The Journal of Transport and Land Use http://jtlu.org
Vol. 15 No. 1 [2022] pp. 463–480

On the empirical association between spatial agglomeration of commercial facilities and transportation systems in Japan: A nationwide analysis

Maya Safira
Bandung Institute of Technology
safiramaya@itb.ac.id

Makoto Chikaraishi
Hiroshima University
chikaraishim@hiroshima-u.ac.jp

Abstract: Understanding the impact of transport systems on the spatial agglomeration of urban facilities is critical for urban and transport planning. Recent studies show three separate mechanisms, including matching, sharing, and trip chaining on the agglomeration of commercial facilities, but little is known about which of these mechanisms is dominant and how its dominance varies across transport systems. Aiming at empirically investigating the mechanisms, we first calculate a simple agglomeration index for 69 Japanese cities and then explore the association between the index and city-level characteristics (including transport) using a decision tree analysis. The results confirm that (1) cities with larger areas and higher train shares experience agglomeration, presumably through matching and/or trip chaining, while cities with smaller areas have less agglomeration despite high train shares; and (2) car-dependent cities experience agglomeration, presumably through sharing, particularly by agglomerating in their residential and roadside areas. These findings indicate that effective agglomeration forces vary across transport systems.

Keywords: Urban agglomeration, city-level characteristics, transport systems, moderation effects, network distance

Article history:
Received: February 10, 2021
Received in revised form: April 27, 2022
Accepted: June 24, 2022
Available online: August 15, 2022

1 Introduction

The relationship between urban form and transport systems is complex, and land use and transport investment decisions are mutually influential (Anderson et al., 1997; Giuliano, 2004). Urban form is the outcome of location choices made by thousands of households, private firms, and public agencies, whereby each agent directly or indirectly considers other agents in their decision making (i.e., some sort of interaction occurs). Urban agglomeration, i.e., the spatial concentration of human activities in an urban area, is an important outcome of those location decisions, which shapes urban form together with transport development. From a policy perspective, exploring the impacts of transportation on agglomeration has gained popularity because these decisions may affect the broader economic impacts of transport infrastructure investment (Graham, 2007; Graham & Gibbons, 2019).

In recent decades, urban agglomeration has been analyzed in the fields of urban economics and geography (e.g., Anas, et al., 1998; Combes et al., 2008; Takayama et al., 2020), particularly following the emergence of the new field of economic geography (Fujita et al., 1999; Krugman, 1991). Establish-

Copyright 2022 Maya Safira & Makoto Chikaraishi
http://dx.doi.org/10.5198/jtlu.2022.1968
ISSN: 1938-7849 | Licensed under the Creative Commons Attribution – Noncommercial License 4.0

The *Journal of Transport and Land Use* is the official journal of the World Society for Transport and Land Use (WSTLUR) and is published and sponsored by the University of Minnesota Center for Transportation Studies.

ing a solid link between theoretical models and empirical results is crucial for improving the development of urban, regional, and national planning, but it is still a challenging task. This is mainly because although agglomeration is essentially rooted in interactions among agents as mentioned above, there are a number of plausible theoretical explanations and mechanisms, including technological spillovers, labor pooling, input–output linkages (Marshall, 1920), and sharing, matching, and learning effects (Duranton & Puga, 2004), where each has different policy implications (Combes & Gobillon, 2015). Accumulating empirical evidence on the agglomeration mechanisms is crucial to properly estimate the broader economic impacts of transport infrastructure investment.

Broadly speaking, there are two major strategies to identify the causal relationship between potential causes (e.g., technological spillover, labor pooling) and the effect (i.e., agglomeration) empirically. The first is the direct use of the statistical causal inference models such as difference-in-differences and regression discontinuity (Baum-Snow & Ferreira, 2015). The second is to use structural estimation methods explicitly to consider interactions among agents (Holmes & Sieg, 2015). Both strategies are useful but less powerful for empirical identification of the causal relationships of agglomeration because (1) it may be difficult to set a social interaction variable as a treatment effect in statistical causal inference models (in addition, a reduced form cannot be analytically derived for many location choice models with interactions) and (2) structural estimation requires an analyst to prespecify the type of interactions; therefore, it is difficult to derive a clear answer concerning which type of interactions is the most plausible. Combes and Gobillon (2015) conduct a comprehensive review on the empirical studies of agglomeration economies and conclude that "most of the literature identifies the overall impact of local determinants of agglomeration economies, but not the role of specific mechanisms that generate agglomeration effects" (p. 338). They also highlight recent studies using well-organized indexes to explore the agglomeration mechanisms. For example, Duranton and Overman (2005) propose a spatial agglomeration index, and Ellison et al. (2010) use the index and empirically confirm that the abovementioned three theories of agglomeration identified by Marshall (1920) are supported, while the most dominant mechanism is input–output linkages. These studies use an explorative strategy to draw empirical inferences rather than the abovementioned well-established strategies (i.e., statistical causal inference and structural estimation methods). We believe that carefully organized explorative studies should be conducted to provide a better understanding of the agglomeration mechanisms and to establish a solid link between the theoretical explanations and empirical findings.

