論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Research on Real-Time Economy:Bottom-up construction of SNA from micro-transaction data
著者(和文)	赤木茅
Author(English)	Kaya Akagi
出典(和文)	学位:博士(理学), 学位授与機関:東京工業大学, 報告番号:甲第11730号, 授与年月日:2022年3月26日, 学位の種別:課程博士, 審査員:三宅 美博,山村 雅幸,石井 秀明,高安 美佐子,小野 功,出口 弘
Citation(English)	Degree:Doctor (Science), Conferring organization: Tokyo Institute of Technology, Report number:甲第11730号, Conferred date:2022/3/26, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

TOKYO INSTITUTE OF TECHNOLOGY

DOCTORAL THESIS

Research on Real-Time Economy: Bottom-up construction of SNA from micro-transaction data

Author:
Kaya AKAGI

Supervisor:
Prof. Yoshihiro MIYAKE, and
Prof. Hiroshi DEGUCHI

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Science

in the

Artificial Intelligence
Department of Computer Science

February 28, 2022

TOKYO INSTITUTE OF TECHNOLOGY

Abstract

School of Computing
Department of Computer Science

Doctor of Science

**Research on Real-Time Economy:
Bottom-up construction of SNA from micro-transaction data**

by Kaya AKAGI

RTE (Real-Time Economy) is a digital ecosystem exchange, record, and utilize, information related to various economic transactions in real-time through digitization and automation. In recent years, RTE has been attracting much attention, especially in Europe, where developed information infrastructures began to discuss how to utilize the data obtained.

Once RTE is realized, it will be possible to obtain various data, especially transaction data such as accounting data and invoices, in real-time, which we have not used in current economic statistics and analytical models. If the type, quality, and update frequency of data changes with RTE, the economic statistics and economic analysis models that handle such data will also need to be changed. In this paper, we outline the concept and history of RTE and then discuss the system's requirements for the realization of RTE. The implementation of RTE requires three layers: information infrastructure, e-service layer, and management layer. This paper mainly discusses the management layer and economic statistics and models to realize that. We then propose the three elemental technologies required to implement the RTE management layers: an automatic economic statistics estimation system, an economic statistics estimation system using accounting data, and an economic analysis method using accounting information of each firm as a state space.

Acknowledgements

I would like to thank my esteemed supervisor – Prof. Hiroshi Deguchi, for his invaluable supervision, support, and tutelage during my Ph.D. degree, and Prof. Yoshihiro Miyake for his appropriate guidance and assistance in completing my doctoral thesis. I would also like to thank Dr. Shuang Chang for her technical support on my study.

My gratitude extends to the Center for TDB Advanced Data Analysis and Modeling [1] for the funding and offering data to undertake my studies at the Department of Computer Science, Tokyo Institute of technology. Additionally, I would like to express gratitude to Dr. Takaya Ohsato for his treasured support, which was influential in shaping my estimation methods and critiquing my results. I also thank Prof. Misako Takayasu, Prof. Masayuki Yamamura, Prof. Isao Ono, and Prof. Hideaki Ishii, and Prof. Hideki Takayasu for their mentorship.

[1] http://www.tdb.dis.titech.ac.jp/en/index.html

Contents

Chapter 1

Introduction

1.1 Real-Time Economy

Recently, the concept of a Real-Time Economy (RTE) is attracting attention in Western countries, especially in Baltic countries such as Estonia, to utilize big data for economic statistics, analysis models and Evidence-Based Policy Making(EBPM)[1].[1] [2]

The concept of current RTE was first proposed as a vision of General Electric in 2002, as a way to manage companies in real-time through computerization and society as a whole to respond to changes in real-time[2]. In this sense, the concept of the real-time economy initially spread in the technological field of monitoring through the computerization of procedures.

Similarly, the RTE has also been attracting attention in accounting, especially in auditing[3]. In accounting, the RTE has been the subject of research on managerial accounting at the Rutgers Accounting Research Center at Rutgers Business School since around 2002. The center describes the real-time economy as "The core objective of the real-time economy is the reduction of latency between and within processes."[4]. The Rutgers Business School has been not used the term real-time economy itself before 2002. Still, they have used the concepts of continuous online monitoring of transactions, continuous online auditing, real-time auditing, and real-time reporting since before 2002[5],[6].

As shown above, RTE started as a concept about the management of companies using information and communication technology and developed as a new auditing technology in accounting. On the other hand, the RTE is not limited to corporate units but covers the economy of an entire country. Therefore, the RTE on the whole economy has recently attracted much attention, especially in Baltic countries like Estonia and Finland.

On the other hand, in Japan, a similar concept concerning the whole country's real-time management existed from 2008. Therefore, Deguchi, Sakaki, and Oonuki started to suggest that the government of Japan apply electric scanning and electric accounting data for national statistics, monitoring, and scenario simulation for policymaking in real-time order[7]. To realize this, their research group developed an algebraic representation of accounting (Exchange algebra) and programing language for accounting data called AADL (Algebraic Accounting Description Language)[8]. Moreover, they formulated the System of National Account (SNA) in the accounting base according to these bases[9].

[1] In January 2020, as a researcher of the Cabinet Office of Japan, we visited Estonia to conduct a hearing survey for an overseas research study on the use of invoice information in the preparation of Supply and Use tables.

Research on RTE covers the entire country in Finland started around 2006 at the Aalto University School of Business. Since 2015, the Real-Time Economy Competence Center has been the core of the research.

Similarly, in Estonia, research is being conducted at Tallinn University of Technology. A survey study conducted by Tallinn University on behalf of the Estonian Ministry of Economic Affairs and Communications (MoEAC) describes RTE's current status in Estonia [10]. The report describes RTE prospects in Estonia based on the definition, benefits, obstacles and risks of RTE. The definition of RTE in the report is as follows.

> Real-Time Economy is a digital ecosystem where transactions between diverse economic actors take place in or near real-time by way of an increasingly automated exchange of digital, structured and machine-readable data in standardized formats. The resulting acceleration of information exchange and improved access to information is expected to reduce process latencies, save resources and transaction costs, increase organizational efficiency and business competitiveness, increase the speed and quality of decision-making, improve transparency, and stimulate economic and social innovation[10].

In Estonia, the deployment of RTE is led by the Estonian Association of Information Technology and Telecommunications (ITL)[2] under the name of Real-time economy EE according to Vision 2030[11]. The ITL will develop a system for real-time transfer of electronic invoicing and receipts between organizations in accordance with European standards in 2018 as the first step in the implementation of RTE, and plans to roll it out over the next 10 years.

The government of Estonia divides RTE into three blocks:

- core technological infrastructure,
- E-service layer, and
- management layer,

and describes the development stages of each block[10]. Of these, the core technological infrastructure and the E-service layer are currently the most developed areas, and there are many previous studies on various technologies and standard formats. On the other hand, the last one, the management layer, mainly focuses on the methodologies and concepts of accounting and auditing on a per-firm basis, and the discussion on a per-nation basis has not progressed much.

Let us look at the management layer concept; how to utilize obtained data in decision making. Although the management layer's primary stakeholders are companies and the government, it mainly developed on the former because of the history of RTE. Companies' purpose for introducing RTE is more efficient resource management, better planning and risk assessment, and faster decision-making. On the other hand, for the government, their current vision below.

> The same corporate data can also be used by state agencies to facilitate automated business reporting, real-time taxation or to compile national statistics without imposing reporting burdens on companies.

[2] https://www.itl.ee/en/

> Furthermore, as tools and technologies for data analytics and machine-learning become increasingly prevalent, governments can make use of real-time data from diverse sources, such as national or third party databases or IoT (‘ internet of things ’) sensors, which would allow governments to build dashboards for continuously monitoring and assessing the country ’ s economic situation and develop predictive models for forecasting economic events (e.g. company failures, changes in tax revenues) based on real-time data. This would allow governments to start providing customized services and feedback to companies (e.g. enabling companies to assess their indicators against their peers operating in the same sector or giving indications of possible risks) and to develop early warning systems for individual companies and the government. [10].

In order to realize the management layer of RTE, we can list the necessary elemental technologies as follows:

- Automatic estimation of real-time economic statistics
- Methods of economic analysis on a national scale on a firm-by-firm basis
- Coordination and application of these methods to decision making

Furthermore, we must do all of these in real-time. As will be discussed in detail later, accounting data handling is essential in realizing these goals. RTE acquires transaction data such as invoices and tax-related accounting data. Since the primary RTE data constitute accounting, we must construct statistics and analytical models using the data based on the accounting data. In particular, to realize each firm's economic analysis on the national scale, it is necessary to construct a method for analyzing firms, i.e., an audit using accounting data, at the macro level. There is no analytical method in existing mainstream economics that combines events at the firm level, such as bankruptcy and revenue decline, with events at the national level, such as tax revenue.

These elemental technologies do not yet exist in the world at present. Indeed, when we visited Statistics Estonia to determine how many concrete measures they planned for implementing such management on the national level, we were told that these measures were still conceptual and had not yet determined concrete methods.These concrete implementation methods will become an important research issue in the future and the production and operation of statistics.

This paper proposes methods and technologies to achieve these requirements of RTE. Chapter 4 describes our method and implementation to automate the compilation of Supply and Use Tables, the one of core statistics of the SNA, of Japan. Chapter 5 shows our new system to compilate Input-Output Table with accounting data. Chapter 6 introduces the Turnpike model extended by accounting microdata as a new method for economic analysis enabled by RTE, i.e., analyzing micro accounting events at the national level.

1.2 Real-Time in Economics

In terms of Economics, before 2002, economists have used the term "real-time economy" in the sense of a 24-hour economy or a night-time economy rather than instantaneous financial information transactions through digitization. For example, an article in 1989 states, "We now have a 24-hour, never-sleeping, real-time economy where managers, employees, and customers can do business any time, any place. Flexible working hours will become standard. Rigid opening hours and fixed places of business are obsolete."[12].

Some economists treat the concept of Real-Time in a context similar to RTE. In Kitchen (2003) [13], real-time is defined as follows.

> For us, "real time" refers to the effort to conduct continuous, contemporaneous analyses of incoming information that allow forecasters to make continual and instantaneous updates to their forecasts as new data become available. The real time forecasting system (RTFS) is the result of our efforts to produce a fluid, data-based forecast of contemporaneous real GDP growth that is subject to continual updating the instant new data become available.

This definition of "real time" is close to the concept of RTE. What they call RTFS here is a regression model between various indicators and GDP. In any case, our interest here is in the immediate updating of data, and this perspective is critical in the relationship between RTE and economic models. This point will be discussed later in connection with the concept of Data Assimilation and Model Re-estimation (DAMR).

In a different context from the Real-Time Economy, there is a research field called "Real-Time Economics", conducted mainly by The Federal Reserve Bank of Philadelphia since around 2001[14]. These studies are concerned with the use of "real-time data" in economic forecasting. However, we need to be careful about the meaning of the term. For example, E. Koenig (2001) argues that the use of real-time-vintage data, which were available at the time of the year of the model from back numbers of the government, rather than vintage data from the latest revised official statistics of the government, is an effective way to improve the performance of economic models[15]. We should note that real-time-vintage data in this context is instead the opposite of the concept of real-time-data in RTE. In a similar vein, Croushore and Stark (1999, 2001), Orphanides (2001), and Dynan and Elmendorf (2001) noted the need to work with real-time-vintage data[16] [17] [18] [19] . Although we need to discuss the problem of using the latest data in its own right, we will not treat this problem in this paper.

Chapter 2

Current situation of RTE

Japan is currently undergoing official statistical reforms based on the "Basic Plan for the Development of Official Statistics"[20]. The central theme of the reform is to improve the estimation accuracy of the System of National Accounts (SNA), which is an international standard for producing economic statistics to grasp the economy of a country in terms of both flows and stocks, and to utilize various kinds of big data for economic statistics as a method to achieve this goal.

"Big data" is defined by Doug Laney, an industrial analyst, as data with three characteristics, referred to as 3V, that is different from conventional data: large Volume (more than 1PB), high Velocity (real-time), and wide variety (text, video, and transaction data)[21].In Japan, the main objective of statistical reform is to improve the comprehensiveness of the data by using different types of data (Variety), such as tax information and transaction information, while worldwide, the aim is to improve the immediacy (Velocity).

Conventionally, the data source for official statistics has been various statistical surveys such as the Economic Census. However, with the increase in the number and complexity of statistics produced, the respondents' burden increases, which led to a decline in the response rate and estimates' accuracy[22]. Against this background, several countries attempt to utilize various kinds of new data with high collection rates and coverage, such as tax information, to improve official statistics' estimation accuracy. Several advanced nations have attempt invoices as the first target as the data source for statistics. The Japanese government will introduce the qualified invoice preservation system until 2023 as a requirement for the purchase tax credit, and the use of the obtained invoices for official statistics is under consideration. In some foreign countries, such use of tax information for official statistics is already underway, but in Japan, utilize tax information for official statistics is not allowed. Therefore, the Economic and Social Research Institute, Cabinet Office, Government of Japan, which is responsible for the estimation of the SNA, conducted overseas survey research to four countries (France, Denmark, Estonia, and Australia) to investigate the cases of other countries in 2020, and the author visited Estonia as a researcher of the survey [4]. In addition to the Estonian Statistical Office, we visited Tallin University, ITL (Estonian Association of Information Technology and Telecommunications), Gofore Plc, and Planetway Corp, which are Finnish companies responsible for the implementation and operation of e-government security infrastructure in Baltic countries.

France and Denmark use VAT (Value Added Tax) data to estimate investment and consumption in the SUT, a core statistic for GDP estimation. Although Estonia and Australia do not directly use VAT data in GDP estimation, they incorporate tax information into the estimation as reference values, such as trend estimation for quarterly estimation. Japan is currently simply aiming to use invoice data for statistical data and implement its infrastructure. However, in other countries, this stage has already been achieved, and reforms based on new concepts such as RTE are in progress. In the following, we introduce Estonia and Finland's efforts to realize the RTE, comparing them with the current situation in Japan.

RTE is a generic term for an economic ecosystem in which information on various applications and transactions are communicated, reflected, stored, and managed in real-time through ICT and automation and promoted in countries such as Estonia, Finland, and Iceland. These countries implemented the data communication infrastructure on X-Road. X-Road is an open-source data communication layer that guarantees communication between white-listed institutions. Finland and Estonia are promoting cooperation between Suomi.fi and X-tee, a communication infrastructure based on X-Road, mainly through NORDIC INSTITUTE FOR INTEROPERABILITY SOLUTIONS, a joint organization of the two countries. The participation of Straumurinn of Iceland is planned in the future[23].

In Finland, Aalto University, and in Estonia, Tallin University is responsible for research on RTE implementation, and ITL plays a central role in promoting it. We could divide RTE divided into three conceptual layers in its implementation: 1) technology infrastructure, 2) e-services, and 3) management. In Japan, the concept of technological infrastructure and e-services is currently at the beginning stage. However, in Estonia and Finland, these have already been completed. Today in these countries, 99% of administrative procedures, including taxation, have been operated on the electronic infrastructure system since 2018. On the other hand, the management layer, i.e., how to use these data for decision making, is still in the conceptual stage of discussion.

To realize the management layer, the technological infrastructure and the e-service layer implemented on it are prerequisites. In the following, we will focus on these. In Estonia and Finland, each citizen and corporation have assigned an electronic ID, and the information necessary for administration is linked to the ID and registered in the National Register as e-residency. Everyone registered in the e-residency can execute almost all administrative procedures such as voting, change of address, and investigation of legal entities on the online management screen (eesti.ee in Estonia). Moreover, they can freely set their access rights to this information, and by allowing various services to access this information, updates to the information, such as a change of address, will be automatically reflected in those services.

On top of such an information technology infrastructure, they design services. We introduce below the services required for statistical estimation and accounting transactions in particular and the technological infrastructure to realize them. As a matter of course, we design technical infrastructure and services according to specific objectives. In particular, in Estonia, the seven principles called Principles of Estonian e-governance are essential as the design philosophy of services[24]. For comparison, we will also list the three digital principles outlined in Japan's Digital Government Action Plan[25].

It is interesting to note that the Japanese Government removes political ideals such as Transparency, Decentralization, and Open platform. From the perspective of statistical surveys, Once-only is particularly important among these principles. Japan's government also includes Once-only in Japan's three digital principles, and it is one of the essential issues for digitization in Japan. As mentioned above, in Japan, primary data for statistical production is based on questionnaire surveys, and the collection rate is decreasing due to the increasing burden on respondents. In particular, each ministry and agency in Japan conducts different surveys, and the information from these surveys is not integrated at all, making it necessary to fill in the same information every year, which is a significant burden. The Keidanren's

TABLE 2.1: Principles of Estonian e-governance

Principle	Description
Decentralisation	There ’ s no central database and every stakeholder, whether a government department, ministry, or business, gets to choose its own system.
Interconnectivity	All system elements exchange data securely and work smoothly together.
Integrity	All data exchanges, M2M communications, data at rest, and log files are, thanks to KSI blockchain technology, independent and fully accountable.
Open platform	Any institution may use the infrastructure and it works as an open source.
No legacy	Continuous legal change and organic improvement of the technology and law.
Once-only	Data is collected only once by an institution, eliminating duplicated data and bureaucracy.
Transparency	Citizens have the right to see their personal information and check how it is used by the government via log files.

TABLE 2.2: The three digital principles of Japan

Principle	Description
Degital First	Each procedure and service is completed consistently on digital.
Once-only	Once the information is submitted, it is not necessary to submit it again.
Connected One-Stop	The system realizes a one-stop for multiple procedures and services, including private services.

request for the government to introduce measures to reduce the burden, such as avoiding duplication, has not yet been achieved[26]. On the other hand, in Estonia, all items surveyed once in any survey are shared in the national register, designed so that respondents do not have to write the same information. Estonia sat the amount of time and cost required by companies to fill out the questionnaire as a KPI, and it is obligatory to decrease this amount year by year. Also, the Estonian government has introduced a method to provide respondents with incentives to respond to the statistical survey. Businesses that respond to the survey can obtain information such as the analysis of the company’s productivity based on statistical information. The management layer mentioned above has a philosophy of providing information that is not ordinarily available to firms alone, such as the position of the firm at the national level or the level of the industry to which it belongs from the perspective of macroeconomic indicators and risk assessment, in return for the provision of information. Although Estonia has not implemented these principles yet, they will expand shortly. Japan also needs to discuss the design of incentives for providing such information in the design of the service layer, but the current plan does not include incentive design.

Also, in statistical surveys, the difference between the data format usually held by companies and the questionnaire increases the response cost. In response to this, Europe has established a data standard for accounting in XML format as the Pan-European Online Public Procurement (PEPPOL), and Estonia provides an API for its reporting system. In Estonia, various accounting software supports this API, and they estimate necessary items automatically from bookkeeping. In Japan, the specification of e-Tax was released in 2021 to enable software developers to link accounting software with e-Tax [27]. Besides, Mynaportal, the portal for administrative procedures, has released APIs for developers to automate social insurance and tax procedures since 2020 [28]. As for invoices, the Electronic Invoice Promotion Council has been established and is developing a standard format[29].

Also, Estonia is pursuing efficiency based on the interconnectivity and once-only principles regarding using the collected data. The national register manages all the information of each corporation and individual by linking them to IDs. Statistic Estonia has full access to the national register information, including tax data, through X-Road, and they can use all the information to estimate statistics. In fact, in Estonia, invoice information is not collected for statistical purposes, but they use aggregate invoice values to grasp quarterly estimates trends. In the Business survey, which is the data source of SNA, survery offices supplement some of the survey items before the responses if they already know them from other surveys, and thus some of the survey items may include the contents of other surveys, including tax information.

Japan's government are planning to register some data preferentially in the Base Registry, which is "a database of basic social data on people, corporations, land, buildings, qualifications, registered and disclosed by public organizations and referred to in various situations, and which is the foundation of society, ensuring accuracy and up-to-dateness". The government has not determined how to handle data not included in the base registry. From the viewpoint of statistical production, the government scheduled to submit a draft law on information coordination for research and statistical work using the My Number System in fields except for the social security, tax, and disaster fields to the ordinary Diet session in 2022. For information coordination, it is essential to standardize data, and Japan has only just begun to work on this point. Conventionally, each ministry, each department, and each person in charge registers data in their way, and although they provide APIs for e-stat, a portal for statistics, their convenience is very low. In response to this, the Ministry of Internal Affairs and Communications (MIC) has just started to establish "Formulation of uniform rules for notation of machine-readable data in statistical tables" at the end of 2020[30].

As an information infrastructure system, the Japanese government envisions a cloud-type government common platform using Amazon Web Service, which is expected to take a different from the open-source distributed system using X-Road in Estonia[31]. The difference between the Japanese government's centralized system and the decentralized system in Estonia and other countries needs to be analyzed. However, since there is no information on implementing the Japanese government's common platform, it is not easy to compare.

Let us summarize the discussion so far, advanced e-government countries have already realized the creation of the concept, the implementation of the information infrastructure technology, and the e-service running on it, and they have started to discuss the utilization of the obtained data such as the RTE. On the other hand, the Japanese government has just started to discuss information infrastructure technology and services, not the new data usage. In particular, they have not formulated the concept for statistics with e-government yet. We expect that discussions on information management on the implemented information infrastructure, e-service layer, will become more critical in the future. This study discusses these issues, especially the management layer's future, in later chapters.

Chapter 3

Data and Model Requirements for RTE

3.1 Data Assimilation and Model Re-estimation (DAMR)

RTE is a concept that all the transaction data are generated in a standardized digital format in real-time. In the management layer, RTE aims to update the statistical and estimating models in real-time and use them for policymaking. This section discusses the concept of Data Assimilation and Model Re-estimation (DAMR) to examine what is possible when data is updated in real-time, especially in an economic model.

In general, we need both models and data for simulation. In this case, we distinguish between the structure of the model, i.e., the setting of parameters and boundary conditions, and the data's values. In general, when the data acquisition shortened its time constant, we try to improve the prediction accuracy by shortening the update frequency of the data in a model. Such a method is called Data Assimilation. Data Assimilation is a methodology that has been primarily used in meteorology, and it is a method to improve the accuracy of predictions by simulations by updating the simulation with data obtained for each finest-grained time scale. Data Assimilation is the first concept we should consider as a way to utilize real-time data.

Next, model re-estimation means updating the structure of the model with the new data. In natural physics, the structure of the observed model is usually invariant. Therefore, a scientist does not use Model Re-estimation in natural science. However, in the social sciences, it is a problem that the structure of the model itself changes. In social science, the division between data and structural parameters depends on the frequency of observation of the data, constrained by cost and labor. For instance, in the System of National Account (SNA), which calculates the GDP and other economic statistics, the parameters of the economic structure expressed in the input-output table are modified in the base year estimation at an interval of about five years. Also, we extend the estimation of various economic indices by using various data obtained every quarter under the stable structure.

Therefore, as data observation frequency approaches real-time with RTE's progress, the division between the economic model's fixed structure and the variable values changes. In economics, sometimes we fixedly treat structural changes, and also we can treat them endogenously, as represented by concepts such as technological changes. The necessity of such endogenization decreases with the frequency of data acquisition and increases with the target forecasting span. Therefore, the model's composition and the simulation depend on the type of data obtained and their respective time constraints. To promote RTE, in which the frequency of data update itself is the axis of the concept, and to make the best use of the obtained data, it is essential to re-design a method for updating the data and structural parameters. We will explain our design of DAMR with the format of data obtained in RTE in the next section.

3.2 Accounting Data in RTE

To re-design parameters and structures in real-time, we have to make significant changes to the current economic statistics and models. There are various statistical methods to express the economic status, and there are also several analytical methods using them. The most representative form of expressing the economy is the SNA, an internationally standardized economic statistic. However, we should update the SNA continuously, and the current 08SNA[32] is also facing various issues. For instance, as the complexity of statistics and the amount of information required has increased, the cost of producing statistics has become a problem worldwide. In Japan, the respondents' burden to collect information for statistics increases year by year. Due to this, the accuracy of the official statistics has been declining due to the low data collection rate[22]. In Japan, statistical reform is underway to improve the accuracy of the SNA estimates, and one of the solutions is to utilize privately owned big data for official statistics[20]. Estonia solved this problem by digitization. In Estonia, there is a principle that no matter what kind of survey, once a respondent answered a question, it will be recorded on the national register and shared among other surveys. It follows one of the principles of Estonian e-governance called Once-only. Estonia measures key performance indicators every year to see how much time and money respondents spend on filling out questionnaires to achieve the principle. The improvement of this KPI is a mandatory goal, and they have been achieving to decrease the time and cost since its implementation.

As mentioned above, the national economy's fundamental representation is the aggregated economic statistics such as the SNA. However, on the other hand, the firms' internal states are recorded by accounting. To estimate economic statistics such as the SNA, we must collect data for the purpose through surveys such as the Economic Census. However, due to the gap of form between accounting and statistics, a considerable burden of converting and collecting the data arises. The conversion cost of the information is an obstacle for utilizing big private data because the same problem also arises when we try to utilize the companies' existing data. European countries set accounting and data standards Pan-Europian Public Procurement OnLine (PEPPOL) and provide API to automate this conversion for accounting software companies to reduce this cost[33][34].

From the perspective of data conversion, it is the most cost-effective to use accounting information held by private companies as is. However, it is not common to use bookkeepings directly in economic expressions and analysis methods. To utilize accounting private-sector big data for official statistics, we have developed an input-output table (IO), a critical statistic in the SNA, using inter-company transaction and accounting data owned by Teikoku Databank (TDB)[35]. Although these studies aim to estimate national statistics with accounting data directly, they also solve several current statistics problems, such as the frequency of data update, regions covered, and granularity of data analysis. In these previous studies, we have extended the existing statistics by making use of accounting big data. This study is also an example of the re-design of DAMR in economic statistics with new data. However, to utilize RTE's economic data, it is also necessary to re-design the DAMR of economic analyses.

Fig 3.1 shows our research scheme to realize RTE virtually. Previous studies are responsible for data-statistics relation: a reflection from accounting data to economic statistics in this figure. This paper focuses on the data-model part: applying accounting data to the economic model and proposing an agent-based simulation model with accounting information. Our study's final objective is to develop an economic simulation model based on TDB's accounting information and inter-firm transactions big data. However, in this paper, we are just focussing on the concept and implementation of the model.

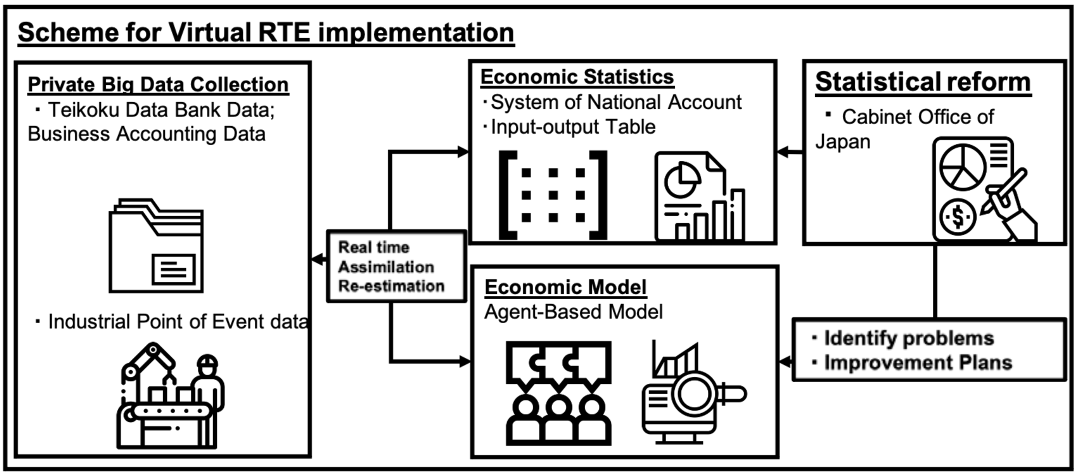

FIGURE 3.1: Scheme for Virtual RTE implementation

3.3 Exchange Algebra

Naturally, the choice of language constrains the space of states that a model can represent or hold. We have already mentioned that bookkeeping is the most common form of information in economic transactions, and there are many data formats for handling bookkeeping and accounting information. There are many data formats for bookkeeping and accounting information, but XBRL (eXtensible Business Reporting Language) is the most common data format for current accounting information. However, XBRL is a format for the representation of data, not for calculations. The concept of accounting has an internationally formulated structure, and the necessary calculations are predefined. Therefore, the necessary information and its computation for handling basic accounting information are fixed and can be defined algebraically or programmatically. By handling accounting information according to algebraically defined data types and systems, only computations allowed in accounting are possible, and safe and efficient processing becomes possible. Therefore, as a data format for bookkeeping information, we use Exchange Algebra, a mathematical formulation of the bookkeeping system, as an algebraic formulation of bookkeeping[9].

We generally define the data type of an Exchange Algebra in the following form.
$value < name, unit, time, subject >$

This data expresses who processed (subject) what (name) in what unit (unit) when (time) and how much (value). Each basis (name,unit,time,subject) has its own basis set, such as $\{apple, orange, banana\} \in name$. For example, $10 < apple, unit, 2015, companyA >$ represents 10 apples in 2015 by company A. Such an algebraic representation makes it possible to maintain mathematically guaranteed robustness in the information processing process.

