Mathematical Formulation of Data Pricing Models: Integrating Economic and Game-Theoretic Principles for a Comprehensive Pricing Approach

A Thesis

submitted to

Indian Institute of Science Education and Research Pune

in partial fulfilment of the requirements for the
BS-MS Dual Degree Programme

by

Durgaprasad C
20191030

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH (IISER), PUNE
Dr. Homi Bhabha Road, Pashan, Pune 411008, INDIA

March, 2025

Supervisor: Prof. Anjula Gurtoo

Certificate

This is to certify that this dissertation entitled "*Mathematical Formulation of Data Pricing Models: Integrating Economic and Game-Theoretic Principles for a Comprehensive Pricing Approach*" towards the partial fulfillment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Durgaprasad C at IISc-Bengaluru under the supervision of Dr. Anjula Gurtoo, Chairperson, Centre for Society and Policy, IISc during the academic year 2024-2025.

-Prof. Anjula Gurtoo

Thesis Assessment Committee(TAC):

Supervisor:
Prof. Anjula Gurtoo
Chairperson,
Centre for Society and Policy,
IISc Bengaluru

प्रो. अंजुला गुर्टू / Prof. Anjula Gurtoo
अध्यक्ष / Chairperson
समाज और नीति केंद्र
Centre for Society and Policy
भारतीय विज्ञान संस्थान / Indian Institute of Science
बेंगलूर / Bangalore - 560 012, भारत / India

Anindya Goswami

Expert member:
Prof. Anindya Goswami
Associate Professor,
Mathematics and Data Science,
IISER Pune

Dedication

I would like to dedicate this work to my parents—my father, **Sri Chandrashekaraiah**, and my mother, **Smt Nagarathna**—whose unwavering support and love have been my greatest strength. I also dedicate it to my sister, **Chinmayi**, and all the members of the **Dasappa family** and the **Nagamma family** whose encouragement and appreciation has always guided me.

Declaration

I hereby declare that the matter embodied in the report entitled "*Mathematical Formulation of Data Pricing Models: Integrating Economic and Game-Theoretic Principles for a Comprehensive Pricing Approach*" are the results of the work carried out by me at the Centre for Society and Policy, Indian Institute of Science, Bengaluru, under the supervision of Prof. Anjula Gurtoo and the same has not been submitted elsewhere for any other degree. Wherever others contribute, every effort is made to indicate this clearly, with due reference to the literature and acknowledgment of collaborative research and discussions.

- Durgaprasad C
20191030

Acknowledgements

I would like to express my deepest gratitude to my supervisor, **Prof. Anjula Gurtoo**, for accepting me as a master's thesis student and for her invaluable support throughout this journey.

I sincerely appreciate the insightful suggestions and feedback from my expert member of my Thesis Advisory Committee (TAC), **Prof. Anindya Goswami**, which have greatly contributed in shaping this work.

My heartfelt thanks to **Ms. Minnu Malieckal** for her continuous guidance and assistance.

I am also grateful to **Mr. Ponnanna**, **Ms. Veena Gururaja**, **Ms. Rani**, **Dr.AVS Durgaprasad**, and other members of the *Centre for Society and Policy* for their assistance and support in various aspects of my thesis and out of it.

In addition, I must mention **Mr. Deep Prakash Ghuge**, a wonderful friend, who has contributed with valuable insights and experience of his, throughout my work.

Their collective contributions have played a significant role in the successful completion of this work.

Contents

List of Tables

List of Figures

Abstract

The rapid expansion of digital platforms and data-driven economies has increased the necessity for effective strategies in data pricing, particularly for Data Exchange platforms (DEPs). As these platforms expand, identifying appropriate pricing methods becomes more challenging due to the diverse characteristics of data, differing needs of consumers, and varying market conditions. This thesis tackles a significant void in existing literature by focusing on the absence of detailed mathematical frameworks for data pricing models. By way of an extensive literature review, various parameters affecting data pricing were identified and fifteen different pricing models were reviewed. Each Formulation is progressively deduced in a stepwise approach, incorporating key economic principles such as cost structures, and data quality along with game theoretic aspects such as data quality, fluctuations in demand, market dynamics and utility theory. A comparative analysis illustrates the advantages, drawbacks, and appropriateness of each model for different types of data and market scenarios. Closing discussion on the analysis presents a decision-making framework for organizations in selecting appropriate pricing strategies. Future work involves a working research paper on the same topic where these models are applied on real world data sets to study their applicability and gather expert feedback to enhance their effectiveness, it also proposes implementation of machine learning techniques to conduct advanced operations and obtain progressive results. The research aims to provide actionable insights for data providers, intermediaries, and policymakers, supporting the development of equitable and efficient frameworks for data pricing.

Chapter 1

Introduction

Data is an intangible asset, akin to a service, yet it can be readily stored and transferred away from its point of origin, similar to a physical good (Coyle et al., 2020). Recently, the swift digitization of the economy and the advancement of the big data market have resulted in an increase in both the availability and volume of data. The world is producing about 2.5 quintillion bytes of data per day, with ninety percent of all data having been produced in just the last two years (PWC Report,2019). This exponential growth of data has led to classification of data from various perspectives; representation, source, informational content, usage, method of generation (Statistics Canada Report,2019; PWC Report,2019; Swedish National Board of Trade,2014). Data possess the potential for generating revenue and hence pricing of data or valuation of data which is the focus of the current literature review has garnered much attention among scholars. Data pricing basically means treating data as an economic good and placing a value on its worth. Formally, data pricing occurs when data owners assign a reasonable price to each dataset to facilitate its entry into digital markets (Liang et al., 2018). If a standardized model for data pricing were to be established—one that takes into account various value aspects such as data age, sample reliability, and other relevant factors—sellers would be able to optimize their pricing strategies in the market, while buyers could make informed comparisons across data service providers to secure fair pricing (Heckman et al., 2015). Three key factors highlight the necessity of pricing or valuing data. The first factor is that establishing a pricing framework for data is essential for creating a robust market or platform for data trading, which will ultimately position data as a tradable asset. Second, valuation of data is critically dependent on a number of factors like origin, quality, frequency of usage etc. Thus, developing a transparent and rigorous model of pricing is of utmost importance. Finally, targeted marketing strategy often collects data from potential customers to ameliorate their performance and often the data that is collected is referred to as personal data . With increasing awareness about personal data, data which were earlier collected free of cost can no longer be acquired in a similar fashion and hence suitable compensation mechanisms need to be designed to compensate the customers for their privacy loss.

1.1 Overview of Data Pricing

Data pricing pertains to assessing data as a valuable economic resource. In essence, it involves data owners determining a financial worth for their datasets to promote their inclusion in digital marketplaces (Liang et al. 2018). Creating a standardized framework for data pricing that considers various value factors, such as the age of the data, the reliability of samples, and other elements, would help sellers refine their pricing approaches. Additionally, this would enable buyers to perform informed comparisons among different data service providers, guaranteeing they pay a reasonable price (Heckman et al. 2015).

The importance of data valuation has surged to the point where a variety of stakeholders are tailoring models to assess the value of data and generate revenue. Players range from corporations gathering data on their products and services, organizations collecting data from the target population and third- party data aggregators. Hence the data markets prevailing nowadays are vertical and restricted. The structure of data markets is crucial in influencing data pricing, and similar to any physical product, the market structures for data can be characterized as competitive, oligopolistic, or monopolistic.

The subsequent phase of data pricing involves determining the suitable pricing strategy. (Muschalle et al. 2012) organized the different types of data pricing strategies into the following six categories – a) Free Data Pricing Strategy b) Usage-Based Pricing Strategy c) Package Pricing Strategy d) Flat Pricing Strategy e) Two-part Tariff Strategy f) Freemium Strategy. Data market structure along with the pricing strategies determine the type of data pricing model to be employed. There exist data pricing models which are classified under two major heads: (i) Economic-Based Pricing Models and (ii) Game Theory-Based Pricing Model (Liang et al.2018). Economic Based Pricing Models are those which are guided by economic principles whereas Game theory Based pricing models functions according to the canons of game theory. Furthermore, based on the economic principles of cost, consumer's perceptions economic-based pricing models can be classified as- Cost Model, Consumer Perceived Value, Supply Model, Demand Model, Differential Pricing and Dynamic Data Pricing. Based on the nature of the market and no of players, game theory- based pricing is subdivided into Non-Cooperative Game, Bargaining Game and Stackelberg Game.

1.2 Research Motivation and Problem Statement

In the ever-expanding digital economy, data has become one of the most valuable assets, but its pricing is a complex topic and an underdeveloped area of specialization, especially for environmental data. Decision-making in industries is increasingly data-driven, but without a proper mathematical framework for pricing data, inconsistencies in valuation creep in, resulting in anomalies in the trade and commercialization of data. This motivates the research so that a systematic and flexible mathematical model for data pricing may be developed, plugging in the gaps in available literature by providing practical frameworks that incorporate key economic factors such as cost, utility, market demand, and risk. By formulating a proper framework, this study will present a fair and transparent data pricing mechanism to be generalized and used in other data markets, ensuring that data providers and consumers engage in fair transactions.

Although data is gaining ground in terms of being recognized as an economic asset, the existing data pricing models are mostly based on heuristic or ad hoc methods to reach a body of trade- offs; they generally lack mathematical rigor. This applies particularly to complex and high-value datasets, such as environmental data, which need sufficiently nuanced pricing mechanisms to account for their worth. Other than this insight, the literature on data pricing is thin, having no adequate frameworks that systematically consider economic principles and market dynamics. This research will address such a gap by giving leeway to mathematical frameworks that strengthen the ability to price data more efficiently with dynamic capabilities. Integrating cost, utility, and risk elements into data pricing models, the current study will provide a comprehensive pricing framework that may be adjusted in accordance with different data characteristics in different market situations, thus contributing to developing a more structured and stable paradigm for data pricing.

1.3 Research Objectives

The objective of this research is to develop and present advanced mathematical formulations for existing data pricing models. By incorporating cost, utility, market conditions, and risk factors, the research aims to provide flexible and adaptive frameworks for diverse data types, with a specific focus on environmental datasets. Additionally, this research seeks to address the current lack of comprehensive mathematical frameworks in data pricing literature and to offer structured approaches for applying these models in both industry and academia.

1.4 Research Contributions

Development of Advanced Mathematical Formulations: The study provides innovative mathematical formulations to existing pricing for data models by taking into account cost, insight, market conditions, and risk.

Flexible and Adaptive Frameworks: The developed models provide flexible and adaptive frameworks to suit various datasets, specifically concentrating on environmental datasets. Addressing Gaps in Literature: This study attempts to fill a gap existing in the literature on data pricing by providing detailed mathematical formulations.

Practical Implications; The research confers structured ways to apply the developed models in both industrial and academic settings.

Chapter 2

Methodology

We conducted the research in three phases. The first phase comprises a bibliometric search and literature review. The next phase is to identify different economic data pricing models and underlying theories. In the third phase we derive mathematical formulations for the identified data pricing models.

2.1 Literature Review and Theoretical Background

The bibliometric search was conducted for literature dated between 1980 to 2024 yielded 76 publications. The key databases approached for the literature search are Google Scholar, ProQuest research, JSTOR, Wiley digital, Taylor & Francis, and other Web of Science publications. The selected key terms were initially obtained from the keyword list of frequently cited publications. After the selection of the first set of papers, the second set of papers is obtained by the application of snowballing approach. Several combinations of broad keywords, such as "pricing model", "data pricing", "market" AND "data pricing", "data" AND "pricing", "pricing models" AND "privacy", "information" OR "models" OR "quality" AND "data pricing" OR "data value", were used in an iterative search strategy. The search scope was limited to academic literature of peer-reviewed articles and included scientific journals, conference proceedings, doctoral thesis, and books in the English language.

Based on the article's scope and the removal of articles that addressed electronic business, wireless networks, mobile data, web-enabled application services, Internet of Things, and e-commerce, the original search of 74 publications was finalized to 48.

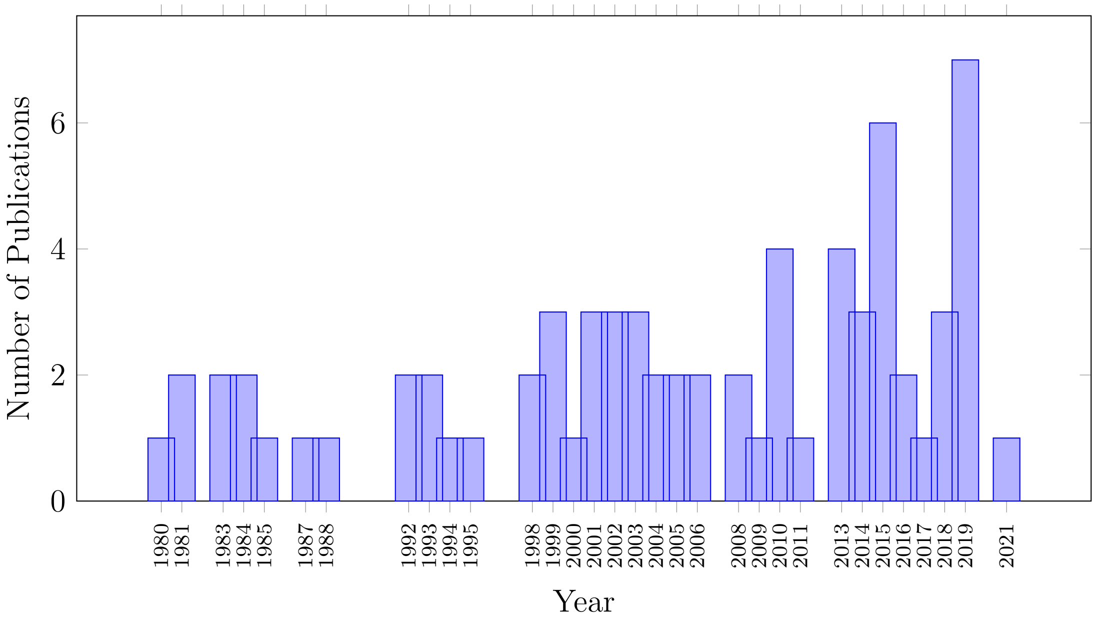

Figure 2.1: Number of Publications per Year

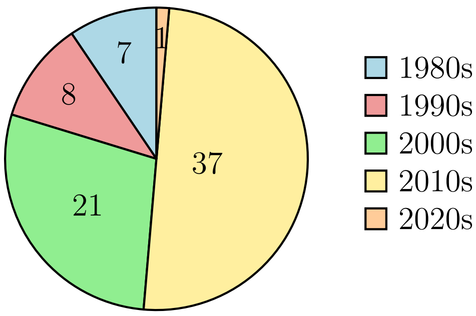

Figure 2.2: Number of Publications per Decade

2.2 Identification of Models and Parameters

2.2.1 List of 15 models extracted from various research papers

Table 2.1: Data Pricing Models, Conceptual References, and Key Parameters

Model	Conceptual References	Key Parameters/ Variables
Cost-Plus Pricing	(Varian, 2014),(Mankiw, 2021) (Laffont & Tirrole, 1993)	$C_i, \mu, U(C_i), \phi(z), F_{fixed}$
Dynamic Pricing	(Varian, 2014),(Basuchaudhary, 2006) (Borenstein & Rose, 1994)	$P_{base}, \beta_t, \gamma_t, D_t, S_t$
Premium Pricing	(Varian, 2014),(Becker et al., 1999) (Goldfarb & Tucker 2019)	$P_{base}, \alpha_i, \theta_i, V_i, Q_i$
Freemium Model	(Varian, 2014),(Becker et al., 1999) Muschalle et al. (2013)	$P_{base}, \alpha_i, U(F_i, Q_i)$
Subscription Pricing	(Becker et al., 1999) (Lambrecht & Skiera, 2006)	$\alpha_i, \gamma_i, V_i, R_i, D_i$
Two-Part Tariff Pricing	(Tirole, 1988) (Henderson & Quandt, 1980)	$\beta_i, Q_i, P_{unit}, \gamma_i, C_i$
Tiered Pricing	(Moorthy, 1984)(Mankiw, 2021) (Anderson & Dana, 2008)	$P_{base,ij}, U_{ij}, Q_{ij}, S_{ij}, T_{ij}$
Utility-Based Pricing	(Varian, 2014),(Kitchin, 2014) (Pindyck & Rubinfeld, 2013)	$U_i, k_i, \alpha_i, \beta_i$
Auction-Based Pricing	(Myerson, 1981),(Klemperer, 1999) (Cramton et al., 2006)	$B_{max}, \gamma, B_{reserve}$
Performance-Based Pricing	(Laffont & Martimort, 2001) (Lazear, 2000),(Bhattacharya et al., 1993)	$B_i, \alpha_i, \beta_i, F_{base}$
Pay-Per-Use Pricing	(Becker et al., 1999) (Tucker, 2019),(mankiw, 2021)	$P_{base}, \alpha_i, U(D_i, R_i), \gamma$
Market-Based Pricing	(Bergemann & Bonatti, 2019) (Tirole, 1988),(Dubé et al., 2010)	$P_{base}, \alpha_i, D_i, S_i, C_i$
Value-Based Pricing	(Becker et al., 1999) (Rochet & Tirole, 2003)	$P_{base}, \alpha_i, V_i, B_i, Q_i, R_i$
Location-Based Pricing	(Y. Zhang et al., 2019),(Varian, 2014) (Lee & Lee, 2015)	$P_{base,i}, \alpha_{G_i}, \beta_{S_i}, \epsilon_{D_i}, \gamma_{I_i}, \delta_{R_i}$
Loyalty-Based Pricing	(Lambrecht & Skiera, 2006) (Becker et al., 1999),(Tirole, 1988)	$P_{base}, \beta_i, F_i, T_i$

2.2.2 Summary of how they function in theoretical and practical contexts

1. **Cost-Plus Pricing**

   Cost-plus Pricing is one of the simplest pricing techniques, which aims at recovering a fixed amount of margin or markup on the cost of production or acquisition of data. The model is grounded in general microeconomic theory focusing on cost recovery and feasibility (Pindyck & Rubinfeld, 2013).