Following Duranton and Overman (2005) and Ellison et al. (2010), this study aims to explore the agglomeration of commercial facilities within a city with special attention to the role of transport systems using a spatial agglomeration index. As Koster et al. (2019) notes, few studies have addressed agglomeration within a city, while a number of empirical studies have been conducted on interregional agglomeration (e.g., Malmberg et al., 2000; Malmberg & Maskell, 2002).

Three different potential mechanisms for the agglomeration of commercial facilities within a city have been identified in the literature: that is, matching, sharing, and trip chaining (Takahashi, 2013; Koster et al., 2019). Takahashi (2013) considers that consumers seek better matching between their preferred product varieties and those sold in each commercial area and confirms that agglomeration occurs when consumers exhibit taste heterogeneity, and the available information is imperfect. Alternatively, agglomeration may also occur when there are positive externalities through sharing, shopping malls being a typical example, i.e., shops in a mall share fixed costs such as facility maintenance and customer acquisition costs (Pashigian & Gould, 1998). Finally, Koster et al. (2019) explain the agglomeration of commercial facilities in shopping streets from the viewpoint of consumer trip-chaining behavior, which is defined as consolidating two or more out-of-home activities at different places in a single departure from home (Greenwald & McNally, 2008). Therefore, consumers can reduce transport and search costs because of the spatial proximity of shops, while shops receive more customers based on the increased numbers of pedestrians passing.

In this paper, we conduct an empirical analysis to explore which agglomeration force is likely to dominate the others. We argue that the dominant agglomeration force depends on the transport system in each city, that is, transport systems embedded within the cities affect agglomeration forces. In most agglomeration studies, transport systems were simply considered to exert a direct agglomeration force (e.g., through saving travel time to the city center) and/or a direct dispersion force (e.g., through congestion increasing travel time) (e.g., Tabuchi, 1998; Glaeser & Kahn, 2004). In this study, we show that transport systems also indirectly influence agglomeration forces by changing the contributions of the three abovementioned agglomeration mechanisms: that is, matching, sharing, and trip chaining. Hereinafter, such indirect effects of transport systems on agglomeration forces are called moderation effects.

These moderation effects can be understood intuitively. Koster et al. (2019) points out that cities that maintain active shopping streets allow consumer trip-chaining behavior, which reduces transport and search costs, and thereby increases the number of customers on shopping streets. However, this is only valid when cities rely on public transit (or nonmotorized) transport. When cities become car dependent, the costs of accessing shopping streets become prohibitively expensive because of the high parking costs and traffic congestion, i.e., dispersion forces increase. For such car-dependent cities, rather than maintaining shopping streets, agglomeration through sharing may be more feasible, for example, by establishing large-scale shopping malls in suburban areas where congestion and parking costs are sufficiently low. The abovementioned difference between transit-dependent and car-dependent cities implies that transport systems would moderate the effectiveness of each agglomeration force.

Although not explicitly mentioning transport systems as moderating factors, narratives similar to the aforementioned are available in the literature (e.g., Glaeser & Kahn, 2004). However, to the authors' knowledge, there is little empirical confirmation of the moderating effects of transport systems on agglomeration forces. As appropriate urban and transport policies depend on the type of the dominant agglomeration/dispersion forces, it is essential for policymakers to have a proper understanding of the role of transport systems in shaping these forces. Particularly in Japan, there is a range of city types, ranging from the very car dependent to the very public transport dependent. Hence, the appropriate urban and transport policy decisions vary between cites.

In our empirical analysis, we first calculate a simple agglomeration index for 69 Japanese cities, and then explore the association between the agglomeration index and city-level characteristics (including transport, socioeconomic and geographical characteristics) using a decision tree analysis to explore the moderation effects of transport systems. We believe that our study makes a significant contribution to the literature. For example, one of the major arguments derived from our empirical analysis is that agglomeration mechanisms seem to be far more complicated than most theoretical models on agglomeration forces would suggest. Of course, theoretical models need not necessarily be realistic. Nonetheless, they matter when we derive policy implications from them. If what the theoretical models indicate is quite different from our empirical observations, policymakers should consider the implications arising from these theoretical models carefully.

This study is structured as follows. Section 2 describes the data used and illustrates the methodology employed to determine the agglomeration index. Section 3 provides the empirical results of the calculated agglomeration index and its association with transport system conditions and other city-level characteristics, followed by the related policy discussions. The study concludes by addressing the essential findings in Section 4.

2 Data and methods

There are two important ideas in our empirical analysis. First, to explore the association between the agglomeration of commercial facilities and transport systems, we conduct a nationwide comparison analysis of 69 Japanese cities varying from car dependent to public transport dependent. Given that other possible confounding factors such as population and area size could affect this association, we conduct an exploratory analysis using a decision tree model including a number of other city-level characteristics. Second, inspired by the work of Duranton and Overman (2005), we use a travel time-based or distance-based agglomeration index where space is treated as continuous, instead of other popular indexes such as the Isard, Herfindahl, and Theil indexes, which require arbitrary geographical units. Specifically, we define the agglomeration index based on the difference between (1) the actual spatial distribution of commercial facilities where various agglomeration forces could apply, and (2) the counterfactual spatial distribution where facilities are randomly distributed, i.e., none of the agglomeration forces are effective.

