the order of action between different subjects (Sherali, 1984; Van Damme & Hurkens, 1999; Van Hoesel, 2008). Previous research has looked into the possibility of utilizing Stackelberg structures to organize carbon trading markets (Hou et al., 2024; Nie et al., 2022), which has prompted us to consider Stackelberg structures as a solution for data markets. However, an inherent challenge when we employ the Stackelberg structure in data markets involving data exchanges is that the traditional leader-follower structure may not be able to reflect the market interaction between data exchanges, data suppliers, and data demanders properly (Jiang et al., 2021; C. Li et al., 2023). To address this problem, we propose a leader-intermediate leader-follower type three-stage Stackelberg data market framework. Further work on intermediate leaders can be found in Fang et al. (2018) and He et al. (2023) and the references therein.

To solve the second difficulty, we present an efficient pricing approach for data markets centered on data exchanges. Through a game theory-based pricing approach, the profits of the data exchange, the data supplier chosen by the demander, and the data demander can all be maximized. Profits earned by data exchange can be enhanced even further by introducing incentives for competitiveness. The game pricing process can be carried out quickly and efficiently, giving advantages, fairness, and incentives to participants.

The major contributions of this study are summarized below.

(1) In this research, we use the Stackelberg model to generate a data market structure centralizing data exchange. To address the difficulties of fairness, information asymmetry,

and efficiency in data exchange-centered data markets, we suggest a three-stage Stackelberg architecture based on leaders, intermediate leaders, and followers.

(2) We suggest an optimal pricing approach for data marketplaces centralized within data exchanges. Notably, we use a Stackelberg game to jointly maximize the profitability of the data exchange, the data supplier selected by the demander, and the data demander. In this game, the data exchange serves as the leader, matching supply and demand and charging a commission for its services. Data suppliers operate as intermediate leaders, determining the unit price of their data goods. The data demander works as a follower, selecting who to trade with and how much data to purchase.

(3) Using backward induction, we first investigate the optimal amount of data to purchase in the third stage. The second stage is to investigate the pricing strategies of data suppliers. Finally, we determine the optimal commission rate to be charged by the data exchange in the first stage. We examine the competition among data providers and propose a Perfectly Competitive Pricing Approach (PCPA). We demonstrate that both the Independent Pricing Approach (IPA) and the PCPA have Stackelberg equilibria.

(4) We provide comprehensive numerical simulations to evaluate the performance of the suggested pricing approaches, and the numerical results show that our proposed PCPA is effective and efficient for data trading in data exchanges.

The remainder of this paper is organized as follows. Section 2 presents related research. In Section 3, we introduce an optimal pricing approach for the data market by establishing a Stackelberg game framework and applying backward induction to solve both the IPA and the PCPA. Section 4 then presents a detailed numerical analysis and performance evaluation of our pricing models. Finally, in Section 5, we conclude with a summary of our research findings.

2. Related work

With the advent of the big data era, data products from diverse sources have become valuable commodities. Although studies on the economics of data products are still in their early stages, numerous scholars are working on issues such as data valuation and trade systems. Several data trading market structures have been developed. Zhao et al. (2019) presented a blockchain-based fair data trading mechanism for big data markets. The mechanism uses techniques such as ring signatures, double authentication to avoid signatures, and similarity learning to ensure transactional data availability, data provider privacy, and data provider-consumer fairness. Yu et al. (2017) employed the prospect theory (PT) model in behavioural economics, with expected utility theory (EUT) as a specific example, to analyze the mobile data transaction problem under the uncertainty of future data demand. However, the above data market model may fail when used in data markets that include data exchanges, which are typically centralized and have high data demand.

Recently, various research has looked into the possibilities of non-financial solutions based on the Stackelberg architecture, centralized organizations, and multi-stage games. Lu et al. (2018) present a data-driven Stackelberg market method for coordinating power dispatch across many virtual power plants. The authors demonstrate the approach's effectiveness using a case study of a renewable energy generation project and a distribution test system in China. Liu and Li (2021) evaluate two community market models: manager-based energy markets and peer-to-peer (P2P) energy trading. The quantitative comparison of social welfare, total payments, and energy transactions reveals that the manager-based energy market supports energy transactions within the community while also allowing for trade with other communities. Dai et al. (2024) provide an three-phase techno-economic framework aimed at capitalizing on the trend of increased integration of renewable energy sources into the power system to create demand for flexible services. The framework seeks to capitalize on the adaptability of demand-side resources such as IoT appliances, battery-integrated rooftop solar panels, and smart charging systems for electric vehicles. These findings indicate that Stackelberg gaming is critical for centralized non-financial solutions. However, there is a lack of effective pricing approaches, particularly for data markets centered on data exchanges.

Currently, various pricing strategies are employed in the global data market. According to Muschalle et al. (2013), these strategies can be grouped into six major categories: free data strategy, usage-based pricing, package pricing (an advanced form of usage-based pricing; see Kantere et al., 2011; Kushal et al., 2012), uniform pricing, two-part tariff, and freemium strategy. Designing an effective pricing model for data commodities requires careful consideration of both the market structure and the specific pricing approaches adopted.

Existing research in data market pricing models has yielded a variety of findings using various data pricing strategies. Xu et al. (2023) systematically outlined the three key issues of data rights, pricing strategy, and privacy calculation to give a theoretical foundation for the development of trustworthy AI systems. On this foundation, Liu et al. (2019) developed a blockchain-enabled edge cloud architecture to effectively solve the transaction execution problem, while Niu et al. (2020) proposed a dynamic pricing model with reserve price constraints and online optimization capability for both linear and nonlinear market scenarios. However, none of the preceding research has fully examined the existence of data exchanges or the rivalry effect among data owners. As a result, this paper presents a market model centered on data exchanges, develops a structured pricing approach based on the Stackelberg game (Xiao et al., 2020), and, for the first time, introduces the profitability model of data exchanges and data supplier market competition into the data trading system, breaking through the limitations of traditional pricing strategies.

To better position our work within the current literature, we clarify its marginal contribution in comparison to the key references. To the best of our knowledge, our study is the first to use the Stackelberg game framework to simulate both data provider competition and data exchange profitability, extending prior static or two-party pricing models. In addition, we address a gap in simulating competitive pricing among data providers in centralized exchanges. To that goal, we provide a mathematical expression for expected utility. This approach improves the operational feasibility of the analysis while retaining its generality.

3. Optimal pricing approaches for the data markets

In the data market, data consumers obtain essential data by paying fees to data providers. The challenge of data pricing lies in determining the optimal price that maximizes the profits of all parties involved in the transaction. This section conceptualizes the pricing problem through the lens of a Stackelberg game framework. Within this approach, the data exchange assumes the role of the leader, establishing the pricing strategy for transaction commissions. Data providers function as intermediate leaders, determining the unit prices of their respective data offerings. Data consumers, as followers, then make purchasing decisions based on these set prices. This structured approach allows for a comprehensive analysis of the strategic interactions among the various stakeholders in the data market, ultimately helping us identify the best pricing strategies.

3.1. Framework for designing game rules

Imagine the Yangtze Data Exchange (YDE) a centralized platform operating in Shenzara's (a hypothetical country) digital economy hub. When a data transaction cycle begins, YDE first establishes its commission strategy. Once the commission structure is published, data providers such as SmartCityTech, GreenGrid Analytics, and HealthStat Solutions list their available datasets on the platform. These datasets include real-time traffic flows, smart meter energy usage, and anonymized hospital data. Each provider sets its own pricing, metadata, access conditions, and data quality indicators. Data demanders such as e-commerce platforms, logistics firms, and urban planning agencies browse the listings and assess what best fits their needs. For instance, a logistics company may purchase both traffic and energy usage data to optimize delivery schedules and reduce operational costs. Based on the metadata and pricing, they decide on a purchasing strategy. As a result of abstraction and simplification, the data market transaction structure (shown in Figure 1) is modelled to include multiple data providers, a representative data consumer, and a data exchange platform. This framework is based on two key considerations. First, non-competition among data consumers stems primarily from data's non-competitive characteristics, i.e., the same data can be used by multiple consumers at the same time without affecting each other, e.g., different enterprises can use the same dataset for analysis and decision-making to improve their competitiveness without reducing the data's value. Second, the monopolistic nature of the data exchange in the data market is caused by technical barriers (Damsgaard & Lyytinen, 1998), economies of scale, and network externalities. Data collection, storage, processing, and analysis require strong technical assistance and significant capital investment, which only a few organizations can provide, resulting in technical barriers and economies of scale. The more clients a data exchange has, the more valuable and appealing its data becomes, making it easier for large data exchanges to acquire customers and data resources while also consolidating their market position. As a result, we assume a monopolistic data exchange to properly examine its impact on the market and resource allocation efficiency.

Let $\mathcal{N}$ represent the number of data providers for a specific data market, defined as the set $\mathcal{N} = \{1, \ldots, N\}$. In this transaction model, the data exchange acts as a platform connecting data providers and the data consumer. It supplies the consumer with information about the data providers and facilitates their transactions. In exchange for this service, the data exchange charges a transaction commission at a rate of τ from the data providers.