   The underlying theory is consistent with cost-based pricing strategies that have gained popularity in traditional industry areas. In reality, the firm must first compute the costs associated with collecting, cleaning, storing, and distributing their data, to which some desired profit margin is then added. This model has a high prevalence in situations with common costs for data, such as environmental data or demographic attributes in a static form.

   Its simplicity makes it suitable for data sets with clearly defined acquisition costs, whereas, it may ignore the value of data which has demand volatility or is strategically important (Muschalle et al., 2013). Cost-Plus Pricing is most appropriate for the structured type of data which has constant flowing in and out cycles and well defined serving costs.

2. **Dynamic Pricing**

   Dynamic Pricing is a model that is characterized by fluctuations depending on demand, market, and market competition. Founded in economic practices of price discrimination and market regulation (Anderson & Dana, 2008), this model incorporates cutting edge algorithms, machine learning, and data analytics to reprice.

   It is mostly applied in e-commerce, online advertising, and usage of cloud services industries where consumer demand fluctuates over time. For example, a merchant of data may set higher prices when demand is high or reduce prices when trying to increase sales during dull periods. This model is advantageous for data which are sensitive to time like current financial information, traffic information, or weather updates.

   While allowing for maximized profit and flexibility, Dynamic Pricing necessitates a robust infrastructure and constant monitoring of the market, which can make it challenging to implement effectively (Belleflamme & Peitz, 2010). It is most appropriate in data markets where there exists demand volatility coupled with fast changing customer needs.

3. **Premium Pricing** In Premium pricing technique, companies charge more than their rivals to access specific data sets, justifying it with the quality, features, and level of exclusivity to outweigh the associated costs. This approach operates on a firm-centered pricing idea that captures value based upon brand prestige and positioning, as explained by Belleflamme and Peitz(2010). The data and information targeted by these businesses tend to be specialized, high-value, proprietary, or very limited in availability. For instance, some economic analysts sell proprietary indicators, while some healthcare organizations sell premium genomic data.

   Premium Pricing is ideal for high-value data markets where the uniqueness, reliability, and quality of the information is essential. It also suits niche data sets that provide competitive or strategic information. While having the utmost revenue potential, the pricing model comes with the caveat of needing strong brand trust, as well as a highly articulated and perceptible value of the data to validate the high price (Becker et al., 1999). Premium Pricing is most suitable for high-value curated datasets intended for specialized uses such as financial modeling, medical research, or business intelligence.

4. **Freemium Model** With the freemium model, basic data can be accessed at no charge. Advanced features, premium datasets, or enhanced services come at a cost. Like most pricing approaches for a digital product, this one originated from a pricing strategy (Goldfarb and Tucker, 2019). The use of freemium content brings in a broad user base, and users are prompted to upgrade when extra value is offered. In data markets, providers may issue outlines of general datasets at no cost and only make high resolution, real-time, or customized data available at a fee.

   This model is useful in areas where financially committing upfront is not an option for many data users who first need to examine the usefulness of the dataset. Open-source research platforms are an example of those which use freemium models by offering baseline data to academics and advanced data services to commercial clients. Dubé et al. (2010) also note that the freemium model works well for the publicly available social media analytic datasets, web traffic data, or public sentiment analysis datasets.

5. **Subscription Model** The Subscription Model keeps the user's data under a locked subscription fee that needs to be paid periodically (monthly, quarterly or annually). It is based on recurring revenue

models (Lambrecht & Skiera, 2006), which guarantee revenue for the provider while offering predictable expenses for the user.

Subscription-based pricing can be found commonly in services that offer data that is often updated, such as financial data, weather information, or market analysis. For example, Farboodi & Veldkamp (2021) use subscription models to provide real-time updates for financial information. This model is more relevant in use with datasets that need active upkeep, frequent refreshes and/or sustained value over time.

Although the Subscription Model guarantees customer retention, the provider is required to continuously offer high quality relevant data in order to maintain engagement. It tends to work in the sectors that depend on continuous data for strategic decision-making like finance, marketing and supply chain management (Muschalle et al., 2013).

6. **Two-Part Tariff** The Two-Part Tariff Model charges a flat rate along with a variable rate per unit of data used. This model is based on price discrimination theory (Anderson & Dana, 2008) as it enables the access and usage value capture. For instance, a cloud data supplier may levy a set fee for accessing the platform and for every data query or download executed, an additional cost is charged.

   This system accommodates revenue security and elasticity in pricing for users with different data consumption levels. Tariffs of this kind are most successful for datasets where both access permissions and elastic utilization are required such as IoT sensor data, energy consumption records, or customer transaction logs.

   This model assists in wider penetration through lower costs of entry while maximizing profit from users with high demand. Nevertheless, to maintain customer goodwill and contentment, trust is achieved by ensuring clear usage metrics alongside transparent billing (Laffont & Tirole, 1993).

7. **Tiered Pricing** The Tiered Pricing Model skillfully balances multiple levels of pricing with feature sets, data quantities, or quality ratings. Based on the product differentiation theory, it helps data providers serve a mix of prospective buyers (Moorthy, 1984). For example, cloud service providers like AWS or Google Cloud tend to offer basic, standard, and premium plans.

   This model is mostly used in areas where consumers have highly different consumption levels which are calculated from customer analytics, satellite imagery, or geospatial information services. Customers get incentives to choose a plan that suits their budget and data needs, which in turn yields revenue increase through segmentation. It works well when the datasets that are for sale are differentiated by volume, detail, or complexity so that providers can sell data (Anderson & Dana, 2008). Effective execution of this approach depends on creating relevant tiers that are preferred by the customers while providing more value at higher price levels.

8. **Utility-Based Pricing** The Utility Based Pricing Model charges clients for the services used. This model is derived from the economic utility theory where costs are covered as per the rewards given (Varian, 2018).

   This model is most often applied by utility companies such as water and electricity supply in which the payment structure has been tailored to data services including cloud computing and IoT platforms. In data markets, this model is especially a good fit in monitored domains integrated with high-frequency and real-time on-demand data environments like everything from environmental monitoring to trading platforms and API data services.

   Utility-Based Pricing pricing structures are convenient for frequent and lower value clients as well as large corporations with enormous data utilization needs. This approach has many benefits, but the main challenge is building strong accountability mechanisms to avoid conflict related to the measurement of consumed data (Farboodi & Veldkamp, 2021).

9. **Auction-Based Pricing** The Auction-Based Pricing Model uses competitive auctions to set the prices for the data being sold. This model stems from auction theory (Myerson, 1981; Klemperer, 1999) which utilizes competition to uncover the actual worth of the data. In practice, this model is particularly favored by data exchange platforms or marketplaces where data appeal stems from either its uniqueness or scarcity. For example, Advertising data for specialized marketing and certain exclusive financial datasets are frequently sold through competitive bidding. The prevailing auction types such as English, Dutch, or sealed-bid bring different outcomes since they all affect the price based on the behavior of the bidders and the prevailing conditions.

   This model is widely used in Google Ads where advertisement placement is dependent on the amount bid for the ad. Auction-Based Pricing Models are particularly effective in datasets where value is not certain, demand is volatile, or the dataset's uniqueness poses a significant competitive challenge (Cramton et al., 2010). There must be effective auction designs in place to restrict collusion and guarantee competition if the model is to be successfully utilized.

10. **Performance-Based Pricing** With the Performance-Based Pricing Model, the price attached to a particular dataset is directly linked to the outcomes or value that was generated after its usage. The concept is backed by incentive theory (Laffont & Tirole, 1993) and performance based contracts (Lazear, 2000), where there is a theoretical guarantee that data vendors get remunerated in relation to their outputs. For instance, marketing agencies that buy predictive analytics data might be charged based on the rise in sales or increase in customer interactions.

    This is also useful in other areas such as healthcare, where the value of diagnostic data might be determined by the number of successful treatments, or in the financial services sector, where the accuracy of data has a bearing on trading results. Performance Based Pricing integrates the rewards for providers with the achievements of clients, which improves responsibility as well as the quality of data. On the downside, its success is highly reliant on having defined performance criteria and reliable tracking systems (Laffont & Martimort, 2001).

    Models which use this concept are user friendly in environments where data that directly influence measurable results is abundant, which makes its adoption in risk averse industries highly common.

11. **Pay-Per-Use Pricing** Using the Pay-Per-Use Pricing Model, users are charged fees based on how frequently they access or consume data. This model is derived from utility-based pricing strategies (Lambrecht and Skiera, 2006) because it allows for flexible payment options that are efficient in cost savings.

    This is a common practice in cloud computing services which charge users for storage and processing data or for bandwidth usage. In such data markets, environmental or IoT sensor data can be bought where consumers pay for certain time intervals or specific pieces of data. Pay-Per-Use Pricing is beneficial for consumers with nominal data needs because they would only pay for what they use. On the contrary, this model can make it challenging to estimate the cost for extensive or high-frequency data consumption. It is most appropriate for data with unpredictable demand or for firms looking to minimize costs on data expenditure (Lee & Lee, 2015). In order for this model to work, accurate usage reporting and clear pricing policies are necessary.

12. **Market-Based Pricing** The Market-Based Pricing Model sets the price of data according to market supply and prices, which is dictated by the demand for data as well as the real time market situation. This model is rooted in microeconomic concepts that focus on market equilibrium and price change (Pindyck & Rubinfeld, 2013). An example is the charging of financial market data, where clients with higher demand for the data and limited availability from the data provider lead to price increases. This specific model is employed more frequently by data marketplaces to price scarce datasets which are in high demand.

    Although very flexible and responsive, Market-Based Pricing has the downside of being very volatile, particularly for data whose relevance is changing rapidly or is difficult to ascertain. This model works very well for the data in highly competitive fields like finance, retail, or advertising, where the market is highly active (Bergemann & Bonatti, 2019). To harness the power of this active pricing technique, much focus has to be placed on providing clear value attribution and an effective trading system.

13. **Value-Based Pricing** The Value-Based Pricing Model classifies prices depending solely on the value offered to the buyer by the data. This approach focuses on the data which should correlate with a customer's value and their preferences within the data (Belleflamme & Peitz, 2010).For example, certain types of predictive analytics data that help improve marketing strategies or operational effectiveness may command a higher price due to their significance.

    Implementing Value-Based Pricing is challenging since it necessitates advanced customer segmentation and a comprehensive understanding of the market, unlike cost-based models. This strategy is advantageous for specialized datasets with higher value, such as forecasts in financial and medical research. While this model has the potential to boost revenue, it requires effective communication of the data's benefits and strong engagement with clients. To ensure that the perceived value of their datasets justifies the cost, data providers must show how their datasets support their clients' strategic goals (Goldfarb & Tucker, 2019). This technique works well in marketplaces where data significantly influences competitive advantages or business results.

14. **Location-Based Pricing**

    The Location-Based Pricing Model alters the pricing strategies of the data according to specific geographic parameters like local demand, economic activity in the area, and costs associated with data harvesting.

The essence of this model is based on concepts from formal spatial economics and behavior of consumers (Pindyck & Rubinfeld, 2013). For instance, there's higher pricing of local traffic datasets or environmental monitoring datasets from urban regions compared to rural regions.

This approach takes advantage of the projected demand for certain data and is relatively easy for logistics, urban planning, and retail industries. While offering lower prices in poorer regions, data suppliers minimize expenses and increase profits in economically developed regions.

However, such methods require extensive geospatial analysis and efficient techniques for segmentation. Location-Based Pricing is optimal for data with lower and higher relevance distributed in various geographic regions, which allows effective pricing (Kitchin, 2014).

15. **Loyalty-Based Pricing** The Loyalty-Based Pricing Model gives repeated customers, (or, long-time subscribers) discounts, incentives, or access to exclusive data. Based on relationship marketing (Lambrecht & Skiera, 2006) this model intends on customer retention through offers of financial rewards contingent on continued use. Data providers could begin using a tiered loyalty system, allowing repeat purchasers to gain access to premier datasets or preferred rates.

    This model is typical in SaaS platforms and data subscription business models, where access to evolving streams of data is king. Loyalty-Based Pricing can help you to create loyal relationships with customers, reduce churn, and grow revenues long term. This is especially helpful to industries that require a need for data on an ongoing basis like Marketing analytics, Financial forecasting, Customer behavior tracking (Dubé et al., 2010). By incentivizing repeat business, data providers create a sustainable revenue stream while ensuring loyal clients gain increasing value over time.

2.2.3 Key Parameters in Data Pricing Models

From literature, along with identifying a list of Data pricing Models, we could prepare a laundry list of key parameters which affect data pricing. We have classified them into following segments and presented them in tables.

Data Characteristics

Parameter	**References**
Age of Data	Lee & Lee, 2015; Muschalle et al., 2013
Volume of Data	Laney, 2001; Kitchin, 2014
Granularity	Zhang et al., 2019
Frequency of Updates	Farboodi et al., 2019
Format of Data	Gandomi & Haider ,2014
Completeness	Kitchin ,2014
Accuracy	Wu et al., 2014
Freshness	Heckman, 2001
Uniqueness	Laney, 2001; Gandomi & Haider, 2015
Relevance to Specific Industries	Varian, 2018
Privacy Sensitivity	Acquisti et al., 2016
Data Annotation Quality	Sheng et al., 2008
Multidimensionality of Data	Laney, 2001
Imputation of Missing Data	Zhang et al., 2019
Contextual Sensitivity of Data	Goldfarb & Tucker, 2019
Resolution of Data (Spatial/Temporal)	Kitchin, 2014
Sample Size Representativeness	Wu et al., 2014
Data Format Conversions	Zhang et al., 2019

Table 2.2: Parameters based on Data characteristics

Market-Related Parameters

Parameter	References
Supply of Similar Data	Muschalle et al., 2013
Demand for Data	Muschalle et al., 2013
Market Power of the Data Holder	Goldfarb & Tucker, 2019
Availability of Free Alternatives	Tucker, 2019
Barriers to Entry for Competitors	Farboodi et al., 2019
Network Effects	Farboodi et al., 2019
Data Resale Value	Ghosh & Roth, 2011
Pricing in Related Markets	Bergemann & Bonatti, 2019
Level of Competition in Data Market	Tucker, 2019
Data Aggregation Possibility	Acquisti et al., 2016
Differentiation in Data Offering	Bergemann & Bonatti, 2019
Elasticity of Data Demand	Tucker, 2019
First-Mover Advantage in Data Market	Varian, 2018
Reputation of Data Supplier	Wu et al., 2014
Strategic Value of Data	Farboodi et al., 2019
Market Opacity	Acquisti et al., 2016
Growth Rate of Data Market	Ghosh & Roth, 2011

Table 2.3: Parameters based on market conditions

Data Ownership and Rights

Parameter	References
Exclusivity of Ownership	Heckman, 2001
Licensing Terms	Goldfarb & Tucker, 2019
Usage Rights	Laney, 2001
Data Ownership Clarity	Acquisti et al., 2016
Restrictions on Sharing	Goldfarb & Tucker, 2019
Intellectual Property Protections	Laffont & Tirole ,1993
Legal Frameworks Governing Data	Acquisti et al., 2016
Exclusivity of Data Usage Rights	Ghosh & Roth, 2011
Negotiability of Ownership Terms	Heckman, 2001
Transferability of Data Rights	Goldfarb & Tucker, 2019
Liability for Data Misuse	Acquisti et al., 2016

