Guilherme Nogueira Queiroz

Data and Control Mechanisms: Efficient Resource Allocation in Dynamic Data Economies (DDEs)

A study submitted in fulfillment of the requirements for the degree of Bachelor of Economics at the School of Economics, Business, Accounting and Actuarial Sciences of the University of São Paulo.

UNIVERSITY OF SÃO PAULO (USP)

SCHOOL OF ECONOMICS, BUSINESS, ACCOUNTING AND ACTUARIAL SCIENCES

Orientador: Prof. Dr. Rodrigo Menon Simões Moita

São Paulo, Brasil

2024

Data and Control Mechanisms: Efficient Resource Allocation in Dynamic Data Economies (DDEs)

July 1, 2024

0.1 Abstract

The text is based on the canonical model from ***"Non-rivalry and the Economics of Data"*** by Charles Jones and Christopher Tonetti, which describes the equilibrium behavior of a data economy. The model's probability kernel, derived from a dynamic system with productive feedback loops, involves firms and households deciding on consumption and output allocations. Including units of work and data in its production function, the model builds on Romer's endogenous growth theory (1990) and incorporates data as an input for producing varieties. While data availability boosts short-term growth and productivity, excessive usage leads to diminishing returns, limiting long-term growth. Data's non-rivalrous nature creates positive externalities, enhancing profitability and encouraging widespread usage in production. Data's non-rivalrous nature results in benefits and positive externalities that promote widespread data use and profitability in provisioning processes.

The research project employs a quantitative methodology to establish probable causal relationships between the components of the data economy model and model identities. The methodology enables experimentation with parameters, distributions, and variables of interest to understand and test the model's behavior in the scenario of efficient dynamic equilibrium, meaning optimal iso-elastic proportions and parametric stability. The project aims to contribute to the understanding of the economic and social implications of the data economy, exploring the potential benefits and challenges of steering corporate bodies towards accelerated growth.

The analysis of a real system describing equilibria behavior over time allows for the examination of different scenarios and the assessment of their impacts on the system. The literature review presents a set of pieces exploring the data economy from both production and consumption perspectives.

Keywords: **Data economies**, **Control Theory**, **Dynamic Equilibrium**, **Growth**.

JEL classification: `C22`, `D02`, `D11`, `D21`, `D24`, `D58`, `D83`, `E22`, `E47`.

- `C22` Time-Series Models • Diffusion Processes
- `D02` Institutions: Design, Formation, Operations, and Impact
- `D11` Consumer Economics: Theory
- `D21` Firm Behavior: Theory
- `D24` Production • Cost • Capital • Total Factor and Multifactor Productivity
- `D58` Computable and Other Applied General Equilibrium Models
- `D83` Search; Learning • Information and Knowledge • Communication • Belief • Unawareness
- `E22` Investment • Capital • Intangible Capital • Capacity
- `E47` Forecasting and Simulation: Models and Applications

0.2 Resumo

O texto é baseado no modelo canônico de ***"Non-rivalry and the Economics of Data"*** de Charles Jones (Stanford Institute for Economic Policy Research) e Christopher Tonetti (National Bureau of Economic Research), Working Paper No. 3716, que descreve o comportamento de equilíbrio de uma economia de dados. Seu núcleo de probabilidade é derivado de um sistema dinâmico com *feedback loops*, na medida em que firmas e famílias decidem alocações de consumo e produção. A função de produção inclui unidades de trabalho e unidades de dados. O modelo é baseado na Teoria do Crescimento Endógeno de Romer (1990) e incorpora dados como um insumo na produção de variedades. Devido a retornos decrescentes, se o uso de dados de uma firma exceder o ótimo, o crescimento de produção a longo prazo é limitado. A disponibilidade de dados é um fator de crescimento econômico a curto prazo, pois a produtividade da firma aumenta com o uso de dados em cadeias de produção. A natureza não-rival dos dados resulta em benefícios e externalidades positivas que promovem o uso generalizado de dados lucratividade nos processos de provisão.

0.3 Contents

0.4 List of Figures

0.5 References

1. **Allen, Beth**. Information as an Economic Commodity. Journal of Information Science, 44(4):475–86, 2018.
2. **Arrow, Kenneth**. The Limits of Organizations. New York: W. W. Norton, 1974.
3. **Avi Goldfarb, and Catherine E. Tucker**. Privacy and Innovation. Journal of Economic Perspectives 25 (2): 63–86, 2011.
4. **Beer, Stafford**. Sketch for a cybernetic factory. Chapter XVI in Cybernetics and Management, English Universities Press, p.142-152, 1959.
5. **Borg, Ingwer; Groenen, Patrick J. F.**. Modern Multidimensional Scaling: Theory and Applications (Springer Series in Statistics). Springer, 2, 2005.
6. **Brynjolfsson, Erik; Rock, Daniel; Syverson, Chad**. Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics. In Economics of Artificial Intelligence (pp. 23-57). University of Chicago Press, 2017.
7. **Carrière-Swallow, Yan; Haksar, Vikram**. The Economics and Implications of Data: An Integrated Perspective. International Monetary Fund Working Paper, 19-116, 2019.
8. **Clark, Norman**. Organization and Information in the Evolution of Economic Systems. In Evolutionary Theories of Economic and Technological Change, edited by P. Paolo Saviotti and Stanley Metcalf. 1991.
9. **Coelho, Antonio Augusto Rodrigues; Jeronymo et. Araújo**. Sistemas Dinâmicos: Controle Clássico e Preditivo Discreto. Editora da UFSC, 2019.
10. **Fajgelbaum, Pablo D.; Schaal, Edouard; Taschereau-Dumouchel, Mathieu**. Uncertainty Traps. Quarterly Journal of Economics 132 (4): 1641–92, 2017.
11. **Farboodi, Maryam; Veldkamp, Laura**. A Growth Model of the Data Economy. 2019.
12. **Farboodi, Maryam; Veldkamp, Laura**. Long-Run Growth of Financial Technology. American Economic Review 110 (8): 2485–523, 2020.
13. **Fluetchiger, Gerald E**. Control, Information, and Technological Change. Economics of Science, Technology and Innovation, 6:1–2, 1995.
14. **Ghosal, Ashitava**. Applied Cybernetics: Its Relevance in Operations Research. Gordon and Breach Science Publishers, 1978.
15. **Hirukawa, Masayuki**. Asymmetric Kernel Smoothing: Theory and Applications in Economics and Finance (JSS Research Series in Statistics). Faculty of Economics, Ryukoku University.
16. **Jones, Charles I.; Tonetti, Christopher**. Nonrivalry and the Economics of Data. American Economic Review, 109(8):2613–42, 2019.
17. **Kandel, Abraham; Langholz, Gideon**. Fuzzy Control Systems. Department of Computer Science and Engineering - University of South Florida; Department of Electrical Engineering - Florida State University; Tel Aviv University, Israel, 1993.
18. **Keith Ansell-Pearson, Joseph McCahery, Nick Land**. Machinic Postmodernism. Published in Association with Theory, Culture and Society Series. 1996.
19. **Kühlkamp, Nilo**. Introdução à Topologia Geral. Editora da UFSC, 2016.
20. **Laney, Douglas B**. Applied Infonomics: Why and How to Measure the Value of Your Information Assets. Infonomics, 1:1–20, 2018.
21. **Langlois, Richard N**. Systems Theory, Knowledge, and the Social Sciences. In The Study of Information: Interdisciplinary Messages, edited by Fritz Machlup and Una Mansfield. New York: John Wiley & Sons, 1983.
22. **Nirei, Makoto; Yamana, Kazufumi**. Value of Data: There's No Such Thing as a Free Lunch in the Digital Economy. BEA Working Paper, 2018-10, 2018.

23. **Romer, Paul M.**. Endogenous Technological Change. The Journal of Political Economy, Vol. 98, No. 5, Part 2: The Problem of Development: A Conference of the Institute for the Study of Free Enterprise Systems. (1990).
24. **Takagi, T.; Sugeno, M**. Derivation of Fuzzy Control Rules from Human Operator's Control Actions. IFAC Symposium on Fuzzy Information Knowledge Representation and Decision Analysis, 55-60, 1983.
25. **Varian, Hal**. Artificial Intelligence, Economics, and Industrial Organization. In The Economics of AI: An Agenda, pages 1–27. National Bureau of Economic Research, 2018.
26. **Veldkamp, Laura; Chung, Cindy**. Data and the Aggregate Economy. Journal of Economic Literature, 58(3):520–79, 2020.
27. **Wiener, Norbert**. Cybernetics: Or Control and Communication in the Animal and the Machine, 1948.
28. **Wiener, Norbert**. Tauberian Theorems. Annals of Mathematics, Second Series, 33 (1): 1–100, 1932.
29. **Yager, R.R.**. Deductive approximate reasoning systems. IEEE Transactions on Knowledge and Data Engineering (Volume: 3, Issue: 4). Page(s): 399 - 414. 1991. Publisher: IEEE. Machine Intelligence Institute, Iona College, New Rochelle, NY, USA.
30. **Zadeh, L. A**. Fuzzy Sets. Information and Control, 8:338-353, 1965.

0.6 Acknowledgments

I would like to thank all of *FEAUSP* for broadening perspectives and sharing the Dismal Science with such attention to detail. I am grateful for the opportunity to learn from one of the most effective, long-tailed knowledge building faculties in Brazil.

1 Introduction

1.1 Literature Review

The Economics of Data encompasses the study of the role of data as an input in value creation. The availability and use of data significantly influence the productivity of firms in global economies.

In the economic literature, the theme is approached from angles pertaining to Information Economics, Growth Theory, and Industrial Organization. The emergence of data markets, in which firms engage in the purchase, sale, and commercialization of data, is relatively new and has generated interest in understanding the principles of data collection, ownership, and custody.

Jones and Tonetti (2018) present an economic growth model based on ***"Endogenous Technological Change"*** by Paul M. Romer, published in the Journal of Political Economy in search of a dynamic solution for *"The Problem of Development"* in an occurrance of the Conference of the Institute for the Study of Free Enterprise Systems.

By inheritance, Romer's model is based on the work of Kenneth Arrow, having introduced the concept of "learning-by-doing" in his 1962 paper ***"Economic Welfare and the Allocation of Resources for Invention"***.

Arrow's model was further developed by Romer, who suggested the concept of endogenous growth, incorporating physical capital and human capital as inputs in the production function

$$Y(H_Y, L, x) = \int_0^{\inf} x_(i)^{1-\alpha=\beta} di$$

with H_Y as human capital, L as labor units, and x the excludable good produced by the firm. This model is, further, reduced to one with a single-state-variable by *"assuming the excludable good A_E that the firm produces intentionally is used in fixed proportions with physical capital"**.

∗: (Vol. 98, No. 5, pp. 71-102)

Jones and Tonetti's model connects firm, household, and optimal price problems for bundles of data. In this economy, the production function

$$Y_i = \left([\alpha x + (1-\alpha)\tilde{x}N]^{\eta}\,\nu\right)^{\frac{1}{1-\eta}}$$

includes units of work and units of data. Due to decreasing returns, if a firm's data usage exceeds the optimum, long-term production growth is limited (**Veldkamp and Farboodi, 2019**) and a boundary of the data bundle is reached. The availability of data is a short-term economic growth factor, as firm productivity increases with data usage in production chains (**Brynjolfsson, Rock, and Syverson, 2017**). Data's non-rivalrous nature, where consumption by one economic agent does not preclude others from using it, results in benefits and positive externalities that promote widespread data use in logistics, marketing, and production chains **(Jones and Tonetti, 2018)**.

According to **Varian (2018)**, firms use applications and software constructs to enhance productive efficiency. These systems promote continuous learning and improvement through feedback loops, positioning data use as a pillar for decision-making and value creation in modern organizations.

Acemoglu et al. (2020) investigate the dynamics of privacy, highlighting externalities where data about one individual can reveal information about others. These externalities can lead to market breakdowns, resulting in inefficiently low prices for data and excessive privacy losses, as individuals sell their data cheaply, anticipating others will do the same. Jones et al. hypothesize these issues are more significant for sensitive data types, such as health or social media information, but less critical for data types such as speech samples (for its recognition), particularly when effective anonymization reduces adverse selection concerns and mild privacy losses.

Allen (1990) discusses recognizing data as economic assets, highlighting the importance of information as a valuable resource to drive economic growth, due to its analysis and intelligent use. The strategic approach to data management, he argues, is essential for firms to leverage data as a competitive advantage in the digital age and is suggested to be considered a production process.

The notions of ownership by individuals and by firms that collect datum are central to the economic implications of data utilization and the subsequent models of data economies. Data ownership and privacy challenges highlight the need for nuanced and suitable approaches to data management, given an agent or firm's preferences for sharing data and the costs, benefits, and pronouncements of data sales.

Firms balance the benefits of data utilization with the ethical considerations of privacy and consent. The economic implications of these considerations are profound, influencing both firm behavior and regulatory frameworks, as per **Carriere-Swallow (2018)**.

In Jones and Tonetti's threat model, the data economy faces challenges related to privacy assurance and consent. Restrictions, of which the most common is the manifestation of the "privacy aversion" effect, are related to individuals' consent to share information, subsequently weighed and interpreted by firms in their allocative decisions. Consumers may be averse to selling their data but may also commercialize it if the price paid for the data bundles is high enough, with constant elasticity of substitution for each unit of shared data.

Beth Allen suggests that information has increasingly been recognized as an economic commodity. With the advancement of the digital age, the availability and access to data have become intrinsic to economic growth. The author emphasizes that the analysis and intelligent use of information bring significant benefits to organizations while providing competitive advantages and fostering innovation, while highlighting the importance of a strategic approach to data management as a valuable resource. Knowledge indexing is modeled as a production process, capable of driving value creation proportional to the quality of the data used, with data having the capacity to scale up knowledge assimilation, thus enhancing profitability in the proportion of data used balanced with the cost of data acquisition, if any.

Gerald E. Fluetinger (1995) explores the relationship between Control, Information, and Technological change. The author suggests that access to information plays a crucial role in the adoption and implementation of new technologies. The dynamic interplay between information access and technological innovation is alluded to be as central to modern economic growth as the availability of data.

The ability to use data efficiently helps companies adapt to technological changes and achieve greater production efficiency. **Makoto Nirei and Kazufumi Yamana (2018)** present a discussion on the value of data in the digital economy, stating there is no such thing as free lunch in data markets, emphasizing data has real economic value, as it is considered an essential resource for companies capable of generating significant economic outcomes through the employment of adaptive technologies, productivity enhancements, and data-driven decision-making.

Brynjolfsson, Rock, and Syverson (2017) address the productivity paradox in the age of artificial intelligence. The authors outline expectations and statistics regarding the impact of AI methods on firm productivity. Although artificial intelligence has the potential to drive productivity, its full realization depends on how data is used and integrated into business operations, they argue. **Varian (2018)** further suggests that companies are increasingly using AI-based applications to enhance efficiency, scale models, and derive cash-flow positive applications from quantitative abstraction. Applications are capable of processing large amounts of data in a performant and timely manner, according to computational learning methods applied in Statistics, Lateral Computing, and Natural Language Processing.