For any given consumer (i.e., the data consumer), each data provider $i \in \mathcal{N}$ establishes an optimal price p_i for their data, while the consumer determines their purchasing strategy, specifically the quantity of data to acquire. Let the unit price and quantity of data purchased from provider i be denoted as p_i and x_i, respectively. The optimal price and quantity of data purchased from provider i are represented as p_i^* and x_i^*, respectively. The optimal commission rate set by the data exchange is τ^*, with an upper limit $\tau_{\max}$ often regulated by government authorities.

This paper denotes the minimum data quantity required by the data consumer as $x_{\min}$. Let the cost per unit of data for data provider i be c_i. Correspondingly, the data consumer sets a maximum price they are willing to pay, denoted as $p_{\max}$. Let $x := (x_1, \ldots, x_N)$ and

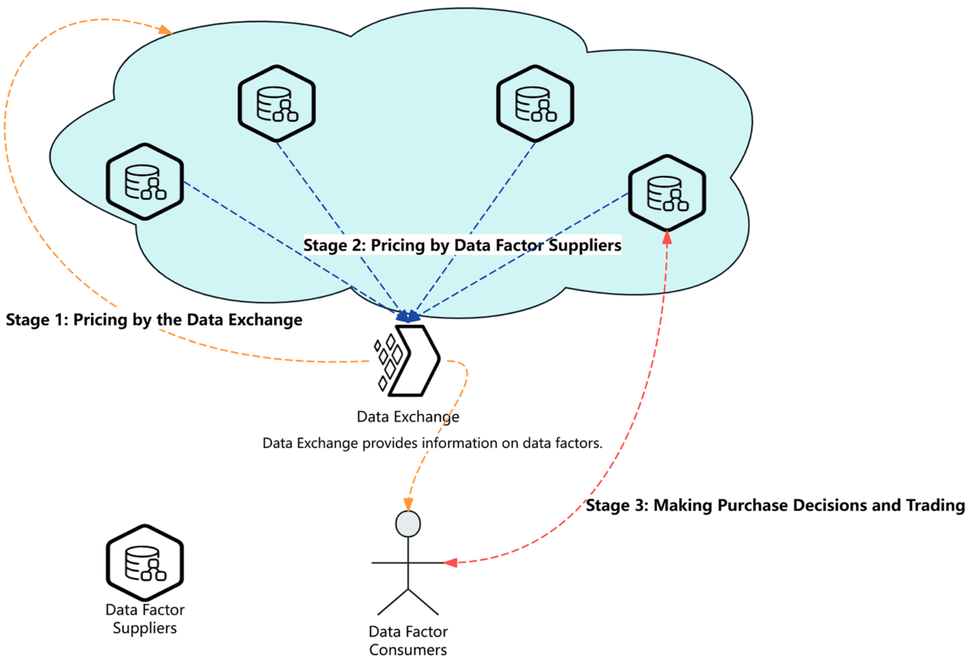

Figure 1. Optimal pricing approach based on Stackelberg game theory.

$x^* := (x_1^*, \ldots, x_N^*)$ represent the overall quantity profile and the optimal quantity profile of data purchased, respectively. Similarly, $p := (p_1, \ldots, p_N)$ and $p^* := (p_1^*, \ldots, p_N^*)$ represent the price profile and the optimal price profile.

In our approach, transactions between data consumers and data providers are facilitated by a data exchange and modelled as a three-stage Stackelberg game. In the first stage, the data exchange sets the commission rate. In the second stage, data providers establish their unit prices for the data. In the final stage, data consumers make purchasing decisions based on these prices. Figure 1 illustrates the structure of this Stackelberg game and outlines the pricing process.

In this paper, we adopt the utility function for data as defined by Jiao et al. (2018) and express it as follows:

$$q(x_i) = \alpha_1 + \alpha_2 \ln(x_i + 1), \tag{1}$$

where α_1 and α_2 (the marginal utility parameter for data) are curve-fitting parameters derived from the empirical approach. Correspondingly, the utility function for the data consumer is formulated by subtracting the price from the data utility function:

$$DU_i(x_i, p_i) = \alpha_1 + \alpha_2 \ln(x_i + 1) - x_i p_i. \tag{2}$$

Here, the data consumer determines x_i, the quantity of data to purchase from data provider i, and selects the provider based on the principle of utility maximization. Using a backward induction approach, we decompose the problem into two subgames. Problem 1 represents the data consumer's decision stage, in which the consumer responds to both the prices set by the data providers and the commission determined by the data exchange. The data consumer aims to maximize their utility by selecting the optimal quantity of data from provider i, subject to a minimum purchase quantity.

Problem 3.1 (Consumer's subgame for provider *i*):

$$\max_{x_i \geq x_{\min}} DU(x,p) = \max_{i \in \mathcal{N}} \left\{ \max_{x_i \geq x_{\min}} DU_i(x_i, p_i) \right\}, \tag{3}$$

where $DU(x,p)$ is the maximum of the function value $DU_i(x_i, p_i)$ for all individuals i in the collection $\mathcal{N}$.

Problem 3.2 corresponds to the data exchange's commission decision stage. In this stage, the data exchange seeks to maximize its utility by choosing an optimal commission rate τ. To be more specific, the utility of the data exchange is defined as its expected economic profit:

$$EU(\tau, x, p, \omega) = \tau \mathbb{E}[xp] - C = \frac{\tau}{N} \left(\sum_{i=1}^{N} \omega_i x_i p_i \right) - C, \tag{4}$$

where C is the operating cost of the data exchange and ω_i is the market competitiveness of data suppliers i. $\omega := (\omega_1, \ldots, \omega_N)$ indicates data providers' market competitiveness, which is determined by customer preferences influenced by supplier factors (reflected in transaction probabilities). To standardize each enterprise's competitiveness for comparison and analysis in a consistent quantitative dimension, we have $\sum_{i=1}^{N} \omega_i / N = 1$. To maximize its utility, the data exchange determines the commission rate τ, leading to a subgame in the optimal pricing approach for the data market.

Problem 3.2 (Exchange's subgame in pricing approach):

$$\max_{\tau \in (0, \tau_{\max}]} EU(\tau, x, p, \omega). \tag{5}$$

In our model, the pricing process unfolds over three interconnected stages. In Stage 3 (Problem 3.1), data consumers maximize their utility by selecting the optimal quantity of data from each provider, subject to a minimum purchase requirement. Their purchasing decisions are influenced by the unit prices that data providers set in Stage 2. In Stage 1 (Problem 3.2), the data exchange – acting as the leader – determines its commission rate to maximize its overall utility, a decision that in turn affects both the pricing strategies of the data providers and the purchasing choices of the consumers.

Notably, this paper examines two optimal pricing approaches that differ based on whether competition among data providers is considered: the Independent Pricing Approach (IPA) and the Perfect Competition Pricing Approach (PCPA). As a result, the optimization problem in Stage 2 varies between these two models. To proceed, we define Problems 3.3 and 3.4 to represent the data providers' pricing decision stage and outline its formulation for each approach.

3.2. *Independent pricing approach (IPA)*

We begin our analysis with the Independent Pricing Approach (IPA). In this framework, data owners, or data element providers, are equally competitive and they establish their pricing strategies independently, without consideration of the strategies employed by other providers. Therefore, competitiveness of data suppliers is equal and we denote $\omega_i = 1$, $i \in \mathcal{N}$.

The second stage of the Stackelberg game is consequently subdivided into a series of subgames involving each data provider and the demand side. The utility for data supplier i is defined as the revenue generated from sales minus the associated costs, which can be mathematically expressed as follows:

$$SU_i(\tau, x_i, p_i) = (1 - \tau)x_i p_i - x_i c_i. \tag{6}$$

To maximize their utility, provider i determines the price p_i, thereby establishing a subgame for providers within the Independent Pricing Approach (IPA) framework. This subgame can be structured as follows.

Problem 3.3 (Data suppliers' subgame in IPA):

$$\max_{p_i \in [c_i, p_{\max}]} SU_i(\tau, x_i, p_i), \quad i \in \mathcal{N}. \tag{7}$$

Problems 3.1, 3.2, and 3.3 collectively constitute the IPA Stackelberg game. The primary objective of this game is to identify the Stackelberg equilibrium, where the utility of the leader is maximized in conjunction with the optimal response strategies adopted by the followers. To proceed, we define the sufficient conditions for Equilibrium Solution (x^*, p^*, τ^*) for IPA model as below to ensure that the equilibrium solution represents a stable state where all parties in the market optimize their respective utilities under the framework of the Independent Pricing Approach.

Definition 3.1 (Independent Pricing Approach Equilibrium Solution): Let (x^*, p^*, τ^*) be defined as the equilibrium solution of the Independent Pricing Approach (IPA) if the following conditions are satisfied for all $i \in \mathcal{N}$.