Unlike other indexes involving aggregation at a specific spatial level, the continuous index values possess the following two properties: (1) they are comparable across spatial scales and (2) unbiased with respect to arbitrary changes to spatial classification (Combes et al., 2008). These properties are particularly important for the current study in which we compare the index values across cities. Other zone-based agglomeration indexes would cause bias, as the spatial scales of administrative boundaries differ by city (e.g., zones in less populated areas are typically larger). Notably, we compare the pairwise distance result in relation to three different distance metrics, including Euclidean distance, network distance, and travel time (by car). By comparing these three metrics, we examine the importance of using travel time and network-based distance measures to reflect the actual distance between facilities accurately considering the geographic restrictions on the developable area. In the remainder of this section, we first introduce the data used followed by details of the analytical procedure.

2.1 Data preparation

To conduct the analysis, we used commercial facility location data from the Census of Commerce conducted by the Ministry of Economy, Trade, and Industry in Japan in 2014. These data include the street addresses, name, and type of the commercial facilities. The commercial facilities in this study are department stores, shops for textiles, clothing, fashion, personal items, food and beverages, building materials, mineral/metal materials, machine tools (e.g., automobiles, bicycles, and equipment), as well as general supermarkets and sellers of other goods (e.g., furniture, fuel, book stationery, sporting goods, and vending machines) and offices for mail order/online shops. Although distinguishing between service types is important in research, particularly when identifying co-agglomeration phenomena (e.g., Kolko, 2007), it lies outside the scope of the present analysis.

One unique characteristic of the Census of Commerce data is that it contains information on whether commercial facilities are located in agglomerated areas in the following five categories: (1) train stations, (2) central business districts (CBDs), (3) residences, (4) roadside and (5) other (i.e., tourism sites, religious sites). We can expect higher transit dependency to increase the number of commercial facilities in the category of "stations," while the shopping streets of European cities would typically be included in CBDs. Conversely, we may observe that additional commercial facilities fall into the "roadside" category in more car-dependent cities.

In the empirical analysis, in addition to the geographical coordinates of commercial facilities, we use public facility location data to obtain the counterfactual distribution. The candidate public facilities include schools (i.e., elementary schools, junior high schools, senior high schools), medical facilities (i.e., hospitals and clinics), community centers, and parks in 2015. The data for public facilities were obtained from the website of the Ministry of Land, Infrastructure, and Transport (available from http://nlftp.mlit.go.jp/ksj/).

We first calculate the agglomeration index, and then analyze the association between the index and city-level characteristics. City-level variables include transport, socioeconomic, and geographical variables. For transport variables, we use the data from the nationwide person-trip survey conducted by the Ministry of Land, Infrastructure, Transport, and Tourism in 2008, which included average trip distance, average travel time, and proportions of travel mode (i.e., the modal share of car, motorcycle, walking, bus, train, and bicycle). For socioeconomic and geographical variables, we use population size, population density, and area, which are obtained from population census data collected in 2015. The selection of 69 Japanese cities corresponds to the availability of other city-level data, particularly the nationwide person-trip survey.

2.2 Analytical procedure

The analytical procedure applied in this study can be divided into four steps: (1) calculating the pairwise network distance between commercial facilities; (2) constructing the counterfactuals; (3) establishing the agglomeration index; and (4) analyzing the association between the agglomeration index and city-level variables through correlations and decision tree analyses.

2.2.1 Calculating the pairwise network distance between commercial facilities

We first calculate the pairwise distance between commercial facilities. While most earlier studies (e.g., Duranton and Overman, 2005) employ Euclidean distances, we also test the network distance and travel time (by car) to reflect the actual travel distance between facilities considering the geographic conditions. More specifically, the pairwise distance $\bar{d}^c$ is defined as: $\sum_i \sum_{j(\neq i)} d_{ij}^c / n_c\,(n_c\text{-}1)$, where n_c is the total number of commercial facilities and d_{ij}^c is the pairwise distance (or travel time) between facility ***i*** and facility ***j***.

2.2.2 Constructing counterfactuals

The locations of public facilities (i.e., schools, medical facilities, community centers, and parks) serve as counterfactuals. Several types of public facilities provide essential services to a community rather than attracting customers from other parts of the city, including police services and schools (i.e., we select the suitable public type based on the empirical comparison of the average distance of public facilities. See Section 3.1 for further detail). It is reasonable to assume that the spatial distribution of these facilities is free from agglomeration forces and that the pairwise distance between public facilities within a city can then serve as a counterfactual. Similar to commercial facilities, the pairwise distance of public facilities $\bar{d}^p$ is defined as $\sum_i \sum_{j(\neq i)} d_{ij}^p / n_p\,(n_p\text{-}1)$, where n_p is the total number of public facilities and d_{ij}^p is the pairwise distance between facility ***i*** and facility ***j***.

2.2.3 Calculating the agglomeration index

We define the agglomeration index as the ratio of the average pairwise network distance between commercial facilities divided by the average pairwise network distance between public facilities as:

$$AI = \frac{\bar{d}^p}{\bar{d}^c} \tag{1}$$

where *AI* is the agglomeration index of a city, with a higher value indicating the greater agglomeration of commercial facilities.

2.2.4 Analysis of the association between the agglomeration index and city-level variables

To explore the association between the agglomeration index and city-level variables, we develop a decision tree model to classify cities into several groups based on the similarity of agglomeration level using their city-level characteristics. Table 1 presents a detailed description of the variables used in the correlation analysis. We append the data of city-level characteristics in the Appendix.