Allen's recognition of information as a critical economic commodity underscores the strategic importance of data management. As firms navigate the complexities of the digital age, the ability to leverage data intelligently becomes a key differentiator. The production process, with information as a crucial input, is adjusted with a myriad of optimization techniques, such as the use of fuzzy logic, to facilitate plant-design and cutting costs of performing self-hosted (or buying external) computation.

Gerald E. Fluetinger's insights into Control Theory and Technological Change provide a framework for understanding how data facilitates technological adaptation and operational improvements. The dynamic interplay between information access and technological innovation is alluded to be as central to modern economic growth as the availability of data.

Nirei and Yamana's (2018) assertion that data has intrinsic economic value reinforces the notion proposed by Allen (2018) that data is not merely a byproduct but a fundamental resource – one with divisibility but non-rivalrous properties.

The productivity paradox outlined by Brynjolfsson *et al* reflects on the relationship between Artificial Intelligence and productivity. The integration of AI into business processes hinges on effective data utilization, as well as effective multiple-input, multiple-output control systems. Furthermore, firms use modules and orchestration tools to deploy different context-intelligent, vertical, lightweight services on the web with causation in physical markets, as publicized, for instance, by Amazon Web Services (AWS) and the suite of integrated services provided by its competitors, able to handle large volumes of data while arranging commerce infrastructure. These services are designed to be scalable, secure, and reliable, thereby confirming the privacy-assimilating mechanism of data management and with firms alledgedly addressing privacy concerns, for purely economic purposes, while competing for market share.

As per what theorists entitle the "Multiplexing of Communication" through decentralized networks of assembled infrastructure and deemed-to-be causal relationships, the corporate world usem *datum corpora* to enhance decision-making processes, diggesting data from multiple sources. Collections of individual data points or datasets are used to train algorithms, analyze trends, and make predictions.

1.2 Consensus

It is considered indisputable that data is a valuable resource in dynamic systems that require efficient resource allocation to ensure optimal productivity and growth.

The intersection of privacy, data ownership, and economic value creation presents a multifaceted challenge for modern economies and is a topic of interest for researchers and policymakers alike.

Ultimately, the integration of advanced data management practices, informed Information Economics, provides a pathway for firms to harness the full potential of data.

It is disputed whether data ownership and utilization can be effectively managed within existing regulatory frameworks, or if new approaches are needed to address emerging challenges.

By addressing the challenges and opportunities associated with data ownership and utilization, firms can position themselves to thrive in an era where information is a key driver of economic success. The strategic use of data not only enhances productivity but also fosters a more informed and responsive approach to market dynamics, enabling for more thoughtful C-level strategic decisions, and, theorethicaly, less structural risk.

1.3 Further Considerations

Laura Veldkamp and Cindy Chung (2020) analyze the relationship between data and the aggregate economy, emphasizing that the availability of data bounds segment-specific growth. Allen, who inspires the duo, argues that the importance of information is self-evident. Based on Robert Solow's work, Allen highlights that economic progress is largely driven by the increase in knowledge, often referred to as "technological progress," which facilitates the transformation of goods through the dissemination of information. Successive entropy-reducing transformations serve as the basis for economic growth, with the iteration of scarce information leading to the creation of innovative offers.

In the Neoclassical system, knowledge is incorporated into production possibilities, with changes occurring exogenously, typically represented by an exponential factor.

While exponential growth is not perpetual, it serves as an approximation. Allen suggests that technical change influences relative prices, having a significant impact on the economy. The author argues that the value of information is determined by its ability to reduce uncertainty, with the value of information increasing as uncertainty decreases. The author also highlights the importance of information in reducing transaction costs.

Through the assessment of the hypotheses in the following section, this research aims to contribute to the understanding of social implications of data.

This emerging research field involves not only questions about the ability to process data but also on the understanding of its informational value and how it can be applied in various contexts to ascertaining the potential benefits and challenges of steering corporate bodies towards possibly accelerated growth.

Control Theory, originating in engineering and mathematics, provides a framework for analyzing and designing dynamic systems to control the behavior of variables of interest. Notable early contributions came from James Clerk Maxwell in the 19th century and the development of automatic control systems in the early 20th century. In the 1950s and 1960s, economists began applying CT to economic systems, particularly in stabilization analysis, with Richard Bellman being a notable figure. Concurrently, Solow and Arrow explored the economics of Technological Change using control-theoretic concepts, laying the foundation for the integration of CT into Economic Theory while exploring how dynamic optimization techniques could model resource allocation to R&D and technology diffusion. Advancements in Stochastic Control Theory during this period allowed economists to better model the uncertainties in technological innovation and adoption. Significant contributions also came from Robert Lucas.

Ashitava Ghosal suggests the formal integration of Control Theory into the economics of technological change began in the 1960s and 1970s. The book "Control, Information, and Technological Change. Economics of Science, Technology, and Innovation" by Fluetinger highlights the relationship between CT and the data economy. The author claimed the application of Information Theory in Economic Sciences is in its early stages but holds significant potential to revolutionize one's understanding of Economics. CT offers a flexible framework for modeling how data can be used to regulate and optimize complex systems.

Fluetinger, additionaly, points out that Information Theory, developed by Claude Shannon, provides powerful mathematical tools for quantifying, storing, and communicating information. These tools are particularly useful in modeling sequential processes, such as economic evolution, where uncertainty and entropy are crucial to the state of the system.

In Economics, *information* typically refers to *"data"*, *"facts"*, or *"knowledge"*, suggesting communication with semantic content. However, for Shannon, the quantity of data relates to the amount of uncertainty associated with a set of events which reveal entropy. Mathematically, the concept of information depends solely on the relative frequencies of events, without presumption of semantic content.

Some economists hoped that the rigor and mathematical richness of Information Theory, as extended by Warren Weaver, could be directly applied to conventional economic theory. However, a detailed analysis of Shannon's work and understanding of the communication problem from an engineering perspective reveals a profound theoretical incompatibility between the optimization and allocation problems of conventional economic theory and the engineering problems of operations research and control theory.

Arrow's comment regarding the definition of information represents conventional wisdom:

> *"The quantitative definition that appears in Information Theory is of limited value for economic analysis, for reasons pointed out by Marschak: different bits of information, equal from the standpoint of information theory, generally have very different benefits or costs."*

The observation highlights the difficulty of directly applying Information Theory to Economics without considering the specific context and differential value of data, as well as how it is used, stored, and communicated.

Fluetinger also proposes that Information Theory provides an appropriate way to model the processes by which control is achieved, being particularly useful for modeling sequential processes. As economists have recently developed evolutionary theories, they have also been drawn to Shannon's work. Examples of how Information Theory can be used to address economic problems can be found in articles by Langlois (1983), Clark & Saviotti (1991). *"Information"* frequently being refered to data and knowledge opposes the different encapsulation of what Communication Theory suggests. The quantity of information is defined by the uncertainty associated with events, in a probabilistic sense, arranging the entropy of the system. The more uncertain the system, the more valuable the information, as complexity is transduced into teleological articulation and virtually abstracted in transactions, as suggested by the authors.

These divergent definitions make it challenging to establish clear generalizations about how Information Theory can be directly applied to conventional Economic Theory. However, when appropriately contextualized, exposures of Systems-theoretic logic to the economic domain can provide a comprehensive understanding of the dynamics of economic systems, particularly in the context of Game Theory.

It has been argued that Information Theory is valuable for modeling processes in both biological and economic fields, and too are part of the same interface of evolutionary processes.

The evolutionary approach has been increasingly explored by economists.

When considering Control and information-transformating structures, it is essential to analyze how the economy operates in terms of institutions and agencies with decentralized responsibilities and limited information. Each controlling agency must form a behavioral strategy based on heuristics atop partial information it receives, thus raising the question of whether there are viable decentralized strategies that can guide the system to a target state, *i.e.*, whether the system is controllable and its map can be tested for the existence of, for example, a martingale, where the conditional expectation of the next value in the sequence is equal to the present value, regardless of all prior values, what provides a stable equilibrium after capital injection and risk-taking activities are spread across the causal network of randomly distributed decisions.

1.4 The Viable System Model

This section provides a brief introduction to the Systems Theoretic description of an organization encapsulated within a single level of the Stafford Beer's Viable System Model (VSM), implemented in England in the early 1970s.

The VSM is a model of organizational structure that identifies the essential functions necessary for an organization to survive and thrive in a complex environment.

Initially developed for the cybernetic analysis and synthesis of enterprises, the VSM provides a qualitative assessment framework for managing spontaneous organizational complexity.

It aids in creating practical tools for operational diagnosis and optimizing decision-making processes. Beer, a pioneer in Managerial Operations Research, developed the VSM by drawing on principles established by Norbert Wiener, Warren McCulloch, and Ross Ashby.

Beer's work paralleled the human organism's functioning with that of an organization. He perceived the human organism as comprising three interdependent components: organs and muscles, nervous systems, and the external environment (body, brain, qualia, and matter). Later, Beer identified five interactive systems he applied to organizational structures:

- System 1: Primary Activities
- System 2: Conflict Resolution
- System 3: Internal Regulation, Optimization, Synergy
- System 4: Adaptation, Action Planning, Strategy
- System 5: Policies, Final Decisions, Identity

A VS comprises five interacting subsystems that can be mapped onto aspects of organizational structure. Broadly, Systems 1 to 3 are concerned with the organization's *"here and now"* operations, System 4 addresses the *"there and then"* by strategizing responses to external, environmental, and future demands, and System 5 balances the *"here and now"* with the *"there and then"* to provide policy directives that maintain the organization as a viable entity.

System 1 contains several primary activities, each of which is itself a VS due to the recursive nature of systems. These activities perform functions that implement at least part of the key transformation of the organization.

System 2 represents the information channels and bodies that allow the primary activities in System 1 to communicate with each other and enable System 3 to monitor and coordinate these activities. It is responsible for scheduling shared resources used by System 1.

System 3 establishes the rules, resources, rights, and responsibilities of System 1 and provides an interface with Systems 4 and 5.

System 4 looks outward to the environment, monitoring how the organization needs to adapt to remain viable.

System 5 makes policy decisions to balance demands from different parts of the organization and steer the organization as a productive entity.

Applying the VSM involves using variety measures to match people, machines, and money to jobs that produce products or services. In a set of processes, some jobs are done by one person, some by many, and often many processes are done by the same person. Throughout the working day, participants may shift their focus between internal and external Systems 1 to 5 as they complete tasks.

The choices or decisions made and their associated costs (or efforts) define the variety and resources needed for a job. In this case, *"variety"* refers to a measure of the complexity and the number of possible states of a provision system.

System 3 operationally manages the processes (Systems 1) by monitoring performance and ensuring (System 2) the flow of products between System 1 and out to users. System 3 can audit (via 3*) past performance, comparing bad times to good times. When performance declines or risk levels increase, System 3 seeks help or consults colleagues for a remedy. This process involves an algedonic alert, which can be automatic when performance falls below capability targets. The autonomic $3-2-1$ homeostatic loop absorbs the problem for resolution within its metasystem, with System 4 (representing Research and Marketing) asked for recommendations.

If additional resources are required, System 5 decides the best option from System 4.

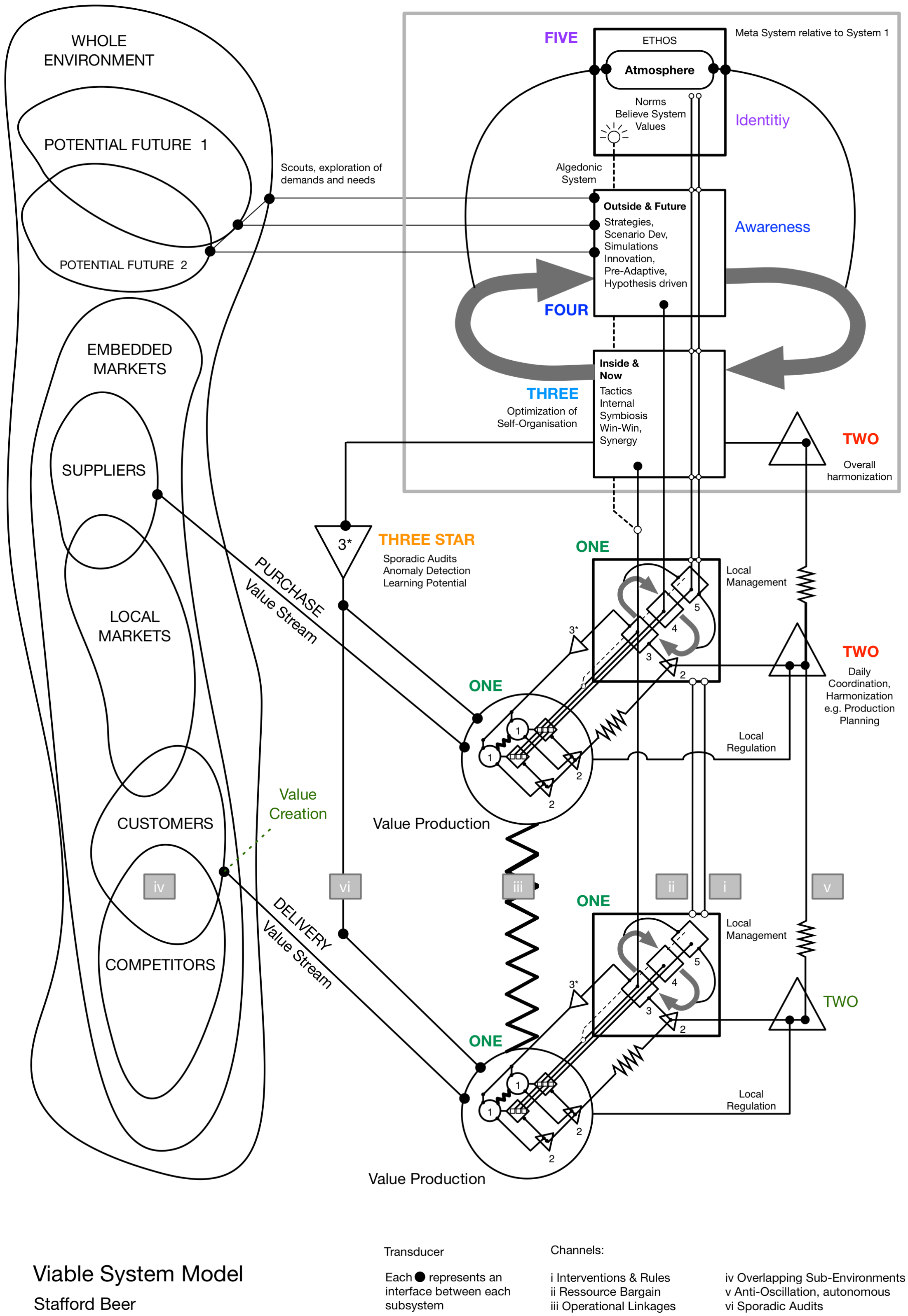
If the solution needs more resources than the current capability or variety can sustain, escalation to higher management is necessary. Similarly, performance-improving innovations can be handled in this manner.