- The expected utility of the data exchange at the equilibrium must be at least as great as any alternative utility derived from different commission rates:

$$EU(\tau^*, x^*, p^*, \omega) \geq EU(\tau, x^*, p^*, \omega).$$

- The utility of data provider i at the equilibrium price and quantity must be greater than or equal to their utility under any other pricing strategy:

$$SU_i(\tau^*, x_i^*, p_i^*) \geq SU_i(\tau^*, x_i^*, p_i).$$

- The utility derived by the data consumer from the equilibrium quantity and price must be at least as high as that derived from any alternative pricing:

$$DU(x^*, p^*) \geq DU(x, p^*).$$

We utilize the backward induction to examine the Independent Pricing Approach (IPA) Stackelberg game, which can be systematically decomposed into a sequence of subgames involving data element demanders, data providers, and the data exchange. Within this framework, the data exchange initiates the process by setting the commission rate. Subsequently, each data provider independently engages in a subgame with demanders, determining their pricing strategies in response to the commission rate. This structure facilitates a detailed analysis of pricing and purchasing decisions, ultimately achieved through the derivation of a Bayesian Nash equilibrium for each subgame. To be more specific, the IPA Stackelberg game is broken down into three distinct stages, starting with consumer choices in Stage 3, then supplier pricing in Stage 2, and finally the commission decision in Stage 1. (All detailed calculations for each stage are included in the Appendix 1).

Through this structured approach, the backward induction analysis offers a comprehensive framework for understanding optimal pricing and purchasing behaviour within the IPA approach, highlighting the strategic interactions between the data exchange, suppliers, and demanders.

3.3. Perfect competitive pricing approach (PCPA)

In contrast to the Independent Pricing Approach (IPA), the Perfect Competition Pricing Approach (PCPA) assumes that data providers operate in a perfectly competitive market. In this setting, individual providers have limited pricing power and must accept the market price determined by overall supply and demand dynamics.

In the data market, pricing rivalry among data suppliers must be considered. There are numerous data suppliers, and their data products or services are substitutable to some degree, so PCPA may better represent the market operation mechanism. The PCPA allows for in-depth analysis of how suppliers set pricing based on costs, market demand, and rivals' strategies, which influences the supply volume and market flow of data. This contributes to the discovery of the data market's price formation law, provides a theoretical foundation for rational pricing, and so promotes the market's healthy and orderly development.

In this pricing framework, competition among data providers is approached through their respective subgames. Each provider independently establishes their price with the objective of maximizing profits while competing with their peers for market share. It is important to note that if a data provider sets a lower price for their data offerings, they can enhance their competitiveness within the market. We define the competitiveness of provider i as follows:

$$\omega_i(p_i) = \frac{\frac{N}{p_i}}{\sum_{j=1}^{N} \frac{1}{p_j}} = \frac{N}{1 + \left(\sum_{j \neq i} \frac{1}{p_j}\right) p_i}; \tag{8}$$

this definition implies that a lower price p_i results in a higher ω_i, indicating greater competitiveness.

The purchasing willingness of data consumers is inherently linked to the competitiveness of data providers. Therefore, the utility of data provider i is defined as the product of their competitiveness and profit:

$$\Phi_i(p_i, p_{-i}, x_i) = (1 - \tau)\omega_i(p_i)x_i p_i - \omega_i(p_i)x_i c_i, \tag{9}$$

where p_{-i} is the data unit price profile of data suppliers other than data supplier i.

In the context of the Perfect Competition Pricing Approach (PCPA), data suppliers determine the unit price p to maximize their utility as defined in Equation (9). This decision process forms the subgame for providers within the PCPA framework, structured as follows.

Problem 3.4 (Data supplier's subgame in PCPA):

$$\max_{p_i \in [c_i, p_{\max}]} \Phi_i(p_i, p_{-i}, x_i), \quad i \in \mathcal{N}. \tag{10}$$

Problems 3.1, 3.2, and 3.4 collectively form the Stackelberg game within the context of the Perfect Competition Pricing Approach (PCPA). The primary objective of this game is to identify a Stackelberg equilibrium, wherein the leader's utility is maximized while the followers adopt their optimal response strategies. Similarly, we define the sufficient conditions for Equilibrium Solution (x^*, p^*, τ^*) for PCPA model as below to ensure that the equilibrium solution represents a stable state where all parties achieve optimal utility in the data market under the framework of the Perfect Competition Pricing Approach.

Definition 3.2 (Perfect Competition Pricing Approach Equilibrium Solution): Let (x^*, p^*, τ^*) be defined as the PCPA if the following conditions are satisfied for all $i \in \mathcal{N}$.

- The expected utility of the data exchange at the equilibrium must be at least as great as any alternative utility derived from different commission rates:

$$EU(\tau^*, x^*, p^*, \omega) \geq EU(\tau, x^*, p^*, \omega).$$

- The utility of data provider i at the equilibrium price and quantity must be greater than or equal to their utility under any other pricing strategy:

$$\Phi_i(p_i^*, p_{-i}^*, x_i^*) \geq \Phi_i(p_i, p_{-i}^*, x_i^*).$$

- The utility derived by the data consumer from the equilibrium quantity and price must be at least as high as that derived from any alternative pricing:

$$DU(x^*, p^*) \geq DU(x, p^*).$$

The conditions outlined above ensure that the equilibrium solution (x^*, p^*, τ^*) represents a stable state where all parties achieve optimal utility in the data market.

This paper uses the backward induction to analyze the PCPA Stackelberg game. Like the IPA model, the game is broken down into subgames involving data consumers, providers, and the data exchange, with refined Bayesian equilibria derived for each. Since the consumer subgame in PCPA mirrors that in IPA – using the same optimal response function (Equation (A3)) – we focus on the pricing strategies in the first and second stages. We present our conclusion by the following theorem.

Theorem 3.1: *There exists a unique subgame-refined Nash equilibrium in the second and first stage.*

Proof: See Appendix 2. ■

4. Numerical analysis and performance evaluation of pricing approaches

In this section, we use comprehensive numerical simulations to evaluate the performance of the proposed optimal pricing models, IPA and PCPA. We compare the performance of these two approaches, focussing especially on the advantages that competition among data suppliers brings under the PCPA model. To demonstrate the impact of various parameters on performance, we examine a scenario with a group of N data suppliers who sell a specific type of data when requested by a consumer.

For our experiments, we use the following default settings: we set $\alpha_1 = 0$ (based on the 'pure market' theoretical paradigm, which isolates the effects of internal market dynamics from external influences) and $N = 10$ (a commonly used number in market simulation focussing on pricing strategies and competition dynamics). We assume that c_i has a normal distribution, $N(10, 9)$. The variance ensures that the cost values lie within the 95% confidence interval of $[4, 16]$. This range captures the diversity seen in real markets–from low-cost automated data acquisition to high-cost manual labelling. To ensure statistical significance, we run the simulation 1000 times for each supplier's unit cost, order the results from smallest to largest to simulate supplier heterogeneity, and then take the average over all simulations.

Additionally, we analyze the parameter α_2, which represents the sensitivity of the data's utility to changes in its quantity, to understand its effect on overall performance. To reflect a range of real-world data demand situations, we change the value of α_2 between 20 and 45 to model demanders' utility sensitivity, ranging from low in typical analysis to high in data-intensive applications. Finally, we compare the performance of IPA and PCPA, highlighting the benefits of competition among data suppliers as demonstrated in the PCPA model.

Numerical simulation experiments show that, as compared to the Independent Pricing Approach (IPA), introducing competition via the Perfect Competition Pricing Approach (PCPA) can enhance transaction volume in the data market, hence, increasing demander utility. At the same time, while competition may result in lower data transaction prices, suppliers' market share will increase as the volume of data grows. Finally, we discover that the Perfect Competition Pricing Approach (PCPA) can significantly increase the profitability of data exchanges while improving the welfare of society. As a result, we feel it is a better pricing model for existing data exchanges.

4.1. Utility analysis of the data demander

We examine the utility of data demanders under both the PCPA and IPA, focussing on the effects of competition among data suppliers and the sensitivity of utility to changes in the data volume. Figure 2 shows a positive correlation between the variance in equilibrium market competitiveness ($\{\omega_i\}_{i\in\{1,2,\ldots,N\}}$) under the PCPA and the number of data suppliers. This indicates that as the number of suppliers increases, the degree of competition rises. Economically, more suppliers mean that market share is more thinly spread, reducing individual profit margins and prompting suppliers to invest more in competitive pricing. We compare the performance of the PCPA – where suppliers compete – with the IPA, in which there is no competition among suppliers.