Chapter 4

Statistics for RTE: Automation

Since RTE aims to achieve real-time economic data communication, register, and utilization with ICT and automation, we need to build an automated estimation system for statistical estimation to utilize the information infrastructure and e-service layer. This section will describe the prototype of the supply table's automated estimation system, a part of SNA, as an example of a necessary automation procedure. Although this chapter mainly focuses on the Japanese statistical system, there are few examples of automatic estimation of SNA globally, and the need for such implementation is expected to increase in the future.

In the current Japanese national accounts (JSNA), we divide estimate procedures into several parts. In the base year, when the Economic Census is conducted about once every five years, the JSNA estimates the Input-Output Table and the Make Table (V Table) based on the Economic Census and other essential statistics and then estimates the Use Table and the Supply Table transferring the concepts and classifications of these tables. In the intermediate years, we extend the Use and Supply tables obtained in the base year estimation based on various statistical surveys' values. In this method, Use and Supply tables are not directly produced, and from the viewpoint of improving statistical accuracy, the government aims to shift to the SUT (Supply and Use Tables) system, which produces Use and Supply tables directly from primary statistical data such as the Economic Census.

"Basic Plan for the Development of Official Statistics, Phase III Basic Plan"[20] plan to estimate SUT, directly by a conceptually seamless method, based on the measured values of primary statistics such as the Economic Census. The concept of "seamless" has various implications, but first, it means direct estimation from the values of basic statistics. In V Table, the economic census accounts for most of the basic statistics. This section describes developing a prototype of an automated system that directly estimates the Supply table using the Economic Census and analyzes the results.

Among the SUTs in Japan, the Supply Table expresses the output value of items by each industry in the matrix format of "industry $\times$ product," but few detailed compilations on compiling the Supply table. As for the current V Table of Japan, "Commodity Input-Output Table by Industry (V Table) [Contents and Views]" published on the website of the Ministry of Internal Affairs and Communications (MIC) and "2011 Input-Output Table, Chapter 7: Types of Ancillary Tables and Their Contents 6: Commodity Input-Output Table by Industry (V Table)" are representative examples of the documents published by the government. However, the former mainly explains the tables' usage and does not refer to specific preparation methods. The later one mentions the method of preparation, but its description is limited to a simple one.

Making V Table (Supply Table) is more straightforward than Use Table or X Table, but it is a complex system that includes various concepts and processes. As far as the author can find, no published material explains the detailed process of creating V tables in Japan in an orderly manner, but we can grasp fragments of it from the materials published in the document of various conferences. In this paper, the author collects and compiles the information on V Table (Supply Table), which will become more critical with the transition to the SUT system, and reproduces the estimation method of V Table to the extent possible using the individual data of the Economic Census, and creates an estimation program using the Economic Census data. We divide the details of the estimation method from this paper because of its enormous volume. The details of the estimation methods were published as a research paper by the Economic and Social Research Institute, Cabinet Office, Government of Japan, in December 2020, entitled "The Methodology for Developing the Supply Table Prototype"[36].

4.1 Estimation method

This section estimates the V Table by reproducing as much as possible the processing method of the Economic Census data based on various published data and materials of meetings and proceedings of the Council of Industrial Relations and the Economic Census Research Committee. The programming code used in the actual estimation is available from the author's GitHub[1], so anyone can apply for the data and perform the estimation in this paper. The programming code and the Japanese explanation of the corresponding estimation method are also available on GitHub.

Fig 4.1 shows the general flow of the steps required to estimate V-tables using the economic census. For details of each procedure, please refer to the research paper mentioned above or the source code. This section gives an overview of each procedure.

Data

In this estimation, we use the Economic Census Activity Survey 2012 and the Economic Census Activity Survey 2016, which the MIC and the METI provide under the provision of Article 33-1 of the Statistics Act. The contents of each survey are as follows.

Economic Census of Japan 2012

> Activity Survey [02-05,12] Single Establishment Questionnaire (Mining, mining, gravel extraction, manufacturing, wholesale, retail, common industry), [13] Enterprise Questionnaire and [17-19] Establishment Questionnaire (Mining, mining, gravel extraction, manufacturing, wholesale, retail)

Economic Census of Japan Activity Survey 2008

> [01] Personal Business Survey, [03-05] Consolidated Business Survey (Mining,, gravel extraction, manufacturing, wholesale, and retail), [11] Industry Questionnaire, [12] Company Questionnaire, and [16-18] Office Questionnaire (Mining, mining, gravel extraction, manufacturing, wholesale, and retail)

Note that due to publication timing for comparison, this paper's results and analysis cover only the data from the 2012 Economic Census - Activity Survey.

[1] https://github.com/yakagika/sna_make_table

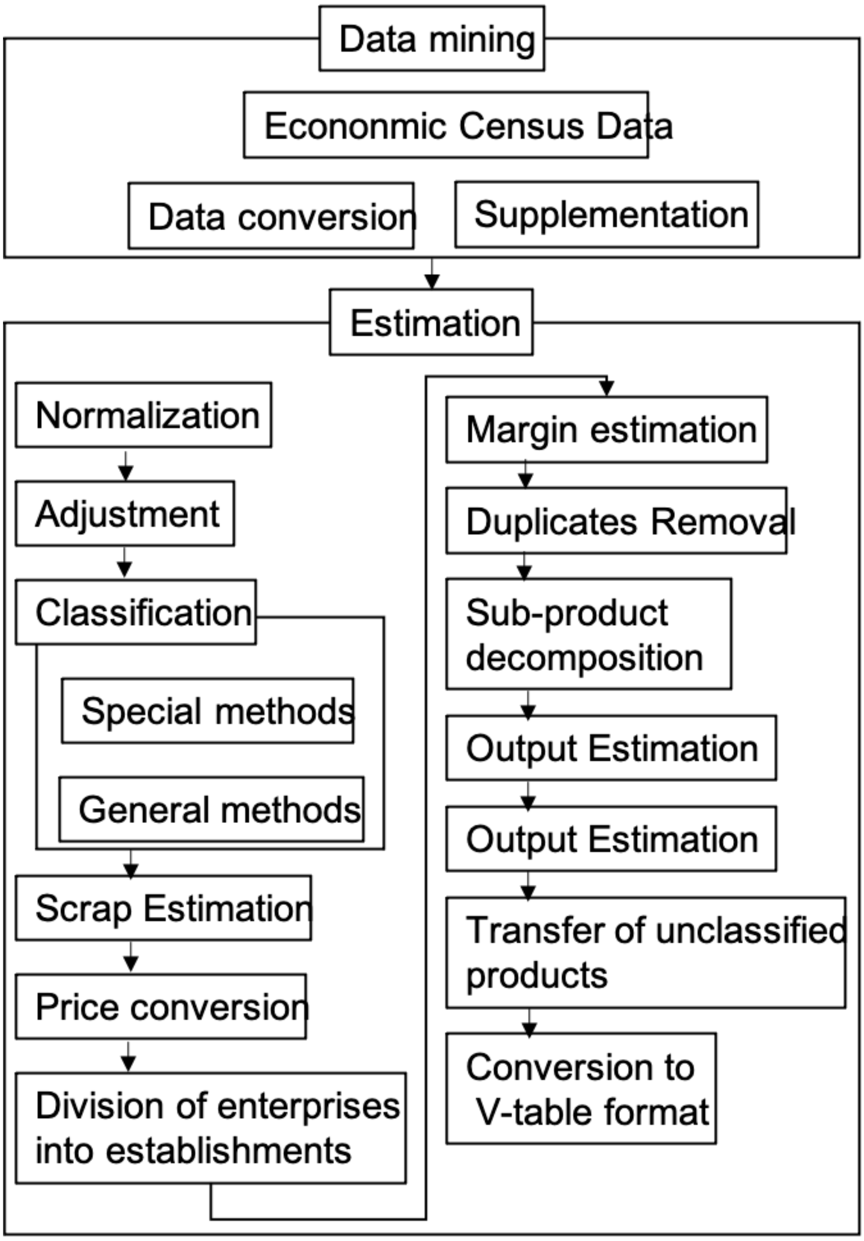

FIGURE 4.1: Procedures for V Table estimation.

Data conversion

The economic census data used in this estimation are provided by the Ministry of Internal Affairs and Communications (MIC) and the Ministry of Economy, Trade, and Industry (METI), but the format of the data differs from ministry to ministry, from survey to survey, and from year to year. Therefore, in order to integrate them into the program, we converted and unified the data format.

Inventory valuation adjustment

The unit of measurement at each intersection in the V Table is the output, which is different from the concept of measurement used in the Economic Census, such as the value of shipments and sales. Therefore, we need to convert the units into output values using the information on inventories.

Classification

V Table aggregates the output of industry $\times$ commodity, so it is necessary to estimate the commodity's output by industry. For this reason, we need to classify the establishments by industry, which is generally called industry classification. MIC and METI classify industries based on "Method of Industry Classification"[37] and "Method of Determination of Establishment Industry" in "Summary of Results"[38]. For general cases, they classify each establishment according to the "general method of industrial rating," which classifies establishments based on the commodity with the highest output value. In some cases, such as in the manufacturing industry, they classify establishments based on the "industrial rating special method," which determines the industry based on raw materials, work processes, machinery, and equipment.

Price conversion

The data included in the Economic Census differ by survey item and by year regarding whether they include or exclude taxes. JSNA assessed the transaction value and output value using the producer price, including tax, so the tax is treated for each item to unify the price assessment.

Division of enterprises

In the Economic Census, some items related to capital investment and expenses for which data are available only in the Enterprise Survey. Since V Table uses establishments as the unit of aggregation with some exceptions, such as the network industry, it is necessary to divide these items proportionally by establishments based on the values of items such as the number of employees.

Margin estimation

The Japanese current V Table record the commercial and transportation margins on each industry's self-intersection[39]. This method of estimating basic prices differs from the international standard in terms of the treatment of margins. In the estimation, the Japanese V Table subtracts the amount of purchase from the number of sales to estimate margin[40].

Sub-product decomposition

The Handout for the 13th Inter-industry Relations Technology Conference[41] states that "sectors that are not available as sideline jobs are excluded from the ratio calculation." However, this exclusion was not made in this study because the actual list was not available.

Output estimation

In the published value of V-Table, few sectors use the output value estimated from the Economic Census as the Control Total, and basically, the public estimation method recombines the row sectors of the X-Table into the column sectors of the V-Table. We estimate that Control Total directly from the Economic Census, so our value differs significantly from the published one. The following sectors are not available in the census.

- Sales of privately owned businesses in agriculture, forestry, and fishery
- Sales of household service establishments

- Sales of national and local governments
- Sales of national and local governments

Our estimation records zeroes at self-intersection of the above sectors. In this process, we delete data on firms, organizations, and establishments in network industries that are not used in the public estimation method. Also, for sectors such as electric power, commerce, and finance/insurance, where the concept of sales and production in the Economic Census is different, the sales-based values obtained from the Economic Census are used and thus differ from the initial estimates of the Ministry of Internal Affairs and Communications.

Compilation by various ministries and agencies
The initial estimates by the MIC are published after review and revision by each ministry. Since these ministries do not disclose the details of the revisions, we could not reproduce these processes. Since our methods are similar to the MIC's initial estimates, our values differ from the published values.

4.2 Estimation result

4.2.1 Structure

Although for reasons of space, we could not show the entire table of estimates with the above method and policy in this paper, that is available on GitHub[2].

Fig.4.2 visualizes the estimated V-table structure by coloring the intersection points where there are values regardless of size, and Fig.4.3 visualizes the public V-Table in the same way. Both figures can not express each row, and the column's details due to the paper's width, so the approximate arrangement is visualized. We can see, in both tables, that most of the self-intersection records values and that the manufacturing and service sectors include many sub-industries.

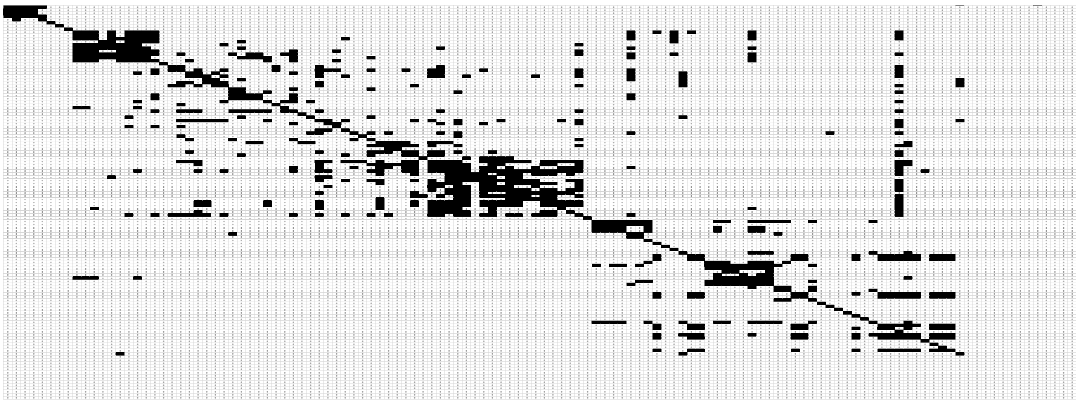

FIGURE 4.2: Estimated V-Table value structure.

The differences between Fig.4.2 and Fig.4.3 are visualized in Fig.4.4 and 4.5. Fig.4.4 shows the cells record values for both published and estimated values. Fig.4.5 represent cells where only the public has values and cells where only the estimated values have values. Also, Fig.4.5 shows the numbered areas where the differences are significant.

[2]https://github.com/yakagika/sna_make_table/blob/master/Result/result.csv

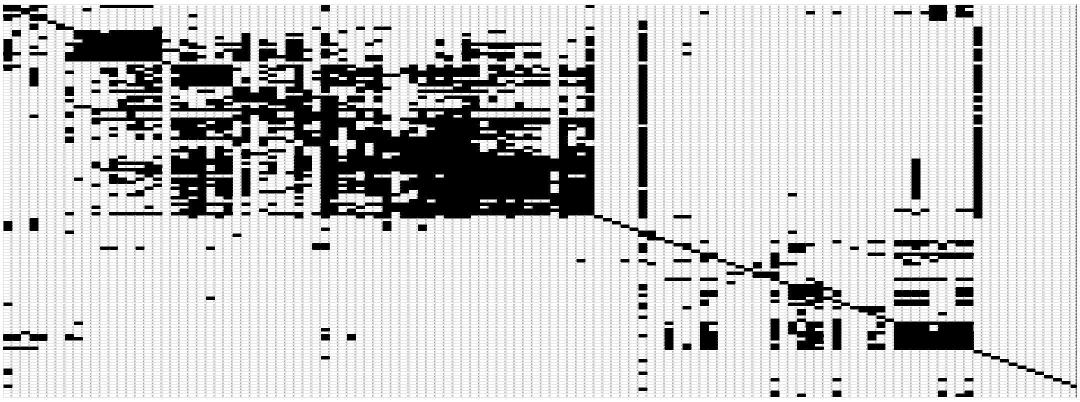

FIGURE 4.3: Public V-Table value structure.

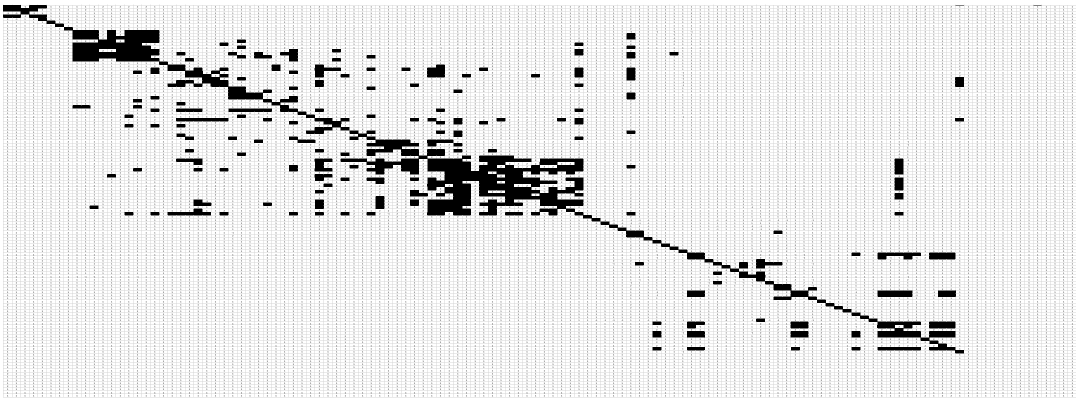

FIGURE 4.4: Location of overlapping values.

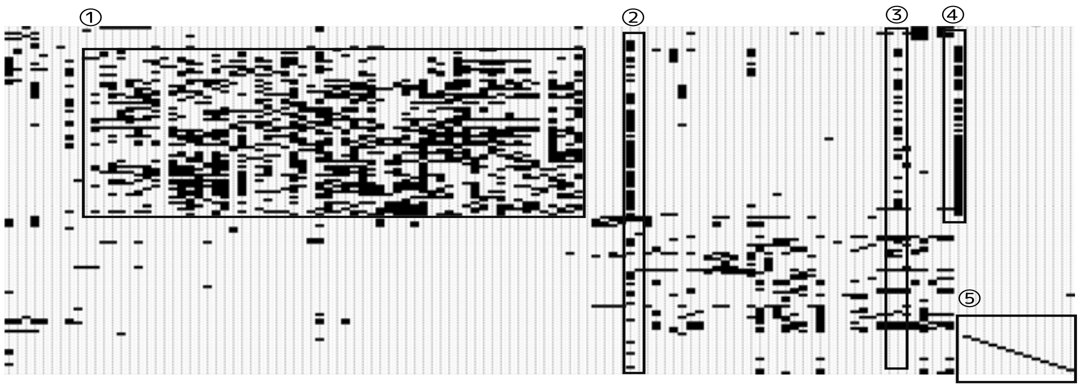

FIGURE 4.5: Structure comparison.

Fig.4.4 shows that the structures of the estimated and published values are consistent for the self-intersection of industries other than government expenditure, as shown in Fig.4.5 (5). Besides, for agriculture, foodstuffs, and manufacturing industries close to the main industries, both the estimated and the published figures show the structure of the calculation. In the following sections, we will explain the causes of each of the differences (1) to (5) in Fig.4.5.

(1) Structural differences in manufacturing sub-products

In particular, the black frame in (1) of Fig.4.5 shows a significant difference. The left end of (1) is "009 Meat" and the right end is "067 Other industrial products", indicating that the estimates do not capture the values on sub-industries within the manufacturing industry identified in the published data.

The first reason is the difference in the method used in the "industry classification special method". Our estimation includes all the shipments and inventories listed on the reverse side of the Economic Census's manufacturing questionnaire. Our estimation does not omit items as far as Economic Census records it. However, the "industry classification special method" classifies some commodities, such as continuous products, based on information such as "work process", "main raw materials", and "machinery and equipment", but our estimation method does not include that. This difference may cause the industry classification to be closer to the self-intersection point. We need to adapt the actual classification method as much as possible, but since the sorting items in the free descriptions is not realistic by the author's hand alone, the incorporation of this method will be limited to the parts that can be automated.

The second cause of the difference in (1) is the difference in data. The "2011 Interindustry Relations Table, General Explanation" (pp.382) mentions the following estimation method of production.

> We estimate the mining and manufacturing sectors' production according to the following four patterns (A to D) and determine the production after checking with the 2011 extended table.
>
> For "semi-products and work-in-progress" in the production, we used the Economic Census's reclassified data.
>
> A. Production estimated using the Survey of Production Dynamics
>
> B. Production quantity estimated using the Survey of Production Dynamics, and production unit price estimated using the reclassified data of the Economic Census - Activity Survey and other industry statistics.
>
> C. Production estimated by using the reclassified total of the Economic Census of Activity.
>
> D. Production estimated using other statistics

METI has not released detailed data processing methods used in the Economic Census classification and other statistics except the above description. Economic Census records only the top eight produced items in the six-digit product categories. On the other hand, METI surveys production for all preregistered products in the Survey of Production Dynamics. Therefore, in the Economic Census, subproducts in the manufacturing industry with small production are likely to be out of the rankings and not included in the questionnaire, and it is challenging to identify sub-products in the same industry.

(2) Electricity

Column (2) in Fig.4.5 represents "073 Electricity", which means that our estimation does not capture electricity output. Since the sales (income) amount surveyed in the Economic Census and the unit of measurement for electricity in the JSNA are different, this value is not used in the estimation of the published value, and they use the Annual Report on Mineral Resources and Energy Statistics published by METI. Also, since the Economic Census surveys only the top 10 production items, they do not record sub-products values, such as the small sale of electricity. Our estimation uses the sales (income) of electricity in the Economic Census, and this difference causes these differences.

(3) Automobile repair

In Fig.4.5, column (3) is for "104 Automobile maintenance/machine repair", and we can see that our estimation records values on "104 Automobile maintenance/machine repair" as sub-products in some manufacturing industries unlike the published data. X-table, thus V-Table, estimates these values mainly based on the White Paper on Automobile Maintenance by the Japan Automobile Service Promotion Association. Thus the difference on (3) is considered to be due to the difference in data sources.

(3) Unknown

In Fig. 4.5, column (4) is the "111 unknown" column. Fig. 3-4 shows that our estimates assign products classified as unknown in the public estimation to some other industry or do not grasp. In the Economic Census, there are some cases where the classification is unknown within a specific industry or product category, such as "unknown within the manufacturing industry". Our estimation classifies industries on the 1st page of the questionnaire as unknown within the classification scale on the 1st page, such as "unknown within the manufacturing industry," because we could not know the details of these industries from the 1st page only. We allocate proportionally products classified as this unknown within some industry to the output of each product in that industry, but public estimation may allocate them to "111 unknown" of the whole.

Since the concepts of V-table classification and census industry/commodity classification are different, various processes are required to convert them. However, our estimation uses a converter prepared by the author, and there may be differences in the processing of "establishments engaged in administrative and auxiliary work" and "unknown."

In particular, the treatment of the unknown has a significant impact on the overall structure. When an industry is classified as unknown, we select candidates to allocate that according to the current estimated output. However, if the current output amounts of candidates are all zero, all the ratios become zero, and our estimate will omit the output of that industry classified as unknown. Therefore, when the current output amounts of the candidates are all zero, our estimation will equally allocate them to the candidates' self-intersections. On the other hand, Fig.4.6 shows the V-table structure allocating the output equally to all candidates instead of the self-intersections if the candidates' existing output amounts are all zero. Fig. 3-5 shows an unrealistic output structure that allocates the values of establishments classified to unknown equally among all the candidates and is thus recorded exhaustively at many intersections. Fig. 4.6 indicates that the estimation results may differ significantly depending on one processing method, such as when the value does not exist. Therefore, to maintain the consistency of estimation among ministries and agencies, between years, and among statistical tables, it is essential to share detailed estimation methods, including the processing of errors and missing values.

(5) Government expenditure

The area (5) in Fig. 3-4 represents the sector related to government expenditure, from "120 school lunches" to "134 other non-profit organization services", which is not recorded in the Economic Census and therefore not included in our estimates.

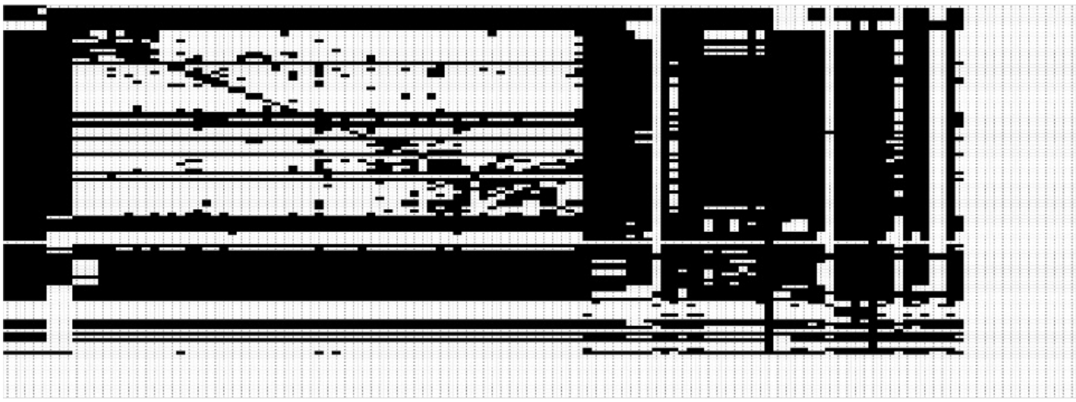

FIGURE 4.6: Structure of the V-table which allocate unkonown equally.

4.2.2 Quantitative comparison of estimates

In the previous section, we compared the structure of our estimation and the public values, and in this section, we compare them by value. Table 4.1 compares our estimation with those in public V-Tables.

TABLE 4.1: Total production.

Total output (million yen)	Estimated	¥928,680,929
	Public	¥929,417,153
Ratio	Estimated / Public	99.92 %

Table 4.1 shows that our direct estimation from the Economic Census captures about 99% of the public estimation's Control Total.

Next, we show the industry and commodity's total output ratio (estimated value/published value) in Table 4.2. and in descending order in Table 4.3. Table 4.4 shows industries and items with ratios close to 100% in descending order of absolute value of the ratio minus 100.

TABLE 4.2: Bottom 30 industries and products with the lowest percentage.

Industry Code	Industry	rate	Commodity Code	Commodity	rate
120	School lunch	0.0%	120	School lunch	0.0%
121	Sewerage	0.0%	121	Sewerage	0.0%
122	Waste disposal	0.0%	122	Waste disposal	0.0%
123	Transportation ancillary services	0.0%	123	Transportation ancillary services	0.0%
124	Public services	0.0%	124	Public services	0.0%
125	Education	0.0%	125	Education	0.0%
126	Research	0.0%	126	Research	0.0%
127	Health and sanitation	0.0%	127	Health and sanitation	0.0%
128	Social insurance and social welfare	0.0%	128	Social insurance and social welfare	0.0%
130	School lunch	0.0%	130	School lunch	0.0%
131	Education	0.0%	131	Education	0.0%
132	Research	0.0%	132	Research	0.0%
133	Social welfare	0.0%	133	Social welfare	0.0%
134	Other non-profit organization services	0.0%	134	Other non-profit organization services	0.0%
70	Construction and repair	2.9%	64	Auto parts and accessories	4.3%
104	Motor vehicle maintenance and mechanical repair	3.4%	81	Housing rentals	5.0%
64	Auto parts and accessories	5.7%	70	Construction repair	5.5%
63	Automobile	6.0%	63	Automobile	6.6%
81	Housing rent	6.7%	75	Water	6.9%
75	Water supply	6.8%	1	Arable agriculture	8.1%
1	Arable agriculture	8.0%	96	Research	9.8%
86	Freight forwarding	8.8%	86	Freight forwarding	13.9%
96	Research	9.7%	71	Public utilities	14.0%
44	Pig iron and crude steel	12.0%	30	Organic chemical industrial products	15.8%
71	Public utilities	12.0%	82	Rail transport	16.6%
109	Entertainment services	13.1%	44	Pig iron and crude steel	17.2%
82	Rail transport	14.8%	35	Petroleum products	18.4%

30	Organic chemical industrial products	16.9%	72	Other civil engineering and construction	22.8%
35	Petroleum products	17.6%	104	Motor vehicle maintenance and machine repair	23.6%
101	Other non-profit organization services	18.7%	33	Pharmaceuticals	24.0%

TABLE 4.3: Top 10 industries and products with the highest percentage.

Industry Code	Industry	rate	Commodity Code	Commodity	rate
39	Tanned leather, fur and similar products	4456%	39	Tanned leather, fur and similar products	4491%
99	Social Insurance and Social Welfare	670%	68	Recycled Resources Recovery and Processing	862 %
43	Other ceramic and stone products	604%	99	Social insurance and social welfare	669 %
67	Other manufacturing and industrial products	510%	43	Other ceramic, earth and stone products	599 %
20	Textile industry products	510%	20	Textile industry products	501 %
23	Furniture and equipment	500%	67	Other manufactured industrial products	500 %
32	Chemical Fibers	454%	23	Furniture and Equipment	488 %
27	Chemical Fertilizers	450%	77	Wholesale	479 %
77	Wholesale	440%	27	Chemical fertilizer	419 %
68	Recycled Resources Recovery and Processing	378%	32	Chemical Fibers	375 %

The sectors underestimated in the current estimation are medical care, automobiles, housing rentals, public works, pig iron, and crude steel, except for the sectors related to government expenditures from "120 School lunches" to "134 Other non-profit organization services". On the other hand, the overestimated sectors are leather, fur and fur products, social insurance, and social welfare. The estimates for plastic products, glass and glass products, and foodstuffs are relatively close to the published values. The details of the reasons for the underestimation or overestimation need to be analyzed individually. However, since it is difficult to analyze all industries and products, we would like to list some of the factors that may have caused the underestimation or overestimation.

Industries for which data are not available

In Table 4.2, data on housing rents, public works, and arable agriculture are not available because they are not included in the census.