In small businesses, all these functions might be performed by one person or shared among participants. In larger enterprises, roles differentiate and specialize. Local conditions, the environment, and the nature of the service or product determine how warehousing, sales, advertising, promotion, dispatch, taxation, finance, and salaries fit into the picture. Not all enterprises charge for their transactions (*e.g.*, some schools, medical services, policing), and voluntary staff may not be paid.

Regardless of the circumstances, all enterprises must be useful to their customers to remain viable. For all participants in the system, the central question is: *"Do I do what I always do for this transaction, or do I innovate?"*

This question is central to System 4's role. The VSM describes the constraints by providing knowledge of past performance and how it can be improved.

Figure 1: Viable System Model



Beer dedicated *“Brain of the Firm”* to his colleagues with the words *“absolutum obsoletum”*, meaning *“If it works, it’s out of date”*. The statement underscores the need for continuous adaptation to retain viability, promote risk-adjusted profitability, and ensure an organization’s long-term survival.

Wiener, considered the founder of Cybernetics and proponent of modern Stochastic Control Theory, developed the concept of feedback control systems. It is theorized that all intelligent behavior results from feedback mechanisms that could potentially be simulated by machines.

The relationship between Information Theory and the Digital Economy is particularly relevant for understanding how data can be used as a resource. It also aids in analyzing how it can affect firm productivity and economic growth, especially concerning data analysis in informed decision-making scenarios.

Economists generally attribute the functioning of the economy to the interaction of various agents, but do so often lacking a comprehensive understanding of the underlying social systems of structures and institutions, falling into the trap of assuming the economy is somehow dissociated from the symbiotic relationship between the agents, the environment, and automatic checks and balances that regulate some of the system’s dynamics, and, by recursion, a portion of the system’s behavior. The structuralist, quantitative perspective proposed offers a holistic view of economic phenomena, often overlooked by mainstream economic theories, but foundational to the understanding of intertwined chains of statistical causality.

2 Hypotheses for the Data Economy Model

a-H, ***Enhanced Firm Productivity through Data Utilization***. Or *"Firms that collect consumer data and use it as an input in their production chains exhibit higher productivity than firms that do not utilize data."*

This hypothesis aims to quantify the productivity gains attributable to data utilization.

b-H, ***Consumer Well-being and Data Sharing*** hypothesis: *"Consumers who share their personal data with firms experience greater well-being than consumers who do not share their data."*

Data sharing can lead to personalized consumer experiences, enhanced service delivery, and better-targeted products, all of which can contribute to higher consumer satisfaction. This hypothesis seeks to evaluate the extent of these benefits.

c-H, ***Data Availability as an Economic Growth Driver***: *"The availability of data is a factor in short-term economic growth."*

Data availability facilitates innovation, efficiency, and scalability within firms, thereby boosting their productivity and contributing to overall economic growth. This hypothesis aims to assess the impact of data availability on economic performance.

3 Justification for this Research, and Implications

The increasing use of algorithms and artificial intelligence in digital economies raises several important questions. One of the key concerns is the potential for market power concentration and the formation of monopolies in the digital space. As firms leverage data and algorithms to enhance their competitive advantage, there is a risk that a few dominant players may emerge, stifling competition and innovation. Tangents on Self-Regulation and Private Policy are also relevant to this discussion.

3.1 Market Power and Monopolies

The implications of market power concentration and the formation of monopolies in the digital economy are significant. Price discrimination, reduced consumer choice, and barriers to entry for new firms are some of the potential consequences of monopolistic practices.

> *"To what extent are digital economies fostering new forms of market dominance, and what measures can be implemented to address these challenges?"*

Is a question that arises from the potential for market power concentration in the digital economy. The emergence of dominant players in the digital space can have far-reaching implications for competition, innovation, and consumer welfare. Addressing these challenges requires a nuanced understanding of the dynamics of digital markets and the regulatory frameworks needed to ensure fair competition and consumer protection.

Understanding the role of data in sustaining economic growth and enhancing firm productivity is a critical area of inquiry in economic research. While the common perception acknowledges the intrinsic value of data in driving innovation and efficiency, it is imperative to delve deeper into how the availability, access, and utilization of data influence the broader economic dynamics of growth and the social calculus of well-being.

3.2 Economic Significance of Data

Data is often heralded as the new oil, a key resource in modern times. Its ability to transform industries, optimize processes, and create new business opportunities makes it a vital component of economic analysis. By studying how data contributes to economic growth, we can identify the mechanisms through which data-driven strategies create capital gains and manifest in local minima and maxima.

3.3 Innovation and Efficiency

Data-driven innovation and efficiency are at the forefront of contemporary business practices. Firms that effectively harness data can streamline operations, reduce costs, and innovate rapidly.

3.4 Privacy and Security Concerns

The rise of data-driven economies brings to the fore significant concerns regarding consumer privacy and data security. Understanding the balance between leveraging data for economic benefits and protecting individual privacy is crucial.

3.5 Policy Implications

Policymakers need insights into how data influences economic growth to craft informed regulations.

3.6 Addressing Popular Perceptions

Incorporating data into dynamic economic models provide a nuanced understanding of growth patterns and productivity shifts. This research will explore how data-driven variables can be integrated into economic models to reflect contemporary economic dynamics accurately. The implications of data on the future of work, including job creation, skill requirements, and workforce dynamics, are of immense interest to researchers and industry stakeholders.

Data-driven technologies already influence labor markets and economic structures, and a compounding effect is expected in the future, not necessarily in the nature or essence of continuity, but in the level, magnitude, and direction of aggregate, marginal changes.

Traditional Economics often utilizes mathematical techniques that, while useful, lack the elegance and depth found in pure mathematics. This elegance is not just aesthetic but functional; it provides a precise and comprehensive framework.

The interdisciplinary application of the Literature Review is indicative of a broader trend where the boundaries between disciplines are becoming increasingly permeable. The historical separation of cited fields has limited the development of a unified approach to understanding complex systems. However, with the realization that these fields share common structures and can inform each other, there is an opportunity to create an integrated approach to Economic Theory.

The resistance to incorporating sophisticated mathematical tools in economics is often due to the political nature of the field. Economics, unlike the hard sciences, is deeply intertwined with political agendas. This political dimension can hinder the adoption of new methods that might disrupt established economic models or challenge prevailing political narratives.

4 Methodology

The study employs a quantitative methodology involving mathematical simulation. By analyzing a real system composed of a set of equations, the numerical resolution of these model equations allows for the examination of different scenarios and the assessment of their impacts on the system.

In a parametric approach, the model is specified in terms of a finite number of parameters. The underlying distribution or process is assumed to follow a specific parametric form. For Wiener-Brownian motion, this typically involves defining the process using parameters such as the drift μ and the volatility σ.

Example:

A parametric form of the Brownian motion with drift can be written as:

$$X(t) = X(0) + \mu t + \sigma W(t)$$

where:

- $X(t)$ is the value of the process at time t,
- $X(0)$ is the initial value,
- μ is the drift parameter,
- σ is the volatility parameter,
- $W(t)$ is a standard Wiener process.

Leveraging the framework proposed in the working paper ***"Non-rivalry and the Economics of Data"***, the primary focus of this project, it is possible to assess dynamic equilibrium in a data market with N varieties.

The stability of the geometric Wiener-Brownian motion (GWBM) is comprised of a continuous-time stochastic process whereby the logarithm of the randomly varying quantity follows a Brownian motion in its classical form.

The GWBM is a generalization of the geometric Brownian motion, which is a stochastic process used to model the dynamics of a quantity whose logarithm is normally distributed. The GWBM is associated to the production equation.

Variety i is produced from labor L_i and data D_i, expressed as:

$$Y_i = D_i^\eta L_i$$

with its unindexed form

$$Y = \left(\int_0^N Y_i^{\frac{\sigma-1}{\sigma}} \, di \right)^{\frac{\sigma}{\sigma-1}} = N^{\frac{\sigma}{\sigma-1}} Y_i$$

evaluating to

$$Y_i = D_i^\eta L_i = D_i^\eta \frac{L}{N} = D_i^\eta \nu$$

where L represents the total amount of labor in the economy, symmetrically allocated across all varieties, and ν represents the firm size. From this, it is interpreter that the bigger the firms are, the more data they use, and the more they produce, with a limit of entrance in the data market, as per the model.

The data equation is

$$D_i = \alpha x_i Y_i + (1-\alpha) B_i = \alpha x_i Y_i + (1-\alpha)\hat{x} N Y_i = [\alpha x + (1-\alpha)\hat{x} N] Y_i.$$

Y_i is the quantity of data generated by variety i.

- x represents the amount of data produced by the firm.
- B represents a bundle of data from other varieties utilized by the firm.
- α represents the proportion of data the firm can use from its own variety.
- $\hat{x}$ is the amount of data generated by other firms, and signifies the importance of a firm's data relative to data bundles from competing firms.
- NY_i denotes the amount of data generated by competing firms.

The allocations x and $\hat{x}$ are endogenously chosen based on consumer privacy considerations.

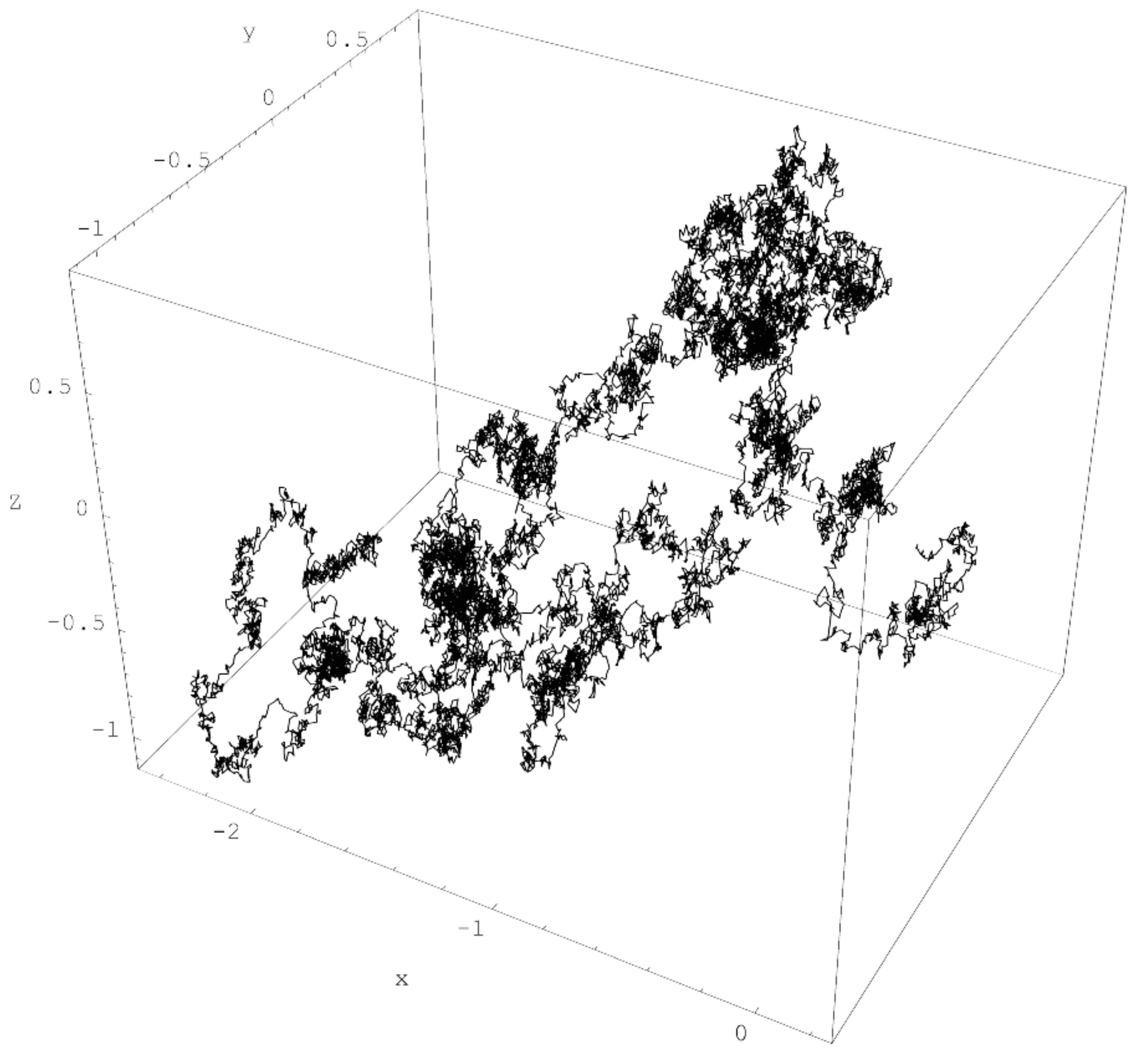
There is a multiplier effect associated with data: as more consumers purchase a product, more data is generated. This increase in data enhances productivity, leading to higher production, consumption, and further data generation, creating a virtuous cycle.

In this economy, a Constant Elasticity of Substitution (CES) aggregator manifests through an aggregate production function, where the elasticity of substitution between data and other inputs is a function of the data price:

$$Y = N^{\frac{\sigma}{\sigma-1}} \left([\alpha x + (1-\alpha)\hat{x} N]^{\eta} \nu \right)^{\frac{1}{1-\eta}}.$$

Y is modeled as a Wiener process, say, $Y(t)$, with a drift term μ, volatility term σ, and a random coordinate $W(t)$.

Figure 2: Random Production Walks modeled as a Wiener Process **



(**) Elaborated by **Gerald E. Fluetinger (1995)**

In an econometric model, the system S represents a network of economic variables and their interactions. For example, in a production system, X could be the set of input factors (labor, capital, data or raw materials), and R would include production functions, cost functions, and constraints. Given an initial state X_0, the system evolves over time according to a set of equations that describe the dynamics of the variables, and its spectral window is defined by the set of possible states the system can reach.

4.1 Example: The demand for new telephone connections, from *"Applied Cybernetics: Its Relevance in Operations Research"*

The demanded quantity at any city depends on a number of economic factors, *e.g.* growth of industries around it, rate of urban development, growth of facilities like educational institutions, hospitals, etc. In some developing countries, however, the total demand (existing plus new) depends on the capacity available in terms of exchanges and line capacity. Let Y, be the total demand at time t, K_t the capacity available then, and X_1, X_2, X_3 say the economic factors affecting the potential demand. It is defined

$$Y_t = f(X_1, X_2, X_3)$$

so long as the demand is very much less than the capacity K_t at time t. When the potential demand (Y_t^p) is significantly greater than K_t, one estimates it through a relation and the effective demand (Y_t) may be governed by a ratio of the form

$$dY_t/dt = \alpha Y_t^r (K_t - Y_t)^s$$

where K_t, is the capacity at time t as planned, α is a parameter, $(K_t - Y_t)$ the unfulfilled demand, and r, s two index parameters satisfying feedback relations of capacity (K_t) on Y_t.

One comes across feedback connections of various types in real-life problems – linear and non-linear, deterministic and stochastic, and mixtures of various types.