Figure 3 illustrates that under the PCPA, market competition among suppliers leads to better outcomes for demanders, including higher utility and more efficient data usage. Suppliers are motivated to improve efficiency and lower costs, resulting in increased transaction volumes and greater benefits for all market participants. To be more specific, Figure 3 consists of two parts, each illustrating the effects of the number of data suppliers N and the marginal utility parameter α_2 on demander utility and transaction volume under the IPA and PCPA. The left part of the figure shows how the optimal utility of data demanders changes with N and α_2. Under the PCPA, demanders achieve higher utility compared to the IPA, indicating

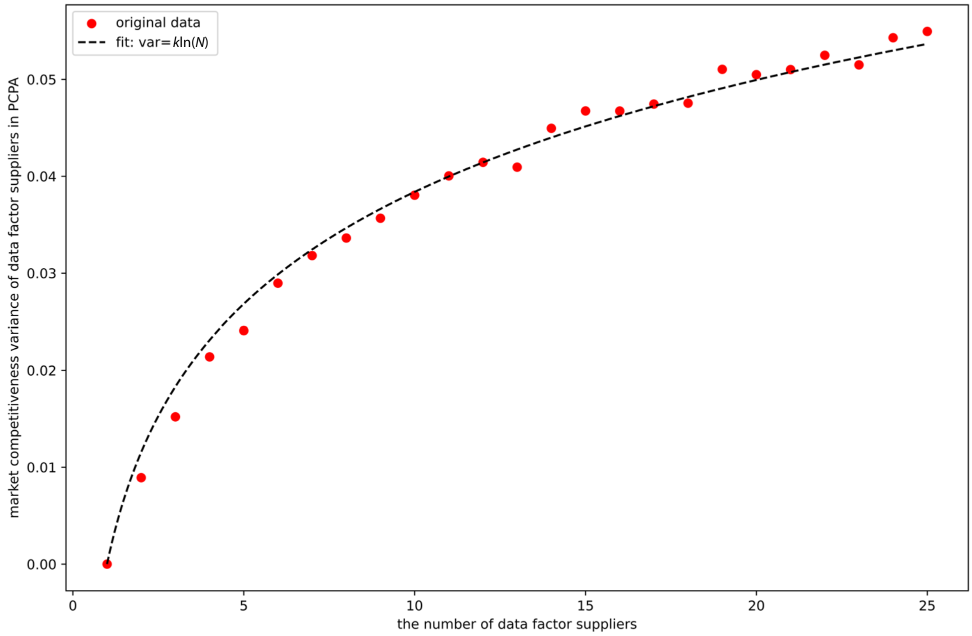

Figure 2. The relationship between the variance of market competitiveness and the number of suppliers.

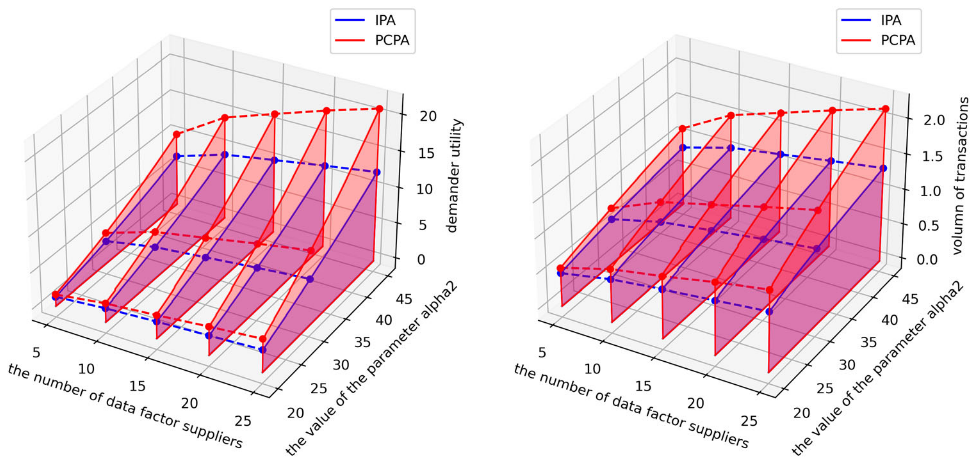

Figure 3. The impact of competition among data suppliers, as well as the marginal effect of data utility, on the data demander's optimum utility and the amount of data purchased.

that competition among suppliers benefits demanders. As α_2, i.e., the elasticity of demand for data, gradually grows, so does the utility of the demander, as data becomes increasingly valuable to them. The right part of the figure highlights the influence of N and α_2 on the total amount of data purchased. Under PCPA, the volume of data purchased increases significantly with N, driven by greater competition among suppliers. In contrast, under IPA, the volume of data remains relatively unchanged with increasing N, as suppliers do not compete. As α_2 increases, the amount of data purchased under the PCPA exceeds that of the IPA. This implies that under the PCPA, competition incentivizes suppliers to deliver more data at a cheaper price, allowing the demand side with high data demand elasticity to use the data more effectively.

4.2. Utility analysis of the chosen data supplier

Next, we look at the supply of the provider selected by the demander under the PCPA and IPA. We employ utility and data unit pricing to evaluate the supply of the provider selected by the demander.

Figure 4 illustrates how the number of data suppliers N and the marginal utility parameter α_2 affect the utility and unit price of the demander-selected supplier under the IPA and PCPA. The figure demonstrates that increased market competition under PCPA drives suppliers to lower prices to compete, benefiting demanders through reduced costs. At the same time, suppliers can achieve higher utility by selling larger volumes of data, leveraging economies of scale. As market demand elasticity α_2 rises, providers can justify higher pricing and increasing profits, highlighting the benefits of competitive market frameworks like PCPA.

The left plot focuses on supplier utility. The utility of the demander-selected supplier increases with N under both IPA and PCPA. However, the utility is significantly higher under PCPA due to enhanced market competition. In the PCPA model, competition forces suppliers to reduce their unit prices while increasing the volume of data sold. The demander-selected supplier, benefiting from greater market competitiveness, can sell more data and achieve higher profits, even at reduced prices. As α_2 increases – which reflects how sensitive demand is to price – the selected supplier gains more utility. This means that when demand becomes more responsive, the market price and the supplier's profit tend to rise as well. On the right, the unit price of data under PCPA is consistently lower than that under IPA. This difference arises because competition in the PCPA compels all suppliers to cut prices to attract demanders. However, as α_2 increases, the selected supplier can still charge a higher price. This happens because greater demand responsiveness enables them to offer data at a price that better reflects its value, capturing more profit per unit sold.

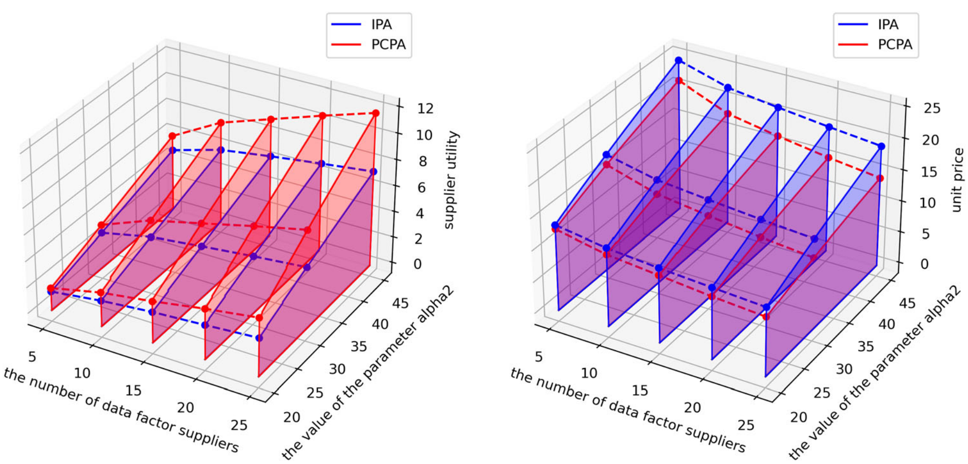

Figure 4. The impact of competition among data suppliers as well as data marginal output on the utility and unit price of the selected data supplier.

4.3. Evaluation of the data exchange's capacity to market-based operation (as measured by profitability)

We then look at how data exchanges can transition to market-based operation under the PCPA and IPA. To assess the potential of the data exchange to market-based operation, we use utility and unit revenue metrics. The unit revenue of a data exchange is defined as the product of the demander's chosen supplier's unit price and the transaction fee rate charged by the exchange, and the utility of the data exchange is computed using Equation (4).

Figure 5 looks at the utility and unit revenues of data exchanges under the IPA and PCPA. The figure demonstrates that under PCPA, data exchanges gain higher utility by adapting to market competition, even though per-unit revenues are lower. In addition, as α_2 increases, the data exchange earns more profit. This indicates that market-based models work especially well when buyers are more responsive to price changes.

The right plot examines the unit revenues of data exchanges, showing that they are lower under PCPA compared to IPA. This reduction reflects the impact of competition, as suppliers in PCPA lower their prices to remain competitive, leading to decreased per-unit revenues for the exchange. However, despite these lower unit revenues, the left plot reveals that data exchanges achieve higher overall utility under PCPA. The utility remains relatively steady with changes in the number of data suppliers under both IPA and PCPA, indicating it is largely unaffected by market competition. Nevertheless, the utility is consistently higher under PCPA, demonstrating that competition benefits the exchange by driving more transactions. As α_2 increases, the utility of the data exchange also rises under both models. This means that exchanges function more effectively in markets where buyers are highly responsive to price, reinforcing the value of adopting a market-based model in such environments. The reason for this paradox is that, under the PCPA, data exchanges can use their informational advantage as the Stackelberg game leader to adapt their commission pricing to the perfectly competitive market, allowing them to profit from competition while mitigating the negative impact of lower per-unit revenues on the data exchange.