Table 2.4: Parameters based on Data rights and Ownership

User-Specific Factors

Parameter	References
End-User Industry	Laney, 2001; Goldfarb & Tucker, 2019
Willingness to Pay by User	Varian, 2018
Number of Data Consumers	Lee & Lee, 2015
Geographical Location of User	Tucker, 2019
Scale of Data Usage	Acquisti et al., 2016
Purpose of Data Use	Muschalle et al., 2013
Dependence on the Data	Laney, 2001
User Expertise Level	Muschalle et al., 2013
Capacity of User to Interpret Data	Wu et al., 2014
User Data Preferences	Laney, 2001
User's Ability to Monetize Data	Varian, 2018

Table 2.5: Parameters based on User-Specific Factors

Data Processing and Infrastructure

Parameter	References
Cost of Collecting Data	Heckman, 2001
Cost of Storing Data	Laney, 2001; Kitchin, 2014
Cost of Processing Data	Kitchin, 2014
Cost of Data Cleaning	Heckman, 2001
Infrastructure Required for Data Access	Laney, 2001
Latency in Data Retrieval	Varian ,2018
Availability of Data APIs	Gandomi & Haider, 2015
Reliability of Data Storage Infrastructure	Laney, 2001
Edge Computing Compatibility	Wu et al., 2014
Cost of Maintaining Data Integrity	Heckman, 2001
Cloud Storage Integration	Wu et al., 2014

Table 2.6: Parameters based on Data Processing and Infrastructure

Economic and Social Factors

Parameter	References
Economic Value of Insights Gained from Data	Varian, 2018
Regulation Impact	Acquisti et al., 2016
Social Good or Harm Potential	Tucker, 2019
Opportunity Cost of Not Owning the Data	Heckman, 2001
Risk of Data Obsolescence	Goldfarb & Tucker ,2019
Externalities Generated by Data Usage	Varian, 2018
Value of Data in Corporate Decision Making	Goldfarb & Tucker, 2019
Impact of Data on Economic Development	Acquisti et al., 2016
Public Perception of Data Monetization Practices	Tucker, 2019
Data's Role in Global Competitiveness	Muschalle et al., 2013

Table 2.7: Parameters based on Economic and Social Factors

Transaction and Contractual Factors

Parameter	References
Transaction Costs for Data Exchange	Heckman, 2001
Contract Duration	Goldfarb & Tucker, 2019
Non-disclosure Agreements	Acquisti et al., 2016
Data Auditing Requirements	Gandomi & Haider, 2015
Service-Level Agreement (SLA) Requirements	Wu et al., 2014
Penalties for Breach of Contract	Goldfarb & Tucker, 2019
Complexity of Contractual Terms	Goldfarb & Tucker, 2019
Renewal Flexibility in Data Contracts	Ghosh & Roth, 2011
Data Access Authorization Levels	Acquisti et al., 2016

Table 2.8: Parameters based on Transaction and Contractual Factors

Technological Factors

Parameter	References
Compatibility with Existing Systems	Heckman, 2001
Scalability of Data Systems	Kitchin, 2014
Integration with Machine Learning Models	Wu et al., 2014
Data Security Measures	Acquisti et al., 2016
Data Encryption Methods	Tucker ,2019
Interoperability of Data	Laney, 2001
Availability of Big Data Processing Tools	Gandomi & Haider, 2015
Automation of Data Collection Systems	Wu et al., 2014
Adaptability of Data to New Technologies (AI, ML)	Gandomi & Haider, 2015
Data Compression Techniques	Tucker ,2019
Data Redundancy Measures	Kitchin, 2014

Table 2.9: Parameters based on Technological Factors

Data Governance and Policy Parameters

Parameter	References
Compliance with GDPR or Other Data Protection Laws	Acquisti et al., 2016
Data Retention Policies	Acquisti et al., 2016
Transparency in Data Collection Methods	Wu et al., 2014
Ethical Concerns Related to Data Use	Tucker, 2019
Regional Data Sovereignty Policies	Acquisti et al., 2016
Fairness in Data Sharing Practices	Wu et al., 2014
Institutional Trust in Data Providers	Laney, 2001
Auditability of Data Transactions	Heckman, 2001

Table 2.10: Parameters based on Data Governance and Policy

Time-Sensitive Parameters

Parameter	References
Data Decay Rate	Heckman, 2001
Time of Data Collection	Muschalle et al., 2013
Seasonal Variations in Data Value	Muschalle et al., 2013
Real-Time Data Availability	Zhang et al., 2019
Urgency of Data for Decision Making	Heckman, 2001
Data Collection Timelines	Muschalle et al., 2013
Time Taken to Access the Data	Zhang et al., 2019

Table 2.11: Parameters based on Time-Sensitivity

Risk Factors

Parameter	References
Risk of Data Breach	Goldfarb & Tucker, 2019
Risk of Misuse	Acquisti et al., 2016
Legal Risks Associated with Data Sharing	Acquisti et al., 2016
Reputational Risk of Data Mismanagement	Tucker, 2019
Uncertainty in Data Accuracy	Wu et al., 2014
Risk of Data Manipulation	Acquisti et al., 2016
Cybersecurity Threats	Goldfarb & Tucker, 2019

Table 2.12: Parameters based on Risk Factors

Environmental and Contextual Factors

Parameter	References
Context of Data Usage	Varian, 2018
Environmental Conditions (e.g., weather data relevance)	Heckman, 2001
Environmental Regulations Impacting Data Collection	Heckman, 2001
Geopolitical Sensitivity of Data (e.g., border or military data)	Acquisti et al., 2016
Climate Change Impact on Data Relevance	Muschalle et al., 2013

Table 2.13: Parameters based on Environmental and Contextual Factors

Behavioral and Psychological Factors

Parameter	References
End-User Trust in Data Source	Acquisti et al., 2016
Perceived Data Credibility	Goldfarb & Tucker, 2019
Perception of Data as Proprietary Knowledge	Wu et al., 2014
Consumer Psychological Ownership of Data	Muschalle et al., 2013
Anchoring Bias in Data Pricing (Initial Price Heuristic)	Ghosh & Roth, 2011
User Satisfaction with Data Quality	Varian, 2018
Framing Effects in Data Presentation	Wu et al., 2014
Cognitive Load in Data Interpretation	Sheng et al., 2008
Fear of Missing Out (FOMO) on Data Insights	Tucker, 2019
End-User Risk Aversion to Data Usage	Acquisti et al., 2016

Table 2.14: Parameters based on Behavioral and Psychological Factors

Miscellaneous

Parameter	References
Historical Data Trends	Muschalle et al., 2013
Potential for Future Data Monetization	Farboodi et al., 2019
Data Visualization Capabilities	Zhang et al., 2019
External Validation or Certification of Data Quality	Laney, 2001
Ability to Anonymize Data	Acquisti et al., 2016
Data Traceability	Laney, 2001
Historical Relevance of Data	Muschalle et al., 2013
Impact of Data on Long-Term Strategic Planning	Varian, 2018
Legacy Data Systems Compatibility	Kitchin, 2014

Table 2.15: Parameters under Miscellaneous

The above identified key parameters play a significant role in configuring data pricing strategies. Each table highlighted various factors influencing the pricing decisions and helps a more structured approach in composing effective pricing models.

In the following section, we combine the above parameters into mathematical ideas to obtain formulations for selected pricing models.

2.3 Formulation of Data Pricing Models

2.3.1 Cost-Plus Pricing

We know Cost-plus pricing involves determining the sale price by adding a specified percentage markup to production costs. The selling price is determined by applying a fixed percentage markup to the production cost(Varian, 2014).Traditional products benefit from cost-plus pricing but this method fails to address all the intricacies in data markets. The value of data to various users and existing market conditions along with potential risks require consideration according to Laffont & Tirrole (1993). The section presents a more advanced cost-plus pricing model that includes these extra elements to handle these challenges.

Base Cost Component

Let C_i denote the cost incurred in production, storage, processing and distribution expenses of the i-th unit of the data (Mankiw, 2021). The total production cost for n units is:

$$C_{\text{total}} = \sum_{i=1}^{n} C_i. \tag{2.1}$$

Markup Component

To ensure profitability, a markup rate μ is applied to the base cost, following the cost-plus principle (Tirole, 1988):

$$C_i^{\text{markup}} = C_i \cdot (1 + \mu). \tag{2.2}$$

Aggregating over all n units:

$$C_{\text{markup}} = \sum_{i=1}^{n} C_i \cdot (1 + \mu). \tag{2.3}$$

Utility-Based Adjustment

We introduce a utility function $U(C_i)$ which accounts for data variation based on contexts by modifying the base cost (Laffont & Tirrole, 1993):

$$C_i^{\text{utility}} = C_i \cdot (1 + \mu) \cdot (1 + U(C_i)). \tag{2.4}$$

Common utility functions include:

- **Linear Utility:** $U(C_i) = \lambda C_i$ (value proportional to cost) (Henderson & Quandt, 1980).
- **Logarithmic Utility:** $U(C_i) = \log(C_i + 1)$ (diminishing marginal value) (Mas-Colell et al., 1995).
- **Exponential Utility:** $U(C_i) = e^{\lambda C_i} - 1$ (steep increases for premium data) (Bergstrom & Varian, 1985).

The total cost including utility effects is:

$$C_{\text{utility}} = \sum_{i=1}^{n} C_i \cdot (1 + \mu) \cdot (1 + U(C_i)). \tag{2.5}$$

Risk Adjustment via Safety Factor

A safety factor $\phi(z)$ is introduced as cost structures are uncertain (Samuelson & Nordhaus, 2009). This function is often modeled using a z-score that accounts for fluctuations in costs. The function $\phi(z)$ is defined as:

$$\phi(z) = \sigma \cdot z, \tag{2.6}$$

where σ is the standard deviation of cost fluctuations and z represents the confidence level (Samuelson & Nordhaus, 2009). Thus the risk-adjusted cost becomes:

$$C_i^{\text{risk}} = C_i \cdot (1 + \mu) \cdot (1 + U(C_i)) \cdot (1 + \phi(z)). \tag{2.7}$$

Summing over all units:

$$C_{\text{total, risk}} = \sum_{i=1}^{n} \left(C_i \cdot (1 + \mu) \cdot (1 + U(C_i)) \cdot (1 + \phi(z)) \right). \tag{2.8}$$

Market Segmentation Adjustment

Market variations require an indicator function $1_{\text{market}(i)}$ that adjusts pricing conditions (Gibbons, 1992). We define the indicator function as:

$$1_{\text{market}(i)} = \begin{cases} 1, & \text{if data is sold in market } i \\ 0, & \text{otherwise.} \end{cases} \tag{2.9}$$

The pricing model incorporating market segmentation becomes:

$$C_{\text{market}} = \sum_{i=1}^{n} C_i \cdot (1 + \mu) \cdot (1 + U(C_i)) \cdot (1 + \phi(z)) \cdot 1_{\text{market}(i)}. \tag{2.10}$$

Fixed Cost Allocation

Fixed costs F_{fixed}, such as infrastructure and regulatory compliance, must also be accounted for ((Hart & Moore, 1988)). We introduce:

$$1_{\text{fixedcost}(j)} = \begin{cases} 1, & \text{if fixed cost applies to } j\text{th component} \\ 0, & \text{otherwise.} \end{cases} \tag{2.11}$$

Thus, total fixed costs become:

$$F_{\text{total, fixed}} = \sum_{j=1}^{m} \left(F_{\text{fixed}} \cdot 1_{\text{fixedcost}(j)} \right). \tag{2.12}$$

Final Cost-Plus Pricing Formula

Combining all components:

$$P_{\textbf{cost-plus}} = \sum_{i=1}^{n} C_i \cdot (1+\mu) \cdot (1+U(C_i)) \cdot (1+\phi(z)) \cdot 1_{\textbf{market}(i)} + \sum_{j=1}^{m} F_{\textbf{fixed}} \cdot 1_{\textbf{fixed cost}(j)} \cdot \tag{2.13}$$

Aspects such as utility adjustments, risk management, market segmentation, and fixed cost recovery have been integrated in the formula to provide a comprehensive pricing framework (Tirole, 1988; Laffont & Tirrole, 1993; Samuelson & Nordhaus, 2009).

2.3.2 Dynamic Pricing

Dynamic pricing represents a strategy that alters prices based on real-time factors including demand changes and competitive market behavior according to Varian (2014). The direct implementation of dynamic pricing models in data markets proves to be challenging despite their successful usage in industries such as airline ticketing and ride-sharing. The unique non-rival quality of data along with its diverse user value assessments and emerging regulatory issues render static pricing models and demand-only pricing solutions inadequate (Laffont & Martimort, 2001). A new dynamic pricing approach emerges from this section by combining utility-based adjustments with time-sensitive changes and risk assessments to enable adaptable market-responsive data valuation methods.

Base Price Component

Let P_{base} be the base price of the data product, which serves as the initial price before any adjustments are applied. The base price may be determined based on cost considerations, historical pricing, or other factors (Mankiw, 2021).

Time-Based Adjustment

Since demand and supply fluctuate over time, a time-based adjustment factor β_t is introduced (Tirole, 1988). The adjusted price incorporating time-based effects is given by:

$$P_t = P_{\text{base}} \cdot (1 + \beta_t) \cdot 1_{\text{time}}(t), \tag{2.14}$$

where $1_{\text{time}}(t)$ is an indicator function that ensures adjustments are applied only during relevant time periods:

$$1_{\text{time}}(t) = \begin{cases} 1, & \text{if time-based adjustment applies at time } t \\ 0, & \text{otherwise.} \end{cases} \tag{2.15}$$

Demand-Based Adjustment

To capture the impact of demand fluctuations, a demand-based adjustment factor γ_t is introduced (Laffont & Martimort, 2001). The price now accounts for demand variations as follows:

$$P_t = P_{\text{base}} \cdot (1 + \beta_t) \cdot 1_{\text{time}}(t) \cdot (1 + \gamma_t \cdot 1_{\text{demand}}(t)), \tag{2.16}$$

where $1_{\text{demand}}(t)$ is an indicator function that triggers a price adjustment when demand exceeds a certain threshold:

$$1_{\text{demand}}(t) = \begin{cases} 1, & \text{if demand at time } t \text{ exceeds threshold} \\ 0, & \text{otherwise.} \end{cases} \tag{2.17}$$

Utility-Based Adjustment

Data products may have different perceived values depending on their quality, uniqueness, and availability (Mas-Colell et al., 1995). To incorporate these effects, we define a utility function $U(D_t, S_t)$ based on demand (D_t) and supply (S_t):

$$P_t = P_{\text{base}} \cdot (1 + \beta_t) \cdot 1_{\text{time}}(t) \cdot (1 + \gamma_t \cdot 1_{\text{demand}}(t)) \cdot (1 + U(D_t, S_t)). \tag{2.18}$$

Common utility functions include:

- **Logarithmic Utility:** $U(D_t, S_t) = \log(D_t + 1)$ (price increases when demand exceeds supply) (Henderson & Quandt, 1980)

- **Linear Utility:** $U(D_t, S_t) = \lambda D_t$ (diminishing price sensitivity to demand changes) (Mas-Colell et al., 1995)
- **Exponential Utility:** $U(D_t, S_t) = e^{\lambda D_t} - 1$ (highly sensitive pricing for extreme demand changes) (Bergstrom & Varian, 1985)

Final Dynamic Pricing Formula

Summing over all time periods, the final dynamic pricing model is:

$$P_{\textbf{dynamic}} = \sum_{t=1}^{T} \left(P_{\textbf{base}} \cdot (1 + \beta_t) \cdot 1_{\textbf{time}}(t) \cdot (1 + \gamma_t \cdot 1_{\textbf{demand}}(t)) \cdot (1 + U(D_t, S_t)) \right), \tag{2.19}$$

where T represents the total number of time periods considered (e.g., hours, days, or months).

This formulation certifies that the pricing process dynamically adapts to changes in demand, chronological fluctuations, and perceived utility, providing a flexible and market-driven approach to pricing digital products (Becker et al., 1999).

2.3.3 Premium Pricing

We earlier mentioned that data products receive premium pricing when they are sold at multiple price tiers based on their additional value and exclusive features. it happens through value-added features coupled with exclusivity and perceived utility according to Becker et al. (1999). Firms can effectively extract consumer surplus through segmentation in subscription-based platforms along with software services and high-value datasets (Mankiw, 2021). Below, we develop a mathematical formulation incorporating its specific components.

Base Price Component

Let P_{base} be the standard base price for the data product. This serves as the benchmark price applied to all users, including those in the *basic* tier.

$$P_{\text{base}} = \text{cost-based or reference-based initial price} \tag{2.20}$$

This base price can be determined based on *production costs, historical pricing, or competitor benchmarks* (varian, 2014).

Premium User Pricing Component

Premium users pay an additional charge beyond the base price, depending on the features they access. Let:

- α_i be the **additional charge** for the i-th premium user based on *enhanced data features.*
- $1_{\text{premium}}(i)$ be an **indicator function** that takes value 1 if the user subscribes to the premium model and 0 otherwise.