The topics of isomorphic and homomorphic systems are discussed in the book. Isomorphic relations between two systems involve a one-to-one correspondence between the elements of the two systems. Homomorphic systems, on the other hand, involve a many-to-one correspondence between the elements of the two systems.

In a stochastic system at least one of the outputs is a random variable, and it may relate isomorphism with respect to their probability distributions.

4.2 Elements of Control Theory

This section considers the mathematical treatment of control systems.

Primarily linear systems have been dealt with, and how the feedback problem can be studied with the help of differential equations have been illustrated. For detailed treatment of CT, one may refer to a standard book, Shinners (1964).

Consider a system in which the input is an m-vector Y.

The system is called linear if an input

$$\sum_{i=1}^{p} a_i X_i$$

produces an output

$$\sum_{i=1}^{p} a_i Y_i$$

where an a_i are constants $(i = 1, \ldots, p)$. Both the input and the output are additive.

Thus we find that the input-output relationship can be described by a linear differential equation.

Suppose $x(t)$ is the system input and $y(t)$ the system output at time t, then we can describe the input-output relation in the form

$$\sum_{i=1}^{m} u_i \frac{d^i y}{dt^i} = \sum_{i=1}^{n} b_i \frac{d^i x}{dt^i}.$$

Likewise we may represent the input-output relation in the form of an integral equation

$$y(t) = \int_{-\infty}^{\infty} r(t,k) z(k) dk$$

where $r(t, k)$ is a function embodying the characteristic R in the system.

Solutions of simple differential equations of linear type can be found on the text by *DiStafano III et al. (1967).* In principle, we solve for an equation in two stages

$$y(t) = y_f(t) + y_g(t)$$

for the free response $y_f(t)$ by putting $x_t = 0$, so that the problem is reduced to a homogeneous equation of order m in y_f.

Solving for the forced response $y_g(t)$, by making $\frac{d^i y}{dt^i} = 0$, $(i = 0, \ldots, m)$, $y_g(t)$ depends only on $y(t)$.

4.3 Nuances on Auto-regressive Models

In the realm of process control, auto-regressive models derived from input-output structures are commonly utilized due to the inherent complexity of phenomena, which often precludes the development of theoretical models. However, these processes frequently exhibit significant non-linearity, resulting in models with excessively high orders that account for past effects unrealistically.

To address this challenge, a strategy involves creating multiple submodels, each simple, comprehensible, and dedicated to specific sub-domains. While the multi-model approach itself is well-established, integrating Fuzzy Set Theory offers a novel method for constructing such multi-models based on input-output forms.

The fuzzy implication inference model employs numerous rules where each rule's premise is defined by a fuzzy proposition, and its consequence is represented by a linear or auto-regressive exogenous model, especially in process systems. The output of a fuzzy model is calculated as a weighted sum of the individual consequences' outputs, with weights determined by the membership grades of inputs to their respective premises.

In the domain of auto-regressive models, the predictive control method has garnered considerable academic interest. This method aims to derive a control sequence that aligns the output sequence closely with a reference trajectory, compensating for model inaccuracies through process measurement. An alternative approach proposes designing such predictive control using a fuzzy auto-regressive exogenous model, where future rule weights are aligned with the reference trajectory.

Measurement data, often termed *"samples"*, implicitly suggest an imaginary population and potential measurement errors. Assuming unlimited nonlinearity or model orders, it is feasible to construct models by approximating samples. However, criteria such as the determination coefficient adjusted for degrees of freedom or Akaike's information criterion penalize degrees of freedom reduction. The group method of data handling addresses this issue by introducing an unbiased criterion. Despite these advancements, many existing methods fail to satisfy modelers dealing with complex, nonlinear systems. System modeling, distinct from pattern recognition, typically seeks direct understanding rather than computer-mediated solutions. The most comprehensible models are linear combinations akin to fuzzy models.

The fuzzy model represents a nonlinear framework comprising rule-based linear models and membership functions, which define rule confidence levels. Fuzzy modeling involves interdependent subproblems, such as fuzzy data space partitioning and membership function identification.

To achieve a satisfactory model amidst infinite possibilities, one must predefine the scope of examination and the criteria for satisfactory results. Establishing a criterion and a search algorithm is logically appealing, as it legitimizes solutions derived from their uses. While Sugeno and Kang's algorithm follows a meticulous procedure, theoretically, it is impossible to achieve an ideal model solely through normative methods. Intuition is crucial for navigating towards a viable model applicable in real-world situations. The Takagi-Sugeno model epitomizes a fuzzy model by merging logic and intuition, contrasting with neural network models in nonlinear modeling.

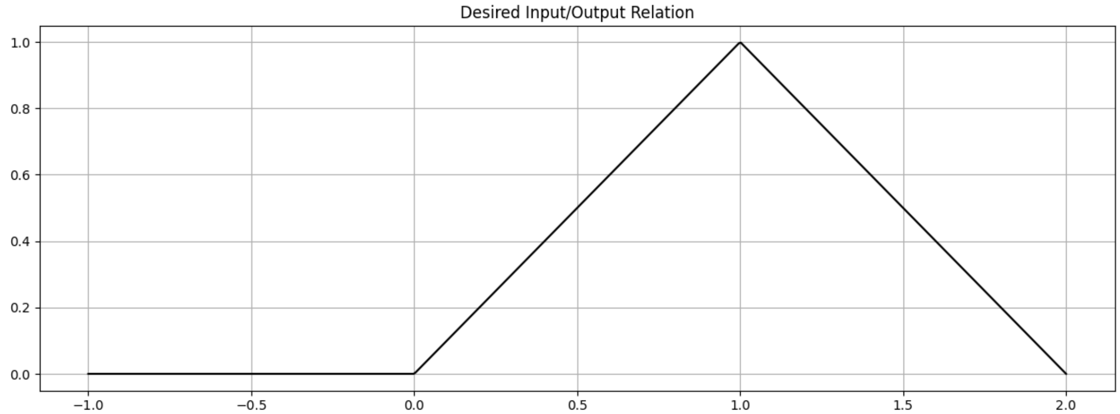
Fuzzy Set Theory transcends mere interpolation techniques for nonlinear systems analysis. A modeling algorithm should strategically enhance the likelihood of discovering superior models through the modeler's judgment. An interactive approach, supported by computational assistance, is recommended. This approach focuses on variable analysis to distill observations to an understandable level.

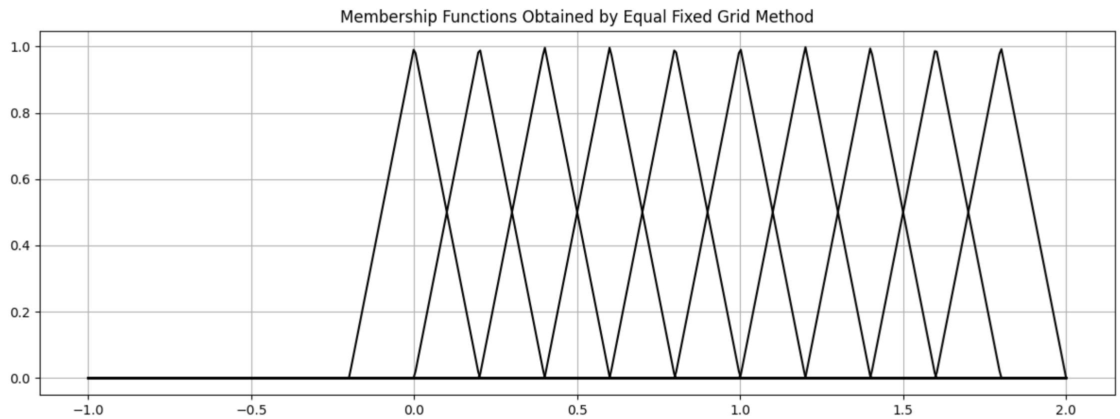
4.4 Knowledge Acquisition of a Fuzzy Controller

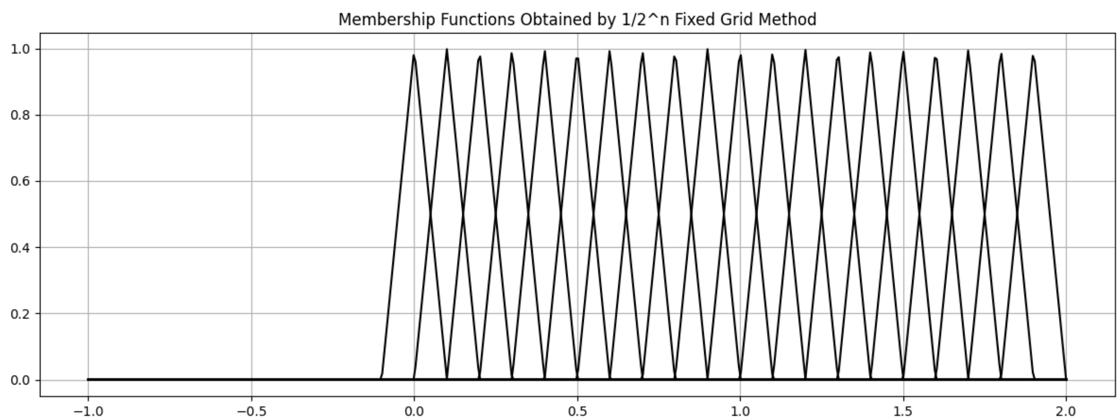
This section examines the application of the hybrid algorithm in the design of a fuzzy control system. One effective approach to designing a fuzzy controller involves linguistically encoding the expertise of skilled operators or control agents using fuzzy rules. However, when operators are unable to linguistically articulate their actions in specific situations, modeling their control actions with numerical data becomes highly beneficial. In such cases, the design of the fuzzy controller is transformed into a fuzzy modeling problem.

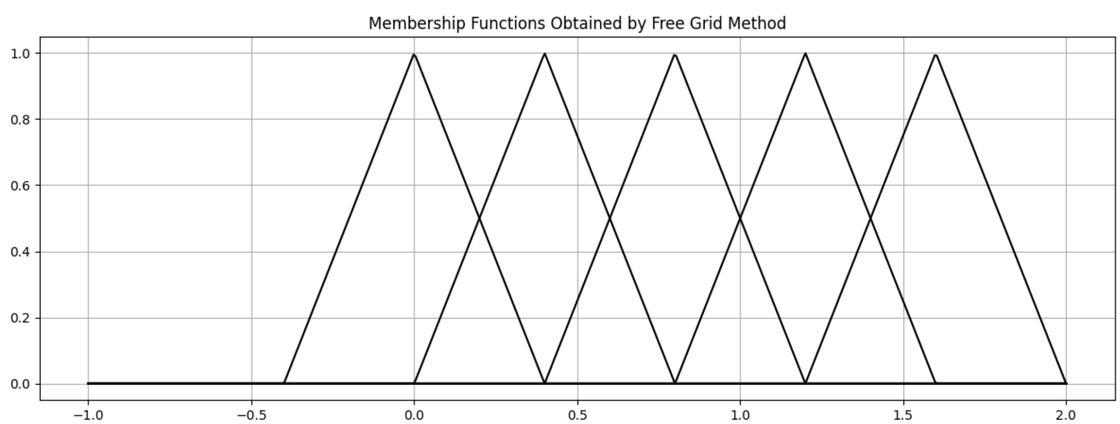
Another significant application of the hybrid algorithm is in the refinement of knowledge within the fuzzy controller pertains to knowledge acquisition issues. The hybrid algorithm excels in achieving specified model errors with a relatively small number of fuzzy rules. Consequently, this method is anticipated to be effective in minimizing the number of optimal fuzzy rules for a feedback fuzzy controller. Consider the model reference of the adaptive fuzzy control system provided in "Fuzzy Control Systems" by Kandel:

Figure 3: Adaptive Fuzzy Control Relations


Desired Input/Output Relation
1.0
0.8
0.6
0.4
0.2
0.0
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0


Membership Functions Obtained by Equal Fixed Grid Method
1.0
0.8
0.6
0.4
0.2
0.0
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0


Membership Functions Obtained by 1/2^n Fixed Grid Method
1.0
0.8
0.6
0.4
0.2
0.0
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0


Membership Functions Obtained by Free Grid Method
1.0
0.8
0.6
0.4
0.2
0.0
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0

Assume that a production facility is a second-order dumping system, whose transfer function $G(s)$ is given by:

$$G(s) = C/(s? + As + B), A = 0.638, B = 0.034, C = 1.228$$

The transfer function $H(s)$ of the reference model is

$$H(s) = 1/(1 + Ts)ST = 1.8.$$

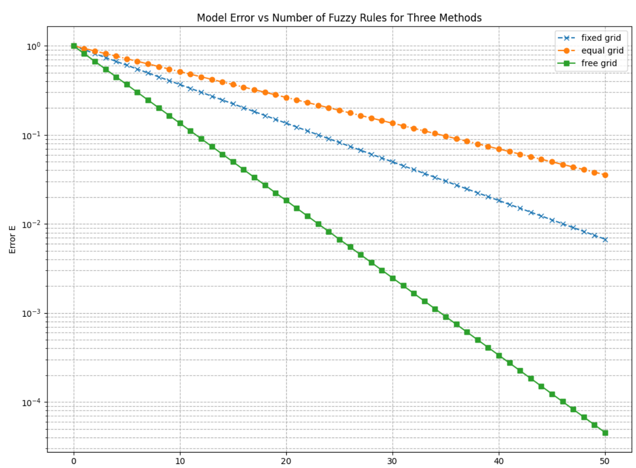
The desired reaching time is $TR = 25$ sec.

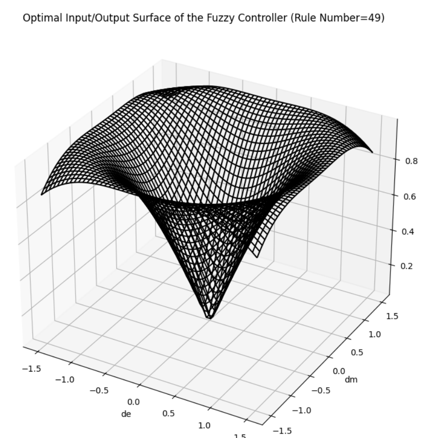
This exercise, altough providing a discrete time solution, is of paramount importance for the development of reverse-engineered models. Econometric models, for instance, can be reverse-engineered to provide a broad understanding of underlying economic dynamics, utilizing a naive approach to the data economy.

The hybrid algorithm is particularly useful in this context, as it can be used to start from a set of data points and derive a model that accurately represents the underlying system, even when a solution is not readily apparent.

Since only the conclusion parts of the fuzzy rules are tuned in adaptive learning gain algorithms, we assume 7×7 membership functions of input. Therefore, the number of fuzzy rules is 49, assuming the errors are distrubuted as follows, the optimal input-output relation is obtained, with the change in error and the derivative of the measured variable as follows:

Figures 4 and 5: Error Magnitude, Optimal I/O Surface





This formulation allows for the derivation of an optimal I/O surface, which is a critical component of adaptive equilibrium as errors are minimized.