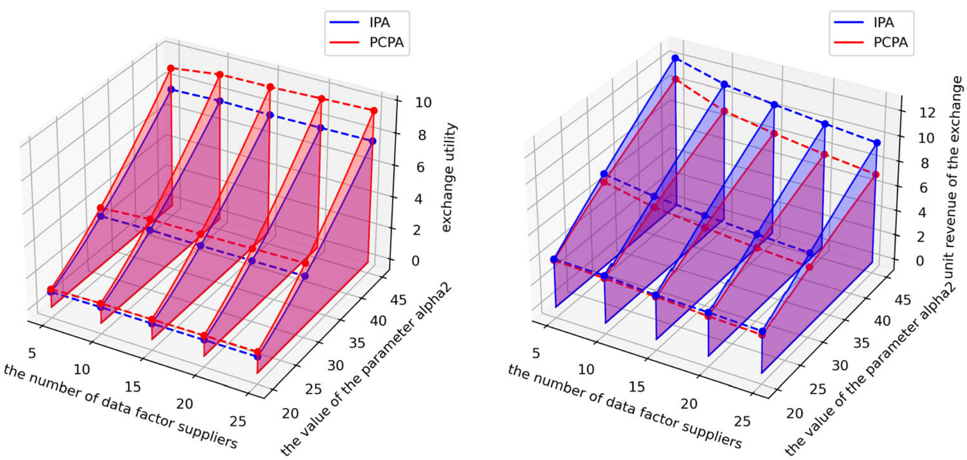

Figure 5. The effect of data supplier competition and the data marginal utility parameter on the data exchange's utility and unit revenue.

4.4. Social welfare analysis

The Social Welfare (SW) measure is an overall evaluation of the data market that reflects the pricing approaches' reliability and performance. We define the SW function for the data market as follows:

$$SW = \alpha_1 + \alpha_2 \ln(x+1) - Qxc, \tag{11}$$

where x and c represent the amount of data purchased by the data demander and the unit data cost of the data supplier selected by the demander, respectively. Q is the parameter describing the effect of cost aversion on social welfare.

Figure 6 illustrates the effect of the marginal utility parameter α_2 and the social welfare priority parameter Q on social welfare (SW) under the IPA and PCPA. SW serves as a comprehensive measure of the data market's performance and reliability, balancing utility, data usage, and cost efficiency. Below are some key observations. For a given Q, irrespective of IPA or PCPA, SW improves as α_2 grows. A greater α_2 implies increased data demand elasticity, which offsets the increased costs associated with more data usage, resulting in an overall improvement in SW. Second, the parameter Q reflects the emphasis on cost aversion in the SW function. As Q rises, the contours reflect a slowing of SW growth as the perceived increase in costs begins to undermine the beneficial impact of higher data usage. However, even for high Q, SW continues to improve with rising α_2. Last but not least, across all values of Q and α_2, SW under PCPA is consistently higher than that under IPA. This demonstrates that the PCPA approach enables more effective data allocation by fostering

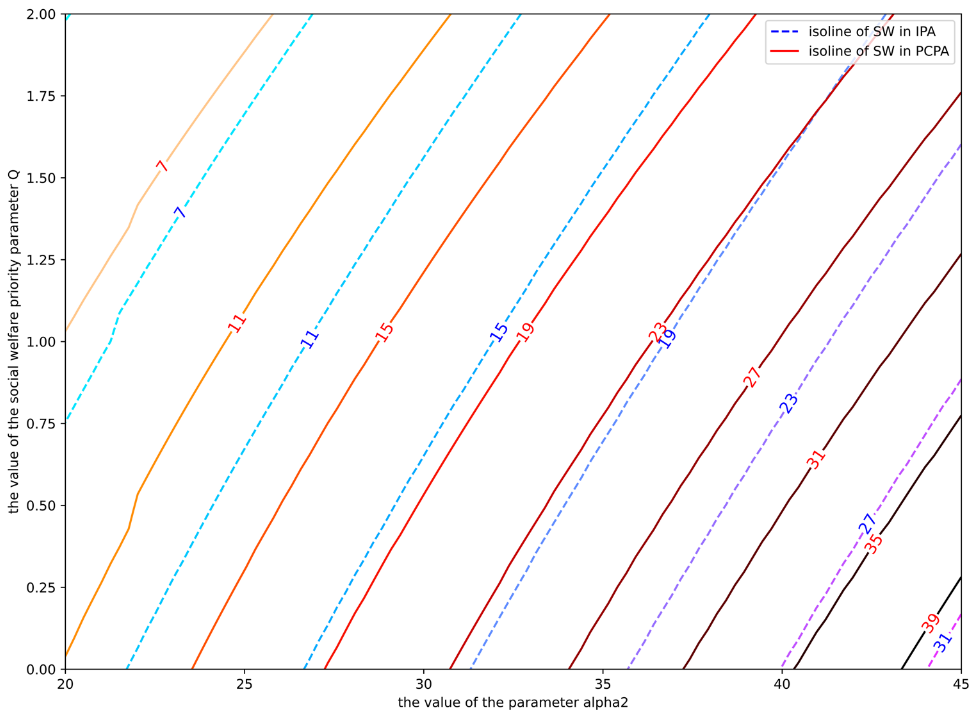

Figure 6. The effect of the data's marginal utility parameter on SW with distinct priorities (high utility or low-cost priority, as measured by Q).

competition among suppliers, which reduces integration costs and maximizes the use of data. Overall, the figure highlights that the PCPA model outperforms the IPA model in terms of social welfare, regardless of the cost aversion parameter Q. This suggests that adopting the PCPA approach can enhance the efficiency and utility of the data market while promoting better societal outcomes through improved data management and allocation.

5. Conclusions and future work

This paper successfully develops an efficient pricing approach for data markets using Stackelberg game theory. The approach seeks to optimize trade behaviour within the data exchange while balancing the interests of suppliers, the demand side, and the data exchange itself. This study, using a well-structured three-stage game framework, comprehensively addresses the issues of data supplier pricing strategies, data demander purchasing behaviours, and data exchange commission strategies, providing a holistic solution for efficiently managing data markets.

There are some limitations to this paper, most notably that it focuses solely on the revenue level of data exchanges without fully accounting for the costs and expenditures invested in developing the data market, which, to some extent, makes the pricing of data exchanges in the research results appear relatively high, with the risk of market distortion. While this error may have an impact on pricing accuracy, it does not call into question the paper's central thesis, which is the viability and general approach of Stackelberg game pricing for data exchanges. Subsequent research will focus on data exchanges' investments in data marketplaces, to offer a more exact and refined pricing approach.

Numerical simulations demonstrate that the PCPA surpasses the IPA in various dimensions. The PCPA not only lowers data prices but also optimizes the competitive environment for data suppliers, encouraging them to improve their market competitiveness and extend their market share. Furthermore, data exchanges can increase their profitability by deliberately altering transaction commission rates, which helps to marketize data exchanges. We should highlight, however, that in early markets with limited supplier rivalry or unclear laws, the IPA may be more appropriate because it assumes isolated pricing behaviour. In contrast, the PCPA becomes increasingly appropriate as markets mature, competition grows, and transparency procedures are implemented.

Based on our findings, we provide the following policy recommendations.

(1) Setting reasonable transaction commission rates: Data exchanges should set reasonable transaction commission rates to maximize profits and promote healthy competition. This will encourage more data suppliers and demanders to enter the market, increasing market efficiency through competition.
(2) Technical innovation: Data suppliers should prioritize technical innovation to lower prices and raise the marginal utility of data. As data becomes an increasingly significant factor in the digital economy, suppliers must continue to innovate to remain competitive.
(3) Selection of data products with high marginal utility: To optimize resource allocation, the demand side of data should prioritize data products with a high marginal utility. Demand-side purchasing decisions have a considerable impact on the healthy growth of the data factor market, and choosing data products with high marginal value can increase factor resource usage efficiency and foster market competition.

(4) Strengthening market regulation: To avoid unfair competition and ensure the stability and healthy development of data factor markets, government regulation must be reinforced. Governments have an important role in guaranteeing fair market behaviour, protecting consumer rights, and supervising market operations to keep markets transparent and fair.

In conclusion, this work offers theoretical direction and practical techniques for data market participants, promoting market health and efficient data resource allocation. This study brings useful knowledge to the management field by providing an optimal pricing approach. It also provides a new perspective and framework for data management and trade in the digital economy.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Natural Science Foundation of China [grant numbers 12171158, 12371474 and 12571510], Fundamental Research Funds for the Central Universities [grant number 2025ECNU-WLJC006].