The price paid by each premium user can be expressed as:

$$P_{\text{premium},i} = (P_{\text{base}} + \alpha_i) \cdot 1_{\text{premium}}(i). \tag{2.21}$$

Aggregating over all premium users (n), the total revenue contribution from premium users is:

$$\sum_{i=1}^{n} (P_{\text{base}} + \alpha_i) \cdot 1_{\text{premium}}(i), \tag{2.22}$$

where the indicator function is defined as:

$$1_{\text{premium}}(i) = \begin{cases} 1, & \text{if user } i \text{ subscribes to the premium model} \\ 0, & \text{otherwise.} \end{cases} \tag{2.23}$$

Exclusivity Surcharge Component

Premium users may also access exclusive data, which incurs an additional price adjustment. Let:

- θ_i be the **exclusivity surcharge**, representing the extra cost premium users pay for accessing *exclusive datasets.*
- $1_{\text{exclusivity}}(i)$ be an **indicator function** that takes value 1 if the user accesses exclusive content and 0 otherwise.

The exclusivity-adjusted price for each premium user is:

$$P_{\text{excl-premium},i} = (P_{\text{base}} + \alpha_i) \cdot 1_{\text{premium}}(i) \cdot (1 + \theta_i \cdot 1_{\text{exclusivity}}(i)). \tag{2.24}$$

Summing over all premium users:

$$\sum_{i=1}^{n} (P_{\text{base}} + \alpha_i) \cdot 1_{\text{premium}}(i) \cdot (1 + \theta_i \cdot 1_{\text{exclusivity}}(i)), \tag{2.25}$$

where the indicator function is:

$$1_{\text{exclusivity}}(i) = \begin{cases} 1, & \text{if user } i \text{ accesses exclusive datasets} \\ 0, & \text{otherwise.} \end{cases} \tag{2.26}$$

Utility-Based Adjustment Component

The perceived value of data is crucial in pricing. Users may be willing to pay a higher price if the data has higher quality (Q) or greater variety (V).

We introduce a **utility function** $U(V_i, Q_i)$, capturing:

- **Variety (V_i)**: The number of unique attributes or diverse sources in the dataset.
- **Quality (Q_i)**: The accuracy, timeliness, and completeness of the dataset.

With above general utility-based adjustment:

$$P_{\text{util-premium},i} = (P_{\text{base}} + \alpha_i) \cdot 1_{\text{premium}}(i) \cdot (1 + \theta_i \cdot 1_{\text{exclusivity}}(i)) \cdot (1 + U(V_i, Q_i)). \tag{2.27}$$

Common utility functions include:

- **Linear Utility Function**:

$$U(V_i, Q_i) = \lambda_1 V_i + \lambda_2 Q_i \tag{2.28}$$

- **Logarithmic Utility Function**:

$$U(V_i, Q_i) = \log(1 + V_i) + \log(1 + Q_i) \tag{2.29}$$

- **Exponential Utility Function**:

$$U(V_i, Q_i) = e^{\lambda_3 Q_i} - 1 \tag{2.30}$$

Basic User Pricing Component

Basic users only pay the base price, with no premium or exclusivity charges. Let:

- m be the **number of basic users**.
- $1_{\text{basic}}(j)$ be an **indicator function** for basic users.

The total revenue contribution from basic users is:

$$\sum_{j=1}^{m} P_{\text{base}} \cdot 1_{\text{basic}}(j), \tag{2.31}$$

where:

$$1_{\text{basic}}(j) = \begin{cases} 1, & \text{if user } j \text{ is a basic user} \\ 0, & \text{otherwise.} \end{cases} \tag{2.32}$$

Final Premium Pricing Formula

Summing the contributions from premium and basic users, the final **premium pricing model** is:

$$P_{\text{premium}} = \sum_{i=1}^{n} \Big((P_{\text{base}}+\alpha_i)\cdot 1_{\text{premium}}(i)\cdot(1+\theta_i\cdot 1_{\text{exclusivity}}(i))\cdot(1+U(V_i, Q_i)) \Big) + \sum_{j=1}^{m} (P_{\text{base}}\cdot 1_{\text{basic}}(j)). \tag{2.33}$$

This formula ensures premium users pay a higher price based on additional features, exclusivity, and perceived utility, while basic users pay a fixed base price.

2.3.4 Freemium Pricing

Digital markets commonly adopt freemium pricing which provides basic product or service access at no cost but charges for advanced functionalities. The freemium pricing model uses user segmentation to increase revenue by turning some free users into paying subscribers (Varian 2014; Becker et al. 1999). In this section, we derive a mathematical formulation for freemium pricing, incorporating factors such as feature differentiation and user utility.

Base Price Component

Let P_{base} represent the base price for premium users, which serves as the fundamental pricing structure for the paid tier. This base price is determined by various economic factors such as marginal cost, market demand, and competitive pricing (Tirole, 1988).

User Segmentation and Indicator Functions

In the freemium pricing model, users are classified into two categories: free-tier users and premium-tier users. To mathematically distinguish between these segments, we introduce the following indicator functions:

$$1_{\text{basic}(i)} = \begin{cases} 1, & \text{if the } i\text{-th user is a free-tier user} \\ 0, & \text{otherwise,} \end{cases} \tag{2.34}$$

and

$$1_{\text{premium}(i)} = \begin{cases} 1, & \text{if the } i\text{-th user has subscribed to the premium tier} \\ 0, & \text{otherwise.} \end{cases} \tag{2.35}$$

Since free-tier users do not contribute to revenue, their term in the revenue function is multiplied by zero:

$$0 \cdot 1_{\text{basic}(i)} \tag{2.36}$$

This ensures that only premium users contribute to the final revenue calculation.

Premium Pricing Adjustments

For premium users, the final price incorporates a markup factor α_i, which reflects the added value from premium features. This factor scales the base price to account for differentiated service levels (Becker et al., 1999). The modified base price for premium users is:

$$P_{\text{base}} \cdot (1 + \alpha_i). \tag{2.37}$$

This ensures that the price charged varies depending on the premium features offered.

Utility-Based Adjustments

Premium users are charged based on the perceived value of the premium features. We introduce a utility function $U(F_i, Q_i)$ that accounts for the richness of features (F_i) and the quality of data (Q_i):

$$U(F_i, Q_i) = f(F_i, Q_i), \tag{2.38}$$

where $f(F_i, Q_i)$ represents the specific utility formulation. Common functional forms include:

- **Logarithmic Utility**: $U(F_i, Q_i) = \log(1 + F_i Q_i)$, capturing diminishing returns (Mas-Colell et al., 1995).
- **Linear Utility**: $U(F_i, Q_i) = \gamma(F_i + Q_i)$, where γ is a sensitivity parameter representing direct proportionality (Henderson & Quandt, 1980).
- **Exponential Utility**: $U(F_i, Q_i) = e^{\delta(F_i+Q_i)} - 1$, emphasizing rapid valuation increases for high-quality services (Borenstein & Rose, 1994).

Final Freemium Pricing Formula

Aggregating all the components while incorporating user segmentation indicators, the total revenue from the freemium pricing model is given by:

$$P_{\text{freemium}} = \sum_{i=1}^{n} \left(0 \cdot 1_{\text{basic}(i)} + (P_{\text{base}} \cdot (1 + \alpha_i) \cdot (1 + U(F_i, Q_i))) \cdot 1_{\text{premium}(i)}\right). \tag{2.39}$$

The derived freemium pricing model captures aspects such as User Segmentation, Feature-Based Pricing and Utility-Driven Pricing which provides a structured approach to optimizing revenue in free-mium business models while considering economic and user-based factors.

2.3.5 Subscription Pricing

Subscription pricing is a business model whereby the user pays for usage or access to data services through regularly paying a charge. While one-time payment, however, does generate steady revenue for business, with a variable number of users and consumption levels. This section builds a personalized subscription pricing scheme which builds on base charges, modifies due to the additional cost incurred, and introduces utility-based adjustments in light of individual user preferences.

Base Subscription Price Component

Let P_{base} be the base subscription price, which serves as the minimum charge for accessing the data. This base price may be determined based on cost structures, historical pricing, or market benchmarks (Tirole, 1988).

Additional Cost-Based Adjustment

To account for premium features and higher-tier subscriptions, an additional cost component C_i is introduced for the user i. The price adjustment is scaled by a parameter γ_i, capturing the cost variations due to premium access (Varian, 1995).

Utility-Based Adjustment

Subscribers derive different levels of utility from data access, which depends on factors such as usage value (V_i), renewal consistency (R_i), and data demand (D_i). To model these effects, we define a utility function:

$$U(V_i, R_i, D_i). \tag{2.40}$$

Common utility functions include:

- **Linear Utility Function:**
$$U(V_i, R_i, D_i) = \lambda V_i + \delta R_i + \theta D_i \tag{2.41}$$
(Becker et al., 1999)

- **Logarithmic Utility Function:**
$$U(V_i, R_i, D_i) = \log(1 + V_i) + \log(1 + R_i) + \log(1 + D_i) \tag{2.42}$$
(Varian, 1995)

- **Exponential Utility Function:**
$$U(V_i, R_i, D_i) = e^{(\lambda V_i + \delta R_i + \theta D_i)} - 1 \tag{2.43}$$
(Tirole, 1988)

where:

- λ represents the sensitivity of the utility function to usage value (V_i).
- δ captures the impact of renewal consistency (R_i) on the utility.
- θ reflects the influence of data demand (D_i) on pricing adjustments.

These formulations ensure that pricing adjustments align with user engagement and data value.

Subscription Indicator Function

To include only active subscribers in the pricing model, we define an indicator function:

$$1_{\text{sub}}(i) = \begin{cases} 1, & \text{if user } i \text{ is an active subscriber} \\ 0, & \text{otherwise.} \end{cases} \tag{2.44}$$

This function ensures that revenue calculations only consider paying users (Becker et al., 1999).

Final Subscription Pricing Formula

Aggregating all users and incorporating the aforementioned factors, the total subscription revenue is given by:

$$P_{\text{sub}} = \sum_{i=1}^{N} \left((\alpha_i P_{\text{base}} + \gamma_i C_i) \cdot (1 + U(V_i, R_i, D_i)) \cdot 1_{\text{sub}}(i) \right), \tag{2.45}$$

where:

- N is the total number of subscribed users.
- α_i is a weighting factor for a user i's subscription level (e.g., basic, premium, enterprise) (Varian, 1995).

This formulation ensures a flexible and utility-driven pricing approach that adapts to user engagement and demand variations.

2.3.6 Two-Part Tariff Pricing

We know that the two-part tariff pricing structure is composed of a fixed access charge plus a usage-dependent variable component. With it, data sellers could earn stable revenues while at the same time setting pricing flexible, as a function of the consumption levels of the user, the different modifications, and the personalized tuning of (Becker et al., 1999; Varian, 2014). This section builds a formulation including base fixed charge, variable usage-dependent costs, and utility-based dynamic adjustment.

Fixed Access Fee Component

Let F_{fixed} represent the fixed access fee, ensuring a baseline revenue stream while filtering out low-value consumers. The value of F_{fixed} can be determined based on cost recovery, competitive benchmarking, or strategic pricing objectives (Tirole, 1988).

Usage-Based Cost Adjustment

The total cost for an individual user i includes a usage-dependent charge, scaled by a weighting factor β_i, to account for variations in consumption patterns:

$$C_{usage,i} = \beta_i Q_i P_{unit}, \tag{2.46}$$

where:

- Q_i represents the quantity of data consumed by user i.
- P_{unit} is the price per unit of data consumed.
- β_i is a weighting factor that adjusts unit pricing based on the user segment (e.g., enterprise users may receive volume discounts) (Mas-Colell et al., 1995).

To incorporate additional costs for premium access or personalized services, we introduce:

$$C_{premium,i} = \gamma_i C_i, \tag{2.47}$$

where:

- C_i represents the additional cost component for premium features or personalized data services.
- γ_i is a scaling parameter adjusting the additional cost based on service levels (Henderson & Quandt, 1980).

Utility-Based Adjustment

To capture variations in willingness to pay, demand elasticity, and loyalty effects, we introduce a utility function:

$$U(V_i, D_i, L_i), \tag{2.48}$$

where:

- V_i represents the perceived value of the data for user i.

- D_i captures demand elasticity, reflecting fluctuations in usage patterns.
- L_i represents the user's loyalty factor, influencing price incentives (Laffont & Martimort, 2001).

Common utility functions include:

- **Linear Utility:**

$$U(V_i, D_i, L_i) = \alpha V_i + \delta D_i + \theta L_i \tag{2.49}$$

- **Exponential Utility:**

$$U(V_i, D_i, L_i) = e^{(\alpha V_i + \delta D_i + \theta L_i)} - 1 \tag{2.50}$$

(nonlinear impact on price changes)

- **Sigmoid Utility:**

$$U(V_i, D_i, L_i) = \frac{1}{1 + e^{-(\alpha V_i + \delta D_i + \theta L_i)}} - 0.5 \tag{2.51}$$

(smooth transition between discounts and surcharges)

Active User Indicator Function

To ensure that only active users are included in revenue calculations, we define an indicator function:

$$1_{active}(i) = \begin{cases} 1, & \text{if user } i \text{ is actively using the service} \\ 0, & \text{otherwise.} \end{cases} \tag{2.52}$$

Final Two-Part Tariff Pricing Formula

Aggregating across all active users and incorporating the previous components, the total revenue from the two-part tariff model is given by:

$$P_{TPT} = F_{fixed} + \sum_{i=1}^{n} \left((\beta_i Q_i P_{unit} + \gamma_i C_i) \cdot (1 + U(V_i, D_i, L_i)) \cdot 1_{active}(i) \right), \tag{2.53}$$

where:

- n is the total number of active users.
- F is the fixed access fee.

This formulation ensures a structured and flexible pricing model that adapts to user engagement, consumption levels, and service personalization.

2.3.7 Tiered Pricing

We established that the different tiers of pricing can be seen as the way of charging for data where consumption, service quality, or commitment of the user is given a particular price. This allows businesses to divide customers into various groups and apply different prices based on willingness to pay and demand elasticity (Becker et al., 1999). This section presents tiered pricing as a base price model including modifications made due to utility consumption and segmentation parameters and also a form of risk adjustment pricing..

Base Price

Let $P_{\text{base},ij}$ be the base price for tier i and customer segment j. This price is set based on initial demand, industry benchmarks, and users' baseline willingness to pay (Moorthy, 1984).

Utility Function Adjustments

The utility function $U_{ij}(Q_{ij}, S_{ij}, T_{ij})$ modifies the base price based on three key elements:

- **Quantity-Based Utility (Q_{ij}):**
  Higher data consumption reduces per-unit pricing due to economies of scale (Mankiw, 2021).

- **Service Level (S_{ij}):**
  Higher-quality datasets or premium features increase the price (Anderson & Dana, 2008).

- **Subscription Length (T_{ij}):**
  Long-term commitments reduce per-month pricing to encourage retention (Moorthy, 1984).

The utility function can take different forms:

Linear Utility Function

$$U(Q_{ij}, S_{ij}, T_{ij}) = \alpha Q_{ij} + \beta S_{ij} + \gamma T_{ij} \tag{2.54}$$

This provides a direct proportional price adjustment.

Logarithmic Utility Function

$$U(Q_{ij}, S_{ij}, T_{ij}) = \log(1 + Q_{ij}) + \log(1 + S_{ij}) + \log(1 + T_{ij}) \tag{2.55}$$

Ensures diminishing returns on pricing benefits for large-scale buyers.

Exponential Utility Function

$$U(Q_{ij}, S_{ij}, T_{ij}) = e^{\alpha Q_{ij} + \beta S_{ij} + \gamma T_{ij}} - 1 \tag{2.56}$$

Magnifies price variations based on customer preferences.

The price with utility considerations will be:

$$P = P_{\text{base},ij} \cdot (1 + U_{ij}(Q_{ij}, S_{ij}, T_{ij}))\,. \tag{2.57}$$

Customer Segmentation with Indicator Function

To ensure correct pricing per segment, we define the indicator function:

$$1_{\text{tier}(i,j)} = \begin{cases} 1, & \text{if customer } j \text{ belongs to tier } i \\ 0, & \text{otherwise.} \end{cases}$$

This function guarantees that price modifications apply only to relevant customer groups (Borenstein & Rose, 1994).

Incorporating this in the pricing gives:

$$P = P_{\text{base},ij} \cdot (1 + U_{ij}(Q_{ij}, S_{ij}, T_{ij})) \cdot 1_{\text{tier}(i,j)}. \tag{2.58}$$

Risk-Adjusted Pricing Component

To account for uncertainties in data quality and market competition, a risk discounting factor is applied:

$$e^{-\lambda R_{ij}},$$

where:

- R_{ij} represents risk factors such as market volatility, data reliability concerns, and competitive pricing pressure (Tirole, 1988).