TABLE 4.4: Top 10 industries and products with the closest percentage 100%.

Industry Code	Industry	rate	Commodity Code	Commodity	rate
040	Glass and glass products	99.7 %	037	Plastic products	99.8 %
037	Plastic products	98.8 %	040	Glass and glass products	99.1 %
025	Paper processed products	102.4 %	080	Real estate brokerage and leasing	102.0 %
088	Transportation ancillary services	96.5 %	088	Transportation ancillary services	95.4 %
012	Grain and flour milling	104.5 %	053	Production machinery	94.1 %
053	Production machinery	93.9 %	110	Other personal services	107.6 %
011	Marine foodstuffs	91.6 %	011	Marine foodstuffs	92.0 %
009	Meat	91.1 %	025	Processed paper products	108.6 %
047	Other iron and steel products	89.3 %	047	Other iron and steel products	86.1 %
103	Advertisement	88.9 %	009	Meat	83.2 %

Classification Methods

Economic Census surveys Medical care in Table 4.2, health and sanitation, social insurance, and social welfare in Table 4.3 by the same questionnaire. Because actual is not published, classification method, our estimation might have misallocated Medical care's shortfall suggested to health and sanitation, social insurance, and social welfare. We could not reproduce the method to classify pig iron, crude steel, and other continuous products in Table 3.2 because METI does not publish their methods.

Transfer method for unknown

The output of tanned leather, fur, and fur products and other ceramic and earthenware products, in Table 4.3, is relatively small in the manufacturing sector and is considered excessive due to the reallocation of the values classified to the unknown within the manufacturing sector. Since this transfer has a significant impact on the results, we need to introduce the actual method as a future issue.

Besides, there might be many possible causes to make differences, such as the margin estimation method and the estimation of intracompany transactions. For instance, the automobile in Table 3.2 is unnatural, but we will leave the case analysis to future work.

4.3 Conclusion

This chapter developed a prototype of the Supply Table based on the V-Table's estimation method from public information such as various conference materials. Furthermore, we estimated the Supply Table using the economic census data according to that method. We have succeeded in creating a system that can estimate some of the current economic statistics from a unified data source, but we have identified several problems related to the automation of various current statistics in the process.

The SNA and the V-tables constitute many procedures based on various concepts, and it is tough to integrate the large and complex overall picture of the estimation methods. This chapter has reproduced the estimation procedures based on the published information. However, the V-table reproduced by this method is, of course, different from the published one in several points.

In Japan's production of official statistics, the Guidelines for Quality Assurance of Official Statistics exist as a standard for promoting quality assurance activities to ensure and improve official statistics' usefulness and reliability. The Guidelines stipulate that "information on the statistical production process and precautions on the use of statistical information should be made clear" as the main element of "interpretability and clarity" in the quality of statistics. And also "information on the statistical production process (methods of collection, processing, accumulation, and publication of statistical data)" as the additional element of "reliability." Our research reproduces the estimation method based on publicly available information, although in a fragmented manner, proof that this "interpretability, clarity" and "reliability" are maintained to some extent. At the same time, however, as in the case of the processing of values classified to unknown, it is a fact that in the production of statistics, detailed processing rules at the level of algorithms and codes have a significant influence on the actual estimation results. However, there is no such information in the published statistical methods. Our estimation differed from the published values mainly due to the opaqueness of the estimation method.

In Japan, the details of the methods used to compile official statistics, especially major economic statistics such as the SNA, are kept secret for citizens. Furthermore, even between ministries and departments, the details of the methods are often unclear. The current situation, where there is no common understanding of a unified production method and reproduction is impossible, is problematic both from statistical quality and management. Our estimation methods and results reveal the problems and issues in implementing a seamless automated estimation system required in the future RTE.

Chapter 5

Statistics for RTE: Compilation of Input-Output Table with Accounting Data

In the previous section, we discussed the automatic economic statistics using existing government data, but RTE's implementation requires developing new methods for producing statistics using various data that have not been utilized so far, especially account data like invoices. This section attempts to estimate the Input-Output Table (IOT), one of the essential statistics in SNA, using new data such as accounting and transaction data, typical data source of RTE.

Since John Maynard Keynes (1883-1946) conceived the System of National Account (SNA) as a statistic to represent a country's economy as a whole, the 68SNA adopted by the 15th United Nations Statistical Commission, the 93SNA revised in 1993, and the 08SNA introduced in Japan in 2016, the national accounts have continued to be modified to overcome the problems of each period. In this way, economic statistics change in response to new technologies and concepts in each era. For example, with the development of information technology, the demand for immediacy and precision in various information, including economic data, has increased. In particular, with the development of new information technologies, real-time information processing using PoE (Point of Event) data, such as PoS (Point of Sales) data, has been introduced in various situations. Under such circumstances, people force national economic statistics to respond flexibly to the demands of new information technologies and big data. RTE is a response to such a concept.

In this section, we design and validate a system to automatically estimate the Input-Output Table(IOT), one of the core statistics in the SNA, using data provided by TDB (Teikoku Data Bank, LTD.), a private rating agency. IOT is a kind of economic statics systematized by Wassily Leontief (1906-1999), and it is presenting the flow of the commodities among industrial departments in a domestic country. In other words, it is a kind of matrix representing the input-output structure of the commodities among inter-industry. Nowadays, IOT is playing the central role of SNA to grasp the whole economic flow in the domestic country as an account[42].

The IOT is a matrix, as shown in Fig.?? [43]. If some output of industry in the left column is consumed to produce another industry's output in the upper row, we record the consumed amount on the intersection. An IOT is a useful tool for understanding the effects of specific government expenditures through ripple effect analysis and being used for GDP estimation in Japan[1]. In the current SNA, it is the primary statistic for understanding the specific industrial structure of a particular

[1]The statistical reform of Japan planned for 2025 decided to shift from the IOT to the SUT(Supply and Use Tables) for GDP estimation.

region. Any deficiency in the contemporary IOT directly affects the analysis and policymaking of that region.

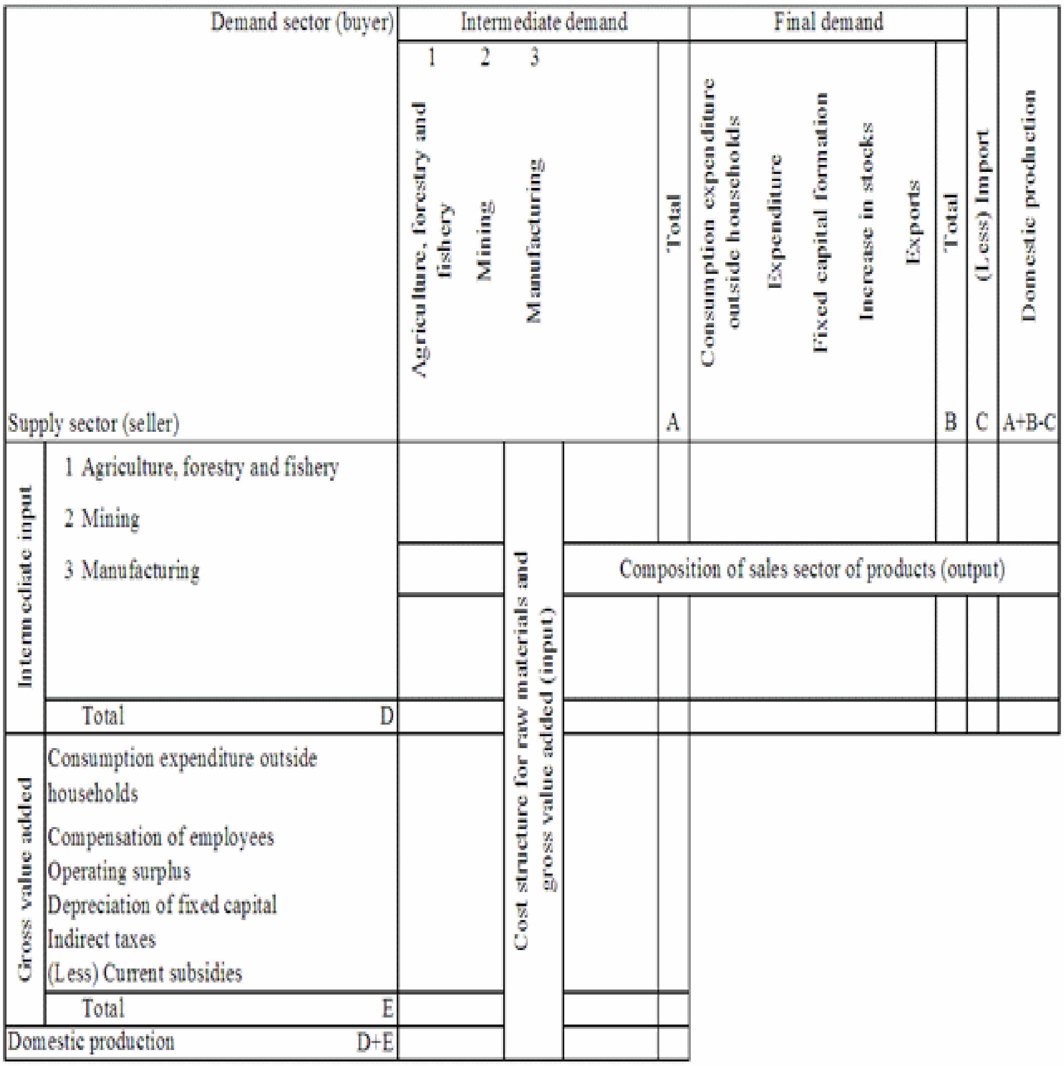

FIGURE 5.1: Image of Input-Output Table

5.0.1 Problems of the current IOT

Recent researchers indicate Current IOT's defects. One of the most significant shortcomings of IOT is the long-term update. Usually, when we do Input-Output Analysis, we fix input coefficients: this indicates that we suppose the structure of the whole industrial transactions and technologies do not change until the following publication of new IOT[42]. The Government of Japan publishes Input-Output Table once every five years [44], hence when we use The Japanese IOT, we suppose the Japanese industrial structure does not change for five years. However, to take the current fluidity of technology and commerce into account, this term to renew the IOT is too long to suppose these structures do not change.

Owing to the compilation method of SNA and IOT, it takes a very long term to renew statistics. As a general rule, the Japanese Government estimates SNA and IOT using various kinds of statistics. Commonly, ministries take two years to collect data and take more than two years for estimation[45]. Although the government updates tables which complement the missing term with a quick estimation method like the RAS method, researchers indicate that these fast estimation methods could not represent the structure of new technologies and industries well[46].

Another problem of the current IOT is the geographic limitation. Recently, due to decentralization progress, the need for input-output analysis between regions is increasing. Nevertheless, the Japanese IOT contains only one sheet, representing the whole economy in Japan. Although prefectures and some large cities compile the regional IOT by themselves, there are very few regional IOT of other small municipalities[47]. Moreover, After the 2011 Tohoku earthquake and tsunami, and during Doshu Special Zone's discussion, the Japanese Government reconfirmed the statistics' necessity, including some administrative districts like inter-regional IOT. However, the Japanese Inter-regional IOT we have divides the whole of Japan into only nine regions[48].

Also, the lack of the standard compilation method of regional IOT is a significant matter of regional statistics. According to the manual[45], the government recommends prefectures and large cities to compile the IOT using the Accumulate Method, which accumulates primary data sources. However, since several municipalities have not estimated several statistics needed to estimate IOT, they make and use distinctive compilation methods. In the case of small cities, they compile regional IOT with the Simple Method that distributes regional IOT of the larger region into smaller by the other statistics like employee number[47]. The regional IOT estimated by the Simple Method usually does not grasp these regions' inter-industrial structure or characteristics. Consequently, this variation of the compilation method of regional IOT makes it challenging to compare regional IOTs between different regions.

As a solution for these problems of IOT, we had developed a new compilation method of the IOT with private business data[35],[49]. TDB is a credit research company that collects financial and business information with questionnaires. TDB renews that information every one to three years, which is shorter than ministries. Also, TDB data includes the statement of accounts(BS, PL) and geographic information. Thereby, we could compile an IOT with these data annually at will for any region or industrial and commodity classification granularity. From now on, we refer to the IOT estimated from this TDB data and our automatic estimation system as TDBIOT. Moreover, we will refer to the existing IOT in SNA as SNAIOT.

Furthermore, the simplicity of the method is also a benefit attributed to our work. The primary obstruction in the compilation of current regional IOT is the time and financial cost due to the complicated method and variation of data types in several statistics to estimate. By contrast, our TDBIOT enables us to compile only with TDB data. Thus, it takes a shorter period and lower labor than SNAIOT. Notably, we made an automatic IOT compiler software with the AADL and FALCONSEED[50]; program language specialized in calculating Exchange Algebra. As a result, we can instantly compile an annual TDBIOT of any region with this system.

5.0.2 Data

Of course, the data TDB collected differs from these of ministries. We should take these characteristics of data into account to create economic statics. Due to its nature as a credit research company, TDB has detailed business data for business companies but few B2C (business-to-consumer) companies and sole proprietors. Moreover, not all data satisfy complete chronological order since the renewal term of each of the data is differ respectively according to their purpose of use.

The total of each companies' profits on TDB data is ¥1,200 trillion, and that of the economic census is ¥1,300 trillion[51]. In other words, the coverage of TDB data compared with economic census measured by the profit overs 90 %.

TDB classifies its data for several categories per business. The details are described on Table.5.1.

TABLE 5.1: Overview of TDB data

Data Name	COSMOS1	COSMOS2	TRD
Data kind	Credit data	Basic profiles of firms.	Trading amount between firms.
Update time	Recorded time	Every year	Every year
Number of firms	73 million firms	145 million firms	5 million firms with traded amount, 70 millions firms with trading partners' name

To compile TDBIOT, we use business trade data from TRD and profit, cost of sales, classification of main business and sub-business from COSMOS2 on the Table.5.1. Since TDB does not collect their Data to estimate IOT, we should note the following tips.

Subject of the survey:
Making SNAIOT, ministries do Input-output structure research[52] into all offices in Japan. On the other hand, TDB researches only head offices since TDB Data includes only head offices, which does not contain the trades between the same firms' offices and does not represent the office's economic activity in different regions to head office.

Price:
Generally, to compile IOT, we differ producer's price table, which counts transport margin on input from transport and commerce industries, and purchaser's price table, which counts transport margin as a price[45]. However, TRD data intermix the producer's and purchaser's price.

Classification:
We use TDB Middle industrial classification[53] to compile our TDBIOT. The industrial department's name on TDB industrial classification represents a tentative department of commodities; thus, these do not represent industries.

5.0.3 Methods

Compiling TDBIOT, we used four procedures; Commodity Classification, Transaction value estimation, Intermediate demand estimation, and IOT estimation. Overviews of these procedures are below.

1. Commodity Classification
   TSCA (Two Side Classification Algorithm) classifies the commodities name written in free text in TRD into the TDB classification.

2. Transaction value estimation
   The TRD contains each companies' transaction amount and items they trade, but the transaction value is only partially available. In Transaction Value Estimation, we estimate the value of each transaction.

   (a) Accuracy improvement by supplementing manufacturing costs
   Before estimating the transaction value, we can improve the IOT's completeness by supplementary data for which no financial information is available and the manufacturing cost is unknown. This paper examines the effect of supplementing the manufacturing cost to estimate the transaction value.

3. Intermediate demand estimation
   This procedure aggregates the estimated transaction values as the input value of each commodity classified by Commodity Classification.

4. IOT estimation
   The data generated from the above three methods are arranged on a matrix to create an IOT.

Commodity classification

There are several kinds of IOT. In this paper, we will compile IOT, which is the matrix of commodities(row) $\times$ commodities(columns) called X table. To create an X table, we should classify and normalize trade data from TRD written in a free format for commodity name to appropriate commodity departments. To classify, we made Two Side Classification Algorithm (TSCA), which enables us to classify free format commodity name to particular classification automatically using data about contractors' main and sub business. This paper will classify free format commodities to 91 departments of TDB industrial classification as a tentative commodity department alternative.

TSCA is constituted with following two heuristics.

a) We create a quotient set if the names of elements in the traded products set are the same, and we associate elements of each quotient class with the elements which appear most frequently in the main or sub-business of a contractor of its product (we will call this set "most frequent appearance department set"). If there are some most frequent appearance business of the contractor producing that product, we chose one that appeared more frequently on the main business than on sub to associate. If the frequency on the main is the same, we selected the early one in the TDB industrial classification code since TDB usually set the earlier one on the upper stream in industrial flow.

b) We mapped the cartesian product made at the procedure (a) by the most frequent appearance department set elements.

The algorithm of TSCA we described above is shown below. This section will describe all algorithms as a set since Exchange Algebra written by AADL and set is compatible.

We defined sets made by TDB data as bellow.

We defined a set representing TDB data as follows.

Transaction count: $I = \{1, 2, ...\}$

Industrial 2-digit classification : $Ind_2 = \{d_i | i = 1, 2....\}$

Commodities :$G = \{\alpha_k | K = 1, 2....\}$

TRD: $TRD = \{< x, y, z, i > | x, y \in Ind_2, z \in G, i \in I\} \subseteq Ind_2 \times Ind_2 \times G \times I$

At this point, we defined the $Selection funnction : G \to I \times Ind_2 \times G$ as

$$\bigcup_{\alpha \in G} Selection[\alpha]$$

Its configuration is as follows.
Exchange pair(an combination of main or sub business and commodity):

$$\begin{aligned} Ex[\alpha] = \{< q, \alpha, i > \quad | \quad & p, q \in Ind_2, i \in I \\ , \quad & (< q, p, \alpha, i > \in TRD \vee < p, q, \alpha, i > \in TRD)\} \end{aligned}$$

Count (counting up the number of commodities and their corresponding industries):

$$\begin{aligned} Count[\alpha] = \{< n, q, \alpha > \quad | \quad & < q, \alpha, i > \in Ex[\alpha] \\ , \quad & Ex[x][y] = \{< y, x, i > | i \in I, < y, x, i > \in Ex[x]\} \\ , \quad & n = |Ex[\alpha][q]|\} \end{aligned}$$

Counted number:

$$CountedNumber[\alpha] = \{x | < x, q, \alpha > \in Count[\alpha]\}$$

First and second count number:

$$\begin{aligned} FirstSecondNumber[\alpha] = \{< x, y > \quad | \quad & x, y \in Count[\alpha] \\ , \quad & x > y \\ , \quad & (\forall z)(z \in Count[\alpha] \wedge y > z)\} \end{aligned}$$

As a result we got follows.

$$\begin{aligned} Selection[\alpha] = \{< x, q, \alpha >, < y, p, \alpha > \quad | \quad & < x, q, \alpha >, < y, p, \alpha > \in ContedNumber[\alpha] \\ , \quad & < x, y > = FirstSecopndNumber[\alpha]\} \end{aligned}$$

Example:
When,

$$TRD = \begin{cases} < x_1, x_2, \alpha_1, 1 > \\ < x_2, x_3, \alpha_2, 2 > \\ < x_1, x_2, \alpha_1, 3 > \\ < x_2, x_3, \alpha_1, 4 > \end{cases}$$

the transformation at each stage will be

$$\bigcup_{\alpha \in G} Ex[\alpha] = \begin{cases} < x_1, \alpha_1, 1 > \\ < x_2, \alpha_1, 2 > \\ < x_2, \alpha_2, 3 > \\ < x_3, \alpha_2, 4 > \\ < x_1, \alpha_1, 5 > \\ < x_2, \alpha_1, 6 > \\ < x_2, \alpha_1, 7 > \\ < x_3, \alpha_1, 8 > \end{cases}$$

$$Ex[\alpha_1] = \begin{cases} < x_1, \alpha_1, 1 > \\ < x_2, \alpha_1, 2 > \\ < x_1, \alpha_1, 3 > \\ < x_2, \alpha_1, 4 > \\ < x_2, \alpha_1, 5 > \\ < x_3, \alpha_1, 6 > \end{cases}$$

Transaction value estimation

In the transaction value estimation, we estimate transaction values between firms based on their financial statement. In this paper, we use the following three methods to evaluate the transaction value:

- Sales proportional division: we divide sales of the ordering companies proportionally based on the manufacturing cost of the ordering companies.
- Manufacturing cost proportional division: We divide manufacturing costs of the ordering companies proportionally based on sales of the ordering companies.
- Weighted estimation: we divide estimation values of "sales proportional division " and "manufacturing cost proportional division " and optimize according to the company's sales scale and industrial classification.

We defined the algorithm for transaction value estimation as follows.
Orderers:

$$O = \{O_i | i = 1, 2, ...n\}$$

Cost of sales on each Orderers:

$$C = \{c_i | i = 1, 2, ...n\}$$

Recievers:

$$R = \{r_j | j\, 1, 2, ...m\}$$

Profit:

$$P = \{p_j | j = 1, 2, ...m\}$$

Exchange Set

$$Ex = \{< O_i, r_j >\} \subseteq O \times R$$

Divided production cost:

$$\begin{aligned} DivC[< o_i, r_j >] = \{< o_i, r_j, dc_k > \quad | & \quad < o_i, r_j > \in Ex \\ , & \quad p_j, p_s \in P \\ , & \quad c_i \in C \\ , & \quad dc_k = c_i * (p_j / \sum_{p_s \in P} c_s) \\ , & \quad < o_s, r_j > \in Ex\} \end{aligned}$$

Transaction value estimated with profit:

$$\begin{aligned} TVp[< o_i, r_j >] = \{< o_i, r_j, pv_k > \quad | & \quad < o_i, r_j > \in Ex \\ , & \quad < o_i, r_j, dc_k > \in DivC[< o_i, r_j >] \\ , & \quad pv_k = p_j * (dc_k / \sum_{< o_s, r_j, dc_q > \in A} dc_q) \\ , & \quad A = \bigcup_{< o_i, r_j > \in Ex} DivC[< o_i, r_j >] \\ , & \quad < o_s, r_j > \in Ex, r_j \in R \\ , & \quad o_s \in O\} \end{aligned}$$

Manufacturing cost proportional division

For each transaction in the TRD, the selling firm's manufacturing cost is prorated based on the divided sales amount of the buying firm. In this case, the divided sales amount means the value obtained by dividing the purchasing firm's sales amount by the manufacturing cost of the selling firm for each firm. Since it is not possible from an accounting point of view for the purchasing company's transaction amount to exceed the sales amount of the selling company, we set the upper limit of the transaction amount as the divided sales amount of the purchasing company.
Orders:

$$O = \{o_i | i = 1, 2, 3, ...n_1\}$$

Cost:

$$C = \{c_i | i = 1, 2, 3, ...n_1\}$$

Receivers:

$$R = \{r_j | j = 1, 2, 3, ...n_2\}$$

Profit:

$$P = \{p_j | j = 1, 2, 3, ...n_2\}$$

Exchange pair:

$$Ex = \{< o_i, r_j >\} \subseteq O \times R$$

Divided sales:

$$\begin{aligned} DivP[< o_i, r_j >] = \{< o_i, r_j, dp_k > \quad | \quad & < o_i, r_j > \in Ex \\ , \quad & c_i, c_p \in C \\ , \quad & p_j \in P \\ , \quad & dp_k = p_j * (c_i / \sum_{c_p \in C} c_p) \\ , \quad & < o_p, r_j > \in Ex\} \end{aligned}$$

Estimated Transaction value with cost:

$$\begin{aligned} TVc[< o_i, r_j >] = \{< o_i, r_j, cv_k > \quad | \quad & < o_i, r_j > \in Ex \\ , \quad & < o_i, r_j, dp_k > \in DivP[< o_i, r_j >] \\ , \quad & cv_k = c_i * (dp_k / \sum_{< o_i, r_s, dp_q > \in A} dp_q) \\ , \quad & A = \bigcup_{< o_i, r_j > \in Ex} DivP[< o_i, r_j >] \\ , \quad & < o_i, r_s > \in Ex \\ , \quad & r_s \in R \\ , \quad & o_i \in O\} \end{aligned}$$

Weighted estimation

We used the value estimated by sales proportional division as the transaction value, but it turned out that the estimated value deviated greatly depending on the sales scale and the industry classification. Therefore, as a measure to adjust these discrepancies, we estimated the transaction value by adding optimized weights to the values estimated by divided sales and manufacturing costs for each industry and sales scale category of firms. In the weighted estimation, the transaction value is estimated by weighting each value estimated by "sales proportional division" and "manufacturing cost proportional division" by the combination of sales scale and industry (TDB industry classification major classification).

Define Estimated value with profit : $X_1 = \{x_{1j} | j = 1, 2, 3, ...n\}$,and Estimated value with cost : $X_2 = \{x_{2j} | j = 1, 2, 3, ...n\}$.

Then, weighted estimation value,W is,
$W = \{w_j | w_j = \alpha_i X_{1j} + \beta_i X_{2j}\}$.
In this case, we α_i, β_i minimize the square of the difference, $\sum_{j=1}^{n}(r_j - w_j)^2$, between the actual transaction value $R = \{r_j | j = 1, 2, 3, ...n\}$ and W for each combination of sales scale and industry classification.

The combination of industries consisted of the TDB industrial classification major categories (A: agriculture, B: forestry/hunting, C: fishing, D: mining, E: construction, F: manufacturing, G: wholesale/retail/restaurants, H: finance/insurance, I: transportation/telecommunications, J: electricity/gas/water/heat supply, K: services, L: public services), and the scale of sales. The scale of sales was divided based on the influence of the sales scale by decision tree analysis.

The decision tree analysis's objective variable is the categorization of R-W obtained by minimizing $\sum_{j=1}^{n}(r_j - w_j)^2$ as follows, and the explanatory variable is the sales amount of the ordering company.

We divided two cases, $(r_j - w_j) \geq 0$ and $(r_j - w_j) \leq 0$, into quartiles. In the case $(r_j - w_j) \geq 0$, We labeled difference as BIGGER, Bigger, Bigger, Pcorrect, starting from the category by scale with the most significant difference between r_j and w_j.

Also, in the case $(r_j - w_j) \leq 0$, we labeled difference as SMALLER, Smaller, Smaller, Mcorrect by scale starting from the category with the most significant difference between r_j and w_j.In both cases, we labeled the In each case, Pcorrect and Mcorrect have the slightest difference between r_j and w_j, which means that the estimated value is close to the actual value.

The decision tree analysis algorithm is CART (Classification and Regression Trees) in the python library scikit-learn[2]. We used the Gini coefficient as an indicator and set the depth of the search to 4. For each category, the number of transactions must be at least two for optimization, and for the categories below that, the optimal value for the entire data, $\alpha = 0.04952, \beta = 0.1302$, was adopted.

Through the decision tree analysis, the four categories of sales amount of 67450,0 (unit: 1 million yen) or more, 67450.0 or less, 374.5 or more, 374.5 or less, 64.5 or more, and 64.5 or less are the categories that strongly affect the result of $(R - W)^2$ in the transaction data of 2014. The α_i, β_i for each industry and sales category in the 2014 transaction data are shown in Table 5.2.

[2]scikit-learn: http://scikit-learn.org/stable/

TABLE 5.2: Weighted estimation with 2014 data.α, β

Sales scale	Classification	α	β
Whole data	Whole data	0.04952	0.1302
6475.0 ∼	A	1.205e-14	0.2366
6475.0 ∼	B	0.04952	0.1302
6475.0 ∼	C	0.04952	0.1302
6475.0 ∼	D	0.1435	0.04578
6475.0 ∼	E	0.1781	0.07848
6475.0 ∼	F	0.04228	0.1655
6475.0 ∼	G	0.03108	0.06387
6475.0 ∼	H	0.1190	1.377e-16
6475.0 ∼	I	0.07438	0.007304
6475.0 ∼	J	0.08080	0.06780
6475.0 ∼	K	0.07555	0.3082
6475.0 ∼	L	0.01296	0.2976
374.5 ∼ 6750.0	A	1.887e-18	0.1296
374.5 ∼ 6750.0	B	0.04952	0.1302
374.5 ∼ 6750.0	C	0.04952	0.1302
374.5 ∼ 6750.0	D	0.06025	0.05749
374.5 ∼ 6750.0	E	0.09050	0.09056
374.5 ∼ 6750.0	F	0.1133	0.1036
374.5 ∼ 6750.0	G	0.07186	0.06738
374.5 ∼ 6750.0	H	0.1730	0.1850
374.5 ∼ 6750.0	I	0.1168	0.1037
374.5 ∼ 6750.0	J	0.1218	0.1120
374.5 ∼ 6750.0	K	0.02249	0.02159
374.5 ∼ 6750.0	L	0.1002	0.07560
64.5 ∼ 374.5	A	0.2200	0.06743
64.5 ∼ 374.5	B	0.04952	0.1302
64.5 ∼ 374.5	C	0.04952	0.1302
64.5 ∼ 374.5	D	0.5419	0.1777
64.5 ∼ 374.5	E	0.07931	0.1974
64.5 ∼ 374.5	F	0.1472	0.1254
64.5 ∼ 374.5	G	0.08341	0.1289
64.5 ∼ 374.5	H	0.03252	0.4776
64.5 ∼ 374.5	I	0.1431	0.1236
64.5 ∼ 374.5	J	0.17611	0.1068
64.5 ∼ 374.5	K	0.04952	0.1302
64.5 ∼ 374.5	L	0.1165	0.09835
0 ∼ 64.5	A	0.04952	0.1302
0 ∼ 64.5	B	0.04952	0.1302
0 ∼ 64.5	C	0.04952	0.1302
0 ∼ 64.5	D	0.04952	0.1302
0 ∼ 64.5	E	1.720e-13	0.3648
0 ∼ 64.5	F	0.1194	0.2533
0 ∼ 64.5	G	0.1543	0.1680
0 ∼ 64.5	H	0.04952	0.1302
0 ∼ 64.5	I	0.2852	0.2647

$0 \sim 64.5$	J	1.686e-18	0.2399
$0 \sim 64.5$	K	0.04952	0.1302
$0 \sim 64.5$	L	0.09017	0.2099

5.0.4 Intermediate demand estimation

At this process, we estimated each department's demanded value on the supply side(columns) on IOT. In general, every customer should also be a producer at another transaction. In other words, we could regard some commodity ordered by some company as demanded commodity for productions of that company. We estimated demand for each product by dividing the transaction value calculated in the previous section.