References

Aaltonen, A., Alaimo, C., & Kallinikos, J. (2021). The making of data commodities: Data analytics as an embedded process. *Journal of Management Information Systems*, *38*(2), 401–429. https://doi.org/10.1080/07421222.2021.1912928

Dai, S., Mansouri, S. A., Huang, S., Alharthi, Y. Z., Wu, Y., & Bagherzadeh, L. (2024). A multi-stage techno-economic model for harnessing flexibility from IoT-enabled appliances and smart charging systems: Developing a competitive local flexibility market using Stackelberg game theory. *Applied Energy*, *373*,123868. https://doi.org/10.1016/j.apenergy.2024.123868

Damsgaard, J., & Lyytinen, K. (1998). Contours of diffusion of electronic data interchange in Finland: Overcoming technological barriers and collaborating to make it happen. *The Journal of Strategic Information Systems*, *7*(4), 275–297. https://doi.org/10.1016/S0963-8687(98)00032-8

Fang, H., Xu, L., Zou, Y., Wang, X., & Choo, K. K.. R. (2018). Three-stage Stackelberg game for defending against full-duplex active eavesdropping attacks in cooperative communication. *IEEE Transactions on Vehicular Technology*, *67*(11), 10788–10799. https://doi.org/10.1109/TVT.2018.2868900

Farboodi, M., & Veldkamp, L. (2023). Data and markets. *Annual Review of Economics*, *15*(1), 23–40. https://doi.org/10.1146/annurev-economics-082322-023244

Große-Bley, J., & Kostka, G. (2021). Big data dreams and reality in Shenzhen: An investigation of smart city implementation in China. *Big Data & Society*, *8*(2),20539517211045171. https://doi.org/10.1177/20539517211045171

Han, Z., Niyato, D., Saad, W., Başar, T., & Hjørungnes, A. (2011). *Game Theory in Wireless and Communication Networks: Theory, Models, and Applications*. Cambridge University Press.

He, W., Yao, H., Mai, T., Wang, F., & Guizani, M. (2023). Three-stage Stackelberg game enabled clustered federated learning in heterogeneous UAV swarms. *IEEE Transactions on Vehicular Technology*, *72*(7), 9366–9380. https://doi.org/10.1109/TVT.2023.3246636

Hou, H., Ge, X., Yan, Y., Lu, Y., Zhang, J., & Dong, Z. Y. (2024). An integrated energy system 'green-carbon' offset mechanism and optimization method with Stackelberg game. *Energy*, *294*,130617. https://doi.org/10.1016/j.energy.2024.130617

Jiang, Y., Zhong, Y., & Ge, X. (2021). IIoT data sharing based on blockchain: A multileader multifollower Stackelberg game approach. *IEEE Internet of Things Journal*, *9*(6), 4396–4410. https://doi.org/10.1109/JIOT.2021.3103855

Jiao, Y., Wang, P., Feng, S., & Niyato, D. (2018). Profit maximization mechanism and data management for data analytics services. *IEEE Internet of Things Journal*, *5*(3), 2001–2014.
Kantere, V., Dash, D., Gratsias, G., & Ailamaki, A. (2011). Predicting cost amortization for query services. *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data. Association for Computing Machinery*, 325-336.
Kushal, A., Moorthy, S., & Kumar, V. (2012). Pricing for data markets. [Online] https://courses.cs.washington.edu/courses/cse544/11wi/projects/kumar_kushal_moorthy.pdf
Li, C., He, A., Wen, Y., Liu, G., & Chronopoulos, A. T. (2023). Optimal trading mechanism based on differential privacy protection and Stackelberg game in big data market. *IEEE Transactions on Services Computing*, *16*(5), 3550–3563. https://doi.org/10.1109/TSC.2023.3242338
Li, T., & Sethi, S. P. (2017). A review of dynamic Stackelberg game models. *Discrete & Continuous Dynamical Systems-Series B*, *22*(1), 125–159. https://doi.org/10.3934/dcdsb.2017007
Li, Y. N., Feng, X., Xie, J., Feng, H., Guan, Z., & Wu, Q. (2020). A decentralized and secure blockchain platform for open fair data trading. *Concurrency and Computation: Practice and Experience*, *32*(7), e5578. https://doi.org/10.1002/cpe.v32.7
Liu, C., & Li, Z. (2021). Comparison of centralized and peer-to-peer decentralized market designs for community markets. *IEEE Transactions on Industry Applications*, *58*(1), 67–77. https://doi.org/10.1109/TIA.2021.3119559
Liu, K., Qiu, X., Chen, W., Chen, X., & Zheng, Z. (2019). Optimal pricing mechanism for data market in blockchain-enhanced internet of things. *IEEE Internet of Things Journal*, *6*(6), 9748–9761. https://doi.org/10.1109/JIoT.6488907
Lu, T., Wang, Z., Wang, J., Ai, Q., & Wang, C. (2018). A data-driven Stackelberg market strategy for demand response-enabled distribution systems. *IEEE Transactions on Smart Grid*, *10*(3), 2345–2357. https://doi.org/10.1109/TSG.5165411
Ma, D., & Zhu, Q. (2022). Innovation in emerging economies: Research on the digital economy driving high-quality green development. *Journal of Business Research*, *145*, 801–813. https://doi.org/10.1016/j.jbusres.2022.03.041
Munappy, A. R., Bosch, J., Olsson, H. H., Arpteg, A., & Brinne, B. (2022). Data management for production quality deep learning models: Challenges and solutions. *Journal of Systems and Software*, *191*,111359. https://doi.org/10.1016/j.jss.2022.111359
Muschalle, A., Stahl, F., Löser, A., & Vossen, G. (2013). Pricing approaches for data markets, *Enabling Real-Time Business Intelligence: 6th International Workshop, BIRTE 2012, Held at the 38th International Conference on Very Large Databases, VLDB 2012, August 27, 2012, Revised Selected Papers 6, Springer Berlin Heidelberg.*
Nguyen, L. D., Leyva-Mayorga, I., Lewis, A. N., & Popovski, P. (2021). Modeling and analysis of data trading on blockchain-based market in IoT networks. *IEEE Internet of Things Journal*, *8*(8), 6487–6497. https://doi.org/10.1109/JIOT.2021.3051923
Nie, Q., Zhang, L., & Li, S. (2022). How can personal carbon trading be applied in electric vehicle subsidies? A Stackelberg game method in private vehicles. *Applied Energy*, *313*,118855. https://doi.org/10.1016/j.apenergy.2022.118855
Niu, C., Zheng, Z., Wu, F., Tang, S., & Chen, G. (2020). Online pricing with reserve price constraint for personal data markets. *IEEE Transactions on Knowledge and Data Engineering*, *34*(4), 1928–1943.
Sherali, H. D. (1984). A multiple leader Stackelberg model and analysis. *Operations Research*, *32*(2), 390–404. https://doi.org/10.1287/opre.32.2.390
Tang, Q., Shao, Z., Huang, L., Yin, W., & Dou, Y. (2020). Identifying influencing factors for data transactions: A case study from Shanghai data exchange. *Journal of Systems Science and Systems Engineering*, *29*(6), 697–708. https://doi.org/10.1007/s11518-020-5473-1
Todaro, D. (2023). 'One City, Two Networks': Big Data and Artificial Intelligence for Smart Public Governance in Shanghai. In *Quo Vadis, Sovereignty? New Conceptual and Regulatory Boundaries in the Age of Digital China* (pp. 91–108). Springer.
Van Damme, E., & Hurkens, S. (1999). Endogenous Stackelberg leadership. *Games and Economic Behavior*, *28*(1), 105–129. https://doi.org/10.1006/game.1998.0687
Van Hoesel, S. (2008). An overview of Stackelberg pricing in networks. *European Journal of Operational Research*, *189*(3), 1393–1402. https://doi.org/10.1016/j.ejor.2006.08.064

Xiao, Z., He, D., & Du, J. (2020). A Stackelberg game pricing through balancing trilateral profits in big data market. *IEEE Internet of Things Journal*, *8*(16), 12658–12668. https://doi.org/10.1109/JIOT.2020.3001010
Xu, J., Hong, N., Xu, Z., Zhao, Z., Wu, C., Kuang, K., Wang, J., Zhu, M., Zhou, J., Ren, K., Yang, X., Lu, C., Pei, J., & Shum, H. (2023). Data-driven learning for data rights, data pricing, and privacy computing. *Engineering*, *25*, 66–76. https://doi.org/10.1016/j.eng.2022.12.008
Yu, J., Cheung, M. H., Huang, J., & Poor, H. V. (2017). Mobile data trading: Behavioral economics analysis and algorithm design. *IEEE Journal on Selected Areas in Communications*, *35*(4), 994–1005. https://doi.org/10.1109/JSAC.2017.2676958
Zhao, Y., Yu, Y., Li, Y., Han, G., & Du, X. (2019). Machine learning based privacy-preserving fair data trading in big data market. *Information Sciences*, *478*, 449–460. https://doi.org/10.1016/j.ins.2018.11.028

Appendices

Appendix 1. Detailed calculations for IPA equilibrium solution

(1) Demand-side purchasing decision in the third stage: Given the unit price of data provider i, i.e. p_i, the consumer maximizes its utility by determining its optimal buying strategy x_i^* to maximize its utility.