4.4.1 Set Point

The set point in a CS refers to the desired value that the system aims to maintain. It is the target or reference point that the controller strives to achieve and stabilize, logically representing the optimal operating condition that the system should follow. The hybrid algorithm assists in fine-tuning the fuzzy rules to ensure that the system's output closely tracks the set point despite any fluctuations or variations in the process.

Achieving accurate set point tracking involves creating fuzzy rules that can interpret and respond to deviations from the set point. The hybrid algorithm enhances this process by optimizing the fuzzy rule set, ensuring precise and consistent control performance.

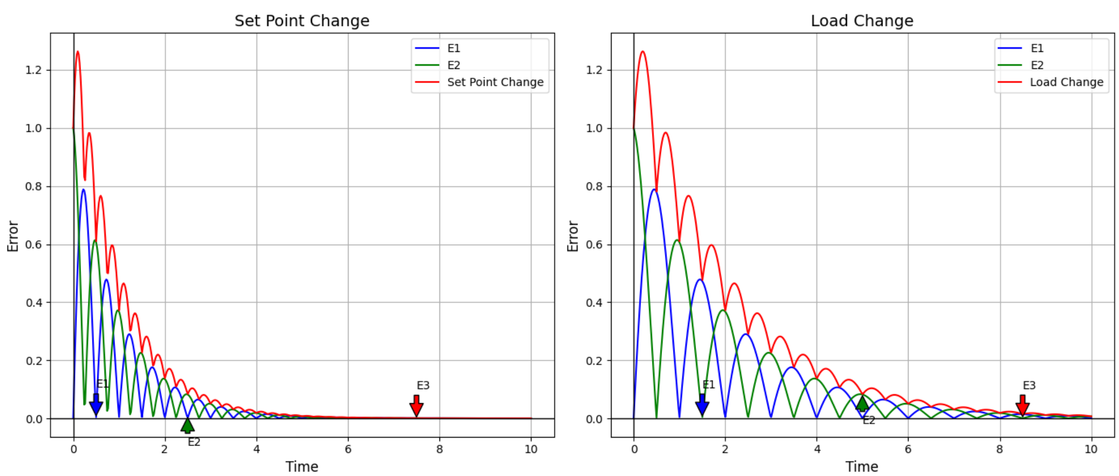
4.4.2 Load Disturbance

Load disturbance refers to unexpected changes or disruptions in the system's operating conditions that can affect the process being controlled. These disturbances can be external factors such as environmental changes, noise, or variations in load demand. Effective control systems must be able to compensate for these disturbances to maintain stability and achieve the desired set point.

The hybrid algorithm plays a pivotal role by refining the fuzzy rules to adapt to these disturbances.

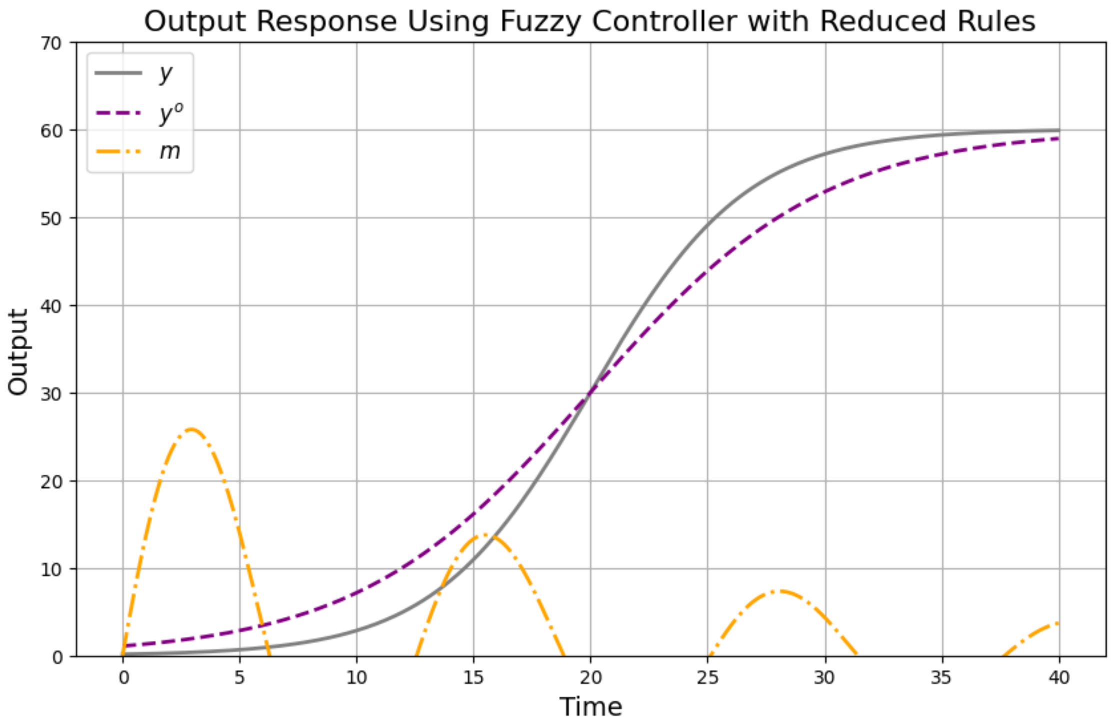
The refinement process involves minimizing the model error and reducing the number of fuzzy rules required to achieve optimal control. By incorporating self-tuning methods such as the adaptive learning gain algorithm, the hybrid approach ensures that the fuzzy controller can dynamically adjust to load disturbances.

Figure 6: Set Point and Load Disturbance Error Responses



Assuming that the system is initially at rest, its output-response to a step input is shown in Figure 6, with y representing the output response, y^0 the initial output, and m the number of fuzzy rules.

Figure 7: Output Response to Step Input



The goal is to design such a fuzzy controller that the output transition of the production for the step change of the setting value r gets closer to that of the reference model as much as possible.

5 The Economic Environment Synthesized

5.1 Utility (A)

$$\int_0^\infty e^{-\rho t} L_t(c_t, x_{it}, \tilde{x}_{it}) dt$$

5.2 Flow Utility (B)

$$u(c_t, x_{it}, \tilde{x}_{it}) = \log c_t - \frac{\mathrm{K}}{2} \frac{1}{N_t^2} \int_0^{N_t} x_{it}^2 di - \frac{\tilde{\mathrm{K}}}{2} \frac{1}{N_t^1} \int_0^{N_t} \tilde{x}_{it}^2 di$$

5.3 Consumption per person (C)

$$c_t = (\int_0^{N_t} c_{it}^{\frac{\sigma-1}{\sigma}} di)^{\frac{\sigma}{\sigma-1}}$$

$$\sigma > 1$$

5.4 Data creation (D)

$$J_{it} = c_{it} L_t$$

5.5 Variety resource constraint (E)

$$c_{it} = \frac{Y_{it}}{L_t}$$

5.6 Firm production (F)

$$Y_{it} = D_{it}^\eta L_{it} G$$

with

$$\eta \in (0, 1)$$

5.7 Data used by firm i (G)

$$D_{it} \leq \alpha x_{it} J_{it} + (1 - \alpha) B_t$$

with

$$x_{it} \in [0, 1]$$

5.8 Data on variety i shared with others (H)

$$D_{sit} = \tilde{x}_{it} J_{it}$$

with

$$\tilde{x}_{it} \in [0, 1]$$

5.9 Data bundle (I)

$$B_t = (N_t^{\frac{1}{\epsilon}} \int_0^{N_t} D_{sit}^{\frac{\epsilon-1}{\epsilon}} di)^{\frac{\epsilon}{\epsilon-1}}$$

with

$$\epsilon \geq 1$$

5.10 Innovation (new varieties) (J)

$$\dot{N}_t = \frac{1}{\mathcal{X}} \times L_{et}$$

5.11 Labor resource constraint (K)

$$L_{et} + L_{pt} = L_t$$

where

$$L_{pt} \equiv \int_0^{N_t} L_{it} di$$

5.12 Population growth (exogenous) (L)

$$L_t = L_0 e^{g_L t}$$

5.13 Aggregate output (M)

$$Y_t \equiv c_t L_t$$

5.14 Creative destruction (N)

$$\delta(\tilde{x}_{it}) = \frac{\delta_0}{2} \tilde{x}_{it}^2$$

The system pertinent to the problem is delineated in terms of its constituent elements and their interrelationships. This holistic view allows for a comprehensive understanding of the system's dynamics. The problem is quantitatively defined, as far as feasible, in relation to the elements of the system. This step involves the translation of the problem into measurable variables and parameters. The system's communication and control mechanisms are examined, considering both deterministic and stochastic elements. This involves understanding how information flows within the system and how different elements respond to changes.

6 The formulation from *"Non-Rivalry and the Economics of Data"*

The classical theory of endogenous growth acknowledges that innovation is fundamental, albeit exogenous to the model. Implicitly, innovation is modeled as a pure public good. Ideas are generated and can be absorbed at no cost by firms, which invest in research and development (R&D) to discover new goods and production methods to sell to their customers. This investment constitutes a fixed cost for the firm; thus, if the newly produced good is excludable through patents, the price of the goods will exceed the marginal cost. Consequently, there will be spillovers from this new technology that the firm does not capture, indicating that firms have an incentive not to invest in R&D.

In this model, data is generated as a byproduct of economic activity and tends to generate more data. Consumers become data generators and factors in the production cycle by participating with personal data. Additionally, data is equally useful across all firms within a given economic sector, allowing firms to purchase data from others to fill knowledge gaps, thereby reinforcing learning and innovation processes.

The characteristics of rivalry and exclusion are related. According to Jones and Tonetti (2019), if there is no rivalry in consumption, there is no reason to exclude, except to raise funds. How non-rival goods are financed determines whether a good becomes a public good or a low-congestion good. Indirect financing of public goods and fees for low-congestion goods are ways to avoid the free-rider problem.

Private and common goods are subject to rivalry in consumption. It is considerably more challenging to clearly define and enforce property rights for common goods, according to **Veldkamp (2019)**.

In this economy, a central planner aims to share data about variety i with firm i, as this increases productivity and output. Similarly, the planner wants to share data about variety i with other firms to leverage the non-rivalry of data, thereby increasing the productivity and output of all firms. Therefore, the question *"What proportion of data should be shared with each firm?"* is considered.

Optimal data sharing decreases with privacy costs and increases with the importance of data. If data is more critical, the economy allocates fewer resources to non-data-producing inputs and more resources to production, which generates data. This represents the privacy cost associated with data about variety i shared with firm i and other firms.

In equilibrium, firms decide whether to sell data. Data is bought and sold through an intermediary, which collects data from all varieties and resells it to individual firms. In this market, data sellers always set prices if they have market power, while data buyers are price takers.

Consumers can sell data to an intermediary and choose how much data to sell to balance income gain against privacy costs, which respond to the phenomenon of privacy aversion.

Additionally, consumer data custody is socially preferable to firm custody, as consumers tend to sell slightly less than the socially optimal amount due to monopoly mark-up distortion. Firms generally value data less than the planner, resulting in an inefficiently low equilibrium price. However, consumer-owned data allocation usually generates higher well-being.

Only when creative force is very weak and privacy concerns are very high might firm data ownership policies be marginally better, as firms are concerned with Schumpeterian creative destruction when granting data access to competing firms.

In an allocation where only firms own the data, they "misuse" the data bundles: the utility cost associated with privacy does not factor into the firm's solution, as firms do not concern themselves with privacy. The fraction of data a firm chooses to sell to other firms when it owns the data depends on its concern over losing data ownership, capturing the role of creative destruction.

For

$$\int_0^\infty e^{-\rho t} L_t u(c_t, x_{it}, \tilde{x}_{it})\, dt$$

and

$$u(c_t, x_{it}, \tilde{x}_{it}) = \log c_t - \frac{\mathcal{K}}{2}\frac{1}{N_t^2} - \int_0^{N_t} x_{it}^2\, di - \frac{\tilde{\mathcal{K}}}{2}\frac{1}{N_t}\int_0^{N_t} \tilde{x}_{it}^2\, di$$

where $\mathcal{K}$ and $\tilde{\mathcal{K}}$ are affine Krylov subspaces for the weighting of privacy versus consumption, representing the concern, as a firm, for being outcompeted by other firms. With N_t varieties, privacy costs are summed across all instances of x and $\tilde{x}$. It is assumed that the utility cost of privacy depends on the average privacy cost across all varieties.

The order-r Krylov subspace generated by an n-by-n matrix A and a vector b of dimension n is the subspace spanned by the images of b under the first r powers of A (starting from $A^0 = I$), that is,

$K_r(A, b) = \text{span}\{b, Ab, A^2b, \dots, A^{r-1}b\}$.

There is a scale of privacy costs of x_{it}, represented by $\frac{1}{N_t}$.

Figure 8: The relationship between Privacy Costs and Creative Destruction

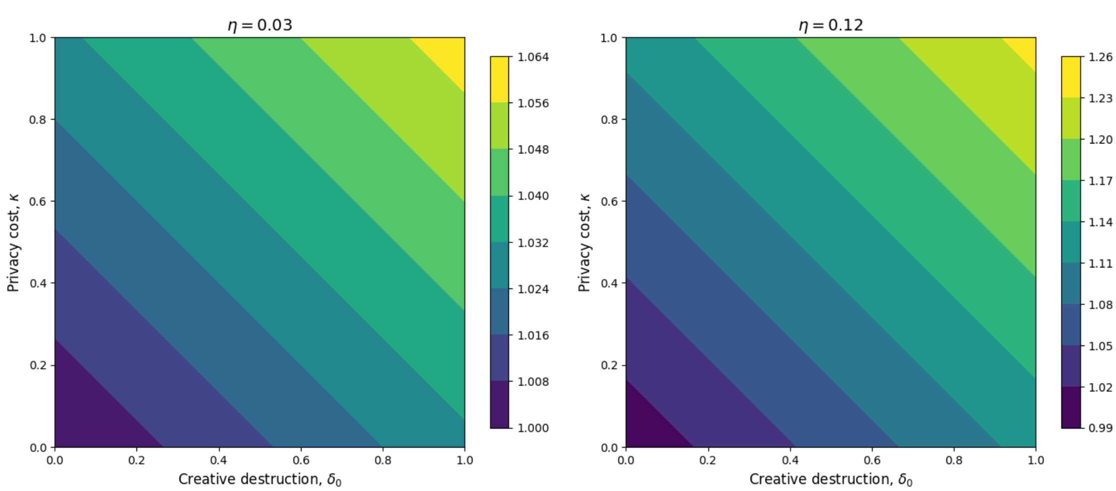


Figure 8 presents a specific depiction of the relationship between privacy costs and creative destruction, as suggested by Jones and Tonetti, and with a simplified manifestation recreated by the author. The so-called *"ratio of consumption-equivalent welfare"*, λ^c/λ^f, for $\mathcal{K} = \tilde{\mathcal{K}}$, defaults to the *"Consumers Own Data"* as a superior scenario as opposed to the *"Firms Own Data"* condition.

With η accounting for the importance of data, the figure illustrates the trade-off between privacy costs K and creative destruction δ_0.