- λ is a sensitivity parameter determining how strongly risk affects pricing.

- Higher R_{ij} values lead to stronger price reductions, ensuring competitive pricing in uncertain markets (Gibbons, 1992).

Final Tiered Pricing Formula

Combining all components, the final tiered pricing equation is:

$$P_{\text{tier}} = \sum_{i=1}^{t} \sum_{j=1}^{n_i} \left(P_{\text{base},ij} \cdot (1 + U_{ij}(Q_{ij}, S_{ij}, T_{ij})) \cdot 1_{\text{tier}(i,j)} \cdot e^{-\lambda R_{ij}} \right), \quad (2.59)$$

where:

- t = Total number of pricing tiers in the model (e.g., Basic, Standard, Premium).
- n_i = Number of customer segments within the i-th tier. Different segments may have distinct data needs, willingness to pay, or risk levels.

And, between double summations:

- Outer Summation $\sum_{i=1}^{t}$ iterates over all tiers t, ensuring each pricing tier is included in the final revenue calculation.
- Inner Summation $\sum_{j=1}^{n_i}$ iterates over all customer segments n_i within each tier. Since each tier may have distinct customer groups, this step ensures that all relevant user segments are accounted for.

This formulation provides a flexible, adaptive approach to tiered pricing, balancing customer segmentation, demand elasticity, and risk mitigation.

2.3.8 Utility-Based Pricing

Utility pricing is a usage-based pricing model where users are charged according to their actual data utilization. It establishes better criteria for the distribution of costs and dynamically corrects itself according to thresholds, elasticities, and variations in demand. In contrast, unlike static pricing models, utility-based pricing adapts to individual consumption patterns, thus being more appropriate for data exchange platforms, where data is a commodity with a wide range of demand and value (Varian, 2014, Pindyck & Rubinfeld, 2013)[1].

We introduce a mathematical formulation to build on utility-based pricing and account for the thresholds(characterizing usage), elasticity-based adjustments, and non-zero base fee.

Base Structure of Utility-Based Pricing

Let P_{utility} represent the total price a customer pays for data usage over n time periods. The core principle of utility-based pricing is that the total price is determined as the sum of charges over each time period, incorporating adjustments based on consumption behavior.

We define:

$$P_{\text{utility}} = \sum_{i=1}^{n} U_i \cdot k_i, \tag{2.60}$$

where:

- U_i represents the amount of data consumed in the i-th time period (e.g., gigabytes, API calls).
- k_i is the price per unit of data for the i-th time period.

This formulation serves as the base pricing structure. However, to account for real-world pricing adjustments, we introduce surcharge factors for excessive usage and discount mechanisms for bulk consumption.

Threshold-Based Surcharge

Many pricing models impose a higher price when usage surpasses a certain threshold, discouraging excessive consumption while ensuring profitability (Pindyck & Rubinfeld, 2013). We introduce a surcharge factor α_i that increases the price when U_i exceeds a predefined threshold.

Define the indicator function:

$$1_{\text{threshold}(i)} = \begin{cases} 1, & \text{if } U_i > U_{\text{threshold}} \\ 0, & \text{otherwise,} \end{cases} \tag{2.61}$$

where $U_{\text{threshold}}$ is the predefined upper limit for normal usage. The surcharge-adjusted price per unit is then given by:

$$P_i = k_i \cdot (1 + \alpha_i \cdot 1_{\text{threshold}(i)}). \tag{2.62}$$

Thus, the total price incorporating threshold-based surcharges becomes:

$$P_{\text{utility}} = \sum_{i=1}^{n} U_i \cdot k_i \cdot (1 + \alpha_i \cdot 1_{\text{threshold}(i)}). \tag{2.63}$$

Elasticity-Based Discount

To encourage higher usage, pricing models often incorporate discounts for bulk data consumption. This concept is derived from demand elasticity, where increased consumption leads to reduced per-unit costs (Mankiw, 2021).

We introduce an elasticity adjustment factor β_i, which decreases the unit price when a user's total consumption reaches a certain level.

Define the indicator function:

$$1_{\text{elasticity}(i)} = \begin{cases} 1, & \text{if total consumption exceeds discount threshold} \\ 0, & \text{otherwise.} \end{cases} \tag{2.64}$$

The elasticity-adjusted unit price is:

$$P_i = k_i \cdot (1 + \alpha_i \cdot 1_{\text{threshold}(i)} - \beta_i \cdot 1_{\text{elasticity}(i)}), \tag{2.65}$$

where:

- β_i represents the discount rate applied when the elasticity condition is met.
- The subtraction term ensures that the per-unit price is reduced under bulk consumption.

Integrating both threshold-based surcharges and elasticity discounts, the refined formula becomes:

$$P_{\text{utility}} = \sum_{i=1}^{n} U_i \cdot k_i \cdot (1 + \alpha_i \cdot 1_{\text{threshold}(i)} - \beta_i \cdot 1_{\text{elasticity}(i)}). \tag{2.66}$$

Fixed Base Fee and Final Formula

A base fee F_{base} is often included in pricing structures to guarantee minimum revenue and cover operational costs (Henderson & Quandt, 1980). This fee is independent of usage but ensures the sustainability of the pricing model.

The final formulation incorporating the base fee is:

$$P_{\text{utility}} = \sum_{i=1}^{n} U_i \cdot k_i \cdot (1 + \alpha_i \cdot 1_{\text{threshold}(i)} - \beta_i \cdot 1_{\text{elasticity}(i)}) + F_{\text{base}}. \tag{2.67}$$

The derived formula for utility-based pricing effectively captures: Threshold-Based Adjustments,Elasticity-Based Discounts and Base Fee Inclusion.

By integrating these elements, this pricing model provides a balanced, fair, and dynamic approach suitable for data exchange platforms.

2.3.9 Auction-Based Pricing

Auction-based pricing is a widely-touted dynamic pricing technique which requires a competitive auction among several buyers to determine the actual selling price. It is customary with fields like digital marketplaces, data exchanges, and auctioning high-value assets. This assures that the data goes to the fiercely competitive who stand to get maximum utility from it, while also allowing sellers to set reserve prices so as not to sell their data too cheaply (Klemperer, 1999; Varian, 2014).Setting a mathematical model for auction-style pricing of data, this section provides for the inclusion of reserve prices and price adjustments for bidding which exceeds a certain limit. The formulation encompasses a market-oriented pricing strategy, allowing for protection of the seller.

Basic Auction Price

In a competitive bidding scenario, multiple buyers submit bids for a dataset. Let B_i be the bid submitted by the i-th buyer, and let there be n total bids in the auction. The auction price is determined by selecting the highest bid:

$$P_{\text{auction}} = \max(B_1, B_2, \ldots, B_n), \tag{2.68}$$

where:

- P_{auction} is the final price at which the data is sold.
- B_i represents the bid submitted by the i-th buyer.
- $B_{\max}$ is the highest bid among all B_i, meaning $B_{\max} = \max(B_1, B_2, \ldots, B_n)$.
- n is the total number of participants in the auction.

This ensures that the data is allocated to the highest bidder, reflecting its true market value based on consumer willingness to pay (Myerson, 1981).

Incorporating the Reserve Price

A reserve price is the minimum acceptable price set by the seller. If the highest bid $B_{\max}$ falls below this threshold, the sale does not take place. We define B_{reserve} as the reserve price,minimum acceptable price set by the seller, below which bids are not considered valid, and use an indicator function to ensure that adjustments only apply when $B_{\max}$ exceeds B_{reserve}.

We introduce the function:

$$1_{B_{\max} > B_{\text{reserve}}} = \begin{cases} 1, & \text{if } B_{\max} > B_{\text{reserve}} \\ 0, & \text{otherwise.} \end{cases} \tag{2.69}$$

This function ensures that price adjustments occur only when the highest bid exceeds the reserve price (Milgrom, 2004).

Price Adjustment and Final Formula

To further refine the pricing mechanism, we introduce a price adjustment factor γ when the highest bid exceeds the reserve price. This adjustment accounts for cases where competitive bidding signals higher willingness to pay, allowing sellers to extract additional value from the transaction.

Thus the final auction price is given by:

$$P_{\text{auction}} = B_{\max} \cdot \left(1 + \gamma \cdot 1_{B_{\max} > B_{\text{reserve}}}\right), \tag{2.70}$$

where:

- γ is a price adjustment factor that increases the price when bids exceed the reserve price.
- The indicator function $\nVdash_{B_{\max} > B_{\text{reserve}}}$ ensures that the adjustment is applied only when the highest bid surpasses the reserve threshold.

This formulation allows price flexibility while ensuring that sellers do not accept bids below their minimum acceptable price. The parameter γ can be determined based on market dynamics, such as auction type, number of bidders, and data demand trends (Mas-Colell et al., 1995).

2.3.10 Performance-Based Pricing

Performance-based pricing includes a total price that depends on how well a data service performs. Unlike static pricing models, this one incorporates incentive structures wherein high performance would be rewarded while poor performance is penalized. It derives a mathematical formulation for performance-based pricing from economic incentives and contract theory principles (Laffont & Martimort, 2001).

Base Price Component

Let B_i be the base price per unit of performance, representing the initial cost before adjustments. This base price serves as the foundation upon which performance-related modifications are applied (Bhattacharya et al., 1993). The total base price for n performance evaluation periods is given by:

$$P_{\text{base}} = \sum_{i=1}^{n} B_i, \tag{2.71}$$

where B_i is determined based on predefined contractual agreements.

Performance-Based Adjustments

To incorporate incentives, the price is modified based on performance evaluation. We consider performance-based adjustment factors that reward or penalize the price based on predefined thresholds.

We introduce α_i as a bonus multiplier that is applied when performance meets or exceeds expectations (Lazear, 2000). To capture this, we define the indicator function for success, denoted as $1_{\text{success}(i)}$, which takes the value 1 if the performance p_i exceeds a predefined threshold τ_{success}, and 0 otherwise:

$$1_{\text{success}(i)} = \begin{cases} 1, & \text{if } p_i \geq \tau_{\text{success}} \\ 0, & \text{otherwise.} \end{cases} \tag{2.72}$$

Similarly, we introduce β_i as the penalty factor applied when performance falls below expectations (Basuchaudhary, 2006). To account for this, we define the failure indicator function $1_{\text{failure}(i)}$, which takes the value 1 if p_i is below a specified threshold τ_{failure}, and 0 otherwise:

$$1_{\text{failure}(i)} = \begin{cases} 1, & \text{if } p_i \leq \tau_{\text{failure}} \\ 0, & \text{otherwise.} \end{cases} \tag{2.73}$$

So, the adjusted price for a given period i is:

$$P_i = B_i \cdot \left(1 + \alpha_i 1_{\text{success}(i)} - \beta_i 1_{\text{failure}(i)}\right). \tag{2.74}$$

Aggregating Over Performance Periods

To obtain the total performance-based price, we sum over all evaluation periods:

$$P_{\text{adjusted}} = \sum_{i=1}^{n} B_i \cdot \left(1 + \alpha_i 1_{\text{success}(i)} - \beta_i 1_{\text{failure}(i)}\right). \tag{2.75}$$

Fixed Base Fee

A fixed base fee F_{base} is included to ensure a minimum guaranteed revenue:

$$P_{\text{total}} = P_{\text{adjusted}} + F_{\text{base}}. \tag{2.76}$$

Final Performance-Based Pricing Formula

Aggregating all components, the final pricing equation becomes:

$$P_{\text{performance}} = \sum_{i=1}^{n} B_i \cdot \left(1 + \alpha_i 1_{\text{success}(i)} - \beta_i 1_{\text{failure}(i)}\right) + F_{\text{base}}. \tag{2.77}$$

This formulation ensures that Performance-based incentives, Risk and uncertainty and Economic efficiency are considered. By applying contract theory principles, this model provides an optimized pricing strategy for data services, aligning provider incentives with buyer expectations.

2.3.11 Pay-Per-Use Pricing

Pay-Per-Use (PPU) pricing is a strategy in which users are charged individually according to their consumption pattern. This model takes into account the variability in data consumption, frequency of demand, and extreme usage patterns(Becker et al., 1999). In the present section, we develop a formal mathematical formulation of the PPU model, where all the aforementioned factors are entered in a systematic way.

Base Price Component

We define P_{base} as the standard base price per transaction, which serves as the foundation upon which adjustments are applied (Mankiw, 2021).

Usage-Based Price Adjustment

In economic theory, pricing strategies for usage-based models often reflect non-linear patterns, with discounts or premiums applied to different levels of usage (Tirole, 1988). To model this, we introduce a scaling factor q_i^{α}, where:

- q_i: Quantity of data used by the i-th customer.
- α: Elasticity parameter controlling how usage affects the price.

The elasticity parameter behaves as follows:

- $\alpha = 1$ represents linear pricing (each unit costs the same).
- $\alpha < 1$ represents discounts for bulk users.
- $\alpha > 1$ represents progressive pricing for increased consumption.

The adjusted price component for each user is then given by:

$$P_{\text{base}} \cdot q_i^{\alpha} \tag{2.78}$$

Utility-Based Adjustment

To reflect variations in user behavior, we introduce a utility function $U(D_i, R_i)$ that depends on two key factors:

- D_i: The frequency of data requests made by the i-th user.
- R_i: The intensity or volume of data requested by the i-th user.

The utility function modifies the pricing as follows:

$$P_i = P_{\text{base}} \cdot q_i^{\alpha} \cdot (1 + U(D_i, R_i)). \tag{2.79}$$

Common forms of the utility function include:

- **Linear Utility:** $U(D_i, R_i) = \lambda D_i + \delta R_i$, where λ and δ are sensitivity parameters.
- **Exponential Utility:** $U(D_i, R_i) = e^{\lambda D_i + \delta R_i} - 1$, for scenarios with aggressive pricing for extreme consumption(Mas-Colell et al., 1995).
- **Logarithmic Utility:** $U(D_i, R_i) = \log(1 + D_i) + \log(1 + R_i)$, emphasizing diminishing sensitivity to frequent usage(Henderson & Quandt, 1980).

High-Usage Surcharge

A surcharge coefficient γ is applied to customers who exceed a pre-defined usage threshold. The indicator function $1_{\text{high-usage}(j)}$ is defined as:

$$1_{\text{high-usage}(j)} = \begin{cases} 1, & \text{if user } j \text{ exceeds the threshold} \\ 0, & \text{otherwise.} \end{cases} \tag{2.80}$$

The surcharge component is calculated as:

$$\gamma \cdot \sum_{j=1}^{m} 1_{\text{high-usage}(j)}, \tag{2.81}$$

where:

- m represents the total number of users that exceed the predefined threshold.

Final Pay-Per-Use Pricing Formula

Combining the above elements, the total price paid by all users is:

$$P_{\text{PPU}} = \sum_{i=1}^{n} P_{\text{base}} \cdot q_i^{\alpha} \cdot (1 + U(D_i, R_i)) + \gamma \cdot \sum_{j=1}^{m} 1_{\text{high-usage}(j)}, \tag{2.82}$$

where:

- n represents the total number of customers.

This formulation combines baseline costs, utility-based adjustments, and high-usage penalties to create a flexible, consumption-driven pricing model that reflects real-world variations in data usage.

2.3.12 Market-Based Pricing

Market-based pricing alters with external elements like competition, changes in demand, and conditions of supply. The model reflects dynamic pricing adjustments to coincide with real-time market changes requiring adaptable and strategic pricing (Becker et al., 1999).

Base Market Price

The starting point is the base price P_{base}, which represents the fundamental value of the data product before market-driven modifications (Varian, 2014). This price reflects internal cost structures, initial valuation, and basic demand conditions.

Introducing Market-Based Adjustment

To account for competition and external trends, we introduce an adjustment factor α_i, where:

- $\alpha_i > 0$: Indicates favorable market conditions, prompting a price increase.
- $\alpha_i < 0$: Reflects competitive pressure or reduced demand, resulting in a price decrease.
- $\alpha_i = 0$: Signifies that there are neutral market conditions — neither favorable nor unfavorable.

The adjusted price component for each customer is:

$$P_{base} \cdot (1 + \alpha_i). \tag{2.83}$$

Utility-Based Pricing Adjustments

Market conditions can fluctuate significantly due to shifts in demand, supply, and production costs. To model these dynamic conditions, we incorporate a utility function $U(D_i, S_i, C_i)$ (Mas-Colell et al., 1995), where:

- D_i: Demand level for the i-th customer.
- S_i: Supply availability for the i-th customer's market.
- C_i: Cost incurred in producing or delivering the data for that customer.