The first and second order derivatives of the demander's utility in Equation (2) with respect to x_i are as follows.

$$\frac{\partial DU_i}{\partial x_i} = \frac{\alpha_2}{1+x_i} - p_i; \tag{A1}$$

$$\frac{\partial^2 DU_i}{\partial x_i^2} = -\frac{\alpha_2}{(1+x_i)^2} < 0. \tag{A2}$$

These derivatives show that $DU_i(x_i, p_i)$ is a strictly upper convex function. Solving for $\frac{\partial DU_i}{\partial x_i} = 0$ can obtain the optimal response of the data demander function as follows:

$$x_i^* = \frac{\alpha_2}{p_i} - 1. \tag{A3}$$

Bringing Equation (A3) into $DU_i(x_i, p_i)$, we obtain the optimal utility of the data demander for i:

$$DU_i(x_i^*, p_i) = p_i - \alpha_2 \ln(p_i) + \alpha_1 - \alpha_2 + \alpha_2 \ln(\alpha_2). \tag{A4}$$

Based on Equation (A4), the utility of the demander under independent pricing is solved as

$$DU(x,p) = \max_{i \in N} [p_i - \alpha_2 \ln(p_i) + \alpha_1 - \alpha_2 + \alpha_2 \ln(\alpha_2)]. \tag{A5}$$

(2) The supplier's second-stage pricing strategy is as follows. Based on the demander's optimal purchase strategy for each data supplier in the third stage, the data supplier provides its utility-maximizing pricing strategy. By entering Equation (A3) into (6), the utility of data supplier i can be rewritten as

$$SU_i(\tau, x_i^*, p_i) = (1-\tau)x_i^* p_i - x_i^* c_i = (1-\tau)\alpha_2 + c_i - \left[(1-\tau)p_i + \frac{\alpha_2 c_i}{p_i}\right]. \tag{A6}$$

$SU_i(\tau, x_i^*, p_i)$'s first and second order derivatives of p_i are as follows:

$$\frac{\partial SU_i}{\partial p_i} = \frac{\alpha_2 c_i}{p_i^2} + \tau - 1, \tag{A7}$$

$$\frac{\partial^2 SU_i}{\partial p_i^2} = -\frac{2\alpha_2 c_i}{p_i^3} < 0. \tag{A8}$$

Therefore, $SU_i(\tau, x_i^*, p_i)$ is a strictly upper convex function. Solving for $\frac{\partial SU_i}{\partial p_i} = 0$ can obtain the optimal pricing strategy for the data supplier as follows:

$$p_i^* = \sqrt{\frac{\alpha_2 c_i}{1-\tau}}. \tag{A9}$$

(3) Data Exchange Pricing Strategy in Stage 1: Based on the demander's optimal response and the supplier's optimal pricing from Stages 3 and 2, the data exchange generates a pricing strategy that maximizes the utility. Filling Equations (A3) and (A9) into (4) can rewrite the utility of the data exchange as follows:

$$\begin{aligned} EU(\tau, x^*, p^*, \omega) &= \tau \mathbb{E}[x^* p^*] - C = \frac{\tau}{N}\left(\sum_{i=1}^{N} x_i^* p_i^*\right) - C \\ &= \alpha_2 \tau - \frac{\sum_{i=1}^{N}\sqrt{c_i}}{N}\sqrt{\frac{\alpha_2}{1-\tau}}\tau - C. \end{aligned} \tag{A10}$$

$EU(\tau, x^*, p^*, \omega)$'s first and second order derivatives of τ are as follows:

$$\frac{\partial EU}{\partial \tau} = \alpha_2 - \frac{\sqrt{\alpha_2}\sum_{i=1}^{N}\sqrt{c_i}}{N} \cdot \frac{1-\frac{\tau}{2}}{(1-\tau)^{\frac{3}{2}}}; \tag{A11}$$

$$\frac{\partial^2 EU}{\partial \tau^2} = -\frac{\sqrt{\alpha_2}\sum_{i=1}^{N}\sqrt{c_i}}{N} \cdot \frac{1-\frac{\tau}{4}}{(1-\tau)^{\frac{5}{2}}} < 0. \tag{A12}$$

In summary, $EU(\tau, x^*, p^*, \omega)$ is a strictly upper convex function. Solving $\frac{\partial EU}{\partial \tau} = 0$ and the optimal pricing strategy for the data exchange is obtained as follows:

$$\begin{aligned} \tau^* &= \sqrt[3]{\sqrt{\frac{27\sigma + 1}{(12\sigma)^3}} - \frac{6^3\sigma^2 + 6^2\sigma + 1}{(12\sigma)^3}} \\ &\quad + \sqrt[3]{-\sqrt{\frac{27\sigma + 1}{(12\sigma)^3}} - \frac{6^3\sigma^2 + 6^2\sigma + 1}{(12\sigma)^3}} + \frac{12\sigma - 1}{12\sigma}, \end{aligned} \tag{A13}$$

where $\sigma = \frac{\alpha_2 N^2}{\left(\sum_{i=1}^{N}\sqrt{c_i}\right)^2}$ is the auxiliary operator.

Appendix 2. Proof of Theorem 3.1

Proof: First, establish the existence of the second-stage subgame-perfect Nash equilibrium. For ease of computation, define the data-assisted unit price profile $b = \left\{b_i = \frac{1}{p_i} \mid i \in \mathcal{N}\right\}$. Correspondingly, there are data optimal auxiliary unit price profiles $b^* = \left\{b_i^* = \frac{1}{p_i^*} \mid i \in \mathcal{N}\right\}$ and the data-assisted unit price profiles of data suppliers other than data supplier i, which is $b_{-i} = \left\{b_j = \frac{1}{p_j} \mid j \neq i\right\}$. Given the demander's optimal purchasing strategy for supply i, i.e. x_i^*, then the utility of supplier i in Equation (9) can be rewritten as

$$\begin{aligned} \Phi_i(b_i, b_{-i}, x_i^*) &= \frac{N}{\sum_{j=1}^{N} b_j}(1-\tau)(\alpha_2 b_i - 1) - \frac{N b_i}{\sum_{j=1}^{N} b_j} c_i(\alpha_2 b_i - 1) \\ &= -\frac{N}{\sum_{j=1}^{N} b_j}\{\alpha_2 c_i b_i^2 - [c_i + (1-\tau)\alpha_2] b_i + 1 - \tau\}. \end{aligned} \tag{A14}$$

Define the b_i's space of strategies as $\left[\frac{1}{p_{\max}}, \frac{1}{c_i}\right]$, which is a nonempty compact subset of the Euclidean space. $\Phi_i(b_i, b_{-i}, x_i^*)$ is continuous in $\left[\frac{1}{p_{\max}}, \frac{1}{c_i}\right]$. Take the $\Phi_i(b_i, b_{-i}, x_i^*)$'s first and second order derivatives of b_i to prove its upper convexity, written as

$$\frac{\partial \Phi_i}{\partial b_i} = -\frac{N}{\left(b_i + \sum_{j\neq i} b_j\right)^2}\left\{\alpha_2 c_i b_i^2 + 2\alpha_2 c_i \left(\sum_{j\neq i} b_j\right) b_i - [c_i + (1-\tau)\alpha_2]\left(\sum_{j\neq i} b_j\right) - 1 + \tau\right\}, \tag{A15}$$

and

$$\frac{\partial^2 \Phi_i}{\partial b_i^2} = -\frac{2N}{\left(b_i + \sum_{j\neq i} b_j\right)^3}\left\{\alpha_2 c_i \left(\sum_{j\neq i} b_j\right)^2 + [c_i + (1-\tau)\alpha_2]\left(\sum_{j\neq i} b_j\right) + 1 - \tau\right\} < 0. \tag{A16}$$

Therefore, $\Phi_i(b_i, b_{-i}, x_i^*)$ is strictly convex with respect to b_i and there exists subgame-refining Nash equilibrium in the second stage. Secondly, prove the uniqueness of the subgame-perfect Nash equilibrium in the second stage. To solve $\frac{\partial \Phi_i}{\partial b_i} = 0$, there is

$$\alpha_2 c_i b_i^2 + 2\alpha_2 c_i \left(\sum_{j\neq i} b_j\right) b_i - \left\{[c_i + (1-\tau)\alpha_2]\left(\sum_{j\neq i} b_j\right) + 1 - \tau\right\} = 0. \tag{A17}$$

Therefore, the optimal auxiliary unit price for the data supplier is

$$b_i^* = \sqrt{\left(\sum_{j\neq i} b_j + \frac{(1-\tau)}{c_i}\right)\left(\sum_{j\neq i} b_j + \frac{1}{\alpha_2}\right)} - \sum_{j\neq i} b_j, \tag{A18}$$

and for the world of definition of b_i, i.e. $\left[\frac{1}{p_{\max}}, \frac{1}{c_i}\right]$, there is an optimal auxiliary pricing function $\mathcal{F}_i(b)$ for i:

$$b_i^* = \mathcal{F}_i(b) = \begin{cases} \frac{1}{c_i}, & \mathcal{E}_i > \frac{1}{c_i}, \\ \mathcal{E}_i, & \mathcal{E}_i \in \left[\frac{1}{p_{\max}}, \frac{1}{c_i}\right], \\ \frac{1}{p_{\max}}, & \mathcal{E}_i < \frac{1}{p_{\max}}, \end{cases} \tag{A19}$$

where $\mathcal{E}_i = \sqrt{\left(\sum_{j\neq i} b_j + \frac{(1-\tau)}{c_i}\right)\left(\sum_{j\neq i} b_j + \frac{1}{\alpha_2}\right)} - \sum_{j\neq i} b_j$. Recall the subgame-perfect Nash equilibrium in the second stage as $b^* = \mathcal{F}(b) = (F_1(b), F_2(b), \ldots, F_N(b))$. By proving that the optimal auxiliary pricing function for i is the standard function (Han et al., 2011), one can prove the uniqueness of the second-stage subgame-perfect Nash equilibrium.