We divided the traded amount to the final demand when a customer of that trade can be consumers; department Education, Laundry beauty and bath service, Amusement, and recreation service, Miscellaneous personal service, Hotels, and other lodgings, and Eating and drinking restaurants.

In some cases, we regard some data as lacking; for instance, no one buys goods from a company that bought something. In these cases, we estimated the demanded amount of that trade using the average ratio of the same combination of traded product and industrial department of customer.

Our algorithm of demand estimation is as below.

Goods:

$$G = \{g_i | i = 1, 2, ...n_1\} = \bigcup_{\alpha \in G} Selection[\alpha], n_1 \in N$$

Orders:

$$O = \{o_j | j = 1, 2, ...n_2\} n_2 \in N$$

Receivings:

$$Re = \{r_k | k = 1, 2, ...n_3\}, n_3 \in N$$

Exchange set

$$Ex = \{< o_j, r_k > | o_j \in O, r_k \in Re, \} \subseteq O \times Re$$

Divided Profits:

$$DivP[< o_j, r_k >] = \{divP_l | o_j \in O, r_k \in R, < o_j, divC_i, r_j, divP_l > \in SelectionProfit\}$$

$$DivP = \{divP_l | divP_l \in \bigcup_{<o_j, r_k> \in Ex} DivP[< o_j, r_k >]\}$$

Set of all Companies:

$$COM = \{co_s | s = 1, 2, ...n_4\} n_4 \in N, G, O \subseteq COM$$

B2C Companies:

$$FC \subseteq COM$$

Buisinsess:

$$\begin{aligned} B = \{< o_j, r_k, g_i, divp_l > \quad | & \quad < o_i, r_k > \in Ex \\ , & \quad g_i \in G \\ , & \quad divp_l \in DivP\} \subseteq Ex \times G \times DivP \end{aligned}$$

Departments of Industry:

$$Ind_2 = \{I_z | z = 1, 2, ...n_1\}$$

Departments of company:

$$DC = \{< Co_s, I_s > | Co_s \in COM, I_s \in Ind_2\} \subseteq COM\ timesInd_2$$

Sellers on the first transaction (Produce):

$$\begin{aligned} P[g_i] = \{p_t \quad | & \quad p_t \in Re \\ , & \quad g_i \in G \\ , & \quad < o_i, r_k > \in Ex \\ , & \quad o_j \in O \\ , & \quad divP_l \in divP\} \end{aligned}$$

Buyers on the first transaction (Demand):

$$\begin{aligned} D[g_i] = \{d_u \quad | & \quad d_u \in O \\ , & \quad g_i \in G \\ , & \quad p_t \in P[g_i] \\ , & \quad < d_u, p_t, g_i, , divP_l > \in B\} \end{aligned}$$

The demanded amount at Secondary Production):

$$\begin{aligned} S[g_i] = \{< g_i, Sg_v, SumSdivP_v > \quad | & \quad g_i, Sg_v \in G \\ , & \quad SdivP_v \in DivP \\ , & \quad d_u \in D[g_i] \\ , & \quad o_j \in O \\ , & \quad < o_j, d_u, Sg_v, SdivP_v > \in B \\ \vee & \quad (< g_i, I_z, Sg_v, SdivP_v > \in averageP[< g_i, I_z, Sg_v >] \\ & \quad \wedge < d_u, I_z > \in DC)\} \end{aligned}$$

$$averageP[< g_i, I_z, Sg_v >] = \{< g_i, Co_s, I_w, Sg_v, aveP[< g_i, I_w, Sg_v >] > \mid g_i, Sg_v \in G, < Co_s, I_w > \in DC\}$$

$$aveP[< g_i, I_w, Sg_v >] = \{ap \mid d_w \in D[g_i] \; , < d_w, I_w > \in DC \; , ap = \sum_{isdP_\alpha \in IndSdivP[<gi,I_w,Sg_v>]} isdP_\alpha \; / |IndSdivP[< gi,I_w, Sg_v >]|\}$$

$$IndSdivP[< gi,I_w, Sg_v >] = \{isdP_\alpha \mid isdP_\alpha \in DivP \; , o_j \in O \; , r_k \in R \; , g_i \in G \; , divp_l \in DivP \; , < o_j, r_k, g_i, divP_l > \in B \; , < o_j, I_w > \in DC \; , O_j \in D[g_i] \; , < o'_j, o_j, Sg_\alpha, SdivP_\alpha > \in B \; , isDP_\alpha = SdivP_\alpha\}$$

$$sumSdivP_v = \sum_{sdivP_w \in SDivP[Sg_w} sdivP_w\}$$

$$SDivP[Sg_w] = \{SdivP_w \mid Sg_w \in G \; , SdivP_w \in DivP \; , o_j|inO \; , d_u \in D[g_i] \; , < o_j, d_u, Sg_w, SdivP_w > \in B \; , d_u \notin FC \vee SdivP_w = 0\}$$

Total productions:

$$SumP[g_i] = \{< g_i, sumP_i > \mid g_i \in G \; , DivP[g_i] = \{divP_l| < o_i, r_k, g_i, divP_l > \in B[g_i]\} \; , sumP_i = \sum_{divP_l \in DivP[g_i]} \}$$

Total Secondary Productions):

$$TS[g_i] = \sum_{sumSdivP_y \in SumSDivP[g_i]} sumSdivP_y$$

$$SumSDivP[g_i] = \{sumSdivP_y| < g_i, Sg_v, sumSdivP_y > \in S[g_i], g_i, Sg_v \in G\}$$

Secondary Production Rate:

$$RS[g_i] = \{< g_i, Sg_z, RS_z > \quad | \quad g_i, Sg_z \in G \quad , \quad RS_z = sumdivP_z / TS[g_i] \quad , 7 < g_i, Sg_z, sumSdivP_x > \in S[g_i]\}$$

Demand for each Secondary products:

$$\begin{aligned} DS[g_i] = \{< g_i, Sg_x, dS_x > \quad | \quad & g_i, Sg_x \in G \\ , \quad & < g_i, Sg_x, sumSdivP_x > \in S[g_i] \\ , \quad & < g_i, sumP_i > \in SumP \\ , \quad & RS_x = SumdivP_x / TS[g_i] \\ , \quad & dS_x = RS_x * sumP_i\} \end{aligned}$$

Then, the demand value is defined as follows.

$$SelectionDemand = \bigcup_{g_i \in g} DS[g_i]$$

5.1 Algorithm validation

To validate the accuracy of TDBIOT compiled with the procedures above, firstly, we verified each of the procedures and, secondary, we compare the TDBIOT with SNAIOT.

5.1.1 Classification

In this subsection, we will inspect the reliability of the TSCA as a method of commodity classification. We checked and counted ineligible products to classify to that department by eyes. The goods that appeared only once and goods which we could not judge, for instance, we exclude "goods," "products," and "other," from the check.

In Table.5.3, we show the top 10 departments with high ineligible rates.

Although the average rate of included ineligible goods are 7%, departments include ineligible goods with a remarkably high rate. Since the rate looks scattering and the characteristics might differ with each of the departments, we dissected characteristics and improved each of the commodities, respectively.

TABLE 5.3: top 10 departments with high ineligible rate.

CLASSIFICATION	INELIGIBLE RATE
WEAPON MANUFACTURE	33%
INSURANCE AGENTS BROKERS AND SERVICES	21%
INVESTMENT COMPANIES	20%
TRANSPORTATION EQUIPMENT MANUFACTURING	20%
GENERAL CONSTRUCTION	19%
AUXILIARY FINANCIAL BUSINESSES	19%
SECURITIES AND COMMODITY EXCHANGE COMPANIES	18%
BANKS AND TRUST COMPANIES	16%
WAREHOUSING INDUSTRY	15%
RAILWAY COMPANIES	15%
AVERAGE	7%

Weapon manufacture:

There are only four transaction records related to weapon manufacturing in TRD. In general, WEAPON MANUFACTURE is the concealing industry, and SNAIOT estimates this department with other statistics. Therefore, to improve estimation, it is better to use other statistics or exclude this department from TDBIOT.

Problems among whole goods:

Since we only have the name of goods to classify goods to the industrial department, using TSCA, there are many cases that we can not judge which is the goods' department, manufacturing, commerce, or service. For instance, when the commodity name is "screw," we can not judge its department; it can be METAL PRODUCTS or COMMERCE. If some manufacturing and retail commodities data show their name in the same spell, TSCA classifies those commodities to the department that appeared more frequently. Thus, it should divide the count of manufacturing and commerce to improve the ineligible rate.

Although we finished analyzing all commodities, we described only two examples due to space limitations in this paper. Therefore, applying these results to classification is left as our future work.

5.2 Estimated volume of transactions

The TRD contains several records of actual transaction amounts. We verified the accuracy of the estimated transaction values by comparing them with the actual values. Table. 5.4 shows the ratio and the correlation coefficient of the total estimated value to the actual value. Fig.5.2, 5.3, and 5.4 show the log plots.

In Fig.5.2, 5.3, and 5.4, the actual values distribute on linear lines since TDB truncated the actual values after the decimal point. The data are for the fiscal year 2014, the unit is ¥1 million, and the red line in the figure is the 45-degree line.

Table. 5.4 shows that a factor of 4 relative overestimates the transaction value to the actual value in the sales and manufacturing cost estimation. However, the overall overestimation is improved to about 0.9 by the weighted estimation.

TABLE 5.4: Ratio and correlation coefficient.

Method	Estimated/Real	correlation
Transaction value estimation	4.380	0.6740
Manufacturing cost proportional division	4.071	0.6625
Weighted estimation	0.9025	0.7368

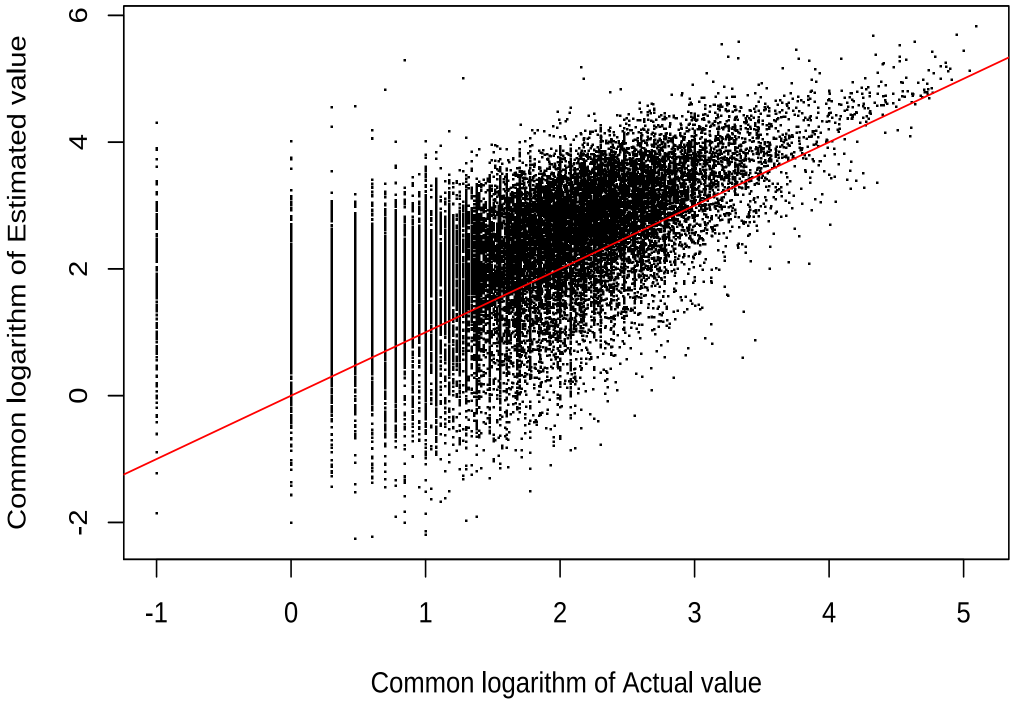

FIGURE 5.2: Transaction value estimation

Fig.5.2, 5.3, and 5.4 show that the weighted estimates are closer to the 45-degree line. However, it is difficult to directly judge the estimation ability of the weighted estimation based on this verification because we optimized the weighted estimation for the actual transaction value used in the verification in this section. For this reason, this paper excludes weighted estimation as a subject of the following verification.

5.3 Comparison with SNAIOT

We inspected the 2 of 3 procedures above sections; product classification and estimation of traded amount. To inspect the rest, estimation of the intermediate demand, we will compare TDBIOT with the SNAIOT using several indexes representing the difference between IOT.

We used "2011 confirmed report of X producer's price Table"[54] as comparison to TDBIOT of same years, and convert these Tables for comparison.

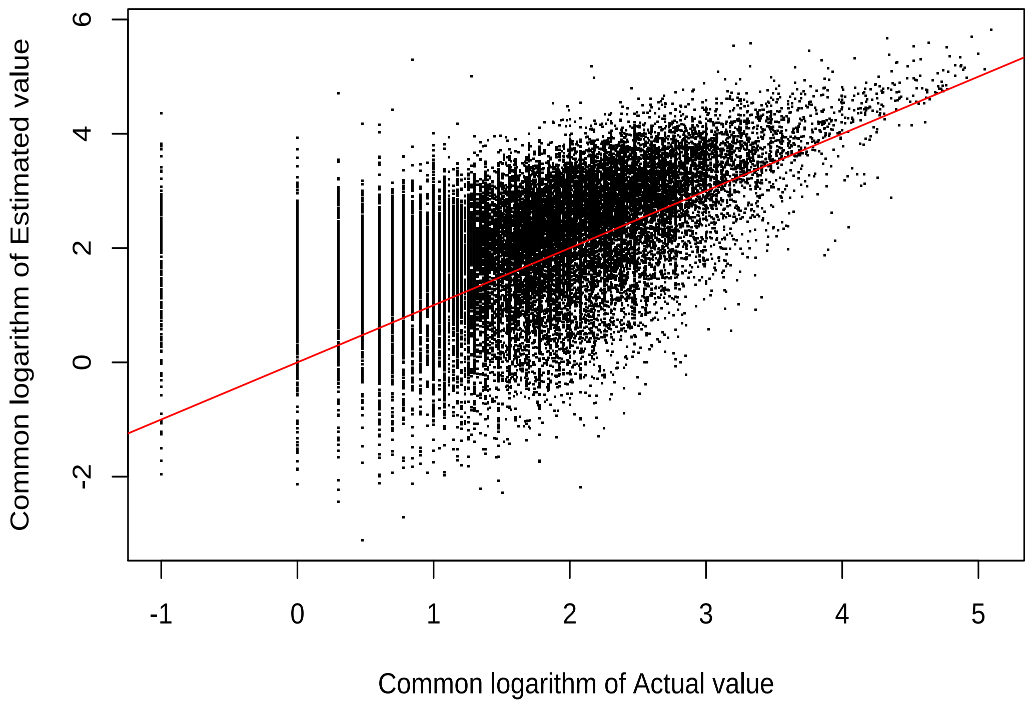

FIGURE 5.3: Manufacturing cost proportional division

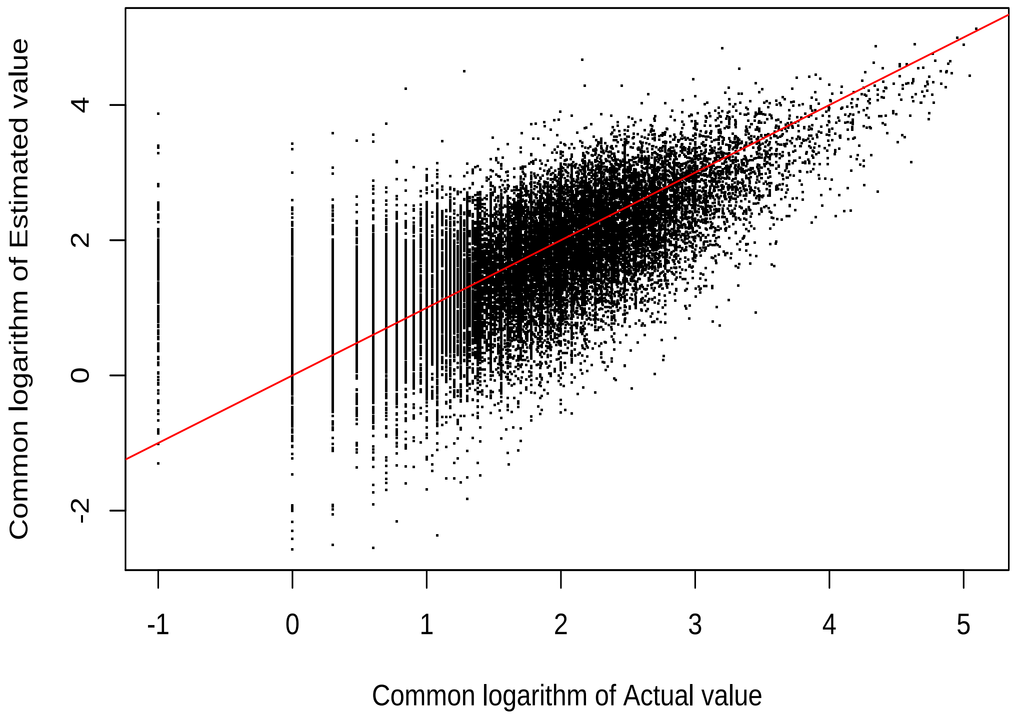

FIGURE 5.4: Weighted estimation

5.3.1 Conversion of classification

Since SNAIOT and TDBIOT's departments are not the same, we converted these tables to a comparable form of a table (hereafter, we refer to as "comparison table").

We compiled a comparison table with 47 departments according to the correspondence tables[55][56][57]. Tables 5.11 and 5.12 are comparative tables for SNAIOT and TDBIOT, respectively. Converting to comparison tables, we exclude the department that does not have a corresponding department to another table or that we need to divide to correspond. Also, if one table's department includes several concepts of another, we combined these departments.

The compilation method of SNA allows the input value on SNAIOT to record minus. By contrast, TDBIOT does not take a minus value. For comparison, according to Ohomori(2012)[58], SNA treats a minus value of SNAIOT as zero. In the rest of this section, we call the comparison table converted from SNAIOT as a SNAIOT, and the comparison table converted from TDBIOT as a TDBIOT.

5.3.2 Comparison of the intermediate demand total

Intermediate demand totals of SNAIOT and TDBIOT are represented on Table.5.5.

TABLE 5.5: Total number of recorded companies and Intermediate demand totals of TDBIOT and SNAIOT

	SNA	TDB	TDB/SNA
Intermediate demand total (¥trillion)	334	433	1.296
Number of recorded companies (million)	412 [59]	20	0.0485
A rough estimate of overabundance of TDBIOT	3.439		

This result shows that the intermediate demand total of the TDBIOT surplus that of SNAIOT. In contrast, the number of recorded companies of SNAIOT is about 20 times as many as that of TDBIOT. In other words, fewer companies of TDBIOT make more trade than that of SNAIOT, which means TDBIOT overestimated the intermediate demand total. However, since the companies recorded in TDB data inclines to be large corporations, the overestimated ratio of TDBIOT to SNAIOT is not 20 times. In Japan, companies with more than 1,000 employers are about 7.59% to whole companies,whereas, about 59% of profit are made by these companies[60]. If we apply this ratio, the rough estimation of the overabundance ratio of TDBIOT is about 3.439.

Transaction value estimation

The intermediate demand total is the sum of the traded amount. Thus, the overestimation of the traded amount affects the intermediate demand total overabundance. The following sections will analyze the causes of overestimation in transaction value estimation.

Use of sales data

Generally, the sum of ordered volume, in other words, shipment, equals the sum of sales. Therefore, the total shipment of whole companies equals that of total sales. In IOT's row, a value on an intersection with some department on column represents a sales amount from the row to the column. Thus, the sum of the intermediate demand, final demand, and export equal the total of sales.

Sales include transactions in the final consumption sector, such as general government expenditure and private consumption expenditure, and export transactions. However, on the other hand, the TRD includes only transactions among domestic firms. Thus, dividing sales on TRD will overestimate the traded amount in proportion to the final demand sector and exports.

In the SNA table before adjustment for comparison, the value of domestic production minus the deduction of imports is about ¥1023 trillion, and the ratio of this value to the SNA total domestic production of ¥463 trillion (SNA total domestic production/(SNA gross domestic production minus deductions)) is 0.4524. However, if we divide the TDB table by the same ratio as in the SNA table (3.439*0.4529), the excess in the TDB table is estimated to be 1.557. Therefore, this adjustment will improve the excess in the TDB table.

According to the estimation above, to deduct the value of final consumption and export from the profit and use that deducted profit as a distribution source might be a simple solution. However, TDB does not have an export and final consumption ratio to the total sales.

Also, using the cost of sales as a source might be another solution. Commonly, on IOT, the following equation holds; "Total Intermediate demand = Total profit -(Total final demand + Export) = Total cost of sales - Import." If we use the total cost of sales as a source, we should know the ratio of import, whereas, in the case of the cost of sales, we only need a ratio of the final demand besides the ratio of the export. This is consistent with the fact that, in the verification of the transaction value estimation in Table. 5.4, the value estimated by the Manufacturing cost proportional division was slightly better than that estimated by prorating sales.

Intermediate demand estimation

Final demand sectors

We regarded the case that some clients are not ordered by another as the lack of the data, and complemented received orders. However, we overlooked cases where these clients bought something and treated it as stock, which they consumed and did not produce something or invested in fixed equipment. In other words, we added up consumption expenditures outside households, expenditure, and increase in stocks, on intermediate demand by mistake. Since sales include these departments, we wrongly added up these sections two times.

5.3.3 Comparison of industrial structures

Although the overabundance of the TDBIOT's total value is the problem, the usage of the IOT is not only a total value but also its representation of the input structures. Thus, if this overabundance is affected equally among all Table levels, the merit of the IOT is verified. Therefore, we will inspect the accuracy of represented structure on the TDBIOT in the following sections.

To compare the industrial structure of the TDBIOT with SNAIOT, we compare the input coefficients of both Tables. Input coefficients represent the minimum unit of the materials required to make some goods on the columns of IOT. An input coefficient of an intersection of row x and column y;$a_{x,y}$, is defined as bellow[45].

$$a_{x,y} = \frac{\text{Value on the intersection of row x and column y}}{\text{sum of row x}}$$

There are some indexes representing similarity between IOT. Euclidean distance[61]and Cosine distance[62] are usually used. In addition, we used also Spearman's rank correlation coefficient as a value of the similarity.

The definitions of these indexes are below. a^T, a^S are representing input coefficients of TDBIOT and SNAIOT respectively, and a_i means the i th input coefficient counting intersection of raw 1 and column 1 to raw n and column n.

Spearman's rank correlation coefficient

$$S_{a^Ta^T} = \sum a_i^{T^2} - \frac{(\sum a_i^T)^2}{n}, S_{a^Sa^S} = \sum a_i^{S^2} - \frac{(\sum a_i^S)^2}{n},$$

$$S_{a^Ta^S} = \sum a_i^T a_i^S - \frac{(\sum a_i^T \cdot \sum a_i^S)}{n},$$

$$r_{a^Ta^S} = \frac{S_{a^Ta^S}}{\sqrt{S_{a^Ta^T} \cdot S_{a^Sa^S}}}$$

Euclidean distance

Euclidean distance is the leveled difference between input coefficients and take 1 at maximum difference and 0 at minimum.

$$d(a_i^T, a_i^S) = \frac{\sqrt{\sum_{i=1}^n (a_i^T - a_i^S)^2}}{n}$$

Cosine distance

To grasp the characteristics of the departments of IOT, we measured cosine distance of the Column ratio: W_j, and row ratio : W_i. Row ratio is the ratio of sum of the values of a raw to the total of intermediate demand, and column ratio means ratio of the sum of the values of a column to a total of intermediate demand.

$$W_i = \frac{\sum_i^n x_{ij}}{\sum_i^n \sum_j^n x_{ij}}, W_i = \frac{\sum_j^n x_{ij}}{\sum_i^n \sum_j^n x_{ij}}$$

Thus cosine distance between TDBIOT and SNAIOT ;$cos(W_k^T, W_k^S)$is,

$$cos(W_k^T, W_k^S) = \frac{\sum_k^n (W_k^S \cdot W_k^S)}{\sqrt{\sum_k^n W_k^{T^2}} \cdot \sqrt{\sum_k^n W_k^{T^2}}}, k = i \vee j$$

Cosine distance takes a value between 0 to 1, and takes 1 when two tables do not differ.

Results and review of the inspection of the industrial structure

The values of these indexes are on Table.5.6.

To understand the meaning of the values, we show the indexes between SNAIOT of 2011 and that of 2005 in parenthesis as a bench mark.

Fig. 5.5 is the plot of the input coefficients of TDBIOT and SNAIOT.

Using these values, we will analyze the distortion of represented industrial structure on the TDBIOT.

TABLE 5.6: Validation of the industrial structure (Index between 2011SNA and 2005SNA)

Spearman's rank correlation coefficient($r_{a^T a^S}$)	0.4633(0.9016083)
Euclidean distance($d(a_i^T, a_i^S)$)	0.118(0.0185)
Cosine distance of row ratio ($cos(W_i^T, W_i^S)$)	0.5563(0.8910)
Cosine distance of column ratio ($cos(W_j^T, W_j^S)$)	0.6077(0.9788)

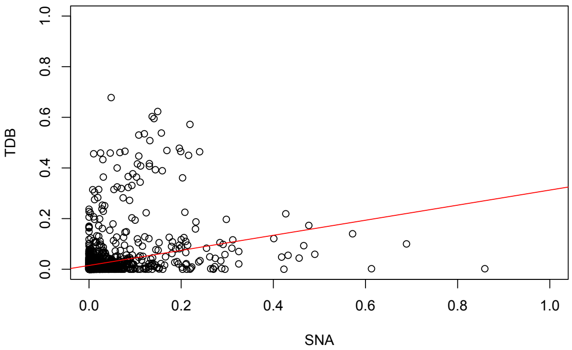

FIGURE 5.5: Plot of the input coefficients of TDBIOT and SNAIOT

Distortion on each of intersections

Although euclidean distance suggests minor differences in most cases, seeing Fig.5.5, some points vastly differ. Regarding the data of SNA as a population, more and more TDB data is collected, more and more TDBIOT asymptote to the SNAIOT. In other words, the separation between tables means the insufficiency of TDB data. Input coefficients show that the ratio of input amounts to the total. Thus high values of some departments on TDBIOT represent that TDB data is partial toward these departments. In contrast, some departments' low values on TDBIOT represent that TDB does not collect data of these departments enough.

Table. 5.7, the intersection that TDBIOT has a higher input coefficient than that of SNA, and the intersection of reverse order is shown.

All contractors of the top 19 intersections that TDBIOT is bigger than SNAIOT are commerce, representing TDBIOT strongly lean to commerce. The reason for this distortion might be the ratio of the final demand to the total profit of commerce. In the above section, we found that we wrongly added the final demand to the intermediate demand section because of the intermediate demand estimation. Naturally, this effect is more significant on the department with a higher final demand ratio. In general, the final demand ratio is high in commerce, construction, and service. Thus because of these high ratios of final demand, commerce is overestimated.

Similarly, according to the bottom of the list, on TDBIOT, the department of energy infrastructures like Crude petroleum and natural gas, Petroleum refinery products, Gas and heat supply, and Electricity are low on the TDBIOT. Besides, regulated industries like Broadcasting are also low. These results are consistent with the fact that TDB is the credit research company, and the demand for information about the companies regarded as commonly credible, like regulated industries or infrastructures, is low.

Similarly, the intersections between the same department like "metals-metals" and "transportation equipment-transportation equipment" are also low because TDB data is not office-based but head office-based. Hence, TDB could not collect the trade data between the same department of the same companies.

If more and more TDB collect data more and more TDBIOT asymptote to the SNAIOT, the difference between the raw or column ratio of SNAIOT and TDBIOT represents the department's distortion on the TDBIOT. To extract the distorted departments, we calculated the differences between $W_k^T and W_k^S$ on Table.5.8.

On both columns and rows, the overabundance of TDNIOT is large in COMMERCE and CONSTRUCTION. These results show that the overabundance of TDBIOT tends to be partial to the department has a high final demand ratio.

Similarly, The department that is not collected enough is written in the Table.5.9.

Distortion on the row and column

According to the cosine distance of row and column ratios, rows are relatively more distorted than columns. One hypothesis is the effect of the distribution of trade type between contractors and clients of each department. For example, suppose the collection ratio of some departments is high. Companies in that department tend to order a little and receive orders like service companies or real estate agencies. In that case, that trade affects rows stronger than columns.

On the other hand, if some company orders a lot and receives a lot, distortion of the collection ratio of that company affects equally among rows and columns. Generally, few companies order a lot and receive less than NPO or NGO. According to the company's type characteristics, distortion of collection ratio is represented strongly on rows than columns.

5.4 Conclusion

We have inspected the accuracy of our IOT compared with SNAIOT. We inspected accuracy by two aspects: the intermediate demand scale and the similarity indexes. According to the discussion above, we summarized the factor that reduces each aspect's accuracy in Table.5.10.

Solving each of the factors in the Table. 5.10 will improve the accuracy of TDBIOT. The exogenous sector's influence in the algorithm and data source of the transaction value estimation is significant in both the scale and the transaction structure.

Originally, structural statistics such as the IOT should be generated directly from primary data as much as possible. However, our current estimation methods, dividing sales and production costs, could not avoid the final consumption, import, and export effect. Therefore, we incorporate weighted estimation as a secondary method to cancel out the exogenous sectors' effects by weighing each sector's data since the data's profits and cost include values for exogenous sectors. However, since the weighted estimation is optimized based on the confirmed actual transaction amount, it is not easy to verify. In addition, it is strongly affected by the data's distribution. For example, the TDB data skews toward large firms, which impacts the estimation.

Also, the transaction path's low coverage substantially impacts estimating the transaction amount. Although TDB data does not record all transaction paths, we do not include the coverage effect as an explanatory variable in the current weighted estimation. The overestimation due to the inclusion of the exogenous sector also impacts the structure. Other causes for structural miss estimation are

- TSCA,
- the coverage rate of TDB data for the total amount, and
- the deviation between rows and sums due to industry characteristics.

Equalizing the overall data coverage by the divergence rate of the row ratio measured after excluding the effect of the exogenous sector contamination will solve the bias in the coverage of the TDB data by item. In addition, the fitting of the number of transactions as an explanatory variable will improve the results. Also, the deviation between rows and sums due to industry characteristics reinforces bias in the data collection distribution. Equalizing the data collection rate will cancel out this deviation.

This chapter explained the estimation method of IOT using transaction data and accounting information offered by private firms. The RTE pays attention to accounting data and transaction data such as invoices as significant data sources and requires incorporating such data into economic statistics. Although the data we used is strictly different in format from the data obtained from invoices and other sources, we believe it is an example of how similar data can create existing economic statistics. Governments do not collect Invoices and accounting data to produce specific economic statistics. However, because of the labor and cost involved in collecting them, future economic statistics will require developing a means of estimating specific economic indicators by appropriately transforming the obtained data, as in this paper. The estimation in this paper also uses different data and solves the problems of current statistics, such as the long revision period and the large granularity. This sense suggests the possibility that new data will improve future economic statistics.

TABLE 5.11: TDBIOT comparative table

CODE	TDB INDUSTRIAL CLASSIFICATION	CODE	CALCULATING TABLE
1	AGRICULTURE EXCLUDING AGRICULTURAL SERVICES	1	AGRICULTURAL PRODUCTS
5	AGRICULTURAL SERVICES	2	AGRICULTURAL SERVICES
6	FORESTRY	3	FORESTRY
7	GAME	3	FORESTRY
8	FISHERY	4	FISHERY
9	AQUACULTURE	4	FISHERY
10	METALLIC ORES	5	METALLIC ORES
11	COAL AND LIGNITE MINING	6	CRUDE PETROLEUM AND NATURAL GAS
12	CRUDE PETROLEUM AND NATURAL GAS DRILLING	6	CRUDE PETROLEUM AND NATURAL GAS
13	NON-METALLIC ORES	7	NON-METALLIC ORES
15	CONSTRUCTION WORK BY OCCUPATION	27	CONSTRUCTION
16	GENERAL CONSTRUCTION	27	CONSTRUCTION
17	FACILITY CONSTRUCTION	27	CONSTRUCTION
19	WEAPON MANUFACTURE	24	BUSINESS ORIENTED MACHINERY
20	FOOD PROCESSING	8	FOOD・FEEDS FOR ANIMAL AND POULTRY
21	TOBACCO MANUFACTURING	9	TABACCO
22	TEXTILE MANUFACTURING EXCEPT CLOTHING	10	TEXTILES
23	CLOTHING AND MISCELLANEOUS TEXTILE PRODUCTS MANUFACTURING	11	WEARING APPAREL AND OTHER FABRICATED TEXTILE PRODUCTS
24	LUMBER AND WOOD PRODUCT, MANUFACTURING EXCEPT FURNITURE	12	LUMBER AND WOODEN PRODUCTS
25	FURNITURE AND EQUIPMENT MANUFACTURING	13	FURNITURE AND EQUIPMENT
26	PULP PAPER AND PAPER PRODUCT MANUFACTURING	14	PAPER PRODUCTS
27	PUBUSHING,PRINTING AND RELATED INDUSTRIES	15	PUBLISHING AND PRINTING
28	CHEMICAL INDUSTRY	16	CHEMICAL PRODUCTS
29	PETROLEUM AND COAL PRODUCT MANUFACTURING	17	PETROLEUM REFINERY PRODUCTS
30	RUBBER PRODUCT MANUFACTURING	18	RUBBER PRODUCTS
31	LEATHER LEATHER PRODUCT AND FUR MANUFACTURING	19	LEATHER AND FUR PRODUCTS
32	CERAMIC STONE AND CLAY PRODUCT MANUFACTURING	20	CERAMIC, STONE AND CLAY PRODUCTS
33	STEEL INDUSTRY AND NONFERROUS METAL MANUFACTURING	21	METALS
34	METALLIC PRODUCT MANUFACTURING	22	METAL PRODUCTS
35	GENERAL MACHINERY MANUFACTURING	23	GENERAL-PURPOSE MACHINERY
36	ELECTRICAL MACHINERY AND EQUIPMENT MANUFACTURING	25	ELECTRICAL MACHINERY
37	TRANSPORTATION EQUIPMENT MANUFACTURING	26	TRANSPORTATION EQUIPMENT
38	PRECISION INSTRUMENT AND MEDICAL INSTRUMENT MANUFACTURING	25	BUSINESS ORIENTED MACHINERY
39	MISCELLANEOUS MANUFACTURING INDUSTRIES	0	EXCEPTION
40	WHOLESALE TRADE(1)	31	COMMERCE
41	WHOLESALE TRADE(2)	31	COMMERCE
42	MERCHANTS AND BROKERS	31	COMMERCE
43	GENERAL MERCHANDISE STORES	31	COMMERCE
44	TEXTILE APPAREL AND ACCESSORY STORES	31	COMMERCE
45	FOOD AND BEVERAGE STORES	31	COMMERCE
46	EATING AND DRINKING RESTAURANTS	100	PRIVATE FINAL CONSUMPTION
47	MOTOR VEHICLE AND BICYCLE DEALERS	31	COMMERCE
48	FURNITURE UTENSIL AND HOUSEHOLD APPLIANCE STORES	31	COMMERCE
49	MISCELLANEOUS RETAIL STORES	31	COMMERCE
50	BANKS AND TRUST COMPANIES	32	FINANCE AND INSURANCE
51	AGRICULTURE FORESTRY AND FISHERY BANKING INDUSTRIES	32	FINANCE AND INSURANCE
52	FINANCIAL INSTITUTIONS FOR SMALL MEDIUM AND PRIVATE ENTERPRISES	32	FINANCE AND INSURANCE
53	AUXILIARY FINANCIAL BUSINESSES	32	FINANCE AND INSURANCE
54	SECURITIES AND COMMODITY EXCHANGE COMPANIES	32	FINANCE AND INSURANCE
55	INSURANCE INDUSTRY	32	FINANCE AND INSURANCE
56	INSURANCE AGENTS BROKERS AND SERVICES	32	FINANCE AND INSURANCE
57	INVESTMENT COMPANIES	32	FINANCE AND INSURANCE
59	REAL ESTATE COMPANIES	33	REAL ESTATE AGENCIES
61	RAILWAY COMPANIES	34	TRANSPORT
62	ROAD PASSENGER TRANSPORT COMPANIES	34	TRANSPORT
63	ROAD FREIGHT TRANSPORT COMPANIES	34	TRANSPORT
64	WATER TRANSPORT COMPANIES	34	TRANSPORT
65	AIR TRANSPORT COMPANIES	34	TRANSPORT
66	WAREHOUSING INDUSTRY	35	STORAGE FACILITY SERVICE
67	SERVICES INCIDENTAL TO TRANSPORT	36	SERVICES RELATING TO TRANSPORT
68	POSTAL SERVICES AND ELECTRONIC COMMUNICATION COMPANIES	37	COMMUNICATIONS AND INFORMATION SERVICES
70	ELECTRICITY	28	ELECTRICITY
71	GAS	29	GAS AND HEAT SUPPLY
72	WATER	30	WATER SUPPLY
73	HEAT SUPPLY	29	GAS AND HEAT SUPPLY
74	GOODS RENTAL AND LEASING	45	GOODS RENTAL AND LEASING SERVICES
75	HOTELS AND OTHER LODGING	100	PRIVATE FINAL CONSUMPTION
76	DOMESTIC SERVICES	100	PRIVATE FINAL CONSUMPTION
77	LAUNDRY BEAUTY AND BATH SERVICES	100	PRIVATE FINAL CONSUMPTION
78	MISCELLANEOUS PERSONAL SERVICES	100	PRIVATE FINAL CONSUMPTION

TABLE 5.11: TDBIOT comparative table

CODE	TDB INDUSTRIAL CLASSIFICATION	CODE	CALCULATING TABLE
79	MOVIE AND VIDEO PRODUCTION	39	IMAGE INFORMATION, SOUND INFORMATION AND CHARACTER INFORMATION PRODUCTION
80	AMUSEMENT AND RECREATION SERVICES	100	PRIVATE FINAL CONSUMPTION
81	RADIO AND TELEVISION BROADCASTING	38	BROADCASTING
82	AUTOMOBILE REPAIR SERVICES AND AUTOMOBILE PARKING	0	EXCEPTION
83	MISCELLANEOUS REPAIR SERVICES	0	EXCEPTION
84	MISCELLANEOUS COOPERATIVE ASSOCIATIONS	44	MISCELLANEOUS NON-PROFIT SERVICES
85	ADVERTISING RESEARCH AND INFORMATION SERVICES	46	ADVERTISING SERVICES
86	MISCELLANEOUS BUSINESS SERVICES	47	MISCELLANEOUS BUSINESS SERVICES
87	MISCELLANEOUS PROFESSIONAL SERVICES	47	MISCELLANEOUS BUSINESS SERVICES
88	MEDICAL SERVICES	41	MEDICAL SERVICE
89	PUBUC HEALTH AND WASTE TREATMENT SERVICES	42	HEALTH AND HYGIENE
90	RELIGION	44	MISCELLANEOUS NON-PROFIT SERVICES
91	EDUCATION	100	PRIVATE FINAL CONSUMPTION
92	SOCIAL INSURANCE AND SOCIAL WELFARE	43	SOCIAL INSURANCE AND SOCIAL WELFARE
93	RESEARCH INSTITUTES FOR SCIENCE	40	RESEARCH
94	POLITICAL BUSINESS AND CULTURAL ORGANIZATIONS	44	MISCELLANEOUS NON-PROFIT SERVICES
95	MISCELLANEOUS SERVICES	0	EXCEPTION
96	FOREIGN GOVERNMENTS AND INTERNATIONAL AGENCIES IN JAPAN	0	EXCEPTION
97	NATIONAL GOVERNMENT SERVICES	0	EXCEPTION
98	LOCAL GOVERNMENT SERVICES	0	EXCEPTION
99	INDUSTRIES NOT ADEQUATELY DESCRIBED	0	EXCEPTION
100	PRIVATE FINAL CONSUMPTION	0	EXCEPTION

TABLE 5.12: SNAIOT Comparative table

CODE	Input-output table basic classification	CODE	calculating table classification
011	CROP CULTIVATION	1	AGRICULTURAL PRODUCTS
012	LIVESTOCK	1	AGRICULTURAL PRODUCTS
013	AGRICULTURAL SERVICES	2	AGRICULTURAL SERVICES
015	FORESTRY	3	FORESTRY
017	FISHERY	4	FISHERY
061	METALLIC ORES	5	METALLIC ORES
062	CRUDE PETROLEUM AND NATURAL GAS	6	CRUDE PETROLEUM AND NATURAL GAS
063	NON-METALLIC ORES	7	NON-METALLIC ORES
111	FOOD	8	FOOD・FEEDS FOR ANIMAL AND POULTRY
112	DRINK	8	FOOD・FEEDS FOR ANIMAL AND POULTRY
113	FEEDS FOR ANIMAL AND POULTRY	8	FOOD・FEEDS FOR ANIMAL AND POULTRY
114	TABACCO	9	TABACCO
151	TEXTILES	10	TEXTILES
152	WEARING APPAREL AND OTHER FABRICATED TEXTILE PRODUCTS	11	WEARING APPAREL AND OTHER FABRICATED TEXTILE PRODUCTS
161	LUMBER AND WOODEN PRODUCTS	12	LUMBER AND WOODEN PRODUCTS
162	FURNITURE AND EQUIPMENT	13	FURNITURE AND EQUIPMENT
163	PULP,FOREIGN PAPER AND JAPANESE PAPER	14	PAPER PRODUCTS
164	CONVERTED PAPER PRODUCTS	14	PAPER PRODUCTS
191	PUBLISHING AND PRINTING	15	PUBLISHING AND PRINTING
201	CHEMICAL FERTILIZERS	16	CHEMICAL PRODUCTS
202	INORGANIC CHEMICAL BASIC PRODUCTS	16	CHEMICAL PRODUCTS
203	PETROCHEMICAL PRODUCTS	16	CHEMICAL PRODUCTS
204	ORGANIC CHEMICAL PRODUCTS	16	CHEMICAL PRODUCTS
205	SYNTHETIC RESIN	16	CHEMICAL PRODUCTS
206	SYNTHETIC FIBERS	16	CHEMICAL PRODUCTS
207	MEDICAMENTS	16	CHEMICAL PRODUCTS
208	FINAL CHEMICAL PRODUCTS (EXCEPT MEDICAMENTS)	16	CHEMICAL PRODUCTS
211	PETROLEUM REFINERY PRODUCTS	17	PETROLEUM REFINERY PRODUCTS
212	COAL PRODUCT	17	PETROLEUM REFINERY PRODUCTS
221	PLASTIC PRODUCTS	0	EXCEPTION
222	RUBBER PRODUCTS	18	RUBBER PRODUCTS
231	LEATHER AND FUR PRODUCTS	19	LEATHER AND FUR PRODUCTS
251	GLASS AND GLASS PRODUCTS	20	CERAMIC, STONE AND CLAY PRODUCTS
252	CEMENT AND CEMENT PRODUCTS	20	CERAMIC, STONE AND CLAY PRODUCTS
253	POTTERY, CHINA AND EARTHENWARE	20	CERAMIC, STONE AND CLAY PRODUCTS
259	MISCELLANEOUS CERAMIC, STONE AND CLAY PRODUCTS	20	CERAMIC, STONE AND CLAY PRODUCTS
261	PIG IRON AND CRUDE STEEL	21	METALS
262	STEEL PRODUCTS	21	METALS
263	CAST AND FORGED STEEL PRODUCTS	22	METAL PRODUCTS
269	MISCELLANEOUS IRON OR STEEL PRODUCTS	22	METAL PRODUCTS
271	NON-FERROUS METALS	21	METALS
272	NON-FERROUS METAL PRODUCTS	22	METAL PRODUCTS
281	METAL PRODUCTS FOR CONSTRUCTION AND ARCHITECTURE	22	METAL PRODUCTS
289	MISCELLANEOUS METAL PRODUCTS	22	METAL PRODUCTS
291	GENERAL-PURPOSE MACHINERY	23	GENERAL-PURPOSE MACHINERY
301	PRODUCTION MACHINERY	23	GENERAL-PURPOSE MACHINERY
311	BUSINESS ORIENTED MACHINERY	24	BUSINESS ORIENTED MACHINERY
321	ELECTRONIC DEVICES	25	ELECTRICAL MACHINERY

TABLE 5.12: SNAIOT Comparative table

329	MISCELLANEOUS ELECTRONIC COMPONENTS	25	ELECTRICAL MACHINERY
331	ELECTRICAL DEVICES AND PARTS	25	ELECTRICAL MACHINERY
332	HOUSEHOLD ELECTRIC APPLIANCES	25	ELECTRICAL MACHINERY
333	APPLIED ELECTRONIC EQUIPMENT AND ELECTRIC MEASURING INSTRUMENTS	25	ELECTRICAL MACHINERY
339	MISCELLANEOUS ELECTRICAL MACHINERY	25	ELECTRICAL MACHINERY
341	HOUSEHOLD ELECTRONICS EQUIPMENT	25	ELECTRICAL MACHINERY
342	ELECTRONIC COMPUTING EQUIPMENT AND ACCESSORY	25	ELECTRICAL MACHINERY
351	PASSENGER MOTOR CARS	26	TRANSPORTATION EQUIPMENT
352	MISCELLANEOUS CARS	26	TRANSPORTATION EQUIPMENT
353	MOTOR VEHICLE PARTS AND ACCESSORIES	26	TRANSPORTATION EQUIPMENT
354	SHIPS AND REPAIR OF SHIPS	26	TRANSPORTATION EQUIPMENT
359	MISCELLANEOUS TRANSPORTATION EQUIPMENT AND REPAIR	26	TRANSPORTATION EQUIPMENT
391	MISCELLANEOUS MANUFACTURING PRODUCTS	0	EXCEPTION
392	REUSE AND RECYCLING	0	EXCEPTION
411	BUILDING CONSTRUCTION	27	CONSTRUCTION
412	REPAIR OF CONSTRUCTION	27	CONSTRUCTION
413	PUBLIC CONSTRUCTION	27	CONSTRUCTION
419	MISCELLANEOUS CIVIL ENGINEERING AND CONSTRUCTION	27	CONSTRUCTION
461	ELECTRICITY	28	ELECTRICITY
462	GAS AND HEAT SUPPLY	29	GAS AND HEAT SUPPLY
471	WATER SUPPLY	30	WATER SUPPLY
481	WASTE MANAGEMENT SERVICE	42	HEALTH AND HYGIENE
511	COMMERCE	31	COMMERCE
531	FINANCE AND INSURANCE	32	FINANCE AND INSURANCE
551	REAL ESTATE AGENCIES AND RENTAL SERVICES	33	REAL ESTATE AGENCIES
552	HOUSE RENT	33	REAL ESTATE AGENCIES
553	HOUSE RENT (IMPUTED HOUSE RENT)	0	EXCEPTION
571	RAILWAY TRANSPORT	34	TRANSPORT
572	ROAD TRANSPORT (EXCEPT SELF-TRANSPORT)	34	TRANSPORT
573	SELF-TRANSPORT	0	EXCEPTION
574	WATER TRANSPORT	34	TRANSPORT
575	AIR TRANSPORT	34	TRANSPORT
576	FREIGHT FORWARDING	34	TRANSPORT
577	STORAGE FACILITY SERVICE	35	STORAGE FACILITY SERVICE
578	SERVICES RELATING TO TRANSPORT	36	SERVICES RELATING TO TRANSPORT
579	POSTAL SERVICES AND MAIL DELIVERY	37	COMMUNICATIONS AND INFORMATION SERVICES
591	COMMUNICATIONS	37	COMMUNICATIONS AND INFORMATION SERVICES
592	BROADCASTING	38	BROADCASTING
593	INFORMATION SERVICES	46	ADVERTISING SERVICES
594	INTERNET BASED SERVICES	37	COMMUNICATIONS AND INFORMATION SERVICES
595	IMAGE INFORMATION, SOUND AND CHARACTER INFORMATION PRODUCTION	39	IMAGE INFORMATION, SOUND AND CHARACTER INFORMATION PRODUCTION
611	PUBLIC ADMINISTRATION	0	EXCEPTION
631	EDUCATION	0	EXCEPTION
632	RESEARCH	40	RESEARCH
641	MEDICAL SERVICE	41	MEDICAL SERVICE
642	HEALTH AND HYGIENE	42	HEALTH AND HYGIENE
643	SOCIAL INSURANCE AND SOCIAL WELFARE	43	SOCIAL INSURANCE AND SOCIAL WELFARE
644	NURSING CARE	41	MEDICAL SERVICE
659	MISCELLANEOUS NON-PROFIT SERVICES	44	MISCELLANEOUS NON-PROFIT SERVICES
661	GOODS RENTAL AND LEASING SERVICES	45	GOODS RENTAL AND LEASING SERVICES
662	ADVERTISING SERVICES	46	ADVERTISING SERVICES
663	MOTOR VEHICLE MAINTENANCE AND MACHINE REPAIR SERVICES	0	EXCEPTION
669	MISCELLANEOUS BUSINESS SERVICES	47	MISCELLANEOUS BUSINESS SERVICES
671	HOTELS	0	EXCEPTION
672	EATING AND DRINKING SERVICES	0	EXCEPTION
673	CLEANING, BARBER SHOPS, BEAUTY SHOPS AND PUBLIC BATHS	0	EXCEPTION
674	AMUSEMENT AND RECREATIONAL SERVICES	0	EXCEPTION
679	MISCELLANEOUS PERSONAL SERVICES	0	EXCEPTION
681	OFFICE SUPPLIES	0	EXCEPTION
691	ACTIVITIES NOT ELSEWHERE CLASSIFIED	0	EXCEPTION

TABLE 5.7: Intersections that difference of input coefficients is bigger / smaller

Rows-Columns (TOP)	TDB	SNA	SNA-TDB
COMMERCE-WATER SUPPLY	0.678	0.048	-0.630
COMMERCE-SOCIAL INSURANCE AND SOCIAL WELFARE	0.623	0.149	-0.474
COMMERCE-AGRICULTURAL SERVICES	0.603	0.137	-0.466
COMMERCE-FISHERY	0.595	0.141	-0.454
COMMERCE-PETROLEUM REFINERY PRODUCTS	0.456	0.010	-0.446
COMMERCE-GAS AND HEAT SUPPLY	0.459	0.025	-0.434
COMMERCE-FORESTRY	0.530	0.108	-0.422
COMMERCE-AGRICULTURAL PRODUCTS	0.535	0.120	-0.415
COMMERCE-CRUDE PETROLEUM AND NATURAL GAS	0.459	0.046	-0.413
COMMERCE-METALS	0.433	0.030	-0.403
Rows-Columns (Bottom)	TDB	SNA	SNA-TDB
CRUDE PETROLEUM AND NATURAL GAS-PETROLEUM REFINERY PRODUCTS	0.002	0.859	0.857
CRUDE PETROLEUM AND NATURAL GAS- GAS AND HEAT SUPPLY	0.002	0.613	0.611
METALS-METALS	0.100	0.689	0.589
TRANSPORTATION EQUIPMENT-TRANSPORTATION EQUIPMENT	0.140	0.572	0.432
METALS-METAL PRODUCTS	0.059	0.490	0.431
CRUDE PETROLEUM AND NATURAL GAS-ELECTRICITY	0.000	0.423	0.423
FORESTRY-FORESTRY	0.044	0.456	0.412
TEXTILES-WEARING APPAREL AND OTHER FABRI- CATED TEXTILE PRODUCTS	0.055	0.432	0.377
MISCELLANEOUS BUSINESS SERVICES-MISCELLANEOUS BUSINESS SERVICES	0.093	0.466	0.373
IMAGE, SOUND AND CHARACTER INFORMATION PRODUCTION-BROADCASTING	0.048	0.418	0.370

TABLE 5.8: The department that column/row ratio of TDBIOT is higher than SNAIOT

Department	Differences of raw ratio
COMMERCE	-0.3004
CONSTRUCTION	-0.0311
CERAMIC, STONE AND CLAY PRODUCTS	-0.0222
REAL ESTATE AGENCIES AND RENTAL SERVICES	-0.0160
MISCELLANEOUS NON-PROFIT SERVICES	-0.0072
METAL PRODUCTS	-0.0072
SERVICES RELATING TO TRANSPORT	-0.0053
GOODS RENTAL AND LEASING SERVICES	-0.0037
TRANSPORT	-0.0033
PUBLISHING AND PRINTING	-0.0026
Department	Differences of column ratio
COMMERCE	-0.3191
CONSTRUCTION	-0.0568
ELECTRICAL MACHINERY	-0.0316
CHEMICAL PRODUCTS	-0.0227
CERAMIC, STONE AND CLAY PRODUCTS	-0.0213
GENERAL-PURPOSE MACHINERY	-0.0200
FOOD FEEDS FOR ANIMAL AND POULTRY	-0.0137
MISCELLANEOUS NON-PROFIT SERVICES	-0.0084
SOCIAL INSURANCE AND SOCIAL WELFARE	-0.0001
TABACCO	0.0000

TABLE 5.9: Departments that column/row ratio of TDBIOT is lower than TDBIOT

Department	Differences of raw ratio
TRANSPORTATION EQUIPMENT	0.0785
METALS	0.0501
PETROLEUM REFINERY PRODUCTS	0.0421
MEDICAL SERVICE	0.0361
FOOD FEEDS FOR ANIMAL AND POULTRY	0.0232
FINANCE AND INSURANCE	0.0214
COMMUNICATIONS AND INFORMATION SERVICES	0.0214
ELECTRICITY	0.0156
CHEMICAL PRODUCTS	0.0125
AGRICULTURAL PRODUCTS	0.0119
Department	Differences of column ratio
MISCELLANEOUS BUSINESS SERVICES	0.0802
METALS	0.0634
CRUDE PETROLEUM AND NATURAL GAS	0.0608
TRANSPORTATION EQUIPMENT	0.0423
PETROLEUM REFINERY PRODUCTS	0.0306
RESEARCH	0.0300
COMMUNICATIONS AND INFORMATION SERVICES	0.0232
AGRICULTURAL PRODUCTS	0.0219
ELECTRICITY	0.0161
ADVERTISING SERVICES	0.0123

TABLE 5.10: Factors reducing accuracy of TDBIOT

Facotr / affected object	Scale	Structure
Products classification		Classification of same name goods on different department
Transaction value	・Coverage of the final demand sectors ・Difference between sum of the profit and total of domestic products・Underestimation on the algorithm	Appropriation of the final demand sectors
Intermediate demand	・Appropriation of the final demand sectors	・Appropriation of the final demand sectors
TDB data	・Contamination of purchaser's price	・Distortion of the collection rate・distortion on the row and column・Head office based data

Chapter 6

Economic Analysis Model for Real-Time Economy

The management layer of RTE aims to analyze corporate accounting on a national scale and decisions making based on the analysis. What are the elemental technologies required to realize these goals? First of all, to analyze a company, it is necessary to treat its internal state in the model. Companies usually record and represent their internal state by bookkeeping accounting, and its analysis method is accounting audit. Secondly, in the case of analyzing specific firms on a national scale, it is necessary to record inter-firm transactions and update each firm's status through these transactions, in addition to auditing, which is the usual method of analyzing individual firms. Thirdly, to analyze the impact of policies on specific firms, it is necessary to analyze the impact of interventions on a firm-by-firm basis. Finally, it should be possible to compare the impacts of those interventions on a national scale in various cases. This chapter explains that existing economic models cannot satisfy these requirements and then propose a new agent-based model as an analytical method to satisfy them.

Data always constrain economic models, and RTE will change these constraints significantly. How can economic analysis be extended by RTE's data, accounting-based real-time data? There are various economic models and their analyses, such as ripple effect analysis and general equilibrium analysis. In terms of the utilization of accounting information, these models have a problem that they cannot deal with micro-state spaces and accounting data. The variables that we can manipulate in an ordinary general equilibrium model are macroscopic variables such as tax rates and interest rates or micro agents' general characteristics such as utility and production functions. Micro agents' inner states, such as stock, liabilities, and inventory, cannot be recorded and changed in the general equilibrium model. Therefore, applying big data about individual agents' states, such as accounting information, to such a model would discard most of that information. As mentioned above, the currently envisioned implementation in the management layer requires analyzing accounting events for each company and analyzing the resulting national-scale impacts. Therefore, a model that cannot grasp and analyze the accounting status at the level of individual firms will not meet the requirements. Moreover, if we need to utilize accounting data with these models, the cost matter of the conversion described above rises.

In the existing model, each agent does not have a detailed internal state. Therefore, with these models, we cannot effectively utilize the new microdata obtained by RTE. Therefore, to solve this problem, we need to develop a new model with an accounting-based state space. In the case of a model that maintains and updates individual accounting-based state spaces, each agent updates its bookkeeping by economic transactions in the same way as in the real economic space. Handling such individual transactions in the model requires the effort to set the constraints individually, but it enables the most effective use of the actual transaction data's granularity. Besides, by incorporating various economic transactions such as production, sales, and depreciation into the model, it becomes possible to incorporate various constraints for each of those transactions. The existing general equilibrium models cannot introduce such detailed constraints on economic transactions. In other words, we cannot analyze the impact of government intervention on a particular company with a general equilibrium model. In analyzing interventions on individual firms, we are interested in the path of the economy as a whole under the conditions of various interventions. A system such as a general equilibrium model, which does not represent individual firms' state space and transactions, cannot measure the impact of interventions on individual firms. Therefore, our analysis focuses on the growth path comparison analysis of the economy under various constraints.

There are several problems with the existing models, especially the general equilibrium model, taking into account the economy's growth path. The RTE aims to monitor and evaluate firms' accounting status and their spillover effects on the nation. Therefore, the detection and prediction of abnormal conditions are the most critical elements. Several economists have pointed out that general equilibrium models cannot represent abnormal states such as financial panic because they are based on a single equilibrium solution[63][64]. The general equilibrium analysis usually finds the equilibrium point for the economy's particular conditioning represented by the simultaneous equations. These characteristics imply the transition from a given initial condition to a single state achieved under the assumption of a rational economy. The growth path of the economy, however, may not be single.

The theory of economic growth, on the other hand, assumes that there are various possible pathways in the economy and analyzes the characteristics of those pathways. There are two kinds of models of growth theory: the two-sector growth model, which explores the relationship between the production sectors of investment goods and consumer goods, and the multisectoral growth model originating from Von Neumann (1973). The multisectoral growth model generally implies a balanced growth path where each industrial sector's output ratio grows constant if each industry's initial production structure is constant returns to scale. Samuelson et al. . argued that the Pareto-optimal growth path in the capital accumulation economy approaches the balanced growth path in this process[65] [66]. The phenomenon of this optimal pathway passing through the balanced growth path is generally called the Turnpike theorem. Turnpike theorem was formulated by Radner[67], Morishima[68] and extended to the Leontief dynamic model by MacKenzie[69] and Tsukui[70]. Initially, Turnpike Theorem was shown by an algebraic proof, but as the complexity of the model increases, it becomes more difficult to prove it, so it is now shown in simulations using input-output analysis[71]. From now on, we call these models as Turnpike model.

Although several patterns exist in the formulation, the simple Turnpike model is solved as a maximization problem for the Leonchev dynamic model.

Such an optimization problem means that there can be both optimized and non-optimized paths in the economy under analysis, and there can be many paths for each objective function. The general equilibrium theory responds to the existence of such multiple equilibria by modifying the model itself so that it has only a single equilibrium from the viewpoint of usefulness. The most prominent of these modifications is restricting the production function to diminishing returns, which is not common in many firms today. Although it is generally known that the production function of individual firms in the real economy is of Leonchev type, only continuous functions such as CES-type functions can be applied to achieve a single equilibrium.

In response to these problems, B.Arthur[72] advocates the necessity of a generative approach to observe the process of generating each equilibrium and analyze the conditions and probabilities of each equilibrium's occurrence. He suggests Agent-Based Modeling as a method to enable the analysis of such micro-processes. On the other hand, the Turnpike model does not have such a problem due to the constraints for achieving a single equilibrium. In this paper, based on these issues, we attempt to extend the agent-based model, an extension of the Turnpike model, based on accounting information.

However, there are many problems with the Turnpike model as well. Firstly, the Turnpike models' analysis usually selects the overall optimal path from the growth potential frontier in each term. However, it is impossible to analyze what changes are necessary to move from one path to another in each selection. The general equilibrium model allows us to discuss policy to move from one equilibrium to another. In contrast, the Turnpike model does not allow us to analyze the effects of changes in parameters such as policy. In other words, although general equilibrium analysis is constrained to obtain a single equilibrium solution, it can analyze the interaction effect of many parameters. In contrast, the Turnpike model cannot investigate the effect of many parameters, such as policy, price, and interest rate. Therefore, although the Turnpike model allows us to discuss the extent to which the current growth path deviates from the optimal growth path, the question of what the bottleneck is can be analyzed only by considering mainly the capital and labor constraints embedded in the production function.

The agent-based model treated in this paper is a Planned Economy Model similar to the Turnpike model. In other words, the model does not seek the equilibrium point as a result of optimizing each agent's utility function but acts based on boundary conditions to optimize the evaluation function of the economy as a whole, such as GDP and total output. This paper calls this kind of model, instead of the Planned Economic Model, a Centralized Model Analysis(CMA). Therefore, this model can represent multiple scenarios depending on the evaluation function to be set up and analyze various cases ranging from the worst to the best, starting from the same state. Furthermore, our model reproduces market transactions made up of various parameters. Therefore, by changing the constraints associated with each transaction, we can observe how a particular growth path changes and identify the elements that are bottlenecks in the growth path. There is no restriction on the type and number of variables and to be included in the model. One of the Agent-Base Modeling features is that there is no limit to the model's expandability to the analysis target. This point of view will be discussed later in the paper regarding the search for constraints and evaluation functions.

6.1 Model Implementation

Fig 6.1 represent whole image of our model implementation. The model proposed in this paper is an extension of the model by Deguchi(1998, 2004)[8][9]. The Deguchi model is a CMA model consisting of nine agents in which each agent records an exchange algebra[9], an algebraic representation of bookkeeping, as an internal state [1]. The exchange algebra in this study consists of a basis with three elements: Unit,

[1] We share the implementation of the exchange algebra in this paper at the following URL. `https://github.com/yakagika/ExchangeAlgebra`

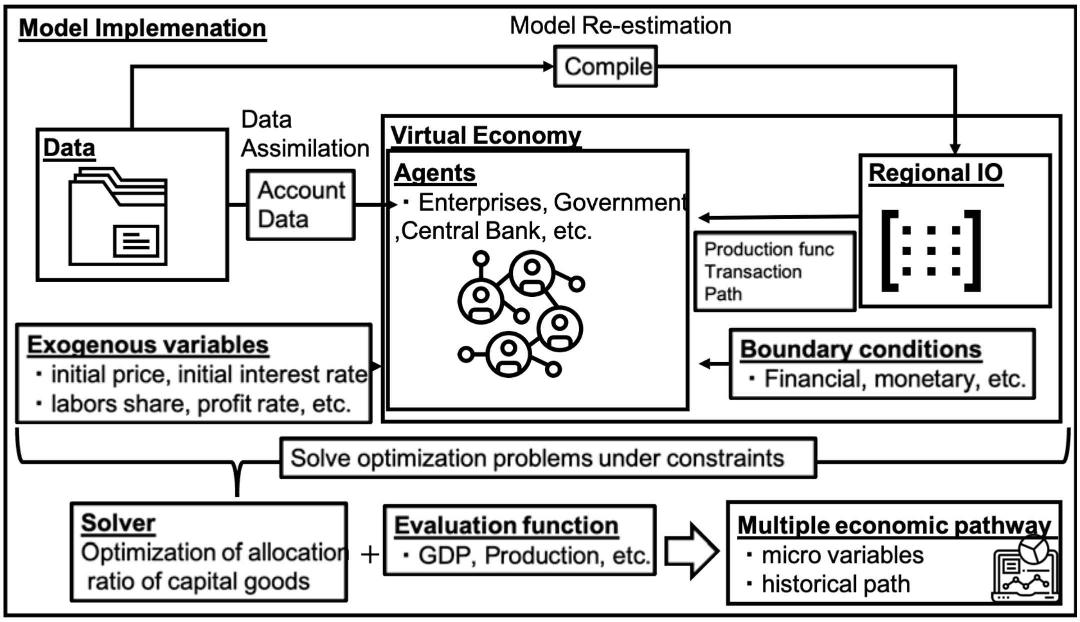

FIGURE 6.1: Model implementation image

Account Title, and Commodity Name. The Unit consists of Yen and Amount, the Account Title is a set of the account titles, and the Commodity Name is the set of commodity names in the market. Therefore, one mining product in quantity unit can be represented as $1 < Amount, Products, MiningC >$.

In each simulation term, each agent trade and its result is recorded as an exchange algebra and updates its internal state. The model seeks the optimal growth path, optimal allocation ratio of investment goods among the agents in each term, solving the turnpike dynamic optimization problem.

Although the trade routes and production functions were set up ad hoc in the previous model, this paper enables creating a matrix of input coefficients from real data using accounting information and transaction information based on the previous study. Some exogenous variables in the last model, such as price and official rates, are made endogenous to solve Turnpike models.

We constructed this model assuming RTE's data, the invoice and accounting data. When we obtaine the new data, the model will re-estimate the economy's transaction structure and production function as IO and assimilate accounting information and price information as each agent's internal state. Implementing the model's re-estimation is entirely dependent on the data, and it also depends on the computational resources and implementation effort in practice. When constructing the model, selecting the most fine-grained one in every unit, such as time, region, and classification is desirable if we do not consider the computational resources. Although the model treated in this paper adopts a very rough number of agents and classifications as a basic model, it can change to any granularity since we treat each company and office separately.

6.1.1 Model usage and analysis methods

As mentioned above, our model is a Concentrated Model Analysis (CMA); in CMA, agents act to achieve the global optimum of a given evaluation function. In other words, agents do not make decisions such as profit maximization or cost minimization in the sense of general equilibrium models. Each agent's decisions are determined only by the overall optimization, and the agents themselves do not act according to any specific objective or decision rule. This setup is the same as the Turnpike model; in the Turnpike model, each agent has constraints such as production function and capital accumulation, but each agent does not behave in a specific way, such as profit maximization.

Each agent takes all possible actions within the given constraints, and the model finally selects the overall optimal action from among them. Therefore, we do not expect agents to take profit-maximizing actions or even take rational actions concerning the outcome. Through the CMA model, we can know an agent's behavior to maximize (minimize) a particular evaluation function, but such behavior is not logically verified or predicted to occur in a real economy. On the other hand, by knowing the outcome that leads to the overall optimum, we can know that it is possible to have a path that leads to a particular optimal outcome by forcing (inducing) agents to behave in a particular way through policies, institutional changes, and interventions.

Therefore, our model has the same inherent problem as the Turnpike model that the results of the analysis only show the existence of some optimal growth path. Samuelson et al. have shown that the optimal growth path is consistent with a balanced growth path to address this problem. As mentioned earlier, it is necessary to examine the design problem to realize the Turnpike path by agent autonomy, i.e., what kind of agent design will realize this path in a real system.

The Turnpike model does not provide a tool to discuss this design problem. As shown later in the simulation results, the optimal growth path that an economy can take varies even with simple constraints. In our model, we construct a state-space with the minimum granularity that constitutes the economy, which allows us to compare various boundary conditions with actual microdata movements. In the later section, as an example, we analyze that we can not achieve the optimal path shown with the conventional very simple Turnpike model under variable prices, including the treatment of depreciation. In this way, we show that conventional simple models do not represent the constraints and conditions and that our model provides tools to discuss these unidentified issues.

In future discussions of evolutionary economics, it is necessary to examine what kind of additional constraints, i.e., other institutions, make it possible to control the entire economy toward a specific optimal path. We need to discuss such verification by looking at the institutional characteristics based on real data by RTE. Now that we can obtain microdata with RTE, we can finally observe and control micro-state space and micro constraints. It is now possible to discuss how to take the growth path in a transaction-based manner. Our model provides a tool to explore the constraints and the evaluation function of the global optimum to achieve the Turnpike path or Laissez-Faire.

The existing Turnpike model uses the integral of the growth path as the evaluation function to optimize, but in the search for the optimal path, the optimal path's meaning differs depending on the evaluation function. In actual policy management, the maximization of GDP and output is not the only target of evaluation, but we need to discuss various targets such as primary balance and environment in a complex and parallel manner. Therefore, there is no consensus on the evaluation function itself for the governance of the entire country. Due to the Turnpike model's simplicity, it is impossible to analyze these evaluation functions in a meta-analysis.

On the other hand, our model can analyze the environment if environmental accounting is incorporated into the model and analyze primary balance if fiscal constraints are incorporated. In such a problem of determining the evaluation function, there is no single solution. However, our model unseals the unidentified problems by looking at the micro details of the optimal growth paths achieved under various evaluation functions and constraints. The identified problems mean that we are adding new evaluation axes to the existing evaluation functions, enabling us to update the evaluation functions from the initial to more desirable ones. We can then search for new evaluation functions and constraints in order to improve the identified problems. In other words, we assume that we can use our model as a gaming simulation tool for consensus building to identify the problems of a particular path and to search for better evaluation functions and constraints.

In the following, we will explain our model in detail. The model we present in this paper is a set of general and basic economic transactions and accounting procedures. Still, due to the amount of effort required for implementation, we programmed only limited boundary conditions. Since our model preserves each agent's state space, readers can easily add various constraints to each transaction. We will release all the models we have used in this paper as OSS[2]. Therefore, the readers could extend the models according to their data and needs, hoping that.

6.1.2 Agent

Our model classified all economic agents into 14 economic sectors; AFF (Agriculture, Forestry, Fisheries, mining, machinery manufacturing, other manufacturing, construction, commerce, real estate, transportation, energy, services, government, a central bank, banks, and households. Each of these classifications has its own set of possible economic activities and boundary conditions, and the behavior of agents changes depending on the category. The government issues government bonds and distributes subsidies to industries and households. The central bank issues currency and adjusts the official rate to keep the interest rate constant. Each industry produces a single good through intermediate consumption according to the input coefficient matrix. Only the machinery, manufacturing, and construction industries produce capital goods. According to the evaluation function, this model optimizes capital goods' allocation rate among agents in each term with a simulated annealing method.

6.1.3 Term

In this model, the agent repeats the specified behavior for a specified term. The number of terms is variable, but we have set it to 20 terms as a default setting, where we can see some degree of change since our purpose is to show its representation ability. In this model, agents revise their prices and close their accounts in one term, so one term's interpretation would be at least a quarter. Depending on the interpretation of this term, various parameters need to be modified. In particular, the population's growth rate, the interest rate, and the production function need to be chosen appropriately according to this term's interpretation. Since we conducted all the simulations in this paper with hypothetical dummy variables, the term's interpretation is arbitrary, but we tentatively choose one year as one term, which is easy to obtain macroeconomic variables. On the other hand, in the RTE context, it

[2] https://github.com/yakagika/VirtualEconomy

is recommended that the term's interpretation be the smallest unit in which we can update the various macro parameters.

6.1.4 Decision maiking rules and Constraints

In our model, each agent determines its behavior in each term according to some constraints shown below.

Production function Each industrial agent determines its output in each term according to the production function defined by the input coefficient matrix and is subject to the constraints of capital goods, intermediate inputs, and labor inputs. We estimated the input coefficient matrix using the Input-Output Table compiled with the method in previous studies[35].

Let the set of products by each industry be Com and set of industry be Ind. Then there exists a bijection $whatCommodity(x) : Ind \rightarrow Com$ and their inverse functions $whatIndustry(x) : Com \rightarrow Ind$.

Let $\mathbb{A}$ be the matrix of input coefficients obtained from IO, and it is expressed in the following way.

$$\mathbb{A} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \dots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \dots & \alpha_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ \alpha_{i1} & \alpha_{n2} & \alpha_{13} & \dots & \alpha_{ij} \end{bmatrix}$$

In $\mathbb{A}$, each of $\alpha_{ij} \in [0,1], i, j \in Com$ represents amount of intermediate demand $j \in Com$ to produce a unit of $i \in Com$. Let coulumn i of $\mathbb{A}$ be $\mathbb{A}_i$. Similarly, the coefficient vectors of labor and capital inputs are set to $\mathbb{L}$ and $\mathbb{K}$, respectively.

$\mathbb{L} = [\beta_i | i \in Ind, \beta_i \in [0,1]]$

$\mathbb{K} = \{\gamma_i | i \in Ind, \gamma_i \in [0,1]]$

In this case, γ_i and β_i are the inputs of capital goods and labor required to produce a unit of good.

Let the quantity of intermediate goods held, the number of workers, and the number of machines (capital goods) held respectively x_{ij}, e_i, and m_i, then the production function of industry i is defined as follows.

$$f(A_i, L, K_i) = min(\{min(\{x_{ij}[t]/a_{ij} | a_{ij} \neq 0\}), e_i[t]/\beta_i, m_i[t]/\gamma_i n\}) \tag{6.1}$$

In this paper, we adopt a Leontief production function which added constraints to the Constant Elasticity of Substitution (CES) function not to allow any substitution between techniques. Since the number of available machines (as a percentage of output) is determined exogenously for each agent to optimize the evaluation function, each firm plans its output in term $t-1$ to maximize the operation of its machines in term t. A firm hires employees and purchases intermediate inputs necessary to realize its production plan in term t. This assumption means that the number of capital goods purchased in $t-1$ term represents the production plan in t term. For this situation, We assumed that each producer decides the next term's production using some information. However, the actual total supply of intermediate inputs and employees is affected by other constraints and may not always be produced as planned.

The supply of an intermediate good is affected by the production of other goods in a spillover manner. Therefore, unlike the general equilibrium model, all variables in our simulation never reach equilibrium simultaneously, and there is always sequence for each action, such as production and purchase. Moreover, these constraints do not converge forever when there are cycles between industries and commodities in the input coefficient matrix.From the viewpoint of introducing detailed constraints, it is desirable to reproduce the procedure in which each agent purchases raw materials for each transaction and produces and sells them to the extent that the raw materials are satisfied. However, these calculations require many computing resources because they reproduce a massive number of transactions for each agent. Therefore, in this paper, we simplify the process.

For this reason, we determined the production rate under the intermediate input constraint with the following algorithm.

1. Obtain the production plan with labor and capital constraints. Calculate the supply and demand of each good according to the production plan and inventory holdings.If the material ratio is more than 100%, agents will not produce.

2. Agent produces goods using the raw materials held. The transaction of this step can be describe as below.

$$\begin{aligned} materialProd[i,t] &= matProd[i,t] < Amount, Product, comName[i] > \\ &+ valueAdded[matProd[i,t], i, t] < Yen, ValueAdded, \# > \\ &+ use[matProd[i,t], t] \end{aligned}$$

   $comName[i]$; returns name of commodity produced by agent i.

   $matProd[i,t]$; returns available amount of production with the raw materials held by agent i at term t.

$$\begin{aligned} valueAdded[x,i,t] &= price[comName[i],t] * x \\ &- materialCost[x, comName[i], t] \end{aligned}$$

   $price[c,t]$; returns price of commodity c at term t.

   $materialCost[x,c,t]$; returns material cost to produce x amount of c at term t.

   $use[x,t]$; returns Exchange Algebra of intermediate input to produce x amount of commodity at term t.

3. Reduce the production plan by the amount of production using inventories and reflect it in quantity supplied and the quantity demanded by goods.

4. If supply and demand are not much, agents reduce the production plan according to the demand quantity, starting from the industry with the most significant demand for materials in the industry where supply is insufficient to meet demand. Calculate the supply and demand again.

5. Repeat step 4 until the whole is satisfied.

6. According to the production plan, agents produce goods and sell and purchase according to the demand. The transaction of this step can be describe as below.

$$prodSellBuy[i,t] = prod[i,t] + sell[i,t] + buy[i,t]$$

$$
\begin{aligned}
prod[i,t] &= plan[i,t] < Amount, Product, comName[i] > \\
&+ valueAdded[plan[i,t], i, t] < Yen, ValueAdded, \# > \\
&+ use[plan[i,t], comName[i], t]
\end{aligned}
\tag{6.2}
$$

$plan[i,t]$; returns production plan of agent i at term t.

$stock[i,t]$; returns amount of stock of agent i at term t.

$totalDemand[c,t]$; returns total demand for commodity c at term t.

$totalSupply[c,t]$; returns total suppuly for commodity c at term t.

$$
\begin{aligned}
s[i,t] &= (totalDemand[comName[i], t] \\
&* ((plan[i,t] + stock[i,t]) / totalSupply[comName[i], t]))
\end{aligned}
$$

$$
\begin{aligned}
sell[i,t] &= s[i,t]\hat{}< Amount, Products, comNamei > \\
&+ s[i,t] * price[comName, t] < Yen, Products, \# >
\end{aligned}
$$

$$
\begin{aligned}
buy[i,t] &= \hat{}use[plan[i,t], t] \\
&+ priceConvert[use[plan[i,t], t], t]
\end{aligned}
$$

$priceConvert[e,t]$: price conversion function.

6.1.5 Population

In each term, the population increases with gr%. Therefore, if the population in term t is $pop[t]$, we have $pop(t+1) = x(t) \times gr/100$.

6.1.6 Labor

In each term, the company hires to maximize the operation of its machines. If the demand for jobs exceeds the supply, we allocate new workers in proportion to its employment plan.

If the propensity to consume is $\mathbb{C} = [c_i | i \in Industry]$, then the consumption of good i by households in each term is $c_i * pop[t]$, but household consumption is within the inventory, excluding intermediate consumption. In this model, we treat the propensity to consume as exogenously. With data, The consumption propensity with the estimation method of Osato[73].

6.1.7 Price and Wage

The price of each good and the wages of each firm vary so that the exogenously given labor share and the rate of ordinary income are held constant. We interpret these constraints as the result of negotiation between the firm and the workers. If there are multiple agents in an industry, we choose the lowest for a price.

Let production plan of product i be $x_i[t] = f(A_i, L, K_i)$, he average wage of the industry j be $w_i[t]$, price of the products i be $p_i[t]$, and labor's share be $LS \in [0,1]$.

Then manufacturing cost, input materials without self-consumption, of $i \in Ind$ is defined as

$$InpWSelf_i[t] = \sum_{j \neq i} p_i[t-1] * a_{ij}[t-1] * x_i[t] \tag{6.3}$$

From the definition of the labor's share, we can derive the following.

$$LS = \frac{w[t] * e_i[t]}{x_i[t] * p_i[t] - InpWSelf_i[t] - x_i[t] * p_i[t] * a_{ii}[t-1]} \tag{6.4}$$

From these equations, the equilibrium wage is

$$w_i[t] = \frac{LS * (x_i[t] * p_i[t] - InpW_i[t] - x_i[t] * p_i[t] * a_{ii}[t-1])}{e_i[t]} \tag{6.5}$$

Let the corporate tax rate in term t be $It[t]$, the interest rate be $r[t]$, financial expenditure be $ie_i[t] * r[t]$, the profit margin be PR. Depreciation be $dep[t]$. If we assume that the market will purchase all the industry products, we can derive the company's profit margin from the corporate tax rate definition as follows.

$$\begin{aligned} PR &= \frac{p_i[t] * x_i[t] - CostWSelf - w_i[t] * e_i[t] - a_{ii}[t] * x_i[t] * p_i[t]}{p_i[t] * x_i[t]} \\ &= \frac{1}{p_i[t] * x_i[t]} \\ &\quad \{p_i[t] * x_i[t] - CostWSelf - a_{ii}[t] * x_i[t] * p_i[t] \\ &\quad -(LS * (x_i[t] * p_i[t] - InpW_i[t] - x_i[t] * p_i[t] * a_{ii}[t-1]))\} \end{aligned} \tag{6.6}$$

$$CostWSelf = InpWSelf_i[t] + dep_i[t-1] + ie_i[t-1] * r[t-1] \tag{6.7}$$

Solve for this and get

$$\begin{aligned} p_i[t] &= \frac{LS * InpWSelf_i[t] - CostWSelf}{x_i[t] * \{PR - 1 + LS(1-a_{ii}) + a_{ii}\}} \\ &= \frac{(1-LS)InpWSelf_i[t] + dep_i[t-1] + ie_i[t-1] * r[t-1]}{x_i[t](1 - PR - LS(1-a_{ii}) - a_{ii})} \end{aligned} \tag{6.8}$$

and substitute $p_i[t]$for $w_i[t]$ to get,

$$\begin{aligned} w_i[t] = & \frac{LS(1-a_{ii})}{e_i[t](1 - PR - LS(1-a_{ii}) - a_{ii})} \\ & \{ \quad dep_i[t] + ie_i[t] * r[t-1] + \frac{LS}{(1-a_{ii})} InpWSelf_i[t]\} \end{aligned} \tag{6.9}$$

In this case, the wages and prices for term t are determined as follows.

$$w_i[t]^* = \begin{cases} w_i[t] & , w_i[t] > 0 \\ w_i[t-1]^* & , otherwise \end{cases} \tag{6.10}$$

$$p_i[t]^* = \begin{cases} p_i[t] & , p_i[t] > 0 \\ p_i[t-1]^* & , otherwise \end{cases} \tag{6.11}$$

In case $x \notin Industry$, it is defined as follows.
$w_{Household}[t] = 0$

$x \in Governments, Bank, Centralbank,$

$w_i[t] = \frac{1}{|Industries|} \sum w_i[t]^*$

Let S be a Set, $|S|$ represents the number of elements in S.

6.1.8 Monetary policy

The central bank determines the official interest rate to reduces the fluctuation of prices. For simplicity, we assume that the short-term interest rate links to the long-term interest rate. The bank provides loans to firms based on the short-term interest rate. The central bank determines the interest rate to cancel out the price changes with the given knowledge of the price elasticity to the interest rate. In this simulation, there is a steps during the same term.

Moreover, we determine the interest rate after the firm's production, the number of jobs, and the price. Therefore, the interest rate is determined after the price is determined. However, companies cannot realize their production plans because the number of employees, the number of producers, the number of sales are subject to their respective constraints.

There are several options for implementing an interest rate policy, and we need to implement the necessary policy for each objective of the analysis. In this paper, we implement a monetary policy similar to Taylor's rule, using the price elasticity for interest rate.

We determined the elasticity of a price for interest rate as follows.

$$\begin{aligned} \epsilon[t] &= \frac{\partial\, p_i[t]^*}{\partial\, r[t]} / \frac{p_i[t]^*}{r[t]} \\ &= \frac{r[t-1] * ie_i[t-1]}{(1-LS)InpWSelf_i[t] + dep_i[t-1] + ie_i[t-1] * r[t-1]} \end{aligned} \tag{6.12}$$

We assume that governments and central banks can observe each variable's average value for a country as a whole. If we denote each market's average value for variable x as $\hat{x}$, then the elasticity assumed by the government and central bank can be defined as follows.

$$\epsilon[t] = \frac{\hat{r}[t-1] * \hat{ie_i}[t-1]}{(1-LS)Inp\hat{W}Self_i[t] + \hat{dep_i}[t-1] + \hat{ie_i}[t-1] * r[t-1]} \tag{6.13}$$

$$\hat{a}_{ii} = \frac{1}{|Ind|} \sum_{i=0}^{|Ind|-1} a_{ii} \tag{6.14}$$

Let elasticity of price be

$$\Pi[t] = \frac{1}{|Ind|} \sum_{i \in Industry} \frac{\hat{p}_i[t]}{\hat{p}_i[t-1]} \tag{6.15}$$

and target inflation rate be $target[t]$.

Then, in order to reduce the fluctuation of the average market price, the central bank changes the official interest rate according to the following

$$r[t] = r[t-1] - \hat{\epsilon}[t] \times ((\Pi[t] - 1) - target[t]) \quad (6.16)$$

As for the monetary policy, we should switch the rules depending on the objective. In this paper, for the sake of simplicity, we adopt the k% rule.

6.1.9 Fiscal constraints

The government determines the output of government services and the total amount of subsidies under budgetary constraints. The government always distributes at a constant ratio to the previous term's tax revenue and decides the government expenditure to keep the primary balance in surplus. If the government spends more than tax revenues, the government issues government bonds, and the Bank of Japan takes over[3].

6.1.10 Financing constraint

In our model, the shortage of cash is transferred to deposits, and the shortage of deposits is automatically transferred to loans. Financial institutions judge loans' availability in the real world, firms that fail to obtain funds bankrupt. However, since the introduction of loan decisions and bankruptcy is very complicated, in this paper, loan decisions are limited only to the purchase of capital goods. More realistic loan conditions and the introduction of bankruptcy, and other factors will be the subject of future work. Therefore, in our model, firms can refinance funds without restriction and never go bankrupt, but there is a restriction on loans for new business expansion when an agent's planned capital investment exceeds the sum of its cash and bank deposits in each term, the agent finances the shortfall with loans. There are various models for making loan decisions. However, here we use three simple conditions: the current ratio as a criterion for stability, the equity ratio as a criterion for profitability, and the current expense to income as a criterion for repayment ability. In this model, if the $current\ ratio \geq 100$, $equity\ ratio \geq 5$, and $currentexpense\ to\ income \geq 100$ are satisfied simultaneously, institutions finance the full amount of the shortfall.

6.1.11 Simulation

We simulate according to the following steps. The exchange algebra describes these transactions, but since the description of all of them would be too large, we omit the description here [4].

[3]The actual underwriting of government bonds in Japan is subject to stricter rules. This point is under consideration for improvement.

[4]Our program code is written in a notation very close to that of exchange algebra. For example, if the exchange algebra for the redemption of government bonds is written as

$$\begin{aligned} f[Government, t] &= redeemNB[Government, t]\hat{} < Yen, ShortTermNationalBondsPayable, \# > \\ &+ redeemNB[Government, t]\hat{} < Yen, Cash, \# > \end{aligned}$$

, the code is as follows.

0) Generate the initial allocation plan.
1) Initialize agents.
2) Develop production plans for each company, employment of workers.
3) Determine prices and wages.
4) Purchase intermediate inputs. Produce goods.
5) Trade final consumer and capital goods. In case of shortage of funds, firms loans from banks. Issue government bonds.
6) Allocate subsidies.
7) Pay income tax and corporate tax.
8) Pay financial expenditure.
9) Transfer Profit and Loss statements to the balance sheet.
10) Repeat steps 2 through 10 until the last term.
11) Calculate the evaluation function and determine the allocation plan for the next term.
12) Repeat steps (1) through (12) until the annealing method satisfies the end condition.

Optimization In this paper, we optimize the allocation plan of capital goods by a simple simulated annealing method. The allocation ratio is updated by randomly selecting the term and agents, and then moving the random allocation ratio from one to the other. We repeat this update in inverse proportion to the elapsed time. Let the set of allocation plan be IS (Investment Schedule), the evaluation function be $f : IS \to \mathbb{R}$ initial and the updated allocation plan be $is, is' \in IS$, and the elapsed time be T, then the adoption rate of rnis is $(f(is) - f(is'))/T$.

In our simulator, we run a given number of parallel calculations and finally obtain the best result among them. In the following simulations, we run eight calculations in parallel from the same initial conditions.

6.2 Result & Discussion

In this section, we simulate several pattern of our model [5].

6.2.1 Exogenous variables setting

This section outlines the settings of the dummy data and exogenous variables used in this simulation. The source code and all the settings used in this study are available here[6].

$$
\begin{aligned}
f\,wld \quad & NationalBondsRedemption\ i\ t \\
= \quad & (redeemNBwldit) < @Hat :< (Yen, ShortTermNationalBondsPayable, (.\#)) \\
<+ \quad & (redeemNBwldit) < @Hat :< (Yen, Cash, (.\#))
\end{aligned}
$$

For details, please refer to the source code.

[5]Environment: MacBook Pro (13-inch, 2018, Four Thunderbolt 3 Ports)
macOS Catalina version 10.15.6
Processor 2.7 GHz Quad-Core Intel Core i7
Memory 16 GB 2133 MHz LPDDR3
stack Version 2.3.1, lts-16.8
compile option : +RTS -N8 -A250M

[6]https://github.com/yakagika/VirtualEconomy

In the next section, we will test the model and analyze its behavior with the dummy data, which we created initially to balance accounting. Since we represent the agent's state as bookkeeping in our model, the agent and the economy must balance each transaction. This feature enables the model to guarantee the accuracy of each transaction. We confirm all the transactions in the simulations performed in this test balanced in an accounting manner.

The initial variables of the simulation have certain restrictions to make the simulation possible. As an extreme example, if the initial population is zero or the population growth rate is −100%, the simulation will be terminated after one term. Similarly, several minimum conditions must be satisfied to run the simulation, and we arbitrarily created the following dummy variables to satisfy these minimum criteria.

For instance, We set initial consumption propensity, and input coefficients according to the output ratio in the Japan Input-Output table for 2011[7]. Although we used the actual input-output table as a guide to obtaining the simulation's stability, our aim in this paper is not to simulate the actual Japanese economy but to show the model's expressive power. The other initial values and exogenous variables are dummy variables adjusted to make the simulation feasible. Therefore, the simulation results do not provide any implications for the analysis of the actual Japanese economy.

We treat all prices and wages as relative prices and set the initial value to 1. We also set Labor and capital productivity to 40 and 50, respectively, using the same values for all industries. We set exogenous variables according to the Table 6.1.

TABLE 6.1: Variables

Parameter	value
Total Term	20
Money Supply Growth Rate	2 %
Loans repayment rate	10 %
Redemption rate	5 %
Initial Population	1200000
Increasing rate of population	2 %
Income tax rate	10 %
Cooprate tax rete	30 %
Inflation taregt	2 %
Subsidy rate	10 %
Official bank rate	0.1%
Target Profit rate	30%
Labors Share	65 %
Redemption rate	20%
National bond rate	1 %
Deposit and loan interest rate differential	0.015
Initial price (all industries)	1
Initial wage (all industries)	10
Productivity of Kapital (all industries)	50
Productivity of Labor (all industries)	40

[7] https://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00200603&tstat=000001073129

6.2.2 Compare constraints

Our model allows us to analyze the possible paths of the economy by changing various constraints. To demonstrate this, we analyze two simulation results in which different price fluctuation conditions. Here, as in the usual Turnpike model, under the evaluation function that maximizes the path integral of output, the price is kept constant in one case, while in the other case, the price changes to keep the profit margin constant. The former is under the same conditions as the usual Turnpike model for price, and the latter shows how the optimal path changed when we added the price constraint to the Turnpike model.

Fig. 6.2 and 6.3 show the path of total production in the economy as a whole with constant and fluctuating prices, respectively. Both of them are in quantity units. In both cases, the final output level is not so different, but in the constant price path, output reaches the output level defined by the labor constraint quickly from the initial state and then increases according to the growth rate of the population, whereas in the price fluctuation path, output reaches that level gradually. This is due to the inclusion of depreciation in the price determination model. When an agent invests a large amount of capital investment at once, the price rises, and the other industry's raw material ratio that uses these goods as intermediate inputs exceeds one. In this model, since agents do not produce under a deficit, stable production cannot be carried out unless price fluctuations are kept within a specific range. Fig.6.4 and 6.5 show the amount of production machinery purchased by each firm, and we can see that agent suppressed investment in the price fluctuation path.

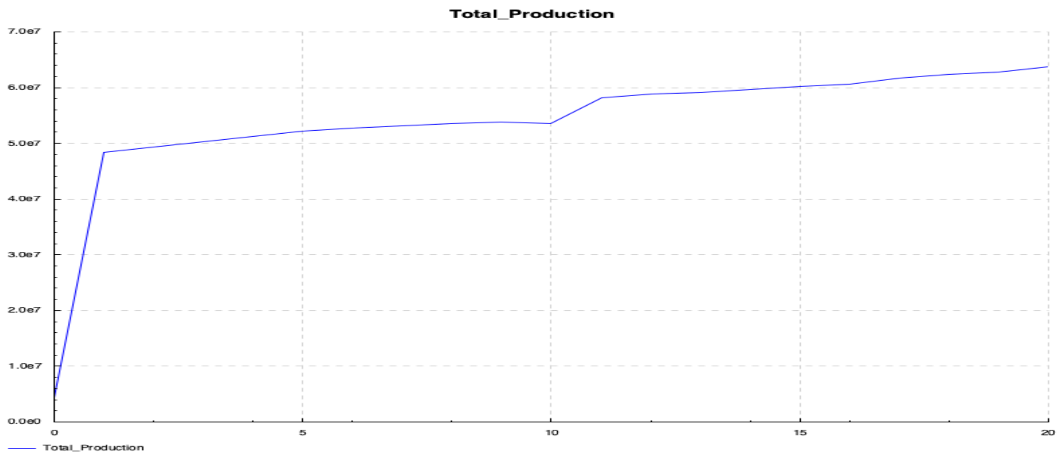

FIGURE 6.2: Total Production path. Fluctuating Price.

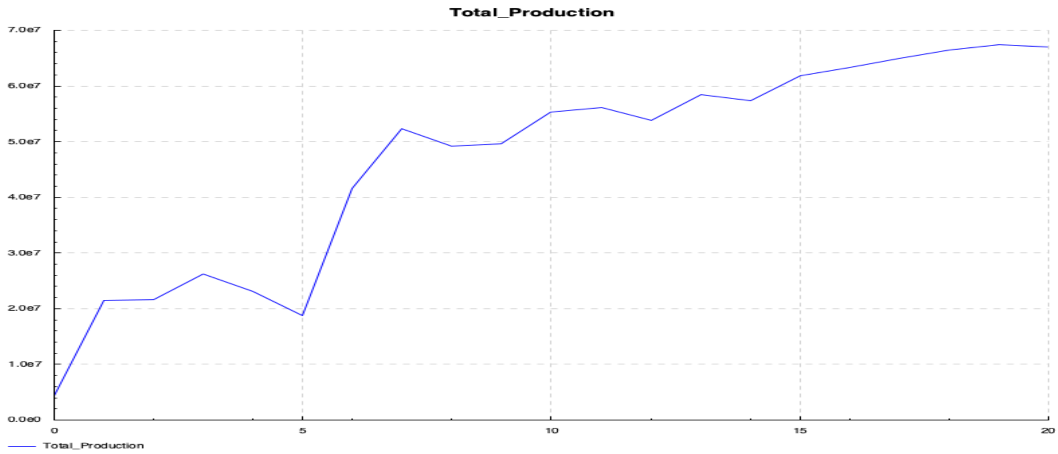

FIGURE 6.3: Total Production path. Fluctuating Price.

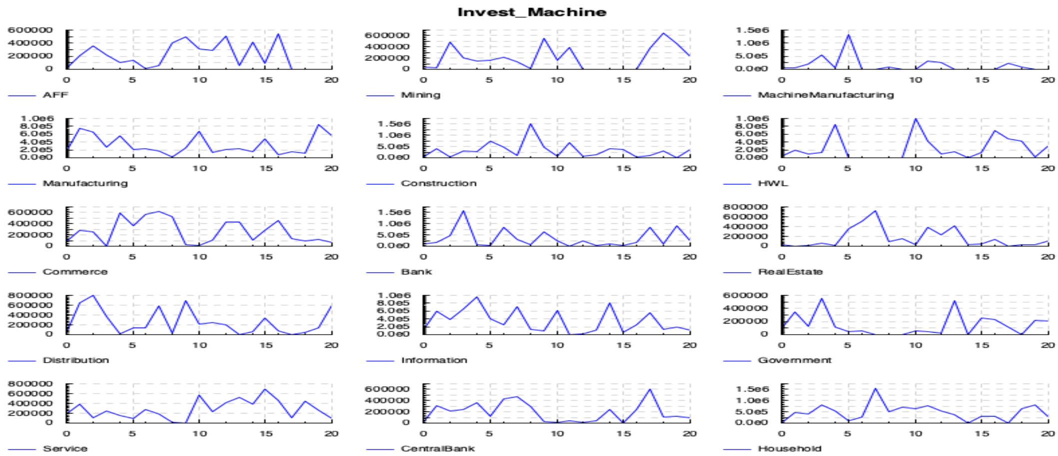

FIGURE 6.4: Number of Investment Machine. Constant Price.

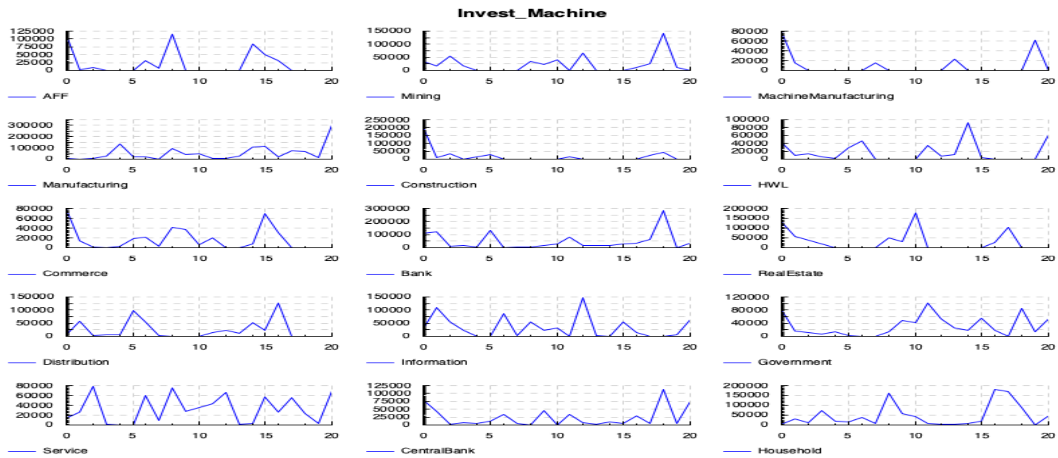

FIGURE 6.5: Total Production path. Fluctuating Price.

Fig.6.6 and 6.7 show the price fluctuations and the raw material ratio for each industry. Fig.6.8 and 6.9 show the raw material ratio and each industry's output when the optimization is deliberately interrupted. We can see that Mining's raw material ratio exceeded one at around the 9th term, and production was terminated, which caused the overall Commodity Flow to be interrupted and production to stop.

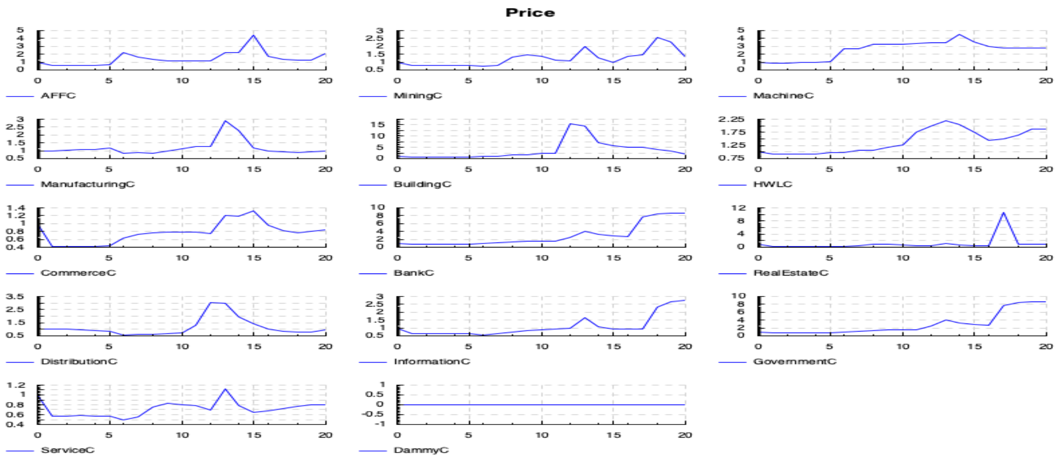

FIGURE 6.6: Price path. Fluctuating Price. Optimized.

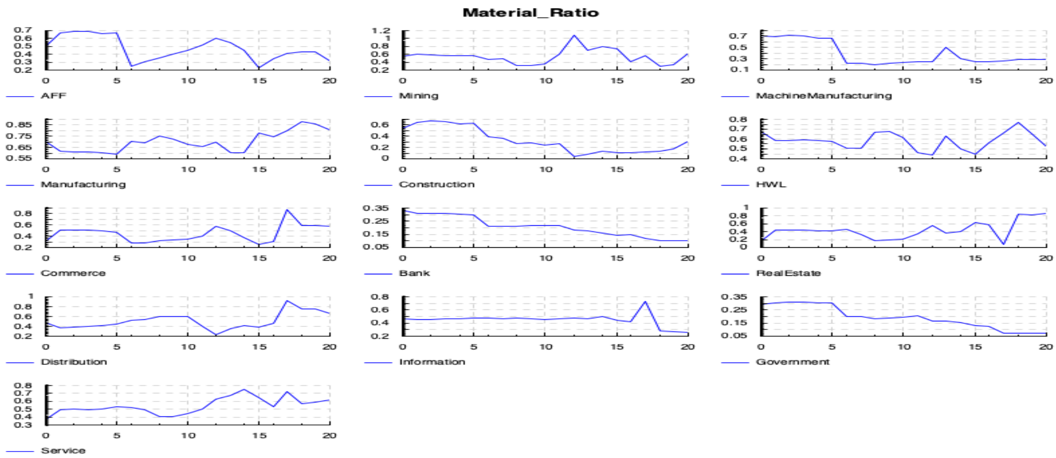

FIGURE 6.7: Raw materials ratio. Fluctuating Price. Optimized.

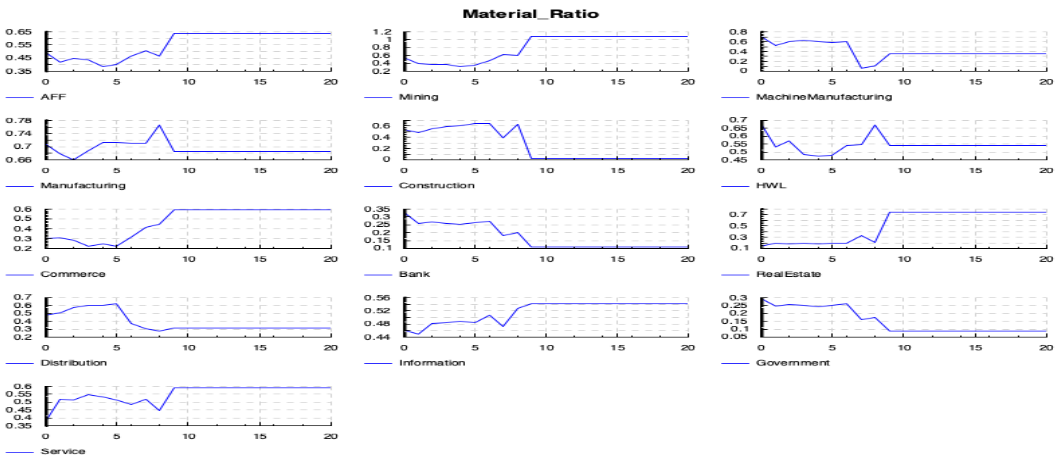

FIGURE 6.8: Raw materials ratio. Fluctuating Price. Not Optimized.

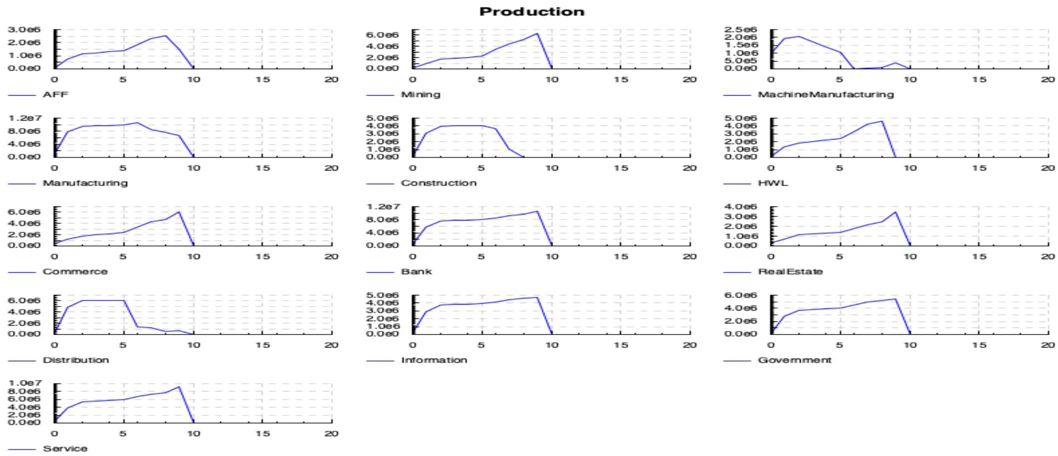

FIGURE 6.9: Production Path. Fluctuating Price. Not Optimized.

These analyses reveal that adding a price constraint to the standard Turnpike Path causes changes in the potential growth rate. Naturally, various constraints affect the potential growth rate, but the existing Turnpike model, which has only a simple state space and constraints, does not allow us to analyze the causes of the real economic path's failure to reach the Turnpike path.

As another example to see the impact of depreciation on output, Fig. 6.10 shows the path when the depreciation period is changed from the current five years to 10 years under fluctuating prices. Fig. 6.10 shows that doubling the depreciation period halves the depreciation cost and significantly reduces the impact of the capital goods constraint on output. This result shows that the price constraint affects output through the constraint on depreciation.

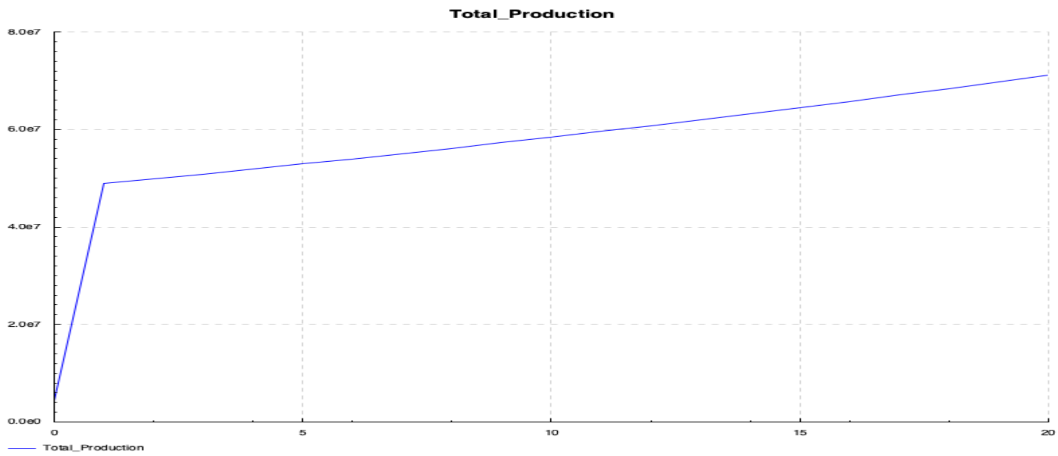

FIGURE 6.10: Total Production path. Fluctuating Price. Depreciation 10 years.

This section analyzes the maximized path changes by varying the price constraint and the capital constraint and shows that the potential growth path and the potential growth rate vary greatly depending on the introduced constraint conditions. We also show an example of a method to find the problem under each constraint condition by outputting various microdata such as the amount of capital investment and the ratio of raw materials.

In the next section, we discuss another aspect of the microdata analysis method proposed in this paper, the search for evaluation functions.

6.2.3 Compare evaluation function

We have already mentioned that our model allows us to explore the constraints and the evaluation function as a means of consensus building or analysis. This section explains the differences in growth paths due to differences in the evaluation function based on several examples. Fig. 6.11 shows the growth path of production when the final point maximization problem is solved instead of the path maximization problem so far.

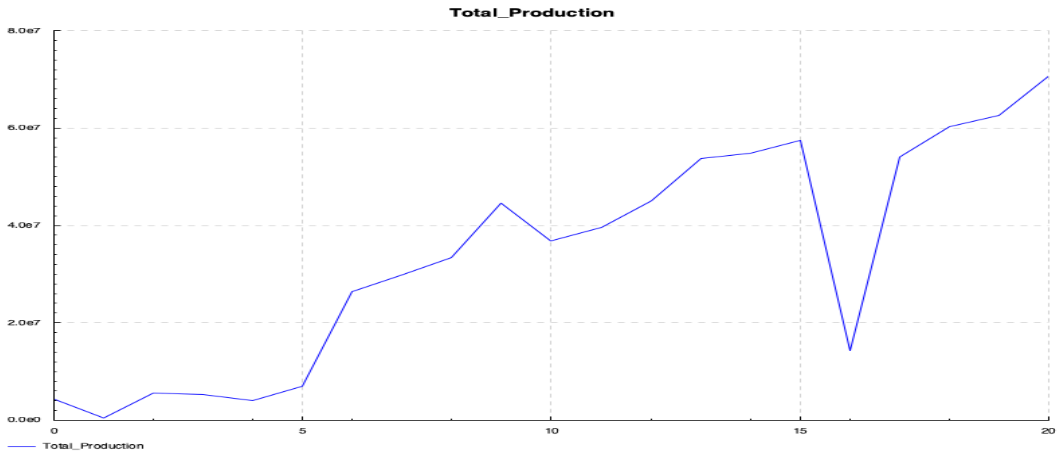

FIGURE 6.11: Total Production path. Fluctuating Price. Last Max.

Fig. 6.11 shows that output increases gradually without maximizing production under the labor constraint in the early term, as shown in Figure B. Besides, a temporary decrease in output appears in the term 16th because the model does not evaluate the intermediate stages. Also, a temporary decrease in output appears in the 16th period because the model does not evaluate the intermediate stages. This decrease is because the ratio of raw materials in the manufacturing industry exceeds 100% in term 16th, as shown in Fig. 6.12. In this case, the final point optimization is superior in terms of both average economic growth rate and final output, but path maximization is preferable from economic stability. The question of which path to choose as a policy objective is a political one. However, we can obtain new indices for analysis and decision making by, for example, adding price stability to the evaluation function of the final point optimization or by adding changes to the constraints on prices or the cost rate.

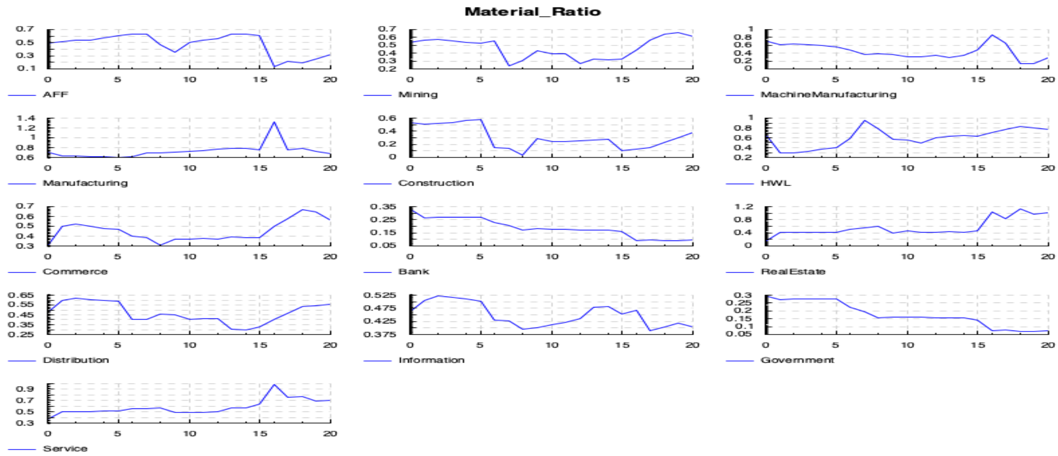

FIGURE 6.12: Raw materials ratio path. Fluctuating Price. Last Max.

Next, Fig. 6.13 and 6.14 show the actual GDP under the path maximization and the actual GDP under the constant price. So far, the analysis has focused on output, but when we shift our attention to GDP, we find that the picture is quite different. In the case of targeting GDP, we find that GDP does not grow much under price fluctuations. These figures show that industries do not generate profit under price fluctuation.

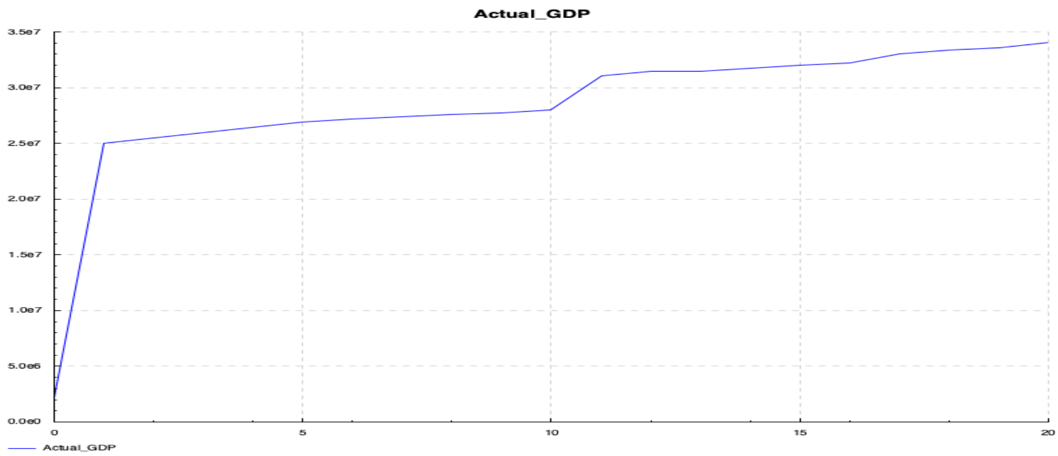

FIGURE 6.13: Actual GDP path. Constant Price. Path Output Max.

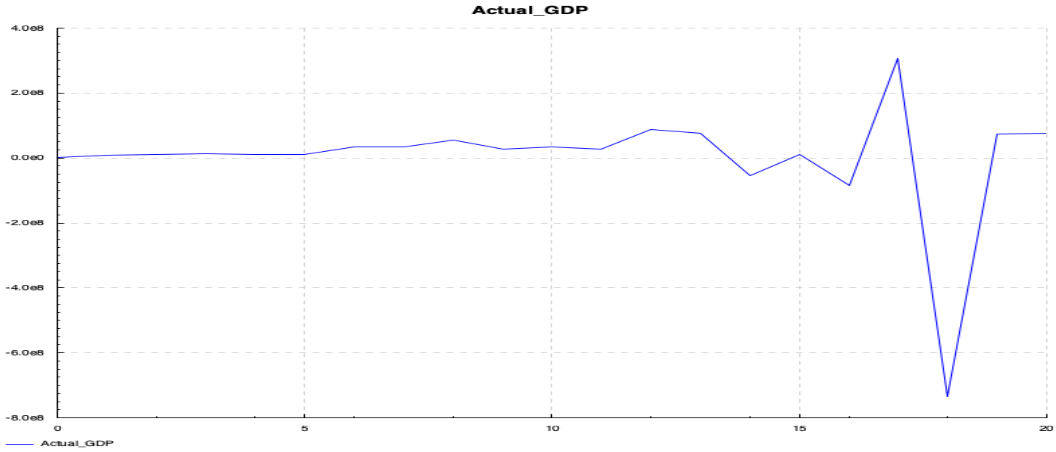

FIGURE 6.14: Actual GDP path. Fluctuating Price. Path GDP Max.