For the firm concerned with creative destruction, $\tilde{x_{it}}$ reflects the costs associated with data usage by all other N_t firms in the economy. There is a difference factor between these costs, with interior solutions along equilibrium growth.

The consumption per person is:

$$c_t = \left(\int_0^{N_t} c_{it}^{\frac{\sigma-1}{\sigma}} \, di \right)^{\frac{\sigma}{\sigma-1}}$$

with $\sigma > 1$.

The firm maximizes expected flow utility with the "Data creation" component:

$$J_{it} = c_{it} L_{it}$$

It is assumed that the firm cannot commit to selling its data to only one company; otherwise, motivated by concerns over creative destruction, firms could make deals with consumers and revert the allocation to when firms owned the data.

There is a "variety resource constraint" for firm i: $c_{it} = Y_{it}/L_{it}$, so that data is produced with the same labor intensity across all varieties, with $D_{it} \leq \alpha x_{it} J_{it} + (1-\alpha) B_t$.

Optimal data sharing is when $D_{sit} = \tilde{x}_{it} J_{it}$.

7 Triangular Norms, Creative Destruction, and the Quantity x_i in the Economy

A triangular norm (or T-norm) is a type of binary operation used in probabilistic metric spaces. It generalizes the concept of intersection in a lattice and conjunction in logic. T-norms are used to generalize the triangle inequality of ordinary metric spaces.

In the context of economic dynamics, particularly within the framework of creative destruction, T-norms can be employed to model the interactions and dependencies between different economic variables. This approach can help us understand how innovation and creative destruction influence the quantity of goods or services x_i produced in the economy.

7.1 Applied to the Economic Environment

Recall the economic environment described by the following key equations:

7.1.1 Utility (A)

$$\int_0^\infty e^{-\rho t} L_t(c_t, x_{it}, \tilde{x}_{it}) dt$$

7.1.2 Flow Utility (B)

$$u(c_t, x_{it}, \tilde{x}_{it}) = \log c_t - \frac{\mathrm{K}}{2} \frac{1}{N_t^2} \int_0^{N_t} x_{it}^2 di - \frac{\tilde{\mathrm{K}}}{2} \frac{1}{N_t^1} \int_0^{N_t} \tilde{x}_{it}^2 di$$

7.1.3 Innovation (New Varieties) (J)

$$\dot{N}_t = \frac{1}{\mathcal{X}} \times L_{et}$$

7.1.4 Creative Destruction (N)

$$\delta(\tilde{x}_{it}) = \frac{\delta_0}{2} \tilde{x}_{it}^2$$

7.1.5 Definition of a T-norm

A T-norm T is a binary operation $T : [0,1] \times [0,1] \to [0,1]$ that satisfies the following properties:

1. **Commutativity**: $T(a,b) = T(b,a)$
2. **Associativity**: $T(a, T(b,c)) = T(T(a,b), c)$
3. **Monotonicity**: If $a \le c$ and $b \le d$, then $T(a,b) \le T(c,d)$
4. **Boundary Condition**: $T(a,1) = a$

7.1.6 Application to Creative Destruction

In our economic model, we can use a T-norm to describe the interaction between the creative destruction process and the quantity of goods or services x_i in the economy. Specifically, we can define a T-norm T that combines the effects of innovation and creative destruction on x_i.

Let x_i represent the quantity of variety i in the economy, influenced by both innovation (which increases x_i) and creative destruction (which decreases x_i). We can express this relationship as:

$$x_i(t+1) = T(\text{Innovation}(x_i(t)), \text{Destruction}(x_i(t)))$$

7.1.7 Example of a T-norm in this Economic Context

Consider the following specific form of a T-norm, often used in fuzzy logic:

$$T(a,b) = \min(a,b)$$

This T-norm can be interpreted as the minimum effect between innovation and destruction on x_i. In this context, if either innovation or destruction is particularly strong, it will dominate the effect on x_i.

7.1.8 Incorporating T-norm into the Economic Model

We can now express the dynamic equation for x_i using the T-norm:

$$x_i(t+1) = T\left(f(\text{Innovation}(x_i(t))), g(\text{Destruction}(x_i(t)))\right)$$

where f and g are functions representing the effects of innovation and destruction, respectively.

For instance, let:

$$f(\text{Innovation}(x_i(t))) = x_i(t) + \beta \dot{N}_t$$

$$g(\text{Destruction}(x_i(t))) = x_i(t) - \delta(\tilde{x}_{it})$$

Then, the update equation becomes:

$$x_i(t+1) = \min\left(x_i(t) + \beta \dot{N}_t, x_i(t) - \frac{\delta_0}{2}\tilde{x}_{it}^2\right)$$

This equation encapsulates the interplay between creative destruction and innovation in determining the quantity x_i.

8 Mathematical Formulations for Testing Model Assumptions

Formally defining the system S can be done as follows:

$$S = \mathcal{X}, \mathcal{R}$$

where $\mathcal{X}$ is equal to the set of elements $x_1, \dots, x_n$ and $\mathcal{R}$ is the set encompassing the behavioral pattern of the system. R includes rules of operations on the set $\mathcal{X}$, as well as interrelationships among x's within $\mathcal{X}$, relationships between x and the external environment, constraints and control phenomena. Relations, *e.g.*, in a steel or digital plant, are required for producing every unit of data. An equipment can be used in a producing unit for only a certain amount of time per day.

$$R = (r_{ij})$$

$$r_{ij} = 1$$

if there is a communication from

x_i to x_j;

$$r_{ij} = 0$$

if not.

The realization of the system requires knowledge of the input-output or stimulus-response mechanism. Given the sets $\mathcal{X}$ and $\mathcal{R}$, one can determine the output vector $Y = (Y_1, \dots, Y_m)$ from the mapping relation:

$$f : \mathcal{X}, \mathcal{R} \rightarrow Y$$

That is, Y is determinable from the knowledge of $\mathcal{X}, \mathcal{R}$ if the map f is known.

For the output Y_{in}, indexed by i and effective at time t, the following relation holds:

$$S = \mathcal{X}, \mathcal{R} \rightarrow Y_{in} = D_{in}^{\eta} L_{in}$$

where the focus on D_{in} can be shifted to the concern of D_{it}, that being the data used by firm i at time t, L_{it} is the labor input, and parametized by η.

The above propoerties hold true both for deterministic and stochastic systems. A deterministic system is one in which a given input X, under the regulation set R and the map f, yields a unique output Y with probability one. In contrast, a stochastic system is one in which the output Y is not uniquely determined by the input X and the regulation set R. The output Y_j is a random vector, and the system is characterized by a probability π_j (discrete):

$$pr(Y_j = y_j | \mathcal{X}, \mathcal{R}, f) = \pi_j; (j = 1, 2, \dots)$$

In the case of Y following a continuous distribution, the probability density function is denoted by

$$pr(Y \leq y|\mathcal{X}, \mathcal{R}, f) = F(y)$$

With

$$F(y) = (F_1(y_1), F_2(y_2), \dots, F_m(y_m)), y \in (0, \infty) \wedge 0 = (0, \dots, 0), \infty = (\infty, \dots, \infty)$$

Thus, the system is defined by a triple $(\mathcal{X}, \mathcal{R}, f)$, and the output Y is determined by the input X and the regulation set R. The problems met with in cybernetic systems are one or more of the following, as per *Ghosal*:

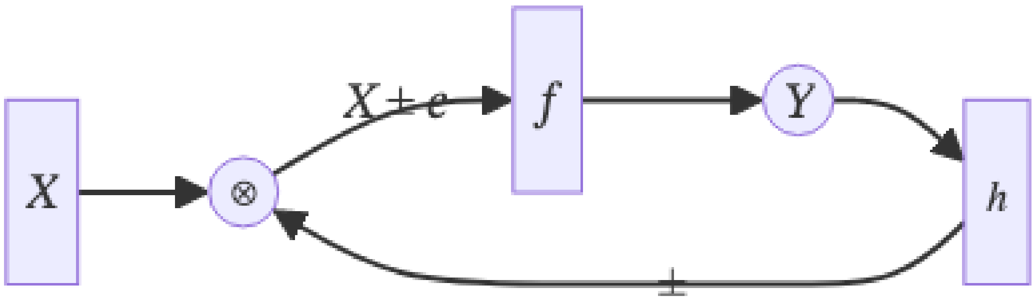
- *(`i`) to identify the system with respect to R and f, from the observations of the input X and the output Y;*
- *(`ii`) to determine the output Y, when the input X, the characteristic set $\mathcal{R}$ and the map fare f given;*
- *(`iii`) to estimate the degree of change in Y for given change in X, R and vice-versa;*
- *(`iv`) to estimate how the system should be changed with respect to one or more of X, R or f so that a desired output Y_0 is obtained;*
- *(`v`) to estimate the feedback mechanism between Y and X, whenever it is present.*

An appraisal of the problems given above with respect to a real life situation helps us in decision-making. It may be remarked here that problems of identification are used in various problems in statistical theory and econometrics are akin to the problems (`i-v`) referred to above. For example, the problem of identification of parameters in Fisher's theory of fiduciary estimation (see Kendall and Stuart, 1967) can be treated as a problem in (`i`).

Tensor products sometimes appear in advanced econometric models when dealing with multi-dimensional arrays of data or complex interactions between multiple sets of variables. The phenomenon of feedback can be explained in terms of elementary CT. Consider the same system X, R with the mapping relation $f : \mathcal{X}, \mathcal{R} \rightarrow Y$.

Let the input vector X_t at time t lead to an output Y_t. The structure of the system is such that the output Y_t is fed back to the system as an input X_{t+1} at time $t + 1$. A closed loop system is thus formed. The feedback mechanism is represented by the relation:

Figure 9: Closed-loop System



The diagram represents a feedback loop where X is an input signal. The summing junction combines X with an error signal ϵ. The result is passed through function f to produce output Y, which is then fed back through function h to provide a feedback signal to the summing junction.

The relationships governing the system are:

$$Y_t = f(X_t)$$

$$X_{t+1} = X_t + e_t$$

and

$$e_t = h(Y_t)$$

With no loss of generality, these relationships may be regarded as forming tbe characteristic set R of the system.

9 Allocating Bundles of Data in the Economic Environment

In the Economic Environment, the allocation of data bundles significantly influences both firm productivity and overall economic growth.

9.1 For hypothesis `a-H`

The firm production function

$$Y_{it} = D_{it}^{\eta} L_{it} G$$

demonstrates that data D_{it} is a critical input in the production process. Firms that effectively collect and use consumer data

$$D_{it} \leq \alpha x_{it} J_{it} + (1 - \alpha) B_t$$

are likely to achieve higher productivity levels.

9.1.1 Proof

Step 1: AR Model An auto-regressive (AR) model describes how the current value of a variable depends on its past values. For data sharing in an economic context, we model the shared data D_{sit} as:

$$D_{sit} = \phi_1 D_{sit-1} + \phi_2 D_{sit-2} + \cdots + \phi_p D_{sit-p} + \epsilon_t$$

where $\phi_1, \phi_2, \ldots, \phi_p$ are parameters and ϵ_t is a white noise error term.

Step 2: Addressing Heteroskedasticity We model the error term ϵ_t as having a time-varying variance σ_t^2:

$$\epsilon_t = \sigma_t \eta_t$$

where η_t is a standard normal error term with mean 0 and variance 1, i.e., $\eta_t \sim N(0, 1)$. The variance σ_t^2 can be modeled using a Generalized Auto-regressive Conditional Heteroskedasticity (GARCH) model:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

The characteristics of the model include:

- **Persistence:** The sum $\alpha_1 + \beta_1$ close to 1 indicates high persistence of volatility.
- **Volatility Clustering:** Periods of high volatility are followed by high volatility and periods of low volatility are followed by low volatility.
- **Leverage Effect:** Asymmetric models can capture the leverage effect where negative shocks increase future volatility more than positive shocks of the same magnitude.

Step 3: Addressing Kurtosis High kurtosis in the error terms indicates the presence of outliers. We address kurtosis by considering a distribution for ϵ_t that allows for heavier tails, such as the Student's t-distribution. The degrees of freedom ν of the t-distribution can be chosen based on the empirical kurtosis of the error terms.

$$\epsilon_t \sim t(\nu)$$

Characteristics of the Student's t-distribution

- **Degrees of Freedom ν:** Lower values of ν increase the heaviness of the tails.
- **Probability Density Function:** The PDF of the Student's t-distribution is given by:

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

where $\Gamma(\cdot)$ is the gamma function, representing a generalization of the factorial function to real and complex numbers.

For $\nu > 2$, the kurtosis is $3 + \frac{6}{\nu-4}$.

Step 4: Innovation Incorporating heteroskedasticity and kurtosis into the innovation equation, we have:

$$\dot{N}_t = \frac{1}{\mathcal{X}} \times f(u(c_t, x_{it}, \tilde{x}_{it})) \times \epsilon_t$$

with ϵ_t exhibiting GARCH-type heteroskedasticity.

The flow utility function

$$u(c_t, x_{it}, \tilde{x}_{it}) = \log c_t - \frac{\mathrm{K}}{2}\frac{1}{N_t^2}\int_0^{N_t} x_{it}^2 di - \frac{\tilde{\mathrm{K}}}{2}\frac{1}{N_t^1}\int_0^{N_t} \tilde{x}_{it}^2 di$$

suggests that consumer well-being is influenced by their consumption and the degree to which they share data. Sharing data can lead to better-targeted products and services, potentially increasing consumption utility $\log c_t$.

However, the cost of privacy loss is also captured, indicating a trade-off that needs to be managed.

The innovation function

$$\dot{N}_t = \frac{1}{\mathcal{X}} \times L_{et}$$

and the aggregate output function

$$Y_t \equiv c_t L_t$$

highlight the role of data in driving economic growth. The data bundle

$$B_t = (N_t^{\frac{1}{\epsilon}} \int_0^{N_t} D_{sit}^{\frac{\epsilon-1}{\epsilon}} di)^{\frac{\epsilon}{\epsilon-1}}$$

provides a mechanism for sharing data across firms, facilitating innovation and productivity improvements. The parameter $\epsilon \geq 1$ controls the elasticity of data substitution, indicating the impact of data availability on economic outcomes.

Let us consider $i = 1$. Let $D_{i0} \in \mathbb{R}_+$ be the initial data for firm i, and suppose that there exist $a, b \in \mathbb{R}_+, a < b$, representing two different levels of data utilization. Let $\varepsilon = b^{1/2} - a^{1/2} > 0$.

From the firm's production function, we have

$$Y_{it} = D_{it}^{\eta} L_{it},$$

where $\eta \in (0, 1)$, D_{it} represents data used by firm i, and L_{it} is the labor input.