Definition A.1: A function $\mathcal{F}(b)$ with the following properties is called a standard function:

(1) Positivity: $\mathcal{F}(b) > 0$;
(2) Monotonicity: If $b \leq b'$, then $\mathcal{F}(b) \leq \mathcal{F}(b')$;
(3) Scale reduction: $\forall \lambda > 1, \lambda\mathcal{F}(b) > \mathcal{F}(\lambda b)$.

The following proof gives that the three properties of standard functions are satisfied for $\mathcal{F}_i(b)$.
Firstly, according to Equation (A18), the $\mathcal{F}_i(b) > 0$ is obvious.
Secondly, if $b \le b'$, then $\sum_{j\neq i} b_j \le \sum_{j\neq i} b'_j$. Therefore, monotonicity can be proved by

$$\frac{\partial \mathcal{F}_i}{\partial \left(\sum_{j\neq i} b_j\right)} = \frac{1}{2}\left(\frac{\left(\sum_{j\neq i} b_j + \frac{(1-\tau)}{c_i}\right)}{\left(\sum_{j\neq i} b_j + \frac{1}{\alpha_2}\right)} + \frac{\left(\sum_{j\neq i} b_j + \frac{1}{\alpha_2}\right)}{\left(\sum_{j\neq i} b_j + \frac{(1-\tau)}{c_i}\right)}\right) - 1 > 0. \quad \text{(A20)}$$

Finally, the proof of scale-reducibility is as follows:

$$\lambda F(b) - F(\lambda b) = \sqrt{\left(\lambda \sum_{j\neq i} b_j + \frac{\lambda(1-\tau)}{c_i}\right)\left(\lambda \sum_{j\neq i} b_j + \frac{\lambda}{\alpha_2}\right)} - \sqrt{\left(\lambda \sum_{j\neq i} b_j + \frac{(1-\tau)}{c_i}\right)\left(\lambda \sum_{j\neq i} b_j + \frac{1}{\alpha_2}\right)} > 0, \quad \forall\, \lambda > 1. \quad \text{(A21)}$$

In summary, the uniqueness of the perfect Nash equilibrium of the second-stage subgame is proved by $\mathcal{F}_i(b)$ satisfying the three properties of the standard function! Its equilibrium solution p^* is satisfied:

$$p_i^* = \begin{cases} p_{\max}, & \mathcal{E}'_i > p_{\max}, \\ \mathcal{E}'_i, & \mathcal{E}'_i \in [c_i, p_{\max}], \quad i \in \mathcal{N}, \\ c_i, & \mathcal{E}'_i < c_i, \end{cases} \quad \text{(A22)}$$

where $\mathcal{E}'_i = \sqrt{\left(\sum_{j\neq i} \frac{1}{p_j^*} + \frac{(1-\tau)}{c_i}\right)\left(\sum_{j\neq i} \frac{1}{p_j^*} + \frac{1}{\alpha_2}\right)} - \sum_{j\neq i} \frac{1}{p_j^*}$.

Finally, the following procedure is utilized to solve the first-stage data exchange optimal pricing problem, and by integrating Equation (A22) into (4), the utility of the data exchange may be represented as

$$\begin{aligned} EU_t(\tau^t, x^t, p^t, \omega^t) &= \tau^t \mathbb{E}[x^t p^t] - C = \frac{\tau^t}{N}\left(\sum_{i=1}^{N} \frac{\frac{N}{p_i^t}}{\sum_{j=1}^{N} \frac{1}{p_j^t}} x_i^t p_i^t\right) - C \\ &= \alpha_2 \tau^t - \frac{N\tau^t}{\sum_{i=1}^{N} \sqrt{\left(\sum_{j\neq i} b_j^{t-1} + \frac{(1-\tau^t)}{c_i}\right)\left(\sum_{j\neq i} b_j^{t-1} + \frac{1}{\alpha_2}\right)} - \sum_{j\neq i} b_j^{t-1}} - C \\ &= \alpha_2 \tau^t - \frac{N\tau^t}{\sum_{i=1}^{N} \sqrt{B_i^{t-1} - A_i^{t-1}\tau^t}} - C, \end{aligned} \quad \text{(A23)}$$

where the auxiliary operator $A_i^{t-1} = \frac{\sum_{j\neq i} b_j^{t-1} + \frac{1}{\alpha_2}}{c_i}$ and the auxiliary operator $B_i^{t-1} = \left(\sum_{j\neq i} b_j^{t-1} + \frac{1}{c_i}\right)\left(\sum_{j\neq i} b_j^{t-1} + \frac{1}{\alpha_2}\right)$. $EU_t(\tau^t, x^t, p^t, \omega^t)$'s first and second order derivatives with respect to τ^t are as follows:

$$\frac{\partial EU_t}{\partial \tau^t} = \alpha_2 - N \frac{\sum_{i=1}^{N} \sqrt{B_i^{t-1} - A_i^{t-1}\tau^t} + \frac{A_i^{t-1}}{2\sqrt{B_i^{t-1} - A_i^{t-1}\tau^t}}}{\left(\sum_{i=1}^{N} \sqrt{B_i^{t-1} - A_i^{t-1}\tau^t}\right)^2}, \quad \text{(A24)}$$

and

$$\frac{\partial^2 EU_t}{\partial \tau^{t^2}} = -\frac{N}{\mathcal{S}_t(\tau^t)^2}\sum_{i=1}^{N}\left[\frac{A_i^{t-1}\left[A_i^{t-1}(1+2\tau^t)-2B_i^{t-1}\right]}{4(B_i^{t-1}-A_i^{t-1}\tau^t)^{\frac{3}{2}}}\right.$$

$$\left.+\frac{A_i^{t-1}}{\sqrt{B_i^{t-1}-A_i^{t-1}\tau^t}}\left(1+\frac{1}{\mathcal{S}_t(\tau^t)}\sum_{j=1}^{N}\frac{A_j^{t-1}}{2\sqrt{B_j^{t-1}-A_j^{t-1}\tau^t}}\right)\right]$$

$$\leq -N\frac{\sum_{i=1}^{N}\left[\frac{A_i^{t-1}(A_i^{t-1}-2B_i^{t-1})}{4B_i^{t-1\frac{3}{2}}}+\frac{A_i^{t-1}}{\sqrt{B_i^{t-1}}}\left(1+\frac{1}{\sum_{j=1}^{N}\sqrt{B_j^{t-1}}}\sum_{j=1}^{N}\frac{A_j^{t-1}}{2\sqrt{B_j^{t-1}}}\right)\right]}{\mathcal{S}_t(\tau^t)^2}$$

$$= -N\frac{\sum_{i=1}^{N}\left[\frac{A_i^{t-1}(A_i^{t-1}+2B_i^{t-1})}{4B_i^{t-1\frac{3}{2}}}+\frac{A_i^{t-1}}{\sqrt{B_i^{t-1}}}\left(\frac{1}{\sum_{j=1}^{N}\sqrt{B_j^{t-1}}}\sum_{j=1}^{N}\frac{A_j^{t-1}}{2\sqrt{B_j^{t-1}}}\right)\right]}{\mathcal{S}_t(\tau^t)^2} < 0, \qquad \text{(A25)}$$

where the auxiliary function $\mathcal{S}_t(\tau^t) = \sum_{i=1}^{N}\sqrt{B_i^{t-1} - A_i^{t-1}\tau^t}$. Therefore, $EU_t(\tau^t, x^t, p^t, \omega^t)$ is a strictly upper convex function. Solving for $\frac{\partial EU_t}{\partial \tau^t} = 0$ obtains the unique optimal pricing strategy for the data exchange. Also in the second stage, there exists a unique subgame-perfect Nash equilibrium, namely $\lim_{t\to\infty} A_i^t$, $\lim_{t\to\infty} B_i^t$ and $\lim_{t\to\infty}\mathcal{S}_t(\tau^t)$ all exist uniquely, and thus $\tau^* = \lim_{t\to\infty}\tau^t$ uniquely exists.

In summary, there is a single perfect Bayesian equilibrium for PCPA. ■