Fig. 6.15 and 6.16 show the changes in value-added by industry under fixed and fluctuating prices. Under the price fluctuation, specific industries' value-added often remains zero or negative, reducing the overall GDP. This is due to the deterioration of raw material ratios caused by price changes, as shown in Fig. 6.7. Under the price constraint in this paper, the maximizing output would result in a large deficit for these firms. From this point of view, evaluating the maximization of the output path under the current price constraint is also problematic from another perspective.

In this paper, we have extracted a few relatively simple indicators from a myriad of variations. However, since our model stores accounting information, we can use any economic or business indicator as to the subject of our analysis. Thus, by comparing all kinds of economic indicators, we can identify various problems that have not been analyzed in the existing models, and thereby we can discuss the constraints of the models and the appropriate treatment of the evaluation function.

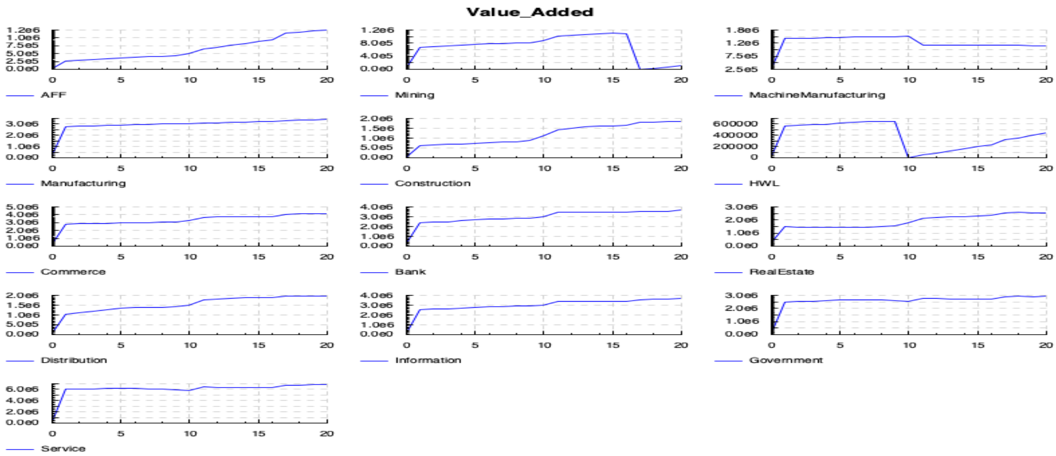

FIGURE 6.15: Value added path . Constant Price. Path Output Max.

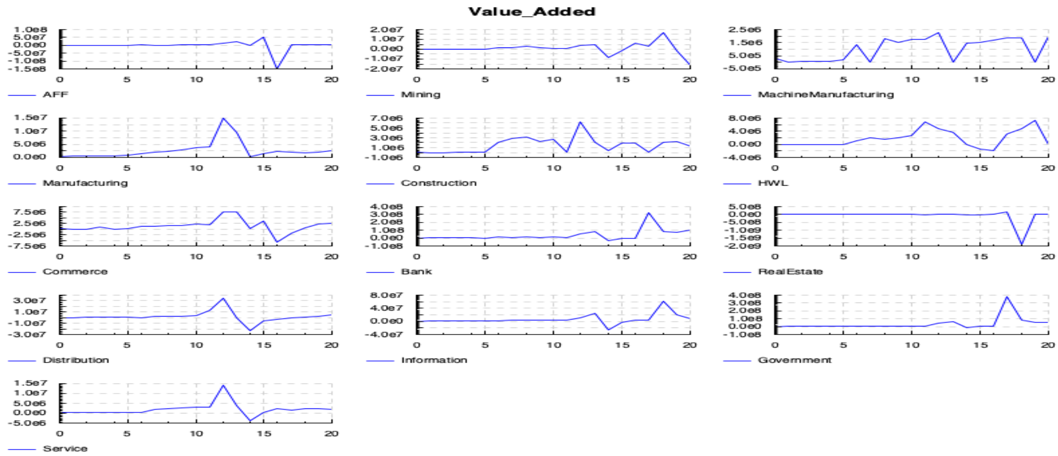

FIGURE 6.16: Value added path. Fluctuating Price. Path GDP Max.

6.3 Conclusions

As RTE progresses, a variety of Big Data will become available for statistical and economic analysis. The new data will enable a variety of analyses that the existing models do not cover. Increasing the frequency and immediacy of data updates is the central theme of RTE. By increasing the frequency of data updates, we can replace constant elements treated as the model's structure as updatable data. Also, because of its historical background, RTE places importance on data related to accounting and transactions, and in this sense, analysis using accounting data will be necessary for the future. In particular, transaction data such as invoices are currently attracting attention, and the use of such data will enable us to directly understand economic structures that we have not grasped so far.

This paper constructed an agent-based model with accounting data as an internal state and reproduced the economic structure based on actual transaction data as a model to handle such new data. This model holds each agent's accounting information as an internal state and can obtain and analyze all economic and management indices. Besides, since our model reproduces and records actual economic transactions, we can add various constraints related to actual economic transactions. Thus, by recording accounting information as state space, we can fully utilize the potential of accounting and trade information, which is not possible in general equilibrium or turnpike models.

We designed our model for Concentrated Model Analysis, which optimizes the whole model like the Turnpike model. As an example of analysis with and without constraints, this paper shows that the optimal path obtained in the same way as the ordinary Turnpike model differs greatly depending on price constraints and depreciation treatment. By changing various constraints like the analysis conducted in this paper, we can analyze various phenomena that we could not treat in the existing models. This function provides a tool to analyze a way to switch from some growth path to another more desirable, which was a lack of the Turnpike model.

In addition to the constraints, this paper's model also allows us to change the valuation function. In real economy management, the decision of the evaluation function is highly political. On the other hand, there have been no tools to analyze what trade-offs will occur if the goals are satisfied. In this paper, we show that, given a specific set of constraints, following the optimal path can lead to a company's management problems. Thus, this paper's model provides a tool to discuss the evaluation function's choice and the economic situation when the chosen evaluation function is optimized. In this regard, we expect to incorporate environmental indicators into the evaluation function, especially in energy accounting and waste cost calculation. When a variety of microdata are obtained and become the target of analysis, the search for constraints and evaluation functions based on various indicators' status is required. In this paper, we discussed the conditions required for the model and propose a model that satisfies them.

The model treated in this paper has only the minimum configuration to reproduce the turnpike model as an agent-based model, without several boundary conditions that we usually assume in reality, and the parameters are pseudo ones. However, we show that the system can handle many parameters and data that the existing models cannot control. It is necessary to enrich the range of boundary conditions, use these models in practice, and select and analyze the changes caused by each condition with actual data. The model in this paper provides a minimal basis for this purpose. Under such an extension, it enables us to perform analyses that were impossible to handle in the past. As mentioned above, we will release all the models in this paper on GitHub as OSS[8]. We strongly hope that this paper's models will be widely used in the same way as many predecessors have extended the general equilibrium models.

[8] https://github.com/yakagika/VirtualEconomy

Chapter 7

Conclusion and Discussion

This paper has focused on the new concept of economic statistics and analysis method, RTE, and discussed its social implementation. RTE is a concept that implies that through ICT and automation, economic data can be communicated and recorded in real-time and used for decision-making at various scales. Several countries have already realized the technological infrastructure and e-services for RTE, and the academic development of economic statistics and economic analysis methods using these technologies is currently underway. On the other hand, the concept of the RTE management layer involves a new paradigm that is quite different from the methods used in existing economics and economic statistics, and many problems need to be solved to realize it.

The digitization of government and new data for statistical estimation and policymaking will inevitably increase its demand soon. This new situation will require new technologies that differ significantly from existing methods. Currently, we are on the eve of the development of such new technologies.

Chapters 1 to 3 discuss the required techniques, and the following chapters describe several implementations. As the elemental technologies required to realize the management layer of RTE, this paper mainly focuses on using accounting data to utilize individual companies' information and on Data Assimilation and Model Re-estimation to realize real-time performance.

Chapter 4 describes the implementation of an automated statistical estimation system to update the obtained accounting data in real-time in the Supply Table, associated with Japan's statistical reform. The automation of statistical estimation is a seemingly simple task, but various limitations make it difficult in practice. In particular, the unification of processing methods and a system for information sharing, which have not been dealt with in-depth so far, need to be considered in future statistical production.Surprisingly, the production of economic statistics is still in the early stage of IT worldwide. Japanese Cabinet Office's survey says that many departments use Excel for estimation, and the systematization has not progressed yet.

In the future, it is necessary to consider the know-how of how to develop and maintain data, processes, programs, and documents in an open manner using technology, such as Git and other sharing tools. This point of view is related to technical issues and convenience and ethical issues in the future utilization of RTE. At present, technologies for a surveillance and management using personal information big data are becoming widespread in the world, and RTE is very close to the implementation of such technologies on a national level. Aside from scoring and monitoring, such as China's centralized personal credit index, there should be no black boxes in the evaluation and analysis methods when using data to monitor and control specific companies. The current methods of producing economic statistics unintentionally create this black box, and there is no institutional guarantee of opportunities for verification. In a fiscal stimulus to a specific company, we should fully clarify methods to culculate the stimulus's value. In particular, fluctuations in numerical values caused by detailed work procedures, such as whether the results are due to the interpolation of specific values or error handling, are trivial problems when measuring macroeconomic indicators at the national level but need to be controlled in microeconomic actions.

The use of accounting data for statistics is also a significant issue. The data acquired by RTE are data on economic transactions, such as invoices, and accounting data, such as tax data. The use of such accounting data has so far been for a single company, but it will be necessary to analyze such accounting data for larger units such as municipalities and nations in the future. In this sense, there is no doubt that the collaboration between accounting, economics, and statistical economics will develop in the future. However, it is not easy to directly apply such indicators to municipalities. It is necessary to develop various indicators and their analysis methods, which are equivalent to conventional accounting audits in units such as municipalities and nations, and incorporate them into the system according to the purpose of cooperation with some existing social accounting methods.

In Chapter 5, we discuss the statistical estimation techniques using the accounting data.The current SNA was designed based on the data and information processing technology available to realize the objectives of the 1960s, namely Keynesian economic policy. Today's information technology and the data available are far more advanced than those of 1960, and the objectives of the SNA are changing dramatically. We need to create new economic statistics that are optimal for the current purposes and environment. What we have done in Chapter 5 is to extend the current statistics with the current technology. Even for the current IOT, it is possible to improve the update interval and information granularity using new data. More governments and economists will promote such efforts to improve economic statistics' convenience in various future statistics. However, what we need to do now is to update the existing statistical methods and produce statistics in the new era.

The IOT that we have targeted in this paper has several significant structural limitations. For example, the IOT or the commodity flow method used to calculate GDP involves aggregation or folding that depends on the computing power at the time of invention. In the real world, the input-output relationship is not an inter-industry relationship but a network among firms and establishments. Such a change in granularity can be partially revised, as was done in this paper, but the fact remains that the network is in effect represented by a matrix of folded and abstracted units. Researchers have developed several methods have to analyze networks directly, and it is possible to take advantage of the characteristics of networks. In this sense, there is a need to incorporate into SNA, not an extension of existing statistics but a new type of economic statistics that overcomes current statistics' structural limitations.

Chapter 6 implements a method of accounting-based analysis of individual firms in the form of an agent-based model, which realizes Model Re-estimation by integrating statistics using accounting data and economic models.To analyze firms and decide governance for them, which is the objective of the management layer of RTE, we need to combine the existing analytical methods of macroeconomics with the analytical methods for micro-entities such as accounting and auditing. Simultaneously, it is necessary to construct a model compatible with the granularity of the various kinds of information obtained by RTE and incorporate real transaction constraints. For this purpose, we have developed an agent-based model that reproduces inter-firm transactions and in which each firm maintains its internal state as accounting information.

Since the model in this paper reproduces each economic agent's transactions and retains all the information associated with the transactions, it is possible to incorporate existing macroeconomic models into the model. In this sense, it will be necessary to analyze our model's implications for existing economics methods. Economists have not addressed methodologically the question of what the existing economic models have missed so far. Thus with our model, we should analyze fluctuations of various indices that have not been the subject of analysis in existing economic models. Furthermore, our model extracts analysis and decision-making of indicators from actual economic transactions and implies the union of economics and statistical economics. Usually, economics only concern using indices, and rarely with the methods of producing economic statistics. However, to utilize obtained data, we should design a consistent process from the statistical estimation method to the analysis, as treated in this paper.

Also, our model is very complex compared to macroeconomic models, but on the other hand, it is too simple to achieve the requirements of RTE. This simplicity is mainly related to the setting of boundary conditions and the number of agents. Our model can incorporate an infinite number of boundary conditions. Choosing the necessary boundary conditions to realize the economy's stability, such as price and employment determination mechanism, is challenging and necessary work. Therefore, the choice of conditions will itself be a major quest to understand the economic phenomena. Our model is, in this sense, an experimental device for clarifying the elements necessary for reproducing the actual economy.

To run our model on a natural system, we need to solve many system implementation problems. The RTE aims to analyze a nation on a firm-by-firm basis, but it is not easy to realize such a calculation in limited computational and administrative resources. Existing economic models are just users of the economic data, and they do not consider data production and management. New economics in RTE should incorporate economic methodologies into the existing ICT system. We plan to apply our model to various actual systems and data, but this will require extensive system design knowledge. In this sense, future economics will need to collaborate with a wide range of existing disciplines, such as ICT system design, accountings, and require significant reforms.

The concepts and elemental technologies we proposed in this paper to realize the management layer of RTE are only a part of the whole system, and all of them have many problems to be solved to run in real-time with actual data on and system. We are currently conducting experiments at the Chiba University of Commerce, together with Gofore inc. of Finland, to implement several systems on X-road, and are aiming to run some of the techniques introduced in this paper on an existing system basis although on a small scale. The system for economic statistics and analysis of a country is very complex and large-scale, and we expect to develop as a research field under the collaboration of industry, government, and academia. The authors hope that this paper will contribute to the development of this field.

Bibliography

[1] "Littrow configuration tunable external cavity diode laser with fixed direction output beam," vol. 72, no. 12, pp. 4477–4479, Dec. 2001. [Online]. Available: http://link.aip.org/link/?RSI/72/4477/1.

[2] L. Siegele. "The real-time economy: How about now?" (2002), [Online]. Available: https://www.economist.com/special-report/2002/02/02/how-about-now.

[3] M. A. Vasarhelyi, R. A. Teeter, and J. Krahel, "Audit education and the real-time economy," *Issues in Accounting Education*, vol. 25, no. 3, pp. 405–423, Aug. 1, 2010, ISSN: 0739-3172, 1558-7983. DOI: 10.2308/iace.2010.25.3.405. [Online]. Available: https://meridian.allenpress.com/iae/article/25/3/405/187425/Audit-Education-and-the-RealTime-Economy (visited on 01/07/2021).

[4] R. A. Web. "The real-time economy: The technological basis for reengineered business reporting." (2020), [Online]. Available: http://raw.rutgers.edu/node/29.html.

[5] K. .A, S. E. .F, and V. M. .A, "Continuous online auditing: A program of research," *Journal of Information Systems*, vol. 13(2), pp. 87–103, 1990.

[6] C. Victoria, L. Qi, Vasarhelyi, and M. A, "The development and intellectual structure of continuous auditing research," *Journal of Accounting Literature; Gainesville*, vol. 33, no. 1, pp. 37–57, Dec. 2014.

[7] H. Deguchi and S. Sakaki. "(in japanese) three proposals for the advanced utilization of statistical systems, document 8-3, working group meeting no. 4, basic planning subcommittee, statistics committee (12th meeting)." (2008), [Online]. Available: https://www.soumu.go.jp/main_sosiki/singi/toukei/2008wg/wg4/wg4_12/wg4_12.html.

[8] H. Deguchi, "(in japanese) the mathematical system for the economical exchange: The axiomatic foundation of exchange algebra and its extension for multi unit, subject and dimensional description," *Fukushima university, commercial science treatises*, 1988.

[9] ——, *Economics as an Agent-Based Complex System: Toward Agent-Based Social Systems Sciences*. Springer-Verlag Tokyo, 2004.

[10] K. Robert, K. Tarmo, A. Art, T. Maarja, S. Ralf-Martin, and S. Carsten, "Real-time economy: Definitions and implementation opportunities," *Tallinn University of Technology*, p. 68,

[11] ITL. "Vision of the estonian association of information technology and telecommunications (itl) of information society in 2030: Smart estonia." (2018), [Online]. Available: https://www.itl.ee/en/vision-2030/.

[12] F. Frank, "Flexi-business: Strategies for success in the roller-coaster decade," *The Globe and Mail*, 1989.

[13] J. Kitchen and R. Monaco, "Real-time forecasting in practice," *Business Economics; Basingstoke*, vol. 38, no. 4, pp. 10–19, Oct. 2003.

[14] Real-time Economics Conference. "Our history." (), [Online]. Available: https://real-time-economics.org/about-the-rte-conference/.

[15] K. Evan, S. Dolmas, and J. M. Piger, "THE USE AND ABUSE OF "REAL-TIME" DATA IN ECONOMIC FORECASTING," *FEDERAL RESERVE BANK OF ST. LOUIS*, p. 49, Aug. 2001.

[16] D. Croushore, "Frontiers of real-time data analysis," *Journal of Economic Literature*, vol. 49, no. 1, pp. 72–100, May 2011.

[17] D. Croushore and T. Stark, "A real-time data set for macroeconomists," *Journal of Econometrics*, vol. 105, no. 1, pp. 111–130, 2001. [Online]. Available: https://search.proquest.com/docview/196629637/772489EBC1C2432DPQ/2.

[18] A. Orphanides, "Monetary policy rules based on real-time data," *The American Economic Review*, vol. 91, no. 4, pp. 964–985, Sep. 2001.

[19] K. E. Dynan and douglas W. Elmendorf, "Do provisional estimates of output miss economic turning points?" *Federal Reserve Board of Governors working Paper*, Nov. 2001.

[20] Ministry of Internal Affairs and Communications. "(in japanese) basic plan for the development of official statistics, phase iii basic plan." (2020), [Online]. Available: https://www.soumu.go.jp/main_content/000690298.pdf.

[21] D. Laney. "3d data management: Controlling data volume, velocity, and variety." (2001), [Online]. Available: http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Varie pdf.

[22] M. Suga, "(in japanese) current status of special surveys (input surveys) for compiling input-output tables," *Business Journal of PAPAIOS*, vol. 17, no. 3, 2009.

[23] N. I. F. I. SOLUTIONS. "Nordic institute for interoperability solutions — x-road implementation models." (Apr. 2020), [Online]. Available: https://www.niis.org/blog/2020/3/30/x-road-implementation-models.

[24] e-estonia.com. "E-estonia guide." (), [Online]. Available: https://e-estonia.com/wp-content/uploads/eas-eestonia-vihik-a5-180404-view.pdf.

[25] Prime Minister's Office. "(in japanese) digital government action plan." (2020), [Online]. Available: https://www.kantei.go.jp/jp/singi/it2/dgov/201225/siryou4.pdf.

[26] Federation of Economic Organizations (Keidanren). "(in japanese) survey report on the actual condition of paperwork burden and improvement measures." (2000), [Online]. Available: https://www.keidanren.or.jp/japanese/policy/2000/018/honbun.html.

[27] National Tax Agency of Japan. "(in japanese) [e-tax] national tax electronic filing and payment system (e-tax)." (), [Online]. Available: https://www.e-tax.nta.go.jp/shiyo/.

[28] Cabinet Office of Japan. "(in japanese) for software developers (providing myna portal api) - cabinet office." (), [Online]. Available: https://www.cao.go.jp/bangouseido/case/business/developer.html.

[29] Computer Software Association of Japan. "(in japanese) electronic invoice promotion council | csaj computer software association of japan." (), [Online]. Available: https://www.csaj.jp/activity/project/eipa.html.

[30] Ministry of Internal Affairs and Communications. "(in japanese) formulation of uniform rules for notation of machine-readable data in statistical tables." (2020), [Online]. Available: https://www.soumu.go.jp/main_content/000723626.pdf.

[31] IT Strategic Planning Office, Cabinet Secretariat and Administrative Management Bureau, and Ministry of Internal Affairs and Communications. "(in japanese) report on cloud service procurement and its contracts in the second phase of government common platform." (2020), [Online]. Available: https://cio.go.jp/node/2704.

[32] European Commission, International Monetary Fund, Organisations for Economic Co-operation and Depelopment, United Nations, and World Bank, "System of national accounts," 2018.

[33] PEPPOL. "Peppol einvoicing is connecting europe." (Apr. 2019), [Online]. Available: https://peppol.eu/peppol-einvoicing-connecting-europe/.

[34] Estonian Banking Association. "Xml b2c & c2b communication messages." (), [Online]. Available: https://pangaliit.ee/settlements-and-standards/xml-b2c-and-c2b-communication-messages.

[35] K. Akagi, T. Ohsato, and H. Deguchi, "Input-output table constructed with private business data and its algebraic description," *IEEE/SICE International Symposium on System Integration*, 2015.

[36] K. Akagi, "(in japanese) the methodology for developing the supply table prototype," *the Economic and Social Research Institute, Cabinet Office, Government of Japan, ESRI Research Note*, vol. 57, 2020. [Online]. Available: https://www.esri.cao.go.jp/jp/esri/archive/e_rnote/e_rnote060/e_rnote057.pdf.

[37] Ministry of Economy, Trade and Industry. "(in japanese) method of industry classification." (), [Online]. Available: https://www.meti.go.jp/statistics/tyo/kougyo/result-4.html#menu06.

[38] Ministry of Internal Affairs and Communications. "(in japanese) summary of the results of the 2016 economic census - activity survey by industry (summary on manufacturing industry) (comparison with the results of the industrial statistics survey)." (2017), [Online]. Available: http://www.stat.go.jp/data/e-census/2016/kekka/pdf/hinsan.pdf.

[39] K. Sakuramoto, "(in japanese) a study on the annual supply use table in japan's national accounts system," *Economic and Social Research Institute, Cabinet Office, Government of Japan, ESRI Working Paper*, vol. 26, pp. 1 –139, 2012.

[40] Ministry of Internal Affairs and Communications. "(in japanse) "2011 industry input-output table." (2015), [Online]. Available: https://www.soumu.go.jp/main_content/000680591.pdf.

[41] ——, "(in japanese) handout (1) estimation of domestic production using data from the economic census - activity survey draft method and notes on decomposing subsector data from the economic census into io sectors i. basic calculation method." (2011), [Online]. Available: https://www.soumu.go.jp/main_content/000288067.pdf.

[42] S. Shishido, *(In Japanese) Handbook of the Input-Output Analysis*. Toyo Keizai, 2010.

[43] Statistics Bureau, Director-General for Policy Planning (Statistical Standards) and Statistical Research and Training Institute. "Outline of the input-output tables for japan." (), [Online]. Available: https://warp.da.ndl.go.jp/info:ndljp/pid/1621354/www.stat.go.jp/english/data/io/outline.htm.

[44] Economic and Social Research Institute and Cabinet Office, Government of Japan. "(in japanese) national accounts estimation methodology manual (annual estimation)." (2017), [Online]. Available: https://www.esri.cao.go.jp/jp/sna/data/reference1/h23/pdf/kaisetsu_a_20190730.pdf.

[45] Ministry of Internal Affairs and Communications. "(in japanese) input-output table manual." (), [Online]. Available: http://www.soumu.go.jp/toukei_toukatsu/data/io/011index.htm.

[46] T. Iwasaki, "(in japanese) input-output analysis and economic forecast," *Hokkaido University Economic Research THE ECONOMIC STUDIES*, vol. 30, no. 1, pp. 121–142, 1980.

[47] Z. Nakazawa, "(in japanese) the construction and its problems of regional input-output table at a local level," *Policy science*, vol. 9, no. 2, 2002.

[48] Minister of Economy,Trade and Industry. "(in japanese) trade and industry :inter-regional input-output table." (2010), [Online]. Available: https://www.meti.go.jp/statistics/tyo/tiikiio/result/result_02.html.

[49] e. a. Takaya Ohsato, "(in japanese) input-output table constructed with private business data and its algebraic description and it's verification," *the society of instrument and control engineers 2015*, 2015.

[50] FALCONSEED. (), [Online]. Available: http://www.soars.jp/category/download/falconseed/.

[51] Minister of Economy,Trade and Industry. "(in japanese) 2012 basic survey of japanese business structure and activities vol 2- table." (), [Online]. Available: http://www.meti.go.jp/statistics/tyo/kikatu/result-2/h24data.html.

[52] Ministry of Internal Affairs and Communications. "(in japanese) research for the input-output structure." (), [Online]. Available: http://www.soumu.go.jp/toukei_toukatsu/data/io/exam.html.

[53] Teikoku Databank, Ltd. "(in japanese tdb industrial classification." (), [Online]. Available: http://www.tdb.co.jp/lineup/pdf/tic.pdf.

[54] Ministry of Internal Affairs and Communications. "(in japanese) 2011 input-output table (confirmed report) producer's price table (108 departments)." (), [Online]. Available: http://www.e-stat.go.jp/SG1/estat/List.do?bid=000001060671&cycode=0.

[55] ——, "(in japanese) 2005 input-output table (confirmed report) producer's price table (108 departments)." (), [Online]. Available: http://www.e-stat.go.jp/SG1/estat/List.do?bid=000001019588&cycode=0.

[56] ——, "(in japanese) 2005 correspondence table of input-output table standard classification - japan standard classification." (), [Online]. Available: http://www.soumu.go.jp/main_content/000286910.pdf.

[57] ——, "(in japanese) 2011 input-output table confirmed table classification code table." (), [Online]. Available: http://www.e-stat.go.jp/SG1/estat/List.do?bid=000001060671&cycode=0.

[58] H. Ohomori, "(in japanese) sna input-output table and fixed input coefficients," *conomic and Social Research Institute, Quarterly SNA*, vol. 148, pp. 101–108, 2012.

[59] National tax agency of Japan. "(in japanese) sample business survey summary." (), [Online]. Available: http://www.nta.go.jp/kohyo/tokei/kokuzeicho/kaishahyohon2011/kaisya.htm.

[60] Ministry of Internal Affairs and Communications. "(in japanese) 2012 economic census for business activity(confirmed report)." (2012), [Online]. Available: http://www.stat.go.jp/data/e-census/2012/kakuho/pdf/gaiyo.pdf.

[61] e. a. K. Matsumoto, "(in japanese) economic growth and change in industrial structure," *Ministry of finance policy research institute, Financial review*, vol. 58, 2001.

[62] E. N. Wolff, "Computerization and structural change," *Review of Income and Wealth*, vol. 48, no. 1, pp. 59–75, 2012.

[63] O. Blanchard, "Do dsge models have a future?" *Peterson Institute for International Economics, Policy Brief*, 2016.

[64] R. A. Braun, L. M. Korber, and Y. Waki, "Some unpleasant properties of log-linearized solutions when the nominal rate is zero," *Bank of England,Staff Working Paperf*, 2012.

[65] R. Dorfman, P. A. Samuelson, and R. M. Solow, *Liner Programming and Economic Analysis*. DOVER PUBLICATIONS, 1958.

[66] P. A. Samuelson, "A catenary turnpike theorem involving consumption and the golden rule," *The American Economic Review*, vol. 55, no. 3, pp. 486–496, Jun. 1965.

[67] R Radner, "Paths of economic growth that are optinal with regard only to final states," *Review of Economics Studies*, Jan. 1961.

[68] M Morishima, "Proof of a turnpike theorem:theno joint production case," *Review of Economic Studies*, Feb. 1961.

[69] Mckenzie, L.W, "Turnpike theorems for a generalized leontief model," *Econometrica*, vol. 31, no. 1-2, Jan. 1963.

[70] J Tsukui, "Turnpike theorem in a generalized dynamic input-output system," *Econometrica*, 1966.

[71] Y. Murakami and e. kazuhiko Tokoyama, "(in japanese) optimal growth path for the japanese economy," *Economic Planning Agency of Japan, Economic Analysis*, no. 30, 1970.

[72] W. B. Arthur, "Out-of-equilibrium economics and agent-based modeling," *SFI WORKING PAPER*, 2005.

[73] T. Ohosato, K. Akagi, and H. Deguchi, "(in japanese) construction of an input-output table that takes into account business-to-consumer transactions using private data," *The 29th Conference of PAPAIOS*, 2018.