For firms utilizing data efficiently, we have:

$$D_{it} \leq \alpha x_{it} J_{it} + (1 - \alpha) B_t,$$

with

$$J_{it} = c_{it} L_t.$$

By substituting D_{it} into the production function:

$$Y_{it} = (\alpha x_{it} J_{it} + (1 - \alpha) B_t)^{\eta} L_{it} G.$$

Now, consider the case where $a^{1/2}$ and $b^{1/2}$ are best-responses to the data allocation decisions. The first-order conditions for optimal data usage are:

$$a^{1/2} = E\left[v_1' \left(\left[a^{1/2} + \sum_{i=2}^{I} (D_i)^{1/2} \right]^2 \right) \cdot \left[a^{1/2} + \sum_{i=2}^{I} (D_i)^{1/2} \right] \middle| D_i \right],$$

$$= a^{1/2} E\left[v_1' \left(\left[a^{1/2} + \sum_{i=2}^{I} (D_i)^{1/2} \right]^2 \right) \middle| D_i \right],$$

$$+ E\left[v_1' \left(\left[a^{1/2} + \sum_{i=2}^{I} (D_i)^{1/2} \right]^2 \right) \cdot \left[\sum_{i=2}^{I} (D_i)^{1/2} \right] \middle| D_i \right].$$

The third term above is nonnegative, thus the equality holds only if

$$E\left[v_1' \left(\left[a^{1/2} + \sum_{i=2}^{I} (D_i)^{1/2} \right]^2 \right) \middle| D_i \right] \leq 1.$$

Since $b > a$ and v_1' is strictly decreasing, we have

$$E\left[v_1'\left(\left[b^{1/2} + \sum_{i=2}^{I}(D_i)^{1/2}\right]^2\right) \cdot \left[\sum_{i=2}^{I}(D_i)^{1/2}\right] \middle| D_i\right],$$

$$< E\left[v_1'\left(\left[a^{1/2} + \sum_{i=2}^{I}(D_i)^{1/2}\right]^2\right) \cdot \left[\sum_{i=2}^{I}(D_i)^{1/2}\right] \middle| D_i\right],$$

and

$$b^{1/2}E\left[v_1'\left(\left[b^{1/2} + \sum_{i=2}^{I}(D_i)^{1/2}\right]^2\right) \middle| D_i\right],$$

$$< a^{1/2}E\left[v_1'\left(\left[a^{1/2} + \sum_{i=2}^{I}(D_i)^{1/2}\right]^2\right) \middle| D_i\right] + \varepsilon.$$

Adding the inequalities and noting that a and b satisfy the first-order conditions, it follows that

$$a^{1/2} + \varepsilon > b^{1/2}.$$

But this contradicts the definition of ε. Hence, the firm's best-response data usage is unique and leads to higher productivity when data is efficiently utilized. This completes the proof. □

9.2 For hypothesis `b-H`

9.2.1 Proof

Given the flow utility function:

$$u(c_t, x_{it}, \tilde{x}_{it}) = \log c_t - \frac{\mathrm{K}}{2}\frac{1}{N_t^2}\int_0^{N_t} x_{it}^2\, di - \frac{\tilde{\mathrm{K}}}{2}\frac{1}{N_t^1}\int_0^{N_t} \tilde{x}_{it}^2\, di$$

When consumers share data, they enable firms to produce more efficiently, increasing the consumption utility $logc_t$. The negative utility from data sharing is captured by the last two terms, representing privacy loss.

To show an increase in overall well-being:

$$\Delta u = \Delta(\log c_t) - \frac{\mathrm{K}}{2}\frac{1}{N_t^2}\int_0^{N_t} x_{it}^2\, di - \frac{\tilde{\mathrm{K}}}{2}\frac{1}{N_t^1}\int_0^{N_t} \tilde{x}_{it}^2\, di$$

For a sufficiently large increase in $\log c_t$ due to efficient production, $\Delta u > 0$, implying an increase in overall well-being for consumers sharing their data.

Given the innovation function:

$$\dot{N}_t = \frac{1}{\mathcal{X}} \times L_{et}$$

And the aggregate output function:

$$Y_t \equiv c_t L_t$$

Substituting c_t:

$$c_t = \left(\int_0^{N_t} c_{it}^{\frac{\sigma-1}{\sigma}}\, di\right)^{\frac{\sigma}{\sigma-1}}.$$

Considering the firm's data use:

$$D_{it} = \alpha x_{it} J_{it} + (1-\alpha) B_t$$

Higher data availability increases D_{it}, enhancing Y_i, and thus Y_t:

$$\frac{\partial Y_t}{\partial D_{it}} > 0.$$

Now, let's expand this proof in the Wiener process with the Neyman-Pearson Lemma, commonly used in hypothesis testing.

Consider the data D_{it} following a GWBM influenced by a Wiener process $W(t)$:

$$D_{it} = D_{i0} \exp\left(\left(\mu - \frac{\sigma^2}{2}\right) t + \sigma W(t)\right)$$

For discrete time, the process can be written as:

$$W(t) = W(t-1) + \mu + \sigma\epsilon_t$$

with ϵ_t i.i.d. $\sim N(0,1), \sigma > 0, \mu \in \mathbb{R}$.

Step 1: Understanding Consumer Utility with Data Sharing Given the flow utility function, with D_{it} as the data input from consumers, L_{it} is the labor input, and G is a productivity parameter. The data D_{it} includes contributions from both x_{it} (direct data sharing by consumers) and $\tilde{x}_{it}$ (data shared with other firms).

By substituting D_{it} in the firm's production function, we get:

$$Y_{it} = (\alpha x_{it} J_{it} + (1-\alpha)B_t)^\eta L_{it} G.$$

Step 2: Data Contribution The direct benefit to consumers from sharing data is reflected in their consumption c_t:

$$c_t = \left(\int_0^{N_t} c_{it}^{\frac{\sigma-1}{\sigma}} \, di \right)^{\frac{\sigma}{\sigma-1}},$$

where $\sigma > 1$ ensures that consumers benefit from a variety of products and services tailored to their preferences.

Sharing data increases firm productivity, which can lead to increased consumption c_t. Higher c_t enhances consumer utility:

$$u(c_t, x_{it}, \tilde{x}_{it}) = \log c_t - \frac{\text{K}}{2} \frac{1}{N_t^2} \int_0^{N_t} x_{it}^2 \, di - \frac{\tilde{\text{K}}}{2} \frac{1}{N_t^1} \int_0^{N_t} \tilde{x}_{it}^2 \, di.$$

Step 3: Impact of Data Sharing on Consumer Well-Being To show an increase in overall well-being, consider the change in utility:

$$\Delta u = \Delta(\log c_t) - \frac{\text{K}}{2} \frac{1}{N_t^2} \int_0^{N_t} x_{it}^2 \, di - \frac{\tilde{\text{K}}}{2} \frac{1}{N_t^1} \int_0^{N_t} \tilde{x}_{it}^2 di.$$

Given that sharing data with firms enhances their productivity, which in turn increases c_t, the term $\Delta(\log c_t)$ is positive. The negative utility from data sharing is captured by the last two terms, representing privacy loss.

Step 4: Equilibrium and Optimal Data Sharing Consumers will share data if the marginal utility gain from increased consumption outweighs the marginal utility loss from reduced privacy:

$$\text{If} \quad \Delta(\log c_t) > \frac{\text{K}}{2} \frac{1}{N_t^2} \int_0^{N_t} x_{it}^2, di + \frac{\tilde{\text{K}}}{2} \frac{1}{N_t^1} \int_0^{N_t} \tilde{x}_{it}^2 \, di,$$

$$\text{then} \quad \Delta u > 0.$$

Step 5: Statistical Confirmation using Proposition From the assumption of complete information, consumers know the benefit of sharing data.

If all consumers share data efficiently, we get an equilibrium where the total utility is maximized. Each consumer i chooses x_{it} and $\tilde{x}_{it}$ such that:

$$v_i'(c_i) \leq 1,$$

with equality holding if $x_{it}, \tilde{x}_{it} > 0$.

Given any hypothesis test with rejection set $\mathcal{R}$, define its statistical power function $\beta_{\mathcal{R}}(\theta) = \Pr(X \in \mathcal{R})$.

Step 6: Existence Given some hypothesis test that satisfies P_α condition, call its rejection region R_{NP} (where NP stands for Neyman-Pearson).

For any level α hypothesis test with rejection region $\mathcal{R}$ we have

$$[1_{R_{\text{NP}}}(x) - 1_{\mathcal{R}}(x)][\rho(x \mid \theta_1) - \eta\rho(x \mid \theta_0)] \geq 0$$

except on some ignorable set A.

Then integrate it over x to obtain

$$0 \leq [\beta_{R_{\text{NP}}}(\theta_1) - \beta_{\mathcal{R}}(\theta_1)] - \eta[\beta_{R_{\text{NP}}}(\theta_0) - \beta_{\mathcal{R}}(\theta_0)].$$

Since $\beta_{R_{\text{NP}}}(\theta_0) = \alpha$ and $\beta_{\mathcal{R}}(\theta_0) \leq \alpha$, we find that $\beta_{R_{\text{NP}}}(\theta_1) \geq \beta_{\mathcal{R}}(\theta_1)$.

Thus the R_{NP} rejection test is a UMP test in the set of level α tests.

Step 7: Uniqueness For any other UMP level α test, with rejection region $\mathcal{R}$, we have from the Existence part,

$$[\beta_{R_{\text{NP}}}(\theta_1) - \beta_{\mathcal{R}}(\theta_1)] \geq \eta[\beta_{R_{\text{NP}}}(\theta_0) - \beta_{\mathcal{R}}(\theta_0)].$$

Since the $\mathcal{R}$ test is UMP, the left side must be zero. Since $\eta > 0$ the right side gives $\beta_{R_{\text{NP}}}(\theta_0) = \beta_{\mathcal{R}}(\theta_0) = \alpha$, so the $\mathcal{R}$ test has size α.

Since the integrand

$$[1_{R_{\text{NP}}}(x) - 1_{\mathcal{R}}(x)][\rho(x \mid \theta_1) - \eta\rho(x \mid \theta_0)]$$

is nonnegative, and integrates to zero, it must be exactly zero except on some ignorable set A.

Since the R_{NP} test satisfies P_α condition, let the ignorable set in the definition of P_α condition be A_{NP}.

$$\mathcal{R} \setminus (R_{\text{NP}} \cup A_{\text{NP}})$$

is ignorable, since for all

$$x \in \mathcal{R} \setminus (R_{\text{NP}} \cup A_{\text{NP}}),$$

we have

$$[1_{R_{\text{NP}}}(x) - 1_{\mathcal{R}}(x)][\rho(x \mid \theta_1) - \eta\rho(x \mid \theta_0)] = \eta\rho(x \mid \theta_0) - \rho(x \mid \theta_1) > 0.$$

Similarly,

$$R_{\text{NP}} \setminus (\mathcal{R} \cup A_{\text{NP}})$$

is ignorable.

Define

$$A_{\mathcal{R}} := (\mathcal{R} \cap R_{\text{NP}}) \cup A_{\text{NP}}$$

$A_{\mathcal{R}}$ is the union of three ignorable sets, thus it is an ignorable set.

Then we have

$$x \in \mathcal{R} \setminus A_{\mathcal{R}} \implies \rho(x \mid \theta_1) > \eta\rho(x \mid \theta_0)$$

and

$$x \in \mathcal{R}^c \setminus A_{\mathcal{R}} \implies \rho(x \mid \theta_1) < \eta\rho(x \mid \theta_0).$$

So the $\mathcal{R}$ rejection test satisfies P_α condition with the same η.

Since $A_{\mathcal{R}}$ is ignorable, its subset $\mathcal{R}\Delta R_{\text{NP}} \subseteq A_{\mathcal{R}}$ is also ignorable. Consequently, the two tests agree with probability 1 whether $\theta = \theta_0$ or $\theta = \theta_1$.

Step 8: Statistical Confirmation Using Neyman-Pearson Lemma We test if the increase in data availability significantly affects economic growth.

Let $X_1, \ldots, X_n$ be a random sample from the $\mathcal{N}(\mu, \sigma^2)$ distribution where the mean μ is known, and suppose we wish to test $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 = \sigma_1^2$. The likelihood for this set of normally distributed data is:

$$\mathcal{L}(\sigma^2 \mid \mathbf{x}) \propto (\sigma^2)^{-n/2} \exp\left\{ -\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2} \right\}.$$

The likelihood ratio is:

$$\Lambda(\mathbf{x}) = \frac{\mathcal{L}(\sigma_0^2 \mid \mathbf{x})}{\mathcal{L}(\sigma_1^2 \mid \mathbf{x})} = \left(\frac{\sigma_0^2}{\sigma_1^2}\right)^{-n/2} \exp\left\{-\frac{1}{2}(\sigma_0^{-2} - \sigma_1^{-2}) \sum_{i=1}^{n} (x_i - \mu)^2\right\}.$$

This ratio depends on the data through $\sum_{i=1}^{n}(x_i - \mu)^2$. By the Neyman-Pearson Lemma, the most powerful test depends only on $\sum_{i=1}^{n}(x_i - \mu)^2$.

If $\sigma_1^2 > \sigma_0^2$, then

$$\Lambda(\mathbf{x})$$

is a decreasing function of $\sum_{i=1}^{n}(x_i - \mu)^2$. We should reject H_0 if $\sum_{i=1}^{n}(x_i - \mu)^2$ is sufficiently large.

By integrating data as a stochastic process in the economic model, we show that increased data availability, characterized by its volatility and impact on firm production, significantly drives short-term economic growth. The Neyman-Pearson Lemma confirms that our test for economic growth influenced by availability is statistically robust within the interval of data volatility.

When consumers share their data, they contribute to the firm's data pool, which in turn enhances the firm's productivity.

Hence, given the positive impact on consumption and controlled privacy loss, sharing data leads to an overall increase in well-being. This completes the proof. □

9.3 For hypothesis `c-H`

9.3.1 Proof

It follows from the assumption of complete information that every firm has a data set with a single element, and that the set of data vectors only has a single element. Thus, to simplify the notation, we use D_i to denote $D_i(\cdot;\theta_i)$, and Y^e to denote $Y^e(\theta)$.

We know that the hypotheses adopted here guarantee that there is a unique efficient provision $Y^e \geq 0$. There are two cases: $Y^e = 0$ or $Y^e > 0$. First, suppose we have $Y^e = 0$. We are going to show that 0 is an equilibrium for firm production. Consider the problem faced by some firm $i \in \mathcal{I}$ when all other firms are contributing zero to the data mechanism:

This can be written as:

$$\max_{D_i \geq 0} v_i \left(\left[D_i^{1/2} + 0 \right]^2 \right) - D_i.$$

Thus,

$$\max_{D_i \geq 0} v_i(D_i) - D_i.$$

The first-order condition for the firm i is that $v_i'(D_i) \leq 1$, with equality holding when $D_i > 0$. But note that, since $Y^e = 0$, $\sum_{j=1}^{I} v_j'(0) \leq 1$. In particular, since v_j is increasing for all $j \in \mathcal{I}$, this implies that $v_i'(0) \leq 1$. Thus, $D_i = 0$ satisfies the first-order condition for i. But since the choice of i was arbitrary, we have that 0 is an equilibrium for data usage.