Table 1. Description of variables used for correlation and decision tree analysis

Variable	Description
Agglomeration index	
AI_TT_SC	Agglomeration index of city based on travel time calculation with schools and community centers as the counterfactual
City-level attributes	
Population	The logarithm of the total city population (number of people)
Area	The total area of the city (km^2)
PopDensity	The city population density (person/km^2)
N_Com	The total number of commercial facilities in the city
N_SCH	The total number of schools in the city
N_CC	The total number of community centers in the city
Dist_Com	The average pairwise network distance between commercial facilities (m)
Dist_SC	The average pairwise network distance between public facilities (km)
Transport variables	
Travel time	The average travel time per trip within the city (minute)
Distance	The average distance per trip within the city (m)
Car	Percentage modal choice, car (%)
Motorbike	Percentage modal share, motorcycle (%)
Bicycle	Percentage modal share, bicycle (%)
Walk	Percentage modal share, walking (%)
Train	Percentage modal share, train (%)
Bus	Percentage modal share, bus (%)

3 Results

In this section, we first introduce the selection results of public facility type for counterfactuals and distance metrics. We then explore the association between the agglomeration level and city-level characteristics through correlation analysis and decision tree analysis.

3.1 Selection of counterfactuals and distance metrics

Using residential distribution at the disaggregate level would be ideal for obtaining counterfactuals. However, we could not obtain data with individual-level address information for this analysis. Therefore, we decided to use the distribution of public facilities to obtain the counterfactuals. Public facility types must be selected carefully because agglomeration occurs for particular types of public facilities. To identify a suitable counterfactual empirically, we compare the average pairwise distances for various public facilities (including schools, medical facilities, community centers, parks, combinations of schools and community centers, and combinations of schools, community centers, and parks), and select the facility type that gives the largest average pairwise distance, assuming that longer pairwise distances indicate less agglomeration. Table 2 shows the results. We confirm that the combination of school and community center has the largest average pairwise distance and thus use it to obtain counterfactuals with the assumption that there is less agglomeration for schools and community centers.

We compare three distance metrics, including Euclidean-based distance, network-based distance, and car travel time, to calculate the average pairwise distance and its agglomeration index. In this study, car travel time was chosen for the subsequent analysis because both shopping destination choices (of consumers) and commercial facility location choices (of suppliers) would be based on travel time rather than distance. However, for some cities or countries, travel distance data were difficult to obtain. Our results (Table 2) confirm that the results with network distance would be quite similar to those with car travel time, while the results with Euclidean distance are considerably different. The dataset of all cities' averages pairwise distances and the agglomeration index by public facility type and distance metric is provided in the Appendix.

Table 2. Average pairwise distance and agglomeration index

Facility type	Average pairwise distance			Agglomeration index		
	Euclidean distance	Network distance	Car travel time	Euclidean distance	Network distance	Car travel time
Commercial	5.07	6.09	14.42	-	-	-
Schools	7.42	8.07	18.16	1.49	1.35	1.29
Medical facilities	5.37	5.90	14.27	1.04	0.96	0.97
Community centers	7.69	8.37	18.05	1.51	1.36	1.26
Parks	6.04	6.67	16.06	1.16	1.08	1.08
School & medical facilities	6.45	6.42	15.21	1.29	1.05	1.05
School & community center	7.75	8.49	20.06	1.59	1.41	1.36
Schools, community centers, and parks	7.56	7.78	17.88	1.51	1.28	1.25

Using the travel time-based distance metric with schools and community centers as counterfactuals, we calculate the average pairwise distance and agglomeration index for all 69 Japanese cities. Table 3 summarizes the findings for cities with the longest and shortest average pairwise distance for both public and commercial facilities, and Table 4 summarizes the most and the least agglomerated cities. We found that Yokohama has the longest average pairwise distance between commercial and public facilities, while Shiogama has the shortest average pairwise distance between public facilities and Hitoyoshi has the shortest average pairwise distance between commercial facilities. We also found significant differences between the average pairwise distance between commercial facilities and that between public facilities (Welch's t-test result = 4.09, which is statistically significant at the one percent significance level).

Table 3. Average pairwise distance of facilities (distance metric: car travel time; public facility type: school & community center)

Travel time (min)	Commercial facility	Public facility
Longest (City)	32.80 (Yokohama)	49.79 (Yokohama)
Shortest (City)	4.97 (Hitoyoshi)	4.94 (Shiogama)
Mean	14.42	20.06
Std. dev.	5.90	9.79
Observations	69	