The utility function ensures pricing adapts to these changing conditions. Common utility function forms include:

1. **Competitive Utility Function** (Tirole, 1988)

$$U(D_i, S_i, C_i) = \frac{\lambda D_i - \sigma S_i + \delta C_i}{1 + \gamma M_i}, \tag{2.84}$$

   Where:

   - λ: Sensitivity of demand to price changes.
   - σ: Sensitivity of supply to price changes.
   - δ: Sensitivity of costs to price changes.
   - M_i: Level of competition in the customer market (higher competition drives prices lower).

2. **Exponential Market Sensitivity Model** ((Henderson & Quandt, 1980))

$$U(D_i, S_i, C_i) = e^{\lambda D_i - \sigma S_i} - 1, \tag{2.85}$$

   This captures sharp price increases during demand surges but stabilizes when supply remains ample.

Market-Based Pricing Indicator Function

The indicator function $\mathbb{1}_{market}(i)$ is an internal condition within the pricing model that determines the presence or absence of market-based price adjustments for the i-th transaction. This function distinguishes between varying levels of market influence ((Dubé et al., 2010)).

It is defined as:

$$1_{market}(i) = \begin{cases} 1, & \text{if additional market-driven factors, such as competition effects or utility-based adjustments, are active} \\ 0, & \text{if no further market-based adjustments are required.} \end{cases} \tag{2.86}$$

This function ensures that internal distinctions within the confirmed market-based pricing model are managed efficiently.

Combining Components for Individual Market Price

The complete formula for the adjusted price for a given customer is:

$$P_{market,i} = P_{base} \cdot (1 + \alpha_i) \cdot (1 + U(D_i, S_i, C_i)). \tag{2.87}$$

This structure integrates both direct market adjustments and utility-based modifications to dynamically adapt the price.

Final Market-Based Formula

To compute the total revenue across all market interactions, we sum the adjusted prices for each customer. The final formulation is:

$$P_{market} = \sum_{i=1}^{n} P_{base} \cdot (1 + \alpha_i) \cdot (1 + U(D_i, S_i, C_i)) \cdot 1_{market}(i) \tag{2.88}$$

,

where:

- n: Total number of market interactions considered.

This comprehensive formula allows businesses to dynamically adjust data prices in response to changing market forces, maximizing competitiveness and revenue potential (Borenstein & Rose, 1994).

2.3.13 Value-Based Pricing

While production costs and market forces count in setting the price, value-based pricing emphasizes to what extent the customer values the data. This section is built to factor in business impact, data quality, risk mitigation, and strategic relevance in determining the proper price (Varian, 2014, Becker et al., 1999, Mankiw, 2021).

Base Value Price

The starting point is the base price P_{base}, representing the fundamental value of the data product before incorporating perceived value adjustments. This price reflects internal cost structures, basic valuation methods, and overall data characteristics (Mas-Colell et al., 1995).

Introducing Value-Based Adjustment

To incorporate customer-specific perceived value, we introduce an adjustment factor α_i, where:

- $\alpha_i > 0$: Indicates the data is highly valuable to the customer, prompting a price increase.
- $\alpha_i < 0$: Reflects reduced perceived value, resulting in a price decrease.
- $\alpha_i = 0$: Signifies neutral perceived value — neither high nor low.

The adjusted price component for each customer is:

$$P_{base} \cdot (1 + \alpha_i). \tag{2.89}$$

Utility-Based Pricing Adjustments

Customer perception varies based on multiple factors. To account for these elements, we introduce a utility function $U(V_i, B_i, Q_i, R_i)$ (Belleflamme & Peitz, 2015), where:

- V_i: Perceived strategic value of the data.
- B_i: Business impact, reflecting cost savings or revenue influence.
- Q_i: Quality of the data (accuracy, completeness, timeliness).
- R_i: Risk reduction factor, measuring the data's role in mitigating uncertainty.

Common utility function forms include:

- **Weighted Linear Combination**

  $$U(V_i, B_i, Q_i, R_i) = \lambda V_i + \delta B_i + \sigma Q_i + \rho R_i \tag{2.90}$$

  This simpler model balances different pricing influences, ideal for general-purpose data pricing strategies (Gibbons, 1992).

- **Exponential Sensitivity Model**

$$U(V_i, B_i, Q_i, R_i) = e^{\lambda V_i + \delta B_i + \sigma Q_i + \rho R_i} - 1 \tag{2.91}$$

This captures sharp price increases when perceived value surges, ensuring rapid response to high-value data.

- **Logarithmic Value Normalization**

$$U(V_i, B_i, Q_i, R_i) = \log(1+\lambda V_i)+\log(1+\delta B_i)+\log(1+\sigma Q_i)+\log(1+\rho R_i) \tag{2.92}$$

This model reflects diminishing returns, ensuring prices grow progressively slower as value increases ((Henderson & Quandt, 1980)).

- **Multiplicative Utility Model**

$$U(V_i, B_i, Q_i, R_i) = (\lambda V_i) \cdot (\delta B_i) \cdot (\sigma Q_i) \cdot (\rho R_i) \tag{2.93}$$

This model reflects strong pricing effects only when all value components are high.

Value-Based Pricing Indicator Function

The indicator function $1_{value}(i)$ determines whether value-based pricing adjustments are applicable for customer i. It is defined as:

$$1_{value}(i) = \begin{cases} 1, & \text{if customer-perceived value factors are active} \\ 0, & \text{if no additional value-based adjustments are required.} \end{cases} \tag{2.94}$$

Final Value-Based Pricing Model

Combining all components, the final formulation for total revenue is:

$$P_{value} = \sum_{i=1}^{n} P_{base} \cdot (1 + \alpha_i) \cdot (1 + U(V_i, B_i, Q_i, R_i)) \cdot 1_{value}(i), \tag{2.95}$$

where

- n is the total number of transactions considered.

This structure enables data providers to adapt prices dynamically based on customer-perceived value, maximizing strategic revenue potential (Rochet & Tirole, 2003).

2.3.14 Location-Based Pricing

The location-based pricing takes into account economic conditions from different regions to assure that prices for data adhere to geographic demand, cost structures, and regulatory frameworks. This formula integrates other factors for a dynamic price adjustment, striking a balance between competitiveness and revenue optimization.

Base Price

The base price, denoted as $P_{\text{base},i}$, represents the fundamental value of the data before regional adjustments are introduced. This base price acts as a reference point, ensuring that data prices do not fall below a predetermined threshold, regardless of geographic variations (Varian, 2014).

Geographic Adjustment

To account for geographic differences, a location-specific adjustment factor αG_i modifies the base price. Here, α represents the *sensitivity coefficient* that determines how strongly geographic differences influence pricing, while G_i is the *geographic pricing index*, capturing factors such as regional economic disparities, infrastructure costs, and taxation differences. For example, urban areas with better digital infrastructure may see lower adjustments, while remote regions with higher data acquisition costs may face increased pricing (Tirole, 1988).

$$\alpha G_i = \text{Geographic Adjustment Factor}, \tag{2.96}$$

where:

- α = Sensitivity coefficient for geographic pricing.
- G_i = Geographic pricing index representing regional economic conditions, infrastructure costs, and taxation differences.

Demand and Supply-Based Adjustment

The model incorporates demand (D_i) and supply (S_i) conditions, ensuring that prices respond to local market dynamics. Here, D_i represents the local demand for data, reflecting usage intensity and data consumption trends, while S_i denotes regional data supply, indicating data availability and distribution. Regions with higher data demand are subject to upward price adjustments, while regions with ample data supply may see prices stabilize or decrease.

The term

$$\beta S_i + \epsilon D_i$$

captures this relationship, where:

- β = Demand-supply elasticity coefficient, determining how pricing responds to shifts in data demand and supply.
- ϵ = Small constant that prevents numerical instability during computation.

This adjustment mechanism ensures that pricing reflects local market conditions, enhancing fairness and responsiveness (Mas-Colell et al., 1995).

Cost of Living and Income Adjustment

Local economic conditions are further accounted for by incorporating the regional cost of living (C_i) and average income levels (I_i). Here, C_i represents the regional cost of living, capturing expenses such as housing, utilities, and essential services, while I_i denotes the average income in the region, ensuring pricing aligns with local purchasing power. These factors ensure data pricing remains sensitive to affordability constraints.

The term

$$\gamma I_i + \epsilon C_i$$

helps balance price adjustments across high- and low-income regions. In this expression:

- γ = Income-cost elasticity coefficient, determining how pricing responds to variations in regional income and living costs.
- ϵ = Small constant that ensures smooth computation and prevents numerical instability.

This adjustment mechanism ensures equitable pricing without disproportionately affecting lower-income areas, promoting data accessibility across diverse economic conditions (Mas-Colell et al., 1995).

Regulatory Impact

To capture regulatory variations, a term $e^{\delta R_i}$ is introduced, where R_i reflects the influence of government-imposed restrictions, regulatory frameworks, or taxation policies in a given region. This ensures pricing adjusts to comply with varying legal and economic conditions across markets.

In this expression:

- R_i = Regulatory impact factor, representing the influence of government policies, restrictions, or taxation on data pricing.
- δ = Regulation sensitivity coefficient, determining how strongly pricing responds to these regulatory conditions.

This adjustment ensures that regions with stricter regulations or higher taxation see appropriate pricing adjustments, maintaining both compliance and market fairness (Tirole, 1988).

Utility Function for Local Conditions

A utility function $U(G_i, D_i, S_i, C_i, I_i)$ refines the pricing model by dynamically adjusting prices based on combined geographic, economic, and data accessibility factors. Possible utility function forms include:

- **Linear Model:**

  $$U(G_i, D_i, S_i, C_i, I_i) = \lambda G_i + \delta D_i - \sigma S_i + \rho C_i \cdot I_i \tag{2.97}$$

  This model directly scales pricing based on regional conditions, balancing costs and income levels (Gibbons, 1992).

- **Logarithmic Model (Diminishing Marginal Effects):**

  $$U(G_i, D_i, S_i, C_i, I_i) = \log(1+\lambda G_i)+\log(1+\delta D_i)-\log(1+\sigma S_i)+\log(1+\rho C_i \cdot I_i) \tag{2.98}$$

  This form accounts for diminishing marginal effects, where higher values in each factor contribute less drastically to pricing ((Henderson & Quandt, 1980)).

- **Multiplicative Interaction Model:**

  $$U(G_i, D_i, S_i, C_i, I_i) = (\lambda G_i) \cdot (\delta D_i) \cdot (1/\sigma S_i + \epsilon) \cdot (\rho C_i \cdot I_i) \tag{2.99}$$

  This model applies stronger price adjustments only when multiple regional factors align, ensuring robust adaptation in varying conditions (Mas-Colell et al., 1995).

Final Location Based Pricing Formula

By combining all the above factors, the complete location-based pricing formula is defined as follows:

$$P_{\text{location}} = \sum_{i=1}^{n} P_{\text{base},i} \cdot (1+\alpha G_i) \cdot (1+\beta S_i + \epsilon D_i) \cdot (1+\gamma I_i + \epsilon C_i) \cdot e^{\delta R_i} \cdot (1+U(G_i, D_i, S_i, C_i, I_i)), \tag{2.100}$$

where:

- n = total number of regions/customers considered

This comprehensive formula integrates geographic conditions, demand-supply dynamics, affordability constraints, and regulatory factors to ensure location-based pricing is flexible yet market-optimized.

2.3.15 Loyalty-Based Pricing

Another pricing strategy that powers discounts or special price structures for repeat customers is called loyalty-based pricing, which is a way to encourage purchase repeat frequency" (Becker et al., 1999). It is a formula-based systematic adjustment to utility that factors in perception-based values of customer retention, purchase frequency, and tenure.

Base Price Component

We define P_{base} as the standard base price of the data product before applying any loyalty-related adjustments. This price serves as the foundation upon which modifications based on customer loyalty and utility considerations are applied(Mankiw, 2021).

Loyalty-Based Discount Adjustment

Economic studies have shown that differentiated pricing strategies, such as loyalty-based pricing, lead to increased customer retention and long-term profitability (Gibbons, 1992; Borenstein & Rose, 1994). Loyal customers often receive discounts based on their purchase history. To model this, we introduce a discount factor β_i for the i-th loyal customer, which reduces the price for returning buyers(Tirole, 1988). The indicator function $1_{\text{loyalty}(i)}$ determines whether a customer qualifies for a loyalty discount:

$$1_{\text{loyalty}(i)} = \begin{cases} 1, & \text{if the customer qualifies for a loyalty discount} \\ 0, & \text{otherwise.} \end{cases} \tag{2.101}$$

The price adjustment due to loyalty-based discounts is applied as a reduction in the base price:

$$P_{\text{base}} \cdot (1 - \beta_i \cdot 1_{\text{loyalty}(i)}), \tag{2.102}$$

where the discount is applied only if the customer meets the criteria for loyalty benefits.

Utility-Based Loyalty Adjustment

The perceived value of loyalty pricing can be further refined using a utility function $U(F_i, T_i)$, which depends on two key factors:

- F_i: The frequency of purchases made by the i-th customer(Mas-Colell et al., 1995).
- T_i: The tenure or duration of the customer's engagement with the service(Lazear, 2000).

The utility function modifies the pricing as follows:

$$P_{\text{base}} \cdot (1 + U(F_i, T_i)). \tag{2.103}$$

Common forms of the utility function include:

- **Logarithmic Utility:** $U(F_i, T_i) = \log(1 + F_i)$, capturing diminishing marginal benefits of loyalty.
- **Linear Utility:** $U(F_i, T_i) = \lambda F_i + \gamma T_i$, where λ and γ are sensitivity parameters measuring the direct effect of purchase frequency and tenure.
- **Exponential Utility:** $U(F_i, T_i) = e^{F_i + T_i} - 1$, emphasizing strong loyalty benefits for highly engaged customers.

Bringing the above components together, the price for each loyal customer would be:

$$P_l = P_{\text{base}} \cdot (1 - \beta_i \cdot 1_{\text{loyalty}(i)}) \cdot (1 + U(F_i, T_i)). \tag{2.104}$$

Price Adjustment for Non-Loyal Customers

Customers who do not qualify for loyalty discounts pay the base price without modifications. This is captured by another indicator function $1_{\text{non-loyalty}(j)}$, defined as:

$$1_{\text{non-loyalty}(j)} = \begin{cases} 1, & \text{if the customer does not qualify for loyalty benefits} \\ 0, & \text{otherwise.} \end{cases} \tag{2.105}$$

The price for each non-loyal customer is:

$$P_{\text{m}} = P_{\text{base}} \cdot 1_{\text{non-loyalty}(j)}. \tag{2.106}$$

Final Loyalty-Based Pricing Formula

Adding the components and summing over all loyal and non-loyal customers, the total price paid is:

$$P_{\text{loyalty}} = \sum_{i=1}^{l} \left(P_{\text{base}} \cdot (1 - \beta_i \cdot 1_{\text{loyalty}(i)}) \cdot (1 + U(F_i, T_i)) \right) + \sum_{j=1}^{m} \left(P_{\text{base}} \cdot 1_{\text{non-loyalty}(j)} \right), \tag{2.107}$$

where:

- l represents the total number of loyal customers.
- m represents the total number of non-loyal customers.

The derived formula for loyalty-based pricing integrates loyalty-based discounts, utility-based modifications, and segmentation of customers. This approach allows companies to optimize pricing strategies while fostering long-term customer retention.

2.4 Comparative Analysis of Models

Identifying an appropriate data pricing model is not just about profit but also about apprehending the specific requirements of exchange platforms, customers, businesses and the data itself. As such data trading platforms are turning more dynamic and competitive, implementation of pricing strategies that balance between value, cost and customer approach becomes crucial for organizations. We can observe how different pricing models perform under diverse conditions such as variations, demand, highly valued data sets or priced accesses by comparing and contrasting them with each other. Gaining understanding from the information that has already been established in economic theories (Varian, 2014; Henderson & Quandt, 1980), this research highlights important elements like complexity, flexibility, and revenue focus. The aim is to provide a clear roadmap that helps companies make informed decisions about the optimum way to monetize their data.The goal is to offer a straightforward guide that assists businesses in making educated choices regarding the best method to profit from their data.

2.4.1 Key Comparison Criteria

It is very important to consider various factors affecting the practicality, performance and overall efficiency of data pricing models while assessing them. The parameters listed below offer an organized approach to evaluate and compare various models:

- Mathematical Complexity: Mathematical structures of these pricing models are significantly different. Intricate functions to factor in parameters like perceived value and risk are usually employed by more advanced models such as utility based pricing, While basic models utilize plain linear adjustments such as cost plus pricing (Henderson & Quandt, 1980).

- Flexibility: Certain models are designed to quickly respond to changing data requirements. For example, dynamic pricing rapidly adjusts prices in response to demand fluctuations, making it suitable for unstable markets (Varian, 2014).