Now, suppose $Y^e > 0$. For all $i \in \mathcal{I}$, let $D_i = (v_i'(Y^e) \cdot (Y^e)^{1/2})^2$. One is able to check $Y^e = \Phi(D) = \left[\sum_{i=1}^{I} D_i^{1/2} \right]^2$. Rearranging, we obtain

$$v_i'(Y^e) = \frac{(D_i)^{1/2}}{(Y^e)^{1/2}}.$$

$$\implies \max_{D_i \geq 0} v_i \left(\left[D_i^{1/2} + 0 \right]^2 \right) - D_i.$$

Thus,

$$\max_{D_i \geq 0} v_i(D_i) - D_i.$$

The first-order condition for the firm i is that $v_i'(D_i) \leq 1$, with equality holding when $D_i > 0$. But note that, since $Y^e = 0$, we have that $\sum_{j=1}^{I} v_j'(0) \leq 1$. In particular, since v_j is increasing for all $j \in \mathcal{I}$, this implies that $v_i'(0) \leq 1$. Thus, $D_i = 0$ satisfies the first-order condition for i. But since the choice of i was arbitrary, we have that 0 is an equilibrium for data usage. Now, suppose $Y^e > 0$. For all $i \in \mathcal{I}$, let

$$D_i = (v_i'(Y^e) \cdot (Y^e)^{1/2})^2.$$

It is easy to check that $Y^e = \Phi(D) = \left[\sum_{i=1}^{I} D_i^{1/2}\right]^2$.

Rearranging,

$$v_i'(Y^e) = \frac{(D_i)^{1/2}}{(Y^e)^{1/2}}.$$

which is precisely the first-order condition for the firm i's optimization problem. We can thus conclude the vector $\mathbf{D} = (D_1, \cdots, D_I)$ as defined is an equilibrium of data usage and its provision is efficient.

Let $u : \mathbb{R}_+ \times \Theta \to \mathbb{R}$ be defined by $u(c_t, x_{it}, \tilde{x}_{it}; \theta) = \sum_{i=1}^{I} u_i(c_t, x_{it}, \tilde{x}_{it}; \theta_i)$, and fix $\theta \in \Theta$. By Assumption 1, we have that $u(\cdot; \theta) \in C^1$ is strictly concave, and that $\lim_{c_t \to \infty} u'(c_t, x_{it}, \tilde{x}_{it}; \theta) = 0$.

We then have two cases: $u'(0; \theta) \leq 1$ or $u'(0; \theta) > 1$. In the first case, it follows from the first order condition that $c_t = 0$ is efficient. Furthermore, since $u(\cdot; \theta)$ is a strictly concave function, we have that $u'(c_t; \theta) < 1$ for all $c_t > 0$ and thus, there can be no efficient provision $c_t > 0$. $c_t^e(\theta) = 0$ is the unique efficient provision.

Now suppose that $u'(0; \theta) > 1$. It follows that $c_t = 0$ is not efficient. Since $\lim_{c_t \to \infty} u'(c_t; \theta) = 0$, there exists $A \in \mathbb{R}$ such that $u'(A; \theta) < 1$. Thus, since $u'(\cdot; \theta)$ is continuous, $u'(0; \theta) > 1$ and $u'(A; \theta) < 1$, it follows from the intermediate value theorem that there exists $0 < B < A$ such that $u'(B; \theta) = 1$. Additionally, since $u'(\cdot; \theta)$ is strictly decreasing, we have that $u'(c_t; \theta) \neq 1$ for all $c_t \neq B$. Therefore, $c_t^e(\theta) = B$ is the unique efficient provision.

For any $\theta \in \Theta$, we then have that in both cases there is a unique efficient provision $c_t^e(\theta) \geq 0$. Thus, we construct a unique function $c_t^e : \Theta \to \mathbb{R}_+$ that maps each profile of types to its efficient consumption.

Now, let us one more time analyze the relationship between Flow Utility and Innovation:

Given the Flow Utility equation:

$$u(c_t, x_{it}, \tilde{x}_{it}) = \log c_t - \frac{\mathrm{K}}{2} \frac{1}{N_t^2} \int_0^{N_t} x_{it}^2 di - \frac{\tilde{\mathrm{K}}}{2} \frac{1}{N_t^1} \int_0^{N_t} \tilde{x}_{it}^2 di$$

And the Innovation equation:

$$\dot{N}_t = \frac{1}{\mathcal{X}} \times L_{et}$$

We aim to show the relationship between $u(c_t, x_{it}, \tilde{x}_{it})$ and $\dot{N}_t$.

By the nature of Flow-Utility, higher levels of consumption c_t and efficient usage of data x_{it} and $\tilde{x}_{it}$ increase the utility. However, data usage also introduces costs, as represented by the terms $\frac{\kappa}{2} \frac{1}{N_t^2} \int_0^{N_t} x_{it}^2 di$ and $\frac{\tilde{\kappa}}{2} \frac{1}{N_t^1} \int_0^{N_t} \tilde{x}_{it}^2 di$.

Innovations, represented by $\dot{N}_t$, are directly influenced by labor allocated to research L_{et}. An efficient allocation of labor to research activities enhances the rate of innovation.

Thus, we propose:

$$L_{et} = f(u(c_t, x_{it}, \tilde{x}_{it}))$$

where f is a function representing how flow utility drives labor allocation towards innovation.

Substituting this into the innovation equation, we get:

$$\dot{N}_t = \frac{1}{\mathcal{X}} \times f(u(c_t, x_{it}, \tilde{x}_{it}))$$

Expanding this relationship, we recognize that $f(u(c_t, x_{it}, \tilde{x}_{it}))$ captures the influence of both consumption and data utilization on the propensity to innovate. Specifically, as $u(c_t, x_{it}, \tilde{x}_{it})$ increases, reflecting higher overall utility, firms are more incentivized to allocate labor towards innovation, thus driving L_{et} higher.

Let us consider the functional form of f. Suppose f is linear for simplicity, then we have:

$$L_{et} = \beta_0 + \beta_1 u(c_t, x_{it}, \tilde{x}_{it})$$

where β_0 is a baseline level of labor allocated to innovation, and β_1 represents the sensitivity of labor allocation to changes in utility. Substituting this into the innovation equation yields:

$$\dot{N}_t = \frac{1}{\mathcal{X}} \times (\beta_0 + \beta_1 u(c_t, x_{it}, \tilde{x}_{it}))$$

Innovation is therefore driven by both a baseline level of research activity and the additional effort spurred by increased flow utility. This formulation highlights the direct linkage between the utility derived from consumption and data usage, and the subsequent impact on innovation rates.

To further analyze this, let us represent this relationship in matrix form. Define the state vector $\mathbf{x}_t = [c_t, x_{it}, \tilde{x}_{it}]^T$ and the utility function as:

$$\mathbf{u}(\mathbf{x}_t) = \log c_t - \frac{\mathrm{K}}{2} \frac{1}{N_t^2} x_{it}^2 - \frac{\tilde{\mathrm{K}}}{2} \frac{1}{N_t^1} \tilde{x}_{it}^2$$

We can write the labor allocation function f as a matrix equation:

$$\mathbf{L}_{et} = \mathbf{B}\mathbf{u}(\mathbf{x}_t)$$

where $\mathbf{B}$ is a matrix of coefficients representing the impact of utility on labor allocation. Substituting this into the innovation equation gives:

$$\dot{\mathbf{N}}_t = \mathbf{A}\mathbf{L}_{et} = \mathbf{A}\mathbf{B}\mathbf{u}(\mathbf{x}_t)$$

Here, $\mathbf{A}$ is a matrix representing the conversion of labor into innovation. Therefore, the overall system can be represented as:

$$\dot{\mathbf{N}}_t = \mathbf{C}\mathbf{u}(\mathbf{x}_t)$$

where $\mathbf{C} = \mathbf{AB}$ is the combined matrix representing the total impact of utility on innovation.

The partial derivatives of $\dot{N}_t$ with respect to the elements of $\mathbf{x}_t$ can be expressed in matrix form. Let $\mathbf{J_u}$ be the Jacobian matrix of $\mathbf{u}$ with respect to $\mathbf{x}_t$:

$$\mathbf{J_u} = \begin{bmatrix} \frac{\partial u}{\partial c_t} & \frac{\partial u}{\partial x_{it}} & \frac{\partial u}{\partial \tilde{x}_{it}} \end{bmatrix}$$

Given the flow utility function:

$$\frac{\partial u(c_t, x_{it}, \tilde{x}_{it})}{\partial c_t} = \frac{1}{c_t}, \quad \frac{\partial u(c_t, x_{it}, \tilde{x}_{it})}{\partial x_{it}} = -\frac{\mathrm{K}}{N_t^2} x_{it}, \quad \frac{\partial u(c_t, x_{it}, \tilde{x}_{it})}{\partial \tilde{x}_{it}} = -\frac{\tilde{\mathrm{K}}}{N_t^1} \tilde{x}_{it}$$

The Jacobian matrix of the flow utility function $\mathbf{u}(\mathbf{x}_t)$ with respect to the state vector $\mathbf{x}_t$ is given by:

$$\mathbf{J_u} = \begin{bmatrix} \frac{1}{c_t} & -\frac{\mathrm{K}}{N_t^2} x_{it} & -\frac{\tilde{\mathrm{K}}}{N_t^1} \tilde{x}_{it} \end{bmatrix}$$

Now, substituting this Jacobian matrix into the innovation equation, we get:

$$\dot{\mathbf{N}}_t = \mathbf{C}\mathbf{J_u}\mathbf{x}_t$$

where $\mathbf{C}$ is the combined matrix representing the total impact of utility on innovation.

To analyze the stability and dynamics of this system, we can compute the eigenvalues of the matrix $\mathbf{C}\mathbf{J_u}$. The eigenvalues will give us insights into the behavior of the system over time, particularly whether the system converges to a steady state, oscillates, or diverges.

The characteristic equation for the eigenvalues λ of the matrix $\mathbf{C}\mathbf{J_u}$ is given by:

$$\det(\mathbf{C}\mathbf{J_u} - \lambda \mathbf{I}) = 0$$

where $\mathbf{I}$ is the identity matrix. Solving this equation for λ will provide the eigenvalues.

If the real parts of all eigenvalues are negative, the system is stable and converges to a steady state. If any eigenvalue has a positive real part, the system is unstable and will diverge. Complex eigenvalues with non-zero imaginary parts indicate oscillatory behavior.

So the relationship between flow utility and innovation can be captured by the matrix equation:

$$\dot{\mathbf{N}}_t = \mathbf{C}\mathbf{J_u}\mathbf{x}_t$$

where:

$$\mathbf{J_u} = \begin{bmatrix} \frac{1}{c_t} & -\frac{\mathrm{K}}{N_t^2} x_{it} & -\frac{\tilde{\mathrm{K}}}{N_t^1} \tilde{x}_{it} \end{bmatrix}$$

The eigenvalues of $\mathbf{C}\mathbf{J_u}$ determine the system's stability and dynamic behavior. By analyzing these eigenvalues, we can understand how changes in flow utility influence the rate of innovation and the overall behavior of the economic system.

Step 1: Firm's Problem Each firm aims to maximize its profit by deciding on the optimal x_{it} and $\tilde{x}_{it}$. The profit function for firm i is given by:

$$\Pi_{it} = p_{it}Y_{it} - w_t L_{it} - p_{DI} D_{it}$$

where p_{it} is the price of output, w_t is the wage rate, and p_{DI} is the price of data.

Step 2: First-Order Conditions The first-order conditions for maximizing profit with respect to x_{it} and $\tilde{x}_{it}$ are given by:

$$\frac{\partial \Pi_{it}}{\partial x_{it}} = p_{it} \eta D_{it}^{\eta-1} L_{it} \frac{\partial D_{it}}{\partial x_{it}} - p_{DI} \frac{\partial D_{it}}{\partial x_{it}} = 0$$

$$\frac{\partial \Pi_{it}}{\partial \tilde{x}_{it}} = -p_{DI} \frac{\partial D_{sit}}{\partial \tilde{x}_{it}} = 0$$

Step 3: Data Sharing Condition Considering the effect of heteroskedasticity and kurtosis in the data distribution, we incorporate these characteristics into the firm's decision-making process. Let the data distribution be characterized by:

- **Heteroskedasticity**: Variance of the data increases with the level of output.
- **Kurtosis**: The data distribution has heavier tails than a normal distribution.

Step 4: Heteroskedasticity Assume the variance of data used by firm i is given by:

$$\sigma_{D_{it}}^2 = \sigma_0^2 + \sigma_1^2 Y_{it}$$

The firm's optimization problem now accounts for the increased uncertainty (risk) associated with higher output levels.

Step 5: Kurtosis Assume the kurtosis of the data distribution is given by the privacy loss parameter $\kappa_{D_{it}}$:

$$\kappa_{D_{it}} = \kappa_0 + \kappa_1 Y_{it}$$

Higher kurtosis implies a higher probability of extreme outcomes, which the firm needs to consider in its risk management strategy.

Step 6: Optimal Data Utilization and Sharing Incorporating heteroskedasticity and kurtosis into the firm's first-order conditions, we obtain:

$$p_{it} \eta D_{it}^{\eta-1} L_{it} \left(\alpha Y_{it} \right) - p_{DI} \alpha Y_{it} - \lambda \sigma_{D_{it}}^2 - \gamma \kappa_{D_{it}} = 0$$

$$-p_{DI} \tilde{x}_{it} J_{it} - \lambda \sigma_{D_{sit}}^2 - \gamma \kappa_{D_{sit}} = 0$$

where λ and γ are parameters representing the sensitivity to variance and kurtosis, respectively.

Step 7: Equilibrium Conditions The equilibrium conditions for data utilization and sharing are determined by solving the above first-order conditions. The optimal values of x_{it} and $\tilde{x}_{it}$ will balance the trade-off between maximizing output and minimizing risk.

Implications for Equilibrium In equilibrium, the following conditions hold:

1. The optimal data utilization x_{it}^* and sharing $\tilde{x}_{it}^*$ are such that the marginal benefit of data utilization equals the marginal cost adjusted for risk.
2. The total data used D_{it}^* and shared D_{sit}^* by each firm are consistent with the market prices for output and data.
3. The overall equilibrium in the data market ensures the aggregate demand for data equals the aggregate supply.

Thus, in all cases, there exists an equilibrium allocation $\mathbf{D}^*$ such that $\Phi(\mathbf{D}^*) = Y^e$, as we wanted to show. $\square$

10 Conclusion

In conclusion, the equilibrium when firms own data is characterized by optimal data utilization and sharing strategies that balance the trade-off between output maximization and risk minimization. The resulting equilibrium conditions ensure that firms' decisions are consistent with market prices and the inherent risk characteristics of the data.

$$\boxed{\begin{array}{l}\dfrac{\partial \Pi_{it}}{\partial x_{it}} = p_{it}\eta D_{it}^{\eta-1} L_{it}\left(\alpha Y_{it}\right) - p_{DI}\alpha Y_{it} - \lambda \sigma_{D_{it}}^2 - \gamma \kappa_{D_{it}} = 0 \\ \dfrac{\partial \Pi_{it}}{\partial \tilde{x}_{it}} = -p_{DI}\tilde{x}_{it} J_{it} - \lambda \sigma_{D_{sit}}^2 - \gamma \kappa_{D_{sit}} = 0\end{array}}$$