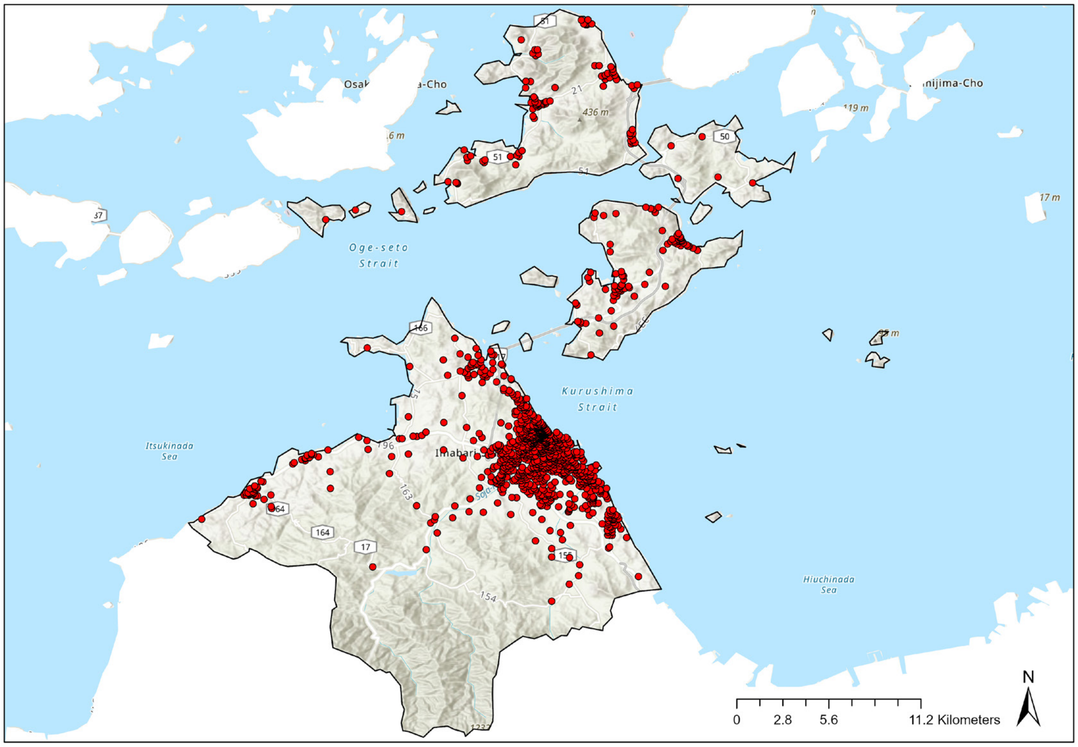
As mentioned above, we took the ratio between the average pairwise distances between public and commercial facilities to obtain the agglomeration index. Table 4 shows that Nara City is the most agglomerated city with an agglomeration index of 1.82, whereas Imabari City is the least agglomerated of 69 Japanese cities with an agglomeration index of 0.77.

Table 4. Summary of agglomeration indexes for 69 Japanese cities

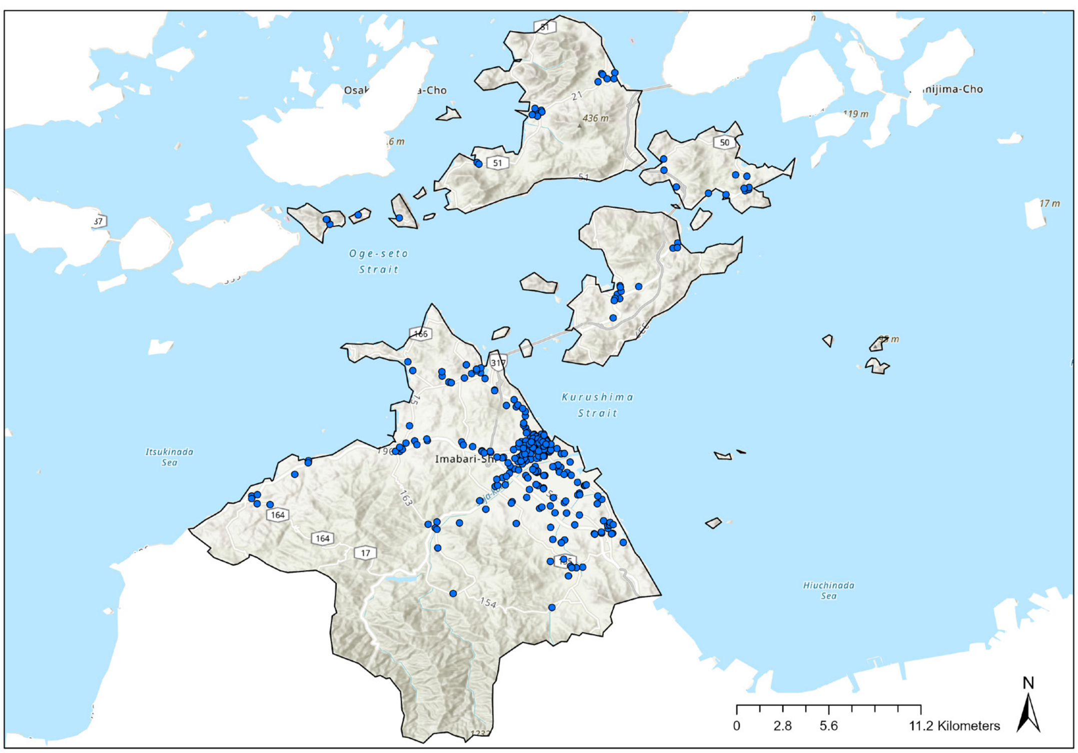
Variable	Agglomeration index
Highest agglomeration index (city)	1.82 (Nara)
Lowest agglomeration index (city)	0.77 (Imabari)
Mean	1.36
Std dev.	0.21
Observations	69

The most agglomerated city is Nara, which is a famous historical tourist destination in Japan. Nara had more than 17 million visitors (including more than 3 million foreign visitors) before the COVID-19 pandemic, where the popular tourism sites are concentrated in the former capital of Japan (Heijo-kyo) during the Nara period (710–784), which is less than one tenth of the current city area size (Nara City, 2022). Figure 1 shows that the commercial facilities are agglomerated in an area connecting Nara Station and the Todaiji Temple. In contrast, there appears to be less agglomeration near public facilities, which is largely because the distribution of public facilities follows the distribution of residents. The least agglomerated city is Imabari (Figure 2), which is a typical regional city in Japan, where population decline has accelerated (down about 20% from its peak). Imabari has a port that has long been a trading center in the region and a large number of shipbuilding and maritime servicing facilities exist along the northern and eastern coastlines of the city (For the details of the city characteristics see Imabari City (2022)).

We also found that of 11 cities with one million or more inhabitants, the agglomeration index exceeds 1.5 for nine cities, implying that the larger cities tend to be agglomerated. In contrast, the five least agglomerated cities (Imabari is the least agglomerated, followed by Shiogama, Urasoe, Tokai, and Dazaifu) are less populated (the populations of these five cities range from 56,256 to 167,872). This result indicates that the city scale could be an important factor affecting the agglomeration level. Given that population typically influences the development of transport systems, we should carefully discuss the impacts of transport systems on agglomeration. Because revealing the actual causal structure may not be possible with cross-sectional data, we explore the association between the agglomeration index and city-level characteristics in the following sections, including scale-related variables such as population and area size, and transport variables such as modal share.

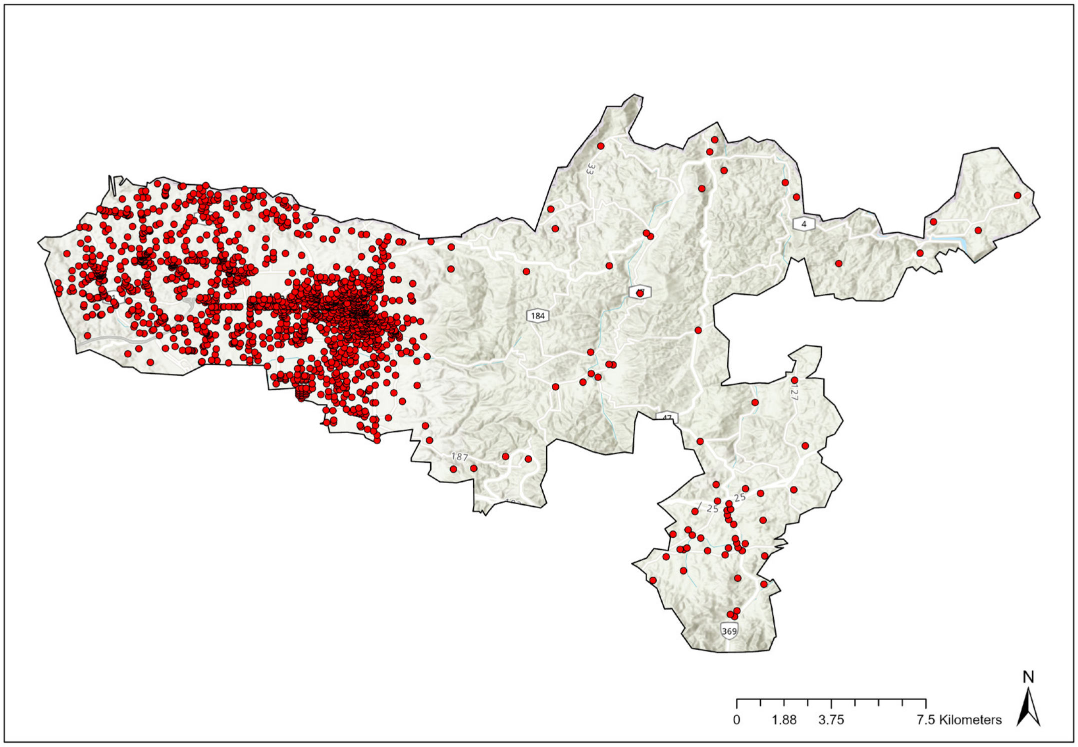


(a) Commercial facility distribution

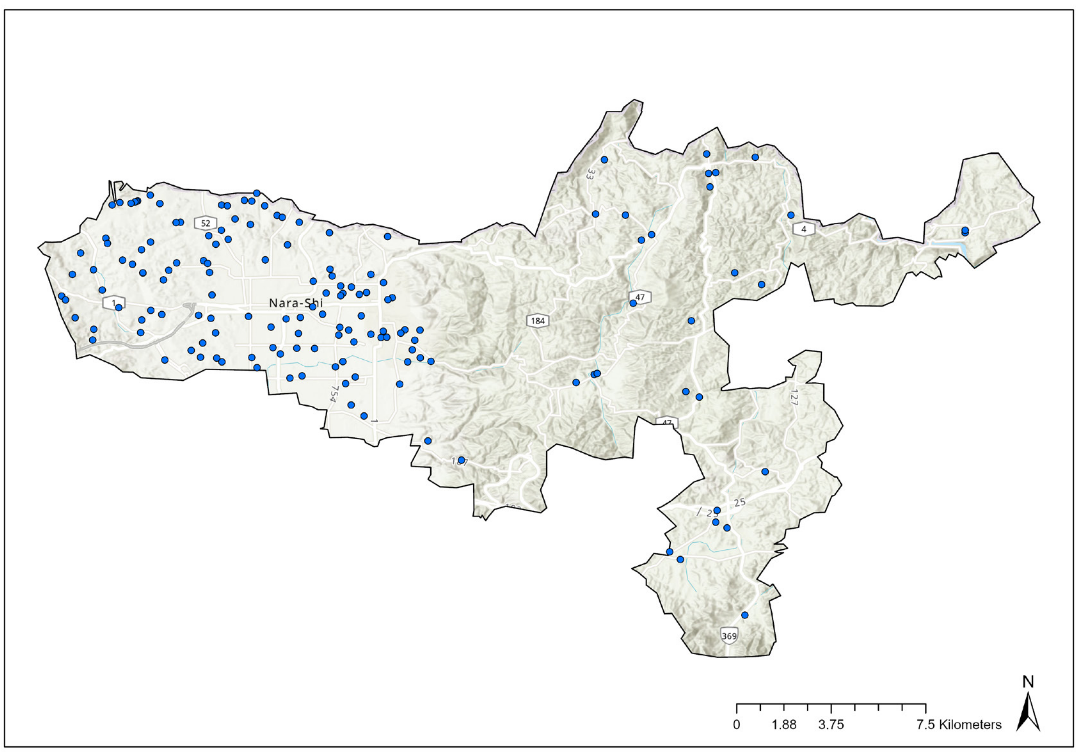


(b) Public facility distribution

Figure 1. Facility distributions in Nara, the most agglomerated city



(a) Commercial facility distribution



(b) Public facility distribution

Figure 2. Facility distributions in Imabari, the least agglomerated city

3.2 Association with city-level characteristics: Correlation analysis

To identify city-level characteristics that are highly correlated with the agglomeration index, we confirm Pearson correlation. Table 5 and Figure 3 show the results. Area size, population density, average travel time, share of car usage and share of train usage are significantly correlated with the agglomeration index. The major findings from the identified correlation coefficients are as follows.

1. Population density and area size have positive associations with the agglomeration index.
2. The average travel time (trip duration) has a negative association with agglomeration level, implying that greater agglomeration emerges together with shorter travel time.
3. Regarding the association between agglomeration level and modal share, more car-dependent and transit-dependent cities are less agglomerated. In contrast, a greater share of walking and bicycle use increases the agglomeration level, though these are not statistically significant. As discussed in Subsection 3.3, the impact of train share is nonlinear; therefore, we could not simply conclude that transit-dependent cities are always less agglomerated.

Overall, the correlation results indicate a significant association between transport systems and agglomeration level. Simultaneously, there are high correlations among the city-level characteristics, implying that it may be difficult to establish a one-to-one relationship between agglomeration level and the state of the transportation system.

Table 5. Correlation coefficients between the agglomeration index and city-level characteristics

Variables	Estimates value	t-value	Sig. sign
Population	4.35E-07	1.58	
Area	4.82E-03	3.08	**
PopDensity	2.61E-02	3.21	**
N_Com	-8.36E-07	-0.08	
N_SCH	–2.16E-05	–1.28	
N_CC	2.27E-04	0.22	
Dist_Com	–5.46E-04	–1.42	
Dist_SC	4.64E-04	1.38	
Travel time	–3.22E-02	–3.11	**
Distance	4.74E-01	0.35	
Car	–5.37E-02	3.49	**
Motorcycle	2.65E-01	0.03	
Bicycle	2.27E-01	0.34	
Walk	3.47E-01	0.31	
Train	–3.71E-02	3.02	**
Bus	3.19E-01	0.03	

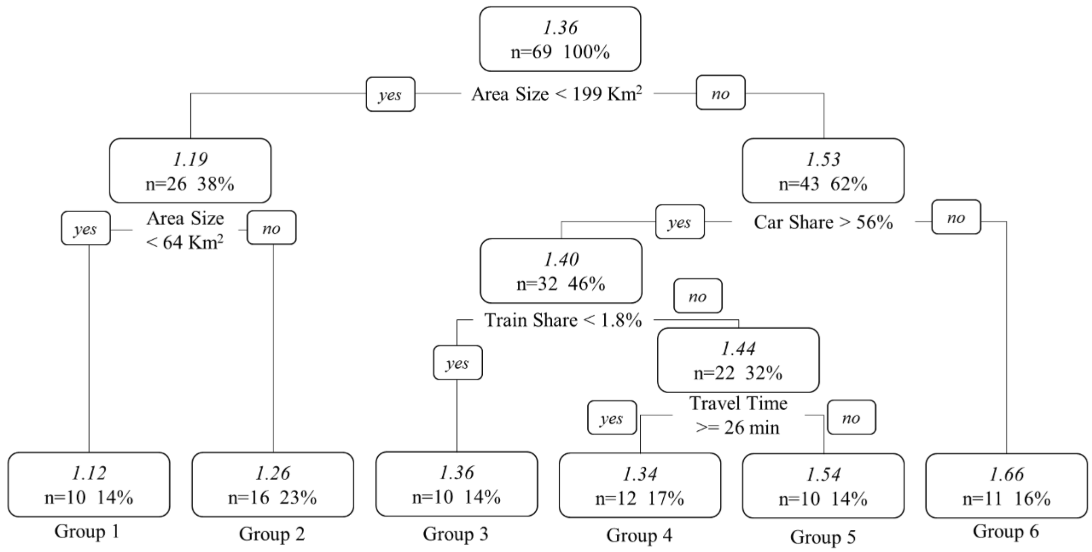
*** Significant at the 0.01 level.*

3.3 Association with city-level characteristics: Decision tree analysis

Based on the correlation analysis in the previous section, we include five significant variables in the decision tree analysis, including area size, population density, average travel time, car share, and train share. We apply a Classification and Regression Trees algorithm (Breiman et al., 1984), which is one of the most widely used algorithms for decision tree analyses, to classify cities into several groups based on their characteristics. For the analysis, we used the "rpart" R package (Therneau et al., 2019), in which the analyses of variance method were chosen to produce a group of cities based on the degree of similarity in their characteristics. We set the minimum split and bucket at 10, ensuring that final nodes (called groups in this study) contain at least 10 cities.

Figure 3 illustrates the results of the decision tree analysis. The analysis divides the 69 cities into six groups. The first split is based on area size (larger or smaller than 199 km^2), confirming that cities with larger area size tend to be agglomerated. One possible reason is that cities with larger area would obtain greater benefits from agglomeration: benefits from *matching* and t*rip chaining* would be larger with increased area.

In the second split, cities with smaller area are further divided into two groups by area (depending on whether they are smaller than 64 km^2), which confirmed that smaller cities tend to be less agglomerated (group 1). Cities with larger areas are further divided into two groups based on car share (depending on whether it is larger than 56%). We confirm that cities with larger areas and lower car share tend to be agglomerated (group 6). Most cities in group 6 are actually metropolitan, including Kyoto, Osaka, Sapporo, Fukuoka, and Hiroshima. Cities with larger areas and higher car share are further divided into two groups by train share (depending on whether it is smaller than 1.8%). Cities with lower train share tend to be less agglomerated (group 3), compared with those with higher train share. Most cities in group 3 are regional, including Hirosaki, Hitoyoshi, and Imabari. Cities with higher train share are further divided into two groups by travel time (depending on whether it is less than 26 minutes). Cities with longer travel time tend to be less agglomerated (group 4) and those with shorter travel time are more so (group 5). To examine further the characteristics of each group, in the remaining part of this section, we conduct an additional aggregation analysis.



Notes: Values in italic font denote the average agglomeration index for each group; n represents the number of cities

Group 1	Akashi, Dazaifu, Inagi, Izumisano, Matsudo, Shiogama, Tokai, Toyonaka, Tsushima, Urasoe
Group 2	Iwata, Kainan, Kameyama, Kasugai, Kawasaki, Nankoku, Odawara, Ome, Omihachiman, Otake, Oyabe, Sakai, Tokorozawa, Tokushima, Toride, Uji
Group 3	Hirosaki, Hitoyoshi, Imabari, Ina, Kochi, Komatsu, Matsue, Nagato, Yasugi, Yuzawa
Group 4	Kagoshima, Kitakyushu, Koriyama, Kumamoto, Kure, Morioka, Sendai, Soja, Takasaki, Toyohashi, Utsunomiya, Yokkaichi
Group 5	Chitose, Gifu, Isahaya, Joetsu, Kanazawa, Matsuyama, Otaru, Shizuoka, Usuki, Yamanashi
Group 6	Chiba, Fukuoka, Hiroshima, Kobe, Kyoto, Nagoya, Nara, Osaka, Saitama, Sapporo, Yokohama

Figure 3. The results of the decision tree analysis

Table 6 shows the share of commercial facilities by type of agglomerated areas for each group and Table 7 presents the average city-level characteristics for each group. From Table 6, we confirm that cities in groups 1 and 2 have more commercial facilities around stations. It is common for the smaller cities in Japan to have more commercial facilities near stations, because stations are often designed or designated to be the central area of a city. This finding is also supported by the train share in groups 1 and 2, which are 15.53% and 11.59%, respectively. Although their train shares are relatively high compared with other groups, their agglomeration levels are the lowest across groups. This is because their area is small. As discussed above, *sharing* and *trip chaining* would be less beneficial for a city with smaller area.

In contrast, cities in group 3 have larger areas, resulting in more agglomeration compared with cities in groups 1 and 2. However, the agglomeration index of group 3 is lower than that of groups 4–6 because of the characteristics of the transport systems in group 3. Group 3 has the highest car share (77.27%), leading to the highest agglomerations in roadside areas (19.16%). This indicates that agglomeration forces for group 3 are different from those for groups 1 and 2. That is, residents in cities belonging to group 3 may access commercial facilities in CBDs and roadside areas by car.

Groups 4 and 5 have similar characteristics in terms of the type of agglomeration areas and modal share. They have higher car share, while agglomeration occurs in residential areas. These findings imply that these cities may have a polycentric urban form, but a further analysis is needed to draw a general conclusion.

In cities belonging to group 6, commercial facilities are agglomerated in station and CBD areas, and they have the highest train share of the six groups. Their average travel distances and travