A detailed method to estimate inter-regional capital flows using inter-firm transaction and person flow big data

Yuki Akiyama¹ · Yohei Yamamoto² · Ryosuke Shibasaki¹ · Hodaka Kaneda³

Received: 29 September 2018 / Accepted: 25 July 2019 / Published online: 31 August 2019
© The Author(s) 2019

Abstract

This study develops a method to estimate inter-regional capital flows in the real economy across Japan, which is important for understating the regional economy with as much detail as possible, and that existing censuses do not monitor adequately. The method can monitor spatial distributions of three kinds of capital flow: inter-firm transactions (F2F), firm-to-consumer flows as salaries (F2C), and consumer-to-firm flows as consumption behavior (C2F) using inter-firm transaction big data (IF data) and person flow big data (PF data). First, we estimated F2F using IF data that reports all capital flows between a company's headquarters. Second, we estimate the home, work place, and consumption locations of all consumers using PF data collected by the auto-GPS function on mobile phones. Third, we estimate the F2C of each consumer by estimating their salaries based on the locations of firms in the IF data. Finally, we estimate the C2F of each consumption area by estimating the consumption of F2C in each consumption location. We develop a dataset that can monitor micro-scale capital flows across Japan. This study is one of the first attempts in any country to develop a dataset that can monitor the spatial distribution and time-series changes of nationwide capital flows at a very high-resolution scale. We expected that our data will contribute to policy-making to address the socio-economic problems in the private sector and among local governments in Japan.

Keywords Big data · Regional economy · Inter-regional capital flow · Inter-firm transaction · Mobile phone · Mass person flow

Yuki Akiyama and Yohei Yamamoto contributed equally to this work.

✉ Yuki Akiyama
aki@csis.u-tokyo.ac.jp

¹ Center for Spatial Information Science, The University of Tokyo, Kashiwa Research Complex 4th Floor, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8568, Japan

² Graduate School of Frontier Science, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8568, Japan

³ ZENRIN-Datacom CO., LTD, Shinagawa Intercity Tower C 6F, 2-15-3 Konan, Minato-ku, Tokyo 108-6206, Japan

JEL Classification C81: Methodology for collecting, estimating and organizing microeconomic data · Data access

1 Introduction

In recent years, Japan's economy stagnated due to depopulation, the declining birth-rate, and population aging, especially in local cities and rural areas. Thus, Japan's national and local governments advanced a policy called "Regional Revitalization", which aims to curb depopulation and to revitalize regional economies. In this policy, one of the national government's key actions is quantitative monitoring of the current situation in the regional economies using big data. For example, the Japanese government launched the Regional Economy and Society Analysis System (RESAS) in 2015 (Cabinet Secretariat and Ministry of Economy, Trade and Industry 2015). RESAS is an online visualization system for national and regional data, such as economic data in the national census, population, number of living convenience facilities, and so on. The Japanese government aims to use RESAS to support policy-making in national and local governments based on various census and GIS data. There are some examples in which the national and local governments failed to enact appropriate policies because experience and intuition currently influences policy-making (Ishida and Matsutani 1994). Therefore, the Japanese government is now considering how to make appropriate policies based on evidence such as census and data. In other words, it aims to adopt evidence-based policy-making (EBPM) by using RESAS and other sources (Cabinet Secretariat 2016; Morikawa 2018). Thus, Japan is beginning to recognize that the EBPM by national and local governments is indispensable to activating local economies.

What data should we monitor to help revitalize local economies? Although there are many factors that seem to affect local economies, this study develops a method to monitor inter-regional capital flows at the lowest level possible, which is very important to understanding the current situation of the real economy, which affects the local economy significantly.

1.1 Inter-regional capital flows

It is difficult to understand the current state of a local economy using regional economic analysis only for a target area because regional economies are currently developing mutual-dependence relations among them by mass and complicated networks. For example, Milgram (1967) proposed the "small world phenomenon", in which we could meet anyone worldwide to connect human networks. In addition, Watts and Strogatz (1998) mention the ability to analyze network structures overall and locally to address connection points as nodes and the relation of connections as links in various kinds of mutual-dependence relations.

These studies note the need for a new dataset to be able to monitor inter-regional capital flow as network data at the lowest level possible to understand today's local economies as being a part of mass networks. The new dataset should contain

micro-capital flow networks: inter-firm capital flow (F2F), firm-to-consumer capital flow (F2C), and consumer-to-firm capital flow (C2F). This data should aid national and regional macroeconomic analyses.

1.2 Capital flow in the real economy

The economy consists of the “real economy” and the “financial economy” (Stiglitz and Walsh 2006). The former includes capital exchanges via the purchase and sale of goods and services, for example, inter-firm transactions and consumer shopping behavior. The latter includes the activities to increase capital using capital, such as the operations of stocks, bonds, insurance, and so on. C2F activities include capital such as consumers’ salary consumption for goods and services, and asset formation through investments in stocks, estates, and so on. C2F in the real economy includes consumption at retail shops or restaurants in the real space and via e-commerce (EC) (Fig. 1). In addition, it is defined in economics as capital flow in the real economy by exchange as deposits of good and services. Generally, the regional economy consists mainly of the real economy in real space (Anselin 1988; Fingar and Aronica 2001; Gökmen 2011). Thus, monitoring capital flows in the real economy in the real space is significant to quantifying the activation level of a regional economy.

1.3 Prior studies on regional economic monitoring

Some censuses and data on inter-regional capital flows exist. The most popular census of Japan is the Inter-Regional Input–Output Table (IRIOT) (Ministry of Economy, Trade, and Industry (MITI) 2011). It is possible to analyze the degree of economic dependence on a specific region and the spillover destination of the economic effect of economic activity in a specific region. For example, the MITI (Ministry of International Trade and Industry 1967) estimates the economic effect of world expositions in the past, and Hiramatsu (2017) estimates the economic effects of foreigners visiting Japan in recent years using the IRIOT. In addition, some related studies outside Japan examine the development of an international IRIOT for use in various countries (De Backer and Yamato 2008) and estimate waste emissions using the IRIOT in the UK (Salemdeeb et al. 2016), and so on. Thus, we use the IRIOT to

Fig. 1 F2F, F2C, and C2F position in the real economy in this study

estimate and clarify the economic activity between certain domestic and international economies and related human activities.

However, the IRIOT is developed only once every few years; in Japan this frequency is once every 5 years. Therefore, immediate regional economic analysis is impossible and it is difficult to make policies to revitalize regional economies and verify the effectiveness of the implemented policies using the IRIOT. Moreover, the Japanese IRIOT is an input–output table for eight regions. Therefore, it is currently difficult to use it for regional economic analysis among prefectures and between municipalities. To address this problem, Nakazawa (2002) compares the methods to create the IRIOT and suggests that pushing detailed investigations to local governments possibly increases the cost to develop the IRIOT and inaccuracies due to a simplified investigation method. However, the IRIOT is currently the only government statistics with estimates based on geographical factors of inter-regional capital flow in Japan. In other words, it is only possible to review an intermediate input tabulated at a very sparse spatial scale in Japan even today.

How can we solve this fundamental problem? If we can monitor capital flows at a finer level than the current eight regions, then we can monitor the capital flow from each firm, from each firm to each consumer, and from each consumer to each firm, thus creating a non-aggregated IRIOT not bound by an aggregation unit. In addition, it is ideal to maintain data at a high update frequency.

1.4 Objective of this study

This study develops capital flow data to clarify the micro-capital flow between regions as much as possible to solve the above problem. In addition, this paper is based on Yamamoto et al. (2017) and Yamamoto (2018) and summarizes and improves a series of their research results so far. Figure 2 shows the structure of this paper. The data integrate existing census data, inter-firm transaction big data (IF data), headquarter and branch big data (HB data), and person flow big data (PF data)

Fig. 2 Flow of this study

based on movement records from mobile phones. Section 2 estimated F2F to utilize IF data. Section 3 estimates F2C by estimating all consumer salaries to utilize HB data, PF data, and census data. Section 4 estimates C2F by estimating the consumption locations based on PF data. Consequently, we estimate the overall capital flow of F2F, F2C, and C2F. Section 5 shows example uses of this data. Finally, Sect. 6 concludes.

2 F2F estimation method

2.1 Inter-firm transaction big data (IF data)

Inter-firm transaction big data (IF data) covers various attributes of a headquarters (capital, number of employees, address, etc.) and transaction information between the headquarters (transaction item, estimated transaction amounts, etc.) maintained by Teikoku Databank, Ltd. (TDB) (Fig. 3). The IF data contains about 5.5 million transactions of about 1.5 million companies for a single fiscal year and is updated annually. This is the largest IF data set in Japan. This data was collected across Japan for the purpose of a corporate credit survey, and nearly all information is collected via interviews with companies by a large number of investigators. In addition, the data also stores spatial distribution information of the headquarters and branches of all firms (HB data) involved in transactions. Table 1 shows the comparison of the HB data headquarters information in 2014 with the company information recorded by the economic census in 2014 (Statistics Bureau, Ministry of Internal Affairs and Communications 2014). It shows that the number of firms by size and their ratios almost match. However, it is necessary to keep in mind that IF data and HB data store information only in Japan and do not store import and export transaction amounts, including overseas enterprises.

Fig. 3 IF data structure

Table 1 Comparison of the number of headquarters in the HB data and the number of companies in the economic census

Number of full-time employees	Economic census (2014)		HB data (2014)		v1/v2
	Number of enterprises (v1)	Rate [%]	Number of enterprises (headquarter) (v2)	Rate [%]	
0–4	992,706	56.72	957,457	57.88	1.0368
5–9	302,876	17.31	292,061	17.66	1.0370
10–19	210,173	12.01	186,121	11.25	1.1292
20–29	80,987	4.68	70,351	4.25	1.1512
30–49	67,498	3.86	59,052	3.57	1.1430
50–99	49,987	2.86	45,268	2.74	1.1042
100–299	32,169	1.84	31,438	1.90	1.0233
300–999	10,133	0.58	9716	0.59	1.0429
1000–1999	1927	0.11	1588	0.10	1.2135
2000–4999	1091	0.06	809	0.05	1.3486
5000 ~	524	0.03	318	0.02	1.6478
Total	1,750,071	100.00	1,654,179	100.00	1.0580

2.2 Transaction value estimation method

IF data does not show the transaction amounts of all transactions. Therefore, we must estimate unknown transaction amounts using known transaction data. We follow Tamura et al. (2012) to estimate the amount of transactions using a gravity model focused on the size of the company and the volume of transactions of two companies that conduct a transaction to resolve this problem. They calculate the estimated transaction amounts of all transactions from corporate sales among business partners by clarifying that the gravity law works for the relationship between transaction amounts and corporate sales.

2.3 Reliability verification of the estimated transaction values

Figure 4 shows the results for comparison of the total amount of transactions included in the IF data with the amount of inter-regional capital flow in the IRIOT. The results indicate that we can estimate 89.15% of the total nationwide capital flow of the IRIOT following Tamura et al. (2012) method. Both values are close, and we can thus conclude that the IF data have high reliability for macroscopically understanding the F2F capital flow in Japan. In addition, we consider that the capital flow of about 11% that the IF data do not include apply to small firms such as individual business owners that do not fall under the corporate credit survey.

Next, we focus on intra-regional transactions in which the transaction amount is particularly large in the IRIOT (Fig. 4b). Both values are close in many regions. We assume that the F2F in the Kanto and Chugoku regions is smaller than the intra-regional transactions in the IRIOT because the IF data do not include the proportion of small-scale firms is higher in these areas than in other areas. Moreover, the

Fig. 4 Comparison of the total amount of capital flows throughout Japan and its regions in the 2011 IRIOT and the estimated total transactions in 2011 obtained from the IF data and the eight regions in the IRIOT

Chugoku region contains many branch offices of large companies, and we assume that many transactions from branch offices actually occur. However, the IF data we use contains only the headquarters' transactions. We expect that this will improve by examining methods to break headquarters transactions into transactions between branches in the future.

In addition, we also validated inter-regional transactions (Fig. 4c). The result shows that IF data can also grasp the amount of inter-regional capital flows relatively accurately. Almost all inter-regional capital flows calculated by IF data are little bit smaller than values of IRIOT. The reason assumed that IF data do not include apply to small firms and it cannot grasp their transactions completely. On the other hand, inter-regional transactions from Kanto to Kinki region and Kinki to Kanto region calculated by IF data are larger than values of IRIOT. This is

because transactions of IF data contains only the headquarters' transactions, and there are many headquarters in Kanto region which includes Tokyo: the world largest metropolitan area and Kinki region which includes Osaka and Keihanshin metropolitan area: second largest metropolitan area of Japan.

For more reliable variations, it is needed to compare much spatially finer statistics and data. However, as we describe in Sect. 1.3, IRIOT is only available data for validation of our data now. If finer statistics and data will be released in the future, we would like to perform further high-definition validation.

Thus, it is possible to estimate the F2F per company across Japan. Figure 5 shows the total amount of capital flows by F2F per 1 km square grid in 2013. Previously, we could only see the capital flows between the eight regions in Fig. 4. On the other hand, we can estimate micro-F2F data, as Fig. 5 shows.

3 F2C estimation method

This section shows the F2C estimation using PF data. We estimate consumers' home, work place, and consumption points using PF data, and then estimate consumers' salaries based on the estimated work places.

Fig. 5 Total amount of capital flows by F2F per 1 km square grid in 2013

3.1 Person flow big data (PF data)

We use the “Konzatsu-Tokei[®]” (direct translation: “Congestion Analysis”) provided by ZENRIN DataCom Co. LTD. (ZDC) to estimate consumers’ home, work place, and consumption points. The data include people flow data collected by individual location data sent from mobile phones with enabled AUTO-GPS functions under users’ consent through the “docomo map navi” service provided by NTT DOCOMO, INC. Those data are processed collectively and statistically to conceal the private information. The original location data are GPS data (latitude, longitude) sent about every 5 min and do not include information specifying an individual’s characteristics such as gender or age. This is a big data for about one million users with about nine billion text records per year. Considering the protection of private information, we received a set of comprehensively processed GPS data from ZDC at the request of NTT DoCoMo. We received only aggregated results.

3.2 Detection of stay points and magnification factor calculation

The person flow big data in this study is not an aggregated data (point data). However, we need to detect stay points, in other words, home, work, and consumption points for calculation of F2C and C2F. Therefore, we detected stay points. First, we calculated all users’ stay points following Horanont (2010) method. Second, we divided all stay points into the locations of home, work place, public transportation, and others following Akiyama et al. (2016), as in Fig. 6. The red and green triangle points are work places and consumption places, respectively. In addition, all stay points have locations (longitude and latitude) as well as the start of the stay and its duration. Finally, we gave each user a magnification factor to estimate the actual

Fig. 6 Method to estimate home, work place, and consumption points using PF data

Japanese population. We determined this proportionally by dividing the residential population of a 1 km square grid unit obtained from the Japanese population census, since each user already has an estimated home location. Equation (1) gives the magnification factor for each user (consumer):

$$M_{ji} = \frac{P_j}{n_j}, \quad (1)$$

where M_{ji} is the magnification factor of user i who has their home in grid j , P_j is the residential population of grid j , and n_j is the number of users with their home in grid j .