- Revenue Optimization: Certain models are designed to maximize revenue by modifying prices based on bidding techniques or competitive strategies such as Auction based pricing and dynamic pricing(Myerson, 1981; Milgrom, 2004). At the same time, statistical models might offer less adaptive revenue streams but are comparatively stable.

- User Incentives and Engagement: Models such as freemium pricing attracts new users by providing cost free threshold access and loyalty based pricing such models offer consistent customers with customized offers like discounts (Mankiw, 2021). Such models effectively encourage customer retention and investments.

- Data Characteristics: The nature of the data is what determines the suitable pricing strategy. Premium or utility based pricing models work well for real time, sensitive and high value data, while subscription or cost plus pricing models handle standardized data in a better way (Che & Gale, 1998).

- Risk Management: Mitigation of uncertainty in data value or demand is essential and specific models such as performance based pricing incorporate mechanisms to do so. There, the risk is significantly reduced for the buyers as revenue is aligned with successful outcomes in such pricing methods.
- Scalability: Tiered pricing and other similar models are comparatively better suited for large scale data trading platforms as they can handle various customer segments effectively in scalable environments (Cramton et al., 2006).
- Implementation Complexity: Models such as freemium pricing or subscription pricing are comparatively easy to implement than other models such as auction based pricing or utility based pricing because of their demand for advanced methodologies to track, bid and assess value (Klemperer, 2004).

2.4.2 Tabular Comparisons

1. Summary Table of Model Characteristics

Table 2.16: Summary Table of Model Characteristics

Model	**Complexity**	**Flexibility**	**Revenue Focus**	**Customer Incentive**	**Best Use Case**
Cost-Plus Pricing	Medium	Low	Cost Coverage	Limited	Stable data costs
Dynamic Pricing	High	High	Maximizing profit	Medium	Fluctuating demand
Premium Pricing	Medium	Medium	Revenue Maximization	Limited	Exclusive datasets
Freemium Model	Low	High	Market Expansion	Strong	Introductory access
Two-Part Tariff	High	Medium	Balanced Profit	Medium	Flexible access tiers
Subscription Model	Medium	High	Stable Revenue	Strong	Regular data delivery
Tiered Pricing	Medium	Medium	Customer Segmentation	Strong	Multiple user types
Utility-Based	High	High	Value-Based Revenue	Strong	Data quality variation
Performance-Based	High	Medium	Reward Efficiency	Medium	Goal-driven outcomes
Auction Pricing	High	High	Maximized Bidding Value	Medium	Scarce datasets
Market-Based	Medium	Medium	Competitive Advantage	Medium	Dynamic competition
Value-Based Pricing	Medium	Medium	Customer-Centric Value	Strong	Specialized insights
Location-Based	Medium	Medium	Geographic Control	Medium	Regional data demand
Pay-Per-Use	Medium	High	Consumption Control	Strong	On-demand data access
Loyalty-Based	High	Medium	Customer Retention	Strong	Encourages repeat users

2. Model Suitability Table

Table 2.17: Model Suitability Table

Model	Ideal Data Type	Best for Data Variability	Market Type	Recommended for
Cost-Plus Pricing	Stable, well-defined data	Low	Static/Stable Markets	Reliable cost recovery
Dynamic Pricing	Real-time, volatile data	High	Competitive Markets	Maximizing profits in fluctuating demand
Premium Pricing	Exclusive, high-value data	Low	Niche/High-end Markets	Luxury insights or limited datasets
Freemium Model	Entry-level, partial datasets	Medium	Expanding Markets	Attracting new customer bases
Subscription Model	Regularly updated data	Medium	Stable Consumer Base	Stable, recurring revenue
Two-Part Tariff	Mixed data with variable access	Medium	Flexible Consumer Base	Users requiring flexible access plans
Tiered Pricing	Data with varying complexity	Medium	Multi-segmented Markets	Businesses with diverse user types
Utility-Based Pricing	Data with variable quality	High	Value-Driven Markets	Specialized clients demanding premium insights
Auction-Based Pricing	Unique, scarce datasets	High	Competitive Bidding Environments	Maximizing revenue through bidding
Performance-Based Pricing	Outcome-driven data	Medium	Project-Based Markets	Ensuring measurable results
Pay-Per-Use Pricing	On-demand, usage-sensitive data	Medium	Flexible Usage Environments	Occasional or sporadic data access
Market-Based Pricing	Competitively available data	Medium	Dynamic Market Conditions	Ensuring competitive pricing
Value-Based Pricing	Data with strategic insights	Medium	Specialized, high-value sectors	Delivering custom insights
Location-Based Pricing	Region-specific data	Medium	Geographically driven markets	Regional data control
Loyalty-Based Pricing	Data tailored for repeat users	Medium	Customer Retention Strategies	Encouraging long-term client relationships

3. Comparative Table of Utility-Based Factors

Table 2.18: Comparative Table of Utility-Based Factors

Model	Perceived Value (V)	Business Impact (B)	Data Quality (Q)	Risk Reduction (R)
Cost-Plus Pricing	Low	Low	Low	Low
Dynamic Pricing	Medium	High	Medium	Medium
Premium Pricing	High	Medium	Medium	Medium
Freemium Model	Low	Low	Medium	Low
Subscription Model	Medium	Medium	Medium	Medium
Two-Part Tariff	Medium	Medium	High	Medium
Tiered Pricing	Medium	Medium	Medium	Medium
Utility-Based Pricing	High	High	High	High
Auction-Based Pricing	Medium	Medium	Medium	Medium
Performance-Based Pricing	Medium	Medium	High	Medium
Pay-Per-Use Pricing	Medium	Medium	Medium	Medium
Market-Based Pricing	Medium	Medium	Medium	Medium
Value-Based Pricing	High	Medium	Medium	Medium
Location-Based Pricing	Medium	Medium	Medium	Medium
Loyalty-Based Pricing	High	High	Medium	Medium

4. Model Strengths and Weaknesses

Table 2.19: Model Strengths and Weaknesses

Model	Strengths	Weaknesses
Cost-Plus Pricing	Simple structure	Limited customer incentive
Dynamic Pricing	Maximizes revenue in demand changes	Requires constant monitoring
Premium Pricing	High-profit margins possible	Limited market size
Freemium Model	Strong customer acquisition	Conversion to paid users may be low
Subscription Model	Predictable revenue stream	May deter occasional users
Two-Part Tariff	Ensures steady revenue	Complex to manage multiple fees
Tiered Pricing	Serves diverse customer needs	Risk of confusing pricing structure
Utility-Based Pricing	Charges based on value gained	Requires detailed usage tracking
Auction-Based Pricing	Maximizes data's true market value	Requires active bidder participation
Performance-Based Pricing	Encourages data efficiency	Complex to define performance benchmarks
Pay-Per-Use Pricing	Fair for occasional users	Unpredictable revenue stream
Market-Based Pricing	Dynamic response to market trends	Requires competitive monitoring
Value-Based Pricing	Aligns pricing with customer value	Requires clear value assessment
Location-Based Pricing	Captures regional value differences	May confuse global buyers
Loyalty-Based Pricing	Encourages repeat purchases	Initial setup can be complex

2.4.3 Flowcharts to Analyze and Model selection

In this section, we compile a few flow charts which provide a basic guideline for an appropriate model selection based on different aspects.

1. **Selection of suitable pricing model based on data characteristics:** Figure 2.3, *page*-61

   In this flowchart, we have designed questions which enquire about the characteristics and features of data, and based on your answer, it suggests a more suitable pricing model for the data.

2. **Selection of suitable pricing model based on business strategy:** Figure 2.4, *page*-62

   In the second flow chart, it again asks questions regarding the alignment of the business strategy. Depending on the answer, it prescribes a model which can produce the best outcome for the case.

3. **Hybrid model selection:** Table 2.20, *page*-63

   If the above flowcharts suggest more than one pricing model, the 3rd part, which is a table, aids you in combining different models to arrive at one hybrid for better performance and profit.

Figure 2.3: Selection of model based on Data Characteristics

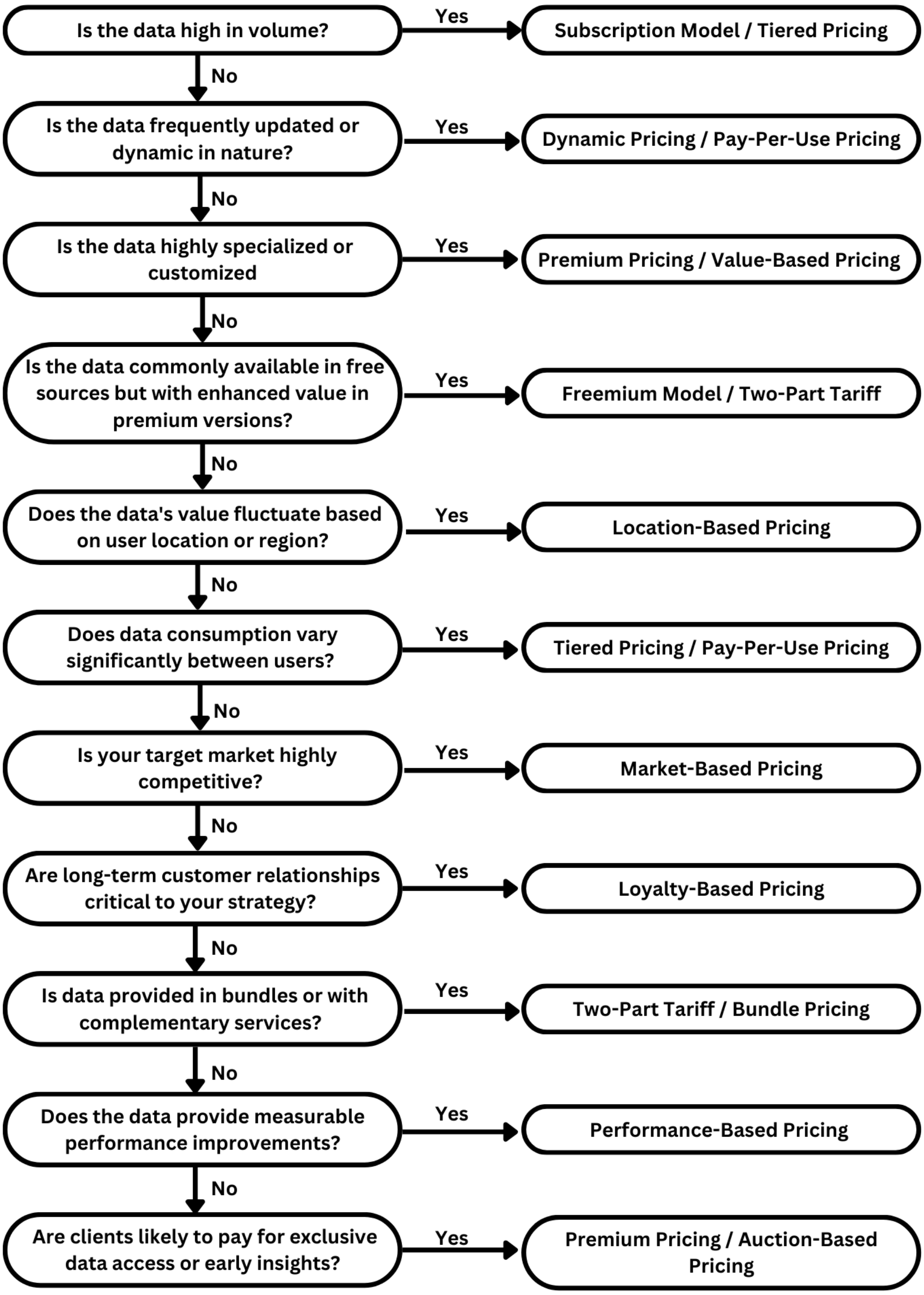

Figure 2.4: Selection of model based on Business Strategy Alignment

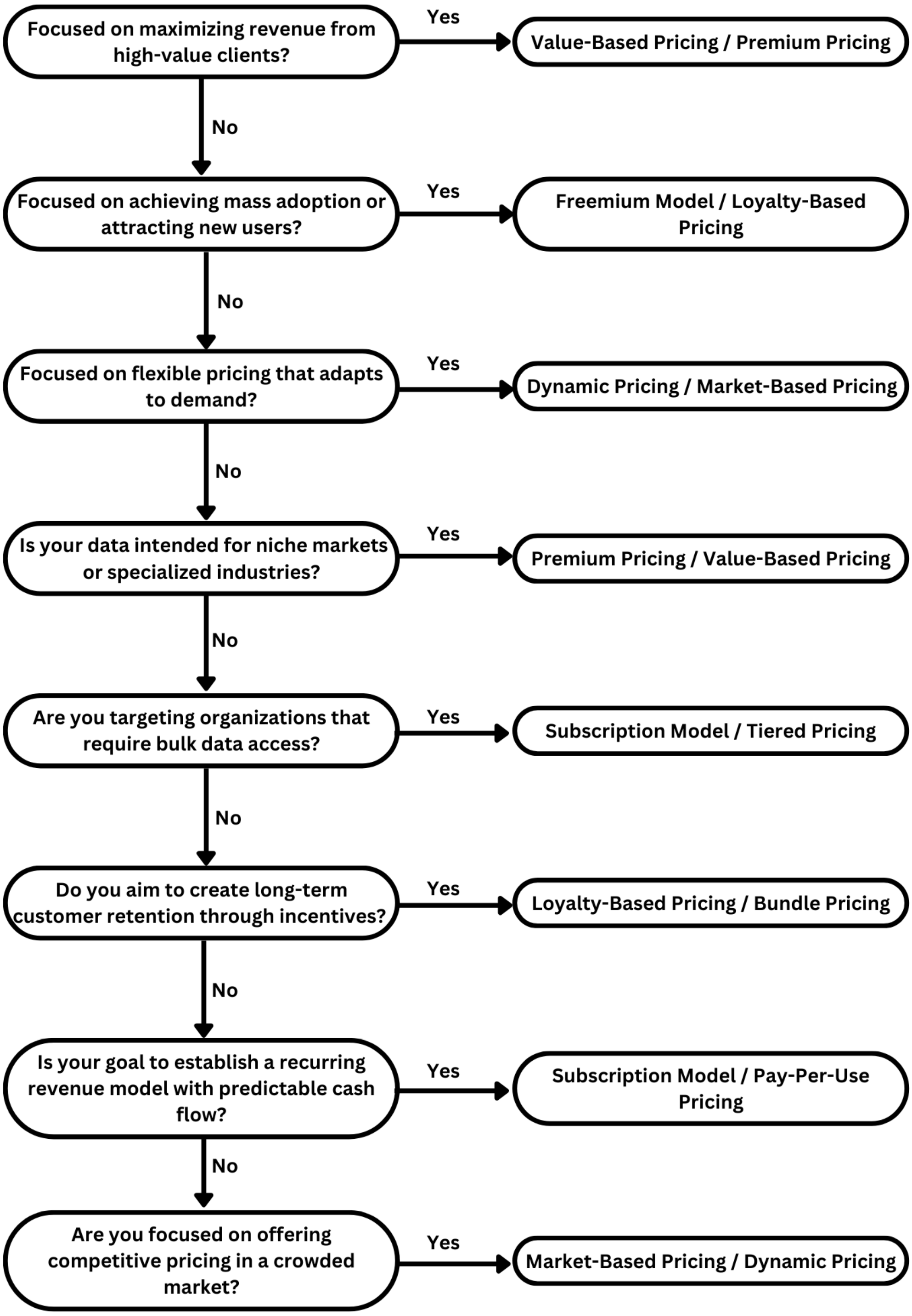

Table 2.20: Hybrid Model Combinations for Enhanced Performance

Model	**Best Hybrid Combinations**
Cost-Plus Pricing	Tiered + Loyalty-Based
Dynamic Pricing	Auction-Based + Utility-Based
Premium Pricing	Market-Based + Loyalty-Based
Freemium Model	Subscription + Tiered
Two-Part Tariff	Utility-Based + Subscription
Subscription Model	Loyalty-Based + Performance-Based
Tiered Pricing	Utility-Based + Pay-Per-Use
Utility-Based	Dynamic + Loyalty-Based
Performance-Based	Auction-Based + Market-Based
Auction-Based Pricing	Dynamic + Value-Based
Market-Based Pricing	Dynamic + Pay-Per-Use
Value-Based Pricing	Utility-Based + Tiered
Location-Based Pricing	Tiered + Subscription
Pay-Per-Use	Utility-Based + Performance-Based
Loyalty-Based Pricing	Subscription + Pay-Per-Use

Thus the above section provides various comparisons and contrasts among the models through tables. Along with that the basic selection guide has been offered through flowcharts. Together these help both data trading platforms and traders to understand and make informed decisions while trading.

Chapter 3

Results and Discussions

This chapter presents the key findings drawn from extensive review, formulation, comparison and analysis of the data pricing models obtained from literature. The results highlight the cumulation of the key parameters and pricing models, practical applications of these models, observe strengths and weaknesses, and offer guidance on model selection depending on varying conditions.

As part of foundational analysis, the research identifies over 140 parameters which affect the data pricing mechanisms and segregates then into various classes such as Data Characteristics, Market-Related Parameters, Data Ownership and Rights, User-Specific Factors, Data Processing and Infrastructure, Economic and Social Factors, Transaction and Contractual Factors, Technological Factors, Data Governance and Policy, Time-Sensitive Parameters, Risk Factors, Environmental and Contextual Factors, Behavioral and Psychological Factors and Miscellaneous.It was essential to recognize these parameters in formulation of the models ensuring their alignment with practical market needs.

Additionally, 15 distinct data pricing models were identified similarly. Each model was analyzed carefully to understand its theoretical basis and practical applicability. The list of models includes widely used methods such as Cost plus pricing, Dynamic pricing, Premium pricing, Freemium pricing, Subscription pricing, Tiered pricing, Pay per use pricing & value based pricing, and less common models like two part tariff pricing, Utility pricing, Performance based pricing, Auction based pricing, market based pricing, Location based pricing & Loyalty based pricing.

By taking the above mentioned parameters as guiding factors and selective integration of them with the pricing models, mathematical formulae for all of them were derived and developed. The derivations capture key aspects such as data characteristics, market conditions, consumer behavior etc. This approach ensured that all the final formulations were comprehensive and aligned closely to the real world dynamics. They highlight various features such as complexity, adaptability and revenue focus. It is evident from the results that certain models are more flexible in adapting to market conditions while some are more stable making them more suitable where the market is predictable. Some are complex yet effective in revenue maximization and some are simple and straightforward without compromising on the market details.

The comparative analysis reveals patterns and insights that indicate the best conditions for each pricing model. The tabulated comparisons unveiled several important trends:

Maximizing Revenue: Models such as Dynamic Pricing and Auction-Based Pricing performed exceptionally well in increasing revenue by adjusting to real-time market fluctuations. These models are well-suited for situations where data demand is frequently changing.

Consistent Revenue Generation: Pricing models like Subscription Pricing and Two-Part Tariff showed significant potential for delivering steady and predictable revenue, making them suitable for long-term data delivery services.

Promoting Customer Participation and Incentives: Models such as Freemium and Loyalty-Based Pricing have proven successful in enhancing customer involvement and encouraging repeat usage. These strategies provide a balance between cost-effectiveness and opportunities for upselling.

Adaptability to Changes in Data Quality: The Utility-Based Pricing model is particularly effective in scenarios where the perceived value of data fluctuates based on its quality, enabling data providers to command higher prices for superior quality data. These insights offer valuable direction for organizations when choosing pricing models aligned with their strategic goals, such as maximizing profits, enhancing customer retention, or streamlining implementation.

The created flowcharts serve as valuable tools for decision-making, assisting businesses in selecting models based on pertinent criteria. By organizing choices around essential factors including data variability, revenue objectives, and customer interaction such as the flowcharts make complex evaluations more straightforward.

For example, companies focusing on high-end datasets with unique insights are directed towards models such as Premium Pricing or Value-Based Pricing. Conversely, for companies operating in volatile market conditions, the flowcharts indicate that Dynamic Pricing or Auction-Based Pricing could enhance revenue more efficiently. The flowcharts also address scalability challenges, identifying models that can effectively scale with substantial data volumes. Approaches like Tiered Pricing and Pay-Per-Use provide structured methods to cater to various customer requirements across different data scales.

A number of significant insights arose from this evaluation:

The versatility of Hybrid Models: Significant potential can be observed when certain models such as Auction based pricing and Utility based pricing are combined with others. For example, Adaptable pricing that shows the worth of the data while also upgrading loyalty of customers can be demonstrated by combining Utility with Loyalty based methods.

Significance of Consistency in Model Selection: Strategies such as Subscription Pricing and Cost-Plus Pricing are more suited to perform in stable income conditions,and also reduce the risks associated with unpredictable demand changes.

Customer-Oriented Approaches: Models like Loyalty-Based Pricing and Freemium pricing encourage customer associations by offering incentives that promote repeated purchases or upgrades to higher-tier data services.

Chapter 4

Conclusions and Future Work

4.1 Conclusions

This research improves our understanding of data pricing by providing an extensive literature review, highlighting essential components, developing mathematical models, and conducting comparative analyses. A comprehensive analysis of existing pricing strategies has formed an organized framework for assessing data pricing models. By pinpointing 140 essential parameters that affect pricing decisions, this study underscores the intricacies involved in valuing data and the various factors that influence pricing outcomes.

The mathematical formulations for all 15 pricing models represent a novel contribution by systematically deriving pricing equations that consider variables such as data quality, demand variability, and customer engagement. These formulations not only clarify the theoretical underpinnings of each model but also offer practical insights into their applicability in real-world situations.

The comparative analysis, organized through tabular assessments and decision-making diagrams, enhances the contributions of this study. By connecting the characteristics, advantages, disadvantages, and applicability of various models to distinct data contexts, the research provides a thorough guide for data exchange platforms (DEPs) and organizations aiming for effective pricing strategies. The hybrid model combinations offer greater adaptability, creating opportunities where the amalgamation of multiple approaches could produce better results.

In conclusion, this study fills a notable void in the current literature by offering detailed mathematical formulations alongside practical advice for selecting models. The structured methodology guarantees that companies can make well-informed decisions when managing stable data sets, responding to demand changes, or striving for strategic revenue objectives.

4.2 Limitations

This research presents important theoretical contributions; however, it is crucial to acknowledge a number of limitations. One major limitation is the lack of validation using real-world data. Although the mathematical models and comparative studies offer robust theoretical foundations, their practical applicability may vary due to elements like data accessibility, market trends, and consumer actions.

Additionally, various assumptions were established during the modeling phase to streamline complicated scenarios. For instance, factors such as customer retention patterns, network influences, or data-sharing behaviors were not thoroughly explored during the model creation. While these assumptions enabled a more focused analysis, they may limit the direct applicability of certain models in environments that are rapidly evolving and highly variable.

Lastly, this research concentrated solely on the development of theoretical models and conducting comparative analyses. Although the provided flowcharts and recommendations for hybrid models are beneficial, additional efforts are necessary to assess their effectiveness in real-life decision-making scenarios.

4.3 Future work

In order to expand this study, we are currently working on a research paper, as further part of which we apply the constructed mathematical formulae to actual real world data sets. We will be able to assess the performances of these models in various real world conditions and observe how resilient they could be in different environments. In particular, publicly accessible yet highly significant datasets such as environmental data. For fields such as agriculture, resource management and climate science, this research could produce applicative insights for all the actors of the data markets, both data providers, consumers, agents etc. Also, with the aim of improving on the models and substantiating their applicability and practicality, we propose to gather feedback from academic researchers and industrial experts. We hope to finish and publish this paper in not more than a couple of months.

Building upon this groundwork, various potential enhancements could increase the relevance and influence of this research. One promising approach is to integrate machine learning techniques to improve dynamic pricing strategies. By leveraging predictive analysis to study changes in data demand, consumer habits, or market competition, businesses can enhance their pricing strategies via adaptive decision-making.

Lastly, developing a tool to aid decision making inspired by the flowcharts pre- sented in this research would provide a practical solution for organizations looking to adopt data pricing strategies. By transform- ing the ideas from the flowcharts into an interactive platform, companies could obtain personalized model recommendations suitable for their data at- tributes and strategic aims. This tool would streamline the pricing decision- making process, making it easier for both technical professionals and business leaders to utilize. In summary, while this research establishes a solid theoretical basis, these suggested enhancements can connect theory with practice, ensuring that data pricing strategies remain effective and adaptable in real-world situations.

In summary, while this research establishes a solid theoretical basis, these suggested enhancements can connect theory with practice, ensuring that data pricing strategies remain effective and adaptable in real-world situations.

Bibliography

1. Acquisti, A., Taylor, C., & Wagman, L. (2016). *The economics of privacy.* Journal of Economic Literature, 54(2), 442–492. https://doi.org/10.1257/jel.54.2.442

2. Anderson, E., & Dana, J. D. (2008). *When is Price Discrimination Profitable?* SSRN Electronic Journal. https://doi.org/10.2139/ssrn.1322528

3. Basuchaudhary, A. (2006). *Contract theory, by Bolton, P. and Dewatripont, M. MIT Press: Cambridge and London. 2005, xvi+724 pp., USD 65.00 (cloth).* Managerial and Decision Economics, 28(1), 81–82. https://doi.org/10.1002/mde.1332

4. Becker, W. E., Shapiro, C., & Varian, H. R. (1999). *Information Rules: A Strategic Guide to the Network Economy.* The Journal of Economic Education, 30(2), 189. https://doi.org/10.2307/1183273

5. Belleflamme, P., & Peitz, M. (2015). *Industrial Organization.* https://doi.org/10.1017/cbo9781107707139

6. Bergemann, D., & Bonatti, A. (2019). *Markets for Information: An Introduction.* Annual Review of Economics, 11(1), 85–107. https://doi.org/10.1146/annurev-economics-080315-015439

7. Bergstrom, T. C., & Varian, H. R. (1985). *When are NASH equilibria independent of the distribution of agents' characteristics?* The Review of Economic Studies, 52(4), 715. https://doi.org/10.2307/2297742

8. Bhattacharya, S., Milgrom, P., & Roberts, J. (1993). *Economics, Organization and Management.* The Journal of Finance, 48(1), 419. https://doi.org/10.2307/2328903

9. Borenstein, S., & Rose, N. L. (1994). *Competition and price dispersion in the U.S. airline industry.* Journal of Political Economy, 102(4), 653–683. https://doi.org/10.1086/261950

10. Cramton, P., Shoham, Y., & Steinberg, R. (2006). *Combinatorial Auctions.* MIT Press.

11. Dubé, J. H., Hitsch, G. J., & Chintagunta, P. K. (2010). *Tipping and Concentration in Markets with Indirect Network Effects.* Marketing Science, 29(2), 216–249. https://doi.org/10.1287/mksc.1090.0541

12. Farboodi, M., Mihet, R., Philippon, T., & Veldkamp, L. (2019). *Big data and firm dynamics.* AEA Papers and Proceedings, 109, 38–42. https://doi.org/10.1257/pandp.20191001

13. Gandomi, A., & Haider, M. (2014). *Beyond the hype: Big data concepts, methods, and analytics.* International Journal of Information Management, 35(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

14. Ghosh, A., & Roth, A. (2011). *Selling privacy at auction.* In Proceedings of the 12th ACM conference on Electronic commerce (pp. 199-208) https://doi.org/10.1145/1993574.1993605

15. Gibbons, R. (1992). *A primer in game theory.* https://espace.library.uq.edu.au/view/UQ:354162

16. Goldfarb, A., & Tucker, C. (2019). *Digital Economics.* Journal of Economic Literature, 57(1), 3–43. https://doi.org/10.1257/jel.20171452

17. Hart, O., & Moore, J. (1988). *Incomplete contracts and renegotiation.* Econometrica, 56(4), 755. https://doi.org/10.2307/1912698

18. Heckman, J. J. (2001). *Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture.* Journal of Political Economy, 109(4), 673–748. https://doi.org/10.1086/322086

19. Heckman, J., Boehmer, E., Peters, E., Davaloo, M., & UC Berkeley School of Information. (2015). *A pricing model for data markets.* In iConference 2015 Proceedings. https://core.ac.uk/download/pdf/158298935.pdf

20. Henderson, J. M., & Quandt, R. E. (1980). Microeconomic Theory: A Mathematical Approach (3rd ed.). McGraw-Hill.

21. Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences.* https://doi.org/10.4135/9781473909472

22. Klemperer, P. (1999). *Auction Theory: a guide to the literature.* Journal of Economic Surveys, 13(3), 227–286. https://doi.org/10.1111/1467-6419.00083

23. Laffont, J., & Martimort, D. (2001). *The Theory of Incentives: the Principal-Agent model.* http://dx.doi.org/10.2307/j.ctv7h0rwr

24. Laffont, J., & Tirole, J. (1993). *A theory of incentives in procurement and regulation.* MIT Press.

25. Lambrecht, A., & Skiera, B. (2006). *Paying Too Much and Being Happy about It: Existence, Causes, and Consequences of Tariff-Choice Biases.* Journal of Marketing Research, 43(2), 212–223. https://doi.org/10.1509/jmkr.43.2.212

26. Laney, D. (2001). *3D Data Management: Controlling Data Volume, Velocity, and Variety.* http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Varie pdf

27. Lazear, E. P. (2000). *Performance pay and productivity.* American Economic Review, 90(5), 1346–1361. https://doi.org/10.1257/aer.90.5.1346

28. Lee, I., & Lee, K. (2015). *The Internet of Things (IoT): Applications, investments, and challenges for enterprises.* Business Horizons, 58(4), 431–440. https://doi.org/10.1016/j.bushor.2015.03.008

29. Liang, F., Yu, W., An, D., Yang, Q., Fu, X., & Zhao, W. (2018). *A survey on Big data market: Pricing, trading and protection.* IEEE Access, 6, 15132–15154. https://doi.org/10.1109/access.2018.2806881

30. Mankiw, N. G. (2014). *Principles of economics.* Cengage Learning.

31. Mas-Colell, A., Whinston, M. D., & Green, J. R. (1995). *Microeconomic Theory.* Oxford University Press.

32. Milgrom, P. R. (2004). *Putting auction theory to work.* http://dx.doi.org/10.1017/cbo9780511813825

33. Moorthy, K. S. (1984). *Market segmentation, Self-Selection, and product line design.* Marketing Science, 3(4), 288–307. https://www.jstor.org/stable/183963

34. Muschalle, A., Stahl, F., Löser, A., & Vossen, G. (2013). *Pricing approaches for data markets*. In Lecture notes in business information processing (pp. 129–144). https://doi.org/10.1007/978-3-642-39872-8_10

35. Myerson, R. B. (1981). *Optimal Auction design*. Mathematics of Operations Research, 6(1), 58–73. https://doi.org/10.1287/moor.6.1.58

36. Pindyck, R. S., & Rubinfeld, D. L. (2013). *Microeconomics*. Prentice Hall.

37. PWC Report. (2019). *Putting a value on data*. https://www.pwc.co.uk/data-analytics/documents/putting-value-on-data.pdf

38. Rochet, J., & Tirole, J. (2003). *Platform competition in Two-Sided markets*. Journal of the European Economic Association, 1(4), 990–1029. https://doi.org/10.1162/154247603322493212

39. Samuelson, P. A., & Nordhaus, W. D. (2009). *Economics*. McGraw-Hill Companies.

40. Sheng, H., Nah, F., & Siau, K. (2008). *An Experimental Study on Ubiquitous commerce Adoption: Impact of Personalization and Privacy Concerns*. Journal of the Association for Information Systems, 9(6), 344–376. https://doi.org/10.17705/1jais.00161

41. Tirole, J. (1988). *The theory of industrial organization*. https://external.dandelon.com/download/attachments/dandelon/ids/DE0108916F5A49262E07 pdf

42. Tucker, C. (2019). *Privacy, algorithms, and artificial intelligence*. NBER. http://www.nber.org/chapters/c14011

43. Varian, H. R. (1995). *Pricing information goods*. https://people.ischool.berkeley.edu/~hal/Papers/price-info-goods.pdf

44. Varian, H. R. (2014). *BEYOND BIG DATA*. Business Economics, 49(1), 27–31. https://doi.org/10.1057/be.2014.1

45. Varian, H. (2018). *Artificial Intelligence, Economics, and Industrial Organization*. https://doi.org/10.3386/w24839

46. Coyle, D., Kay, L., Diepeveen, S., Tennison, J., Julia Wdowin, Bennett Institute, University of Cambridge, & Open Data Institute. (2020). *The value of data. In Policy Implications* [Report]. Bennett Institute for Public Policy, Cambridge in partnership with the Open Data Institute. https://www.bennettinstitute.cam.ac.uk/wp-content/uploads/2020/12/Value_of_data_Policy_Implications_Report_26_Feb_ok4noWn.pdf

47. Wu, N. X., Zhu, N. X., Wu, N. G., & Ding, N. W. (2014). *Data mining with big data*. IEEE Transactions on Knowledge and Data Engineering, 26(1), 97–107. https://doi.org/10.1109/tkde.2013.109

48. Zhang, Y., Xu, C., Lin, X., & Shen, X. (2019). *Blockchain-Based Public Integrity Verification for Cloud Storage against Procrastinating Auditors*. IEEE Transactions on Cloud Computing, 9(3), 923–937. https://doi.org/10.1109/tcc.2019.2908400