Figure 7a shows the comparison of residential population by the population census in 2010 and the number of home locations by PF data in 2012, and Fig. 7b shows the number of employees by the economic census in 2012 and magnified work place population by PF data in 2012. They show that PF data is possible to express the trend of nighttime and daytime population distribution in Japan fairly accurately although the data is about one million people data (about 0.76% of the population of Japan). Therefore, PF data is sufficiently representative in expressing Japanese demographics.

3.3 Estimating the annual salary of each firm

We estimated the average annual salaries for each company using Yamamoto et al. (2017) method. First, we collected information about average annual salaries by prefecture, industry type, and employment type in 2012 from the Basic Survey on Wage Structure by the Ministry of Health, Labor, and Welfare (2012). The census is the

Fig. 7 Comparison of residential population by the population census in 2010 and the number of home locations by PF data (a) and the number of employees by the economic census in 2012 and magnified work place population by PF data (b)

fundamental statistics for each year and aims to investigate and clarify the actual wage structure in Japan. Based on the census information, we prepared the average annual salaries of regular and non-regular employees by prefecture and business category. However, the census does not include the average annual salaries of government employees. Therefore, we collect this data from the Survey on the Actual Salaries of National Government Employees in 2012 by the National Personnel Authority (2012) and that for local government employees in each prefecture from the Survey on the Actual Salaries of Local Government Employees in 2012 by the Local Administration Bureau, Ministry of Internal Affairs and Communications (2012).

Then, we combined the annual salary and HB data. Since the business categories of the Basic Survey on Wage Structure and HB data do not necessarily agree, we matched them based on the Japan Standard Industry Classification (Ministry of Internal Affairs and Communications 2013). The HB data contains the number of regular and non-regular employees for each company. Therefore, we estimated the average annual salary of each company using Eq. (2):

$$S_{pi} = \frac{\text{Pr}_i sB_p + \text{Pn}_i sb_p}{\text{Pr}_i + \text{Pn}_i}, \quad (2)$$

where S_{pi} is average annual salary of firm i in prefecture p , Pr_i is the number of regular employees of firm i , sB_p is the average annual salary of business category B corresponding to a regular employee of company i in prefecture p , Pn_i is the number of non-regular employees of firm i , and sb_p is the average annual salary of business category b corresponding to a non-regular employee of company i in prefecture p . Using Eq. (2), we estimated the average annual salary of all headquarters and branches in the HB data.

3.4 Estimating consumers' annual salaries

Finally, we estimate the annual salaries of all consumers. We combine the estimates of consumers' work places and the location and average annual salary information of all headquarters and branches from the HB data and accumulate these to a 1-km square grid. Equation (3) thus gives the annual salary of a consumer.

$$u_j = \frac{\sum_{k=1}^m \{Sj_k (\text{Pr}_k + \text{Pn}_k)\}}{\sum_{k=1}^m (\text{Pr}_k + \text{Pn}_k)}, \quad (3)$$

where u_j is the estimated annual salary of consumers who work within mesh j , m is the number of firms (headquarters and branches), and sj_k is average annual salary of firm k in grid j . We can thus estimate the salary, that is, the F2C, of each consumer in Japan. Figure 8 shows the total average annual salary, that is, the F2C of 1-km square grids across Japan for 2012. We can use non-aggregated person-scale data to estimate micro-F2C, as Fig. 8 shows.

Fig. 8 Total amount of F2C (average annual salary) in the residential area per 1 km square grid in 2012

In addition, we validated the F2C. Statistics that can obtain income information at the highest spatial resolution in Japan is the “Investigation of the municipal taxation situation (1975–2013)” (Ministry of Internal Affairs and Communications 2015). Using the statistics, we can know the average taxable income per person per municipality (per ward as far as Tokyo prefecture). Therefore, we compared the average taxable income for each municipality (2012) obtained from this statistics with the average salary for each municipality estimated from the F2C data (Fig. 9). Correlation coefficient is 0.7334 and adjusted *R*-square is 0.6379. It means that there is a clear positive correlation between them, and the F2C can estimate consumers’ annual salaries with relatively high reliability. However, in many municipalities, the F2C tended to be larger than the statistical value. The first reason is considered to be that the firm data used in this study contains only the headquarters’ office. If it is possible to use the location data of the branch offices in the future, it is expected that it will be possible to allocate the salary taking into consideration the difference between the headquarters’ and branch offices. In addition, the salary of a large company, which is expected to have a large number of employees in each prefecture, may be affecting the average salary of the prefecture because the salary allocated to firms in Sect. 3.3 was the average by industry by prefecture. It is expected to estimate average salaries of each industry for each municipality by weighting average salaries of each industry by prefecture and distributing them to the municipality

Fig. 9 Comparison of the average taxable income per municipality in 2012 and the average salary per municipality estimated by F2C

unit using the statistics used in this section and the population for each municipality obtained from population census, etc. It is expected that more reliable salary estimation will be possible by using this estimation result.

4 C2F estimation method

We next estimate the spatial distribution of C2F based on the spatial distributions of the consumption points for all consumers using the PF data. Using the consumers' annual salaries from the previous section, we can estimate the spatial distribution of C2F if we can clarify the locations and amounts of salary consumption. Salaries are capital that becomes the source of personal consumption for consumers. However, we do not turn all wages into consumption. Not all salaries are consumed and consumers would save a part of it. In addition, consumption behavior includes EC, so we must estimate C2F considering the influence of EC. In other words, to estimate the realistic spatial distribution of C2F, we need to know the amount of the estimated salaries used for consumption and the amount consumed in consumption areas in real spaces such as shops and restaurants.

4.1 Estimating annual consumption in the real space of each consumer

First, we calculated consumption expenditures and EC consumption according to the income of each consumer using The Family Income and Expenditure Survey by the Statistics Bureau (2018), which measures consumption trends in households via a sample survey four times annually for households across Japan. Therefore, we can also know the seasonality, which greatly affects consumption in Japan. Table 2 shows the annual consumption expenditure by annual income and annual EC consumption in 2012 from this survey. We obtain the estimated annual consumption expenditure in the real space of consumer i using Eq. (4) and the values in Table 2:

Table 2 Average annual EC consumption from the Family Income and Expenditure Survey in 2012

Annual income [million yen]	Average consumption [yen]					Annual e-commerce consumption [yen]
	Jan.–Mar.	Apr.–Jun.	Jul.–Sep.	Oct.–Dec.	Annual total	
~ 2.0	409,458	354,321	364,740	436,935	1,565,454	23,016
2.0–2.5	570,015	665,598	606,234	554,751	2,396,598	42,336
2.5–3.0	657,009	591,084	606,021	727,221	2,581,335	
3.0–3.5	656,067	698,697	614,211	641,217	2,610,192	62,040
3.5–4.0	688,026	724,572	694,476	732,339	2,839,413	
4.0–4.5	737,862	749,322	770,439	737,364	2,994,987	81,288
4.5–5.0	858,777	769,338	734,691	748,509	3,111,315	
5.0–5.5	806,307	765,567	767,265	814,089	3,153,228	83,880
5.5–6.0	912,744	821,301	830,988	833,721	3,398,754	
6.0–6.5	855,897	846,027	811,194	918,984	3,432,102	91,836
6.5–7.0	983,004	967,491	937,320	948,279	3,836,094	
7.0–7.5	1,025,508	951,573	893,826	981,261	3,852,168	101,700
7.5–8.0	1,039,863	1,036,104	1,072,029	1,097,739	4,245,735	
8.0–9.0	1,071,879	1,060,296	1,097,904	1,110,882	4,340,961	125,856
9.0–10.0	1,164,939	1,173,891	1,126,020	1,147,632	4,612,482	134,184
10.0–12.5	1,320,018	1,260,588	1,260,933	1,443,297	5,284,836	158,400
12.5–15.0	1,453,701	1,390,278	1,560,750	1,424,283	5,829,012	169,296
15.0 ~	1,834,560	1,654,746	1,502,607	1,882,245	6,874,158	265,236

$$CR_i = S_g - E_g, \quad (4)$$

where CR_i is the estimated annual consumption in the real space of consumer i , S_g is the average annual consumption of annual income category g , and E_g is the average annual EC consumption of the annual income category g . By applying the calculation to about one million consumers in the PF data, we estimated CR_i for all consumers.

4.2 Estimating the consumption at each consumption point for each consumer

Second, we need to clarify where and how much salary is consumed in the real space. The PF data already clarifies the consumption points for each consumer. Therefore, we estimated the consumption expenditures of all consumption points by dividing the salary consumed proportionally between the real space by the stay durations of each consumption point as a weight. Equation (5) gives the consumption expenditure at consumption point j for consumer i :

$$cr_{ij} = CR_i \frac{t_j}{\sum_{k=1}^n t_k}, \quad (5)$$

where cr_{ij} is the consumption expenditure at consumption point j for consumer i , t_j is the stay duration at consumption point j , and n is the number of annual consumption points of consumer i .

Furthermore, both in Japan and in other countries, there is a strong correlation between staying behavior and consumption behavior (Brockmann et al. 2006; Nakagawa et al. 2012). Therefore, we expect that the method to estimate the consumption expenditure at a consumption point according to Eq. (5) is valid.

4.3 Estimating C2F in the real space on an arbitrary aggregation unit

Finally, we calculate the total amount of expenditure in the real space by consumers in arbitrary spatial units such as municipalities and grids. Equation (6) gives $C2F_i$ in spatial unit i in a certain period:

$$C2F_i = \sum_{k=1}^n M_k cr_k, \quad (6)$$

where n is the number of consumption points in spatial unit i in a certain period, M_k is the magnification factor of the consumer at consumption point k , and cr_k is the estimated consumption expenditure in spatial unit i in a certain period.

It thus becomes possible to estimate the spatial flow from consumers to firms for each consumer unit by estimating the capital flow from consumers earned salaries (F2C) to C2F in the real space. Figure 10 shows the annual C2F in the real space for every 1-km square grid in Japan from 1 January to 31 December in 2012. By using non-aggregated person-scale data, we can estimate micro-C2F, as in Fig. 10.

4.4 C2F reliability verification

We verified the reliability of C2F by comparing it to existing statistics, specifically the commercial statistics of 2014 per 1 km square grid from the Statistics Bureau (2014), from which we can obtain the retail sales value per 1 km square grid. Although C2F includes consumption besides that in retail stores, there are no statistics in Japan covering the spatial distribution of values similar to our estimated C2F. However, we estimate that consumption activities occur in regions with high C2F and that consumption in the retail industry is vigorous. Therefore, we compared our C2F to the retail sales listed in the commercial statistics.

Figure 11 shows the comparison of C2F in 2012 and the retail sales in 2014 for every 1-km square grid throughout Japan. We find a positive correlation between the two values. In other words, C2F was large in areas with active consumption and the spatial distribution of C2F has some degree of reliability. Many of the areas in which we overestimate C2F (A in Fig. 11) are grids with large railway terminals located in urban areas. Large railway terminals provide fewer consumption opportunities relative to the stay duration. Specifically, there are few opportunities to purchase expensive goods but many opportunities for consumption by eating and drinking, so it reflects the result. On the other hand, many areas in which we underestimate C2F

Fig. 10 Total C2F per 1 km square grid in 2012

Fig. 11 Comparison of C2F and annual retail sales value in the commercial statistics per 1 km square grid across Japan

(B in Fig. 11) are grids located in business districts and mixed land use areas that include residential and shopping areas. In these areas, we should be able to improve the method of PF data processing because it is possible that the stay points of consumption behavior were incorrectly determined as stay points at homes or work

places. However, since we find only retail sales amounts in the commercial statistics, future work should verify the reliability using statistics and data that cover sales in the food service and other service industries in the future.

5 Applications

This section introduces examples of applications of the capital flow data developed in this study.

5.1 Support of regional economic promotion focused on consumption and annual income

We expect that many consumers visit areas with high C2F or that high-income consumers frequently visit these areas. Figure 12 shows the relationship between the annual C2F for 1-km square grids in 2012 and the total annual income of consumers who visited each grid in 2012. They show a strong positive correlation. However, in some areas, the C2F was too small or too large for the regression line. Representative examples of underestimated areas were transport hubs such as Tokyo terminal, Osaka terminal, and Haneda Airport (Tokyo International Airport); business districts such as Otemachi; and administrative districts such as Kasumigaseki. This is because many people often use these areas for commuting or on business trips, and have short stay durations although these areas have many visitors. In addition, it also shortens the stay duration, while they have fewer commercial facilities where consumers can buy expensive goods. On the other hand, representative examples of overestimated areas were large-scale entertainment districts such as Kabukicho

Fig. 12 Comparison of the C2F and annual incomes of visitors per 1 km square grid during 2012

and Susukino, and sightseeing spots such as Tokyo Disney Resort and Odaiba. These have long stay durations per consumer and presumably high consumption per consumer.

From this result, we see the policy-making potential of this data to vitalize consumption in regional economies. We can consider two methods to increase consumption expenditures in regional economies: increasing the absolute number of consumers by increasing the number of visitors and increasing the consumption expenditure per visitor by extending their stay durations. Therefore, the possibility of vitalizing consumption increases in underestimated areas by implementing policies to increase visitors' stay durations, while it could be possible to increase capital consumed in overestimated areas by implementing policies to increase the number of visitors.

5.2 Time-series analysis of C2F

There are many analyses of the economic effects of establishing commercial and transportation facilities and so on in specific areas. However, there are few cases analyzing the situations before and after establishing such facilities in time-series. On the other hand, it is also possible to see the time-series change of capital flows in such areas and to analyze the economic effect before and after using the data we developed in this study.

Figure 13 shows representative areas of 1-km grid aggregation where C2F is large every quarter from January 2011 to June 2013. The overall trends show a peak C2F in many areas in the January–March quarter of 2012, and C2F falls in many areas in the October–December quarter of 2012. The Economic and Social Research Institute of Cabinet Office (2012) monitors the turning points of major economic indicators in the economic cycle (economic peaks and troughs) to determine the phase of the economic cycle and to compare the economic activity in each phase following Bry and Boschan's (1971) method. It shows that the peak in the 15th cycle in Japan is in March 2012 and

Fig. 13 Representative areas of high C2F each quarter from January 2011 to June 2013

the trough is in November 2012. This coincides with the peak and trough of C2F in Fig. 13. In other words, the C2F we obtained here reflects the macroeconomic impact of the turning point of the economic cycle.

Next, we focus on areas with characteristic time-series changes. The first example is the Tokyo Sky Tree, a radio tower that opened on May 22, 2012. It is a complex facility that hosts Japan's highest observatory, large-scale commercial facilities, an aquarium, and so on. C2F was growing greatly before and after the April–June quarter of 2012. The C2F of the quarter before its opening was about 7 billion yen and the quarter after opening was about 20 billion yen. Therefore, C2F increased by about 13 billion yen in the quarter after opening and it increased by about 52 billion yen in 1 year after opening. The economic effect of the Tokyo Sky Tree was expected to be about 47.3 billion yen in the 2006 forecast before opening (Dai-ichi Life Group 2006). Although our estimate was 10% higher than the forecast, it was very close. The second example is the Tokyo Disney Resort. It showed a remarkable peak of C2F that does not appear in other areas in the October–December 2011 quarter because 2011 was the 10th anniversary of Tokyo DisneySea opening and the park hosted many anniversary events since September of that year.

By following the time-series change in C2F using the method we propose in this study, it is possible to not only monitor the macro-scale change in the economy, but also to see and analyze the economic effect of the opening of new facilities and holding events in specific areas.

6 Conclusions

In this study, we develop a method to estimate inter-regional capital flows in the real economy as precisely as possible, which is important in understanding regional economies. By using IF data, PF data, and multiple existing sources of statistics, we developed a data set that spatially estimates capital flows and circulation for F2F, F2C, and C2F. In addition, we obtained several results by applying the data set and analysis to real examples. First, it is possible to estimate spatial capital flows microscopically and spatiotemporally and to compare regional economies and conduct a time-series analysis of regional economies. Second, policymakers can use the data could to create policies for regional economic promotion. These results may also be useful in area marketing and real estate development. Furthermore, the data set we develop here is non-aggregated data at the firm and consumer levels, which can be aggregated into arbitrary spatial units or by period according to the purpose. Therefore, this data can be used to set KPIs in policy-making for regional revitalization in a specific region by national and local governments, and to verify the effectiveness of the implemented policies. These data can also contribute to the introduction of EBPM led by the Japanese government.

6.1 Future tasks and prospects

First, it is necessary to refile the IF data spatially. In some regions and business categories, we find differences between the estimated transaction value of the IF data

and the value of the IRIOT because the transaction information in the IF data we use includes only transactions between headquarters. Currently, we are considering a method to decompose head office transactions into transactions between business sites, which will improve this issue in the future. Second, our F2C estimate does not account for the incomes of corporate managers and executives. However, we confirmed that the TDB is also conducting surveys of salaries and including this detail for some firms. Therefore, we expect that developing a salary model using this data will further improve the reliability of F2C. Third, we proportionally distribute consumers' annual C2F to each consumption point based on the stay duration of the consumption point. However, they could differ depending on the attribute of the stay point, such as whether it is a commercial area, business area, sightseeing spot, residential area, and so on. Therefore, we plan to refine the method to estimate the attributes of consumption points further by integrating existing statistics and GIS data. Finally, it is also possible to estimate C2F at specific facilities (commercial facilities, amusement parks, sightseeing spots, etc.) to monitor the number of visitors using PF data. Therefore, we plan to consider developing a C2F estimation model using sales data and C2F at multiple specific facilities.

Acknowledgements The authors would like to thank Teikoku Databank, Ltd. (TDB) for providing the inter-firm transaction big data, ZENRIN DataCom Co. LTD (ZDC) for providing Congestion Analysis[®], Editage (<http://www.editage.jp>) for English language editing, and Center for Spatial Information Science (CSIS), The University of Tokyo (joint research No. 448). In addition, this study was supported by JSPS KAKENHI (Grant number: 15K12453).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Akiyama Y, Ueyama S, Shibasaki R, Adachi R (2016) Event detection using mobile phone mass GPS data and their reliability verification by DMSP/OLS night light image. *ISPRS Ann Photogramm Remote Sens Spat Inf Sci* 3(2):77–84. <https://doi.org/10.5194/isprs-annals-iii-2-77-2016>
- Anselin L (1988) *Spatial econometrics: methods and models—studies in operational regional science*. Kluwer Academic Publishers, Dordrecht
- Brockmann D, Hufnagel L, Geisel T (2006) The scaling laws of human travel. *Nature* 439:462–465. <https://doi.org/10.1038/nature04292>
- Bry G, Boschan C (1971) *Cyclical analysis of time series: selected procedures and computer programs*. NBER, New York
- Cabinet Secretariat (2016) Study group on various issues of economic statistics corresponding to the needs of EBPM. https://www.cas.go.jp/jp/seisaku/ebpm_kenkyukai/index.html. Accessed 15 Sept 2018
- Cabinet Secretariat and Ministry of Economy, Trade and Industry (2015) Regional Economy Society Analyzing System (RESAS). <https://resas.go.jp/>. Accessed 16 Sept 2018
- Dai-ichi Life Group (2006) Economic ripple effect of 2nd Tokyo Tower. http://group.dai-ichi-life.co.jp/dlri/rashinban/pdf/et06_031.pdf. Accessed 16 Sept 2018 (in Japanese)
- De Backer K, Yamato N (2008) The measurement of globalization using international input-output tables. In: OECD science, technology and industry working papers 2007/08, OECD Publishing, Paris. <https://doi.org/10.1787/242020221356>
- Economic and Social Research Institute, Cabinet Office (2012) Turning points of economic cycle. <http://www.esri.cao.go.jp/jp/stat/di/150724hiduke.html>. Accessed 16 Sept 2018 (in Japanese)
- Fingar P, Aronica R (2001) *Death of “e” and the birth of the real new economy: business models, technologies and strategies for the 21st century*. Meghan-Kiffer Press, Tampa

- Gökmen A (2011) Virtual business operations, e-commerce and its significance and the case of Turkey: current situation and its potential. *Electron Commer Res* 12(1):31–51
- Hiramatsu T (2017) The economic impact of inbound international tourism to Japan: the interregional input–output analysis. *J Int Stud* 6(3):51–62
- Horanont T (2010) A study on urban mobility and dynamic population estimation by using aggregate mobile phone sources. Dissertation, University of Tokyo
- Ishida M, Matsutani H (1994) A study on the case of decision making by means of the analytic hierarchy process. *Ann IPSJ SIG* 25(1993-IS-048):69–76
- Local Administration Bureau, Ministry of Internal Affairs and Communications (2012) Survey on the actual salary of local government employees. http://www.soumu.go.jp/main_sosiki/jichi_gyousei/c-gyousei/kyuuyo/kyuuyo_jc.html. Accessed 16 Sept 2018 (in Japanese)
- Milgram S (1967) The small world problem. *Psychol Today* 2:60–67
- Ministry of Economy, Trade and Industry (2011) Inter-regional input-output table. <http://www.meti.go.jp/english/statistics/tyo/tiikio/index.html>. Accessed 16 Sept 2018
- Ministry of Health, Labor and Welfare (2012) Basic survey on wage structure. <http://www.mhlw.go.jp/english/database/db-l/wage-structure.html>. Accessed 16 Sept 2018
- Ministry of Internal Affairs and Communications (2013) Japan standard industrial classification. http://www.soumu.go.jp/english/dgpp_ss/seido/sangyo/index.htm. Accessed 16 Sept 2018
- Ministry of Internal Affairs and Communications (2015) Population, economic relations data according to municipalities. https://www5.cao.go.jp/keizai-shimon/kaigi/special/future/keizai-jinkou_data.html. Accessed 29 June 2019
- Ministry of International Trade and Industry (1967) Measurement of the economic effect of the World's Fair by the interregional input–output table in 1960. MITI, Tokyo
- Morikawa S (2018) Attempts and challenges of evidence-based policy making in Japan. <https://wc2018.ipsa.org/sites/default/files/ipsa-events/wc2018/papers/paper-103075-2018-07-01t071604-0400.pdf>. Accessed 16 Sept 2018
- Nakagawa T, Kuwano M, Zhang J (2012) Analysis of tourist behavior based on a joint model of stay time and expenditure with heterogeneity. *Pap City Plan* 47(3):649–654. <https://doi.org/10.11361/journalcpij.47.649>
- Nakazawa J (2002) The construction and its problems of regional input-output table at a local level. *Policy Sci* 9(2):113–126
- National Personnel Authority (2012) Survey on the actual Salaries of national government employees. http://www.jinji.go.jp/toukei/0312_koumuinkyuuyo/0312_ichiran.htm. Accessed 16 Sept 2018 (in Japanese)
- Saleemdeen R, Tabbaa A, Reynolds C (2016) The UK waste input–output table: linking waste generation to the UK economy. *Waste Manag Res* 34(10):1086–1094
- Statistics Bureau, Ministry of Internal Affairs and Communications (2014) 2014 Economic census for business frame. <http://www.stat.go.jp/english/data/e-census/index.html>. Accessed 16 Sept 2018
- Statistics Bureau, Ministry of Internal Affairs and Communications (2018) The family income and expenditure survey. <http://www.stat.go.jp/english/data/kakei/index.html>. Accessed 16 Sept 2018
- Stiglitz JE, Walsh CE (2006) Principles of microeconomics, 4th edn. W. W. Norton & Company, New York
- Tamura K, Miura W, Takayasu M, Takayasu H, Kitajima S, Goto H (2012) Estimation of flux between interacting nodes on huge inter-firm networks. *Int J Mod Phys* 16:93–104. <https://doi.org/10.1142/S2010194512007805>
- Watts D, Strogatz S (1998) Collective dynamics of small-world networks. *Nature* 393:440–442. <https://doi.org/10.1038/30918>
- Yamamoto Y (2018) Estimation inter-regional money flow using big data of inter-firm transaction and people flow. Masters Thesis, Graduate School of Frontier Science, The University of Tokyo
- Yamamoto Y, Akiyama Y, Shibasaki R (2017) Estimation of inter-regional money flow process using inter-firm transaction big data and people flow big data. In: CUPUM2017 conference proceedings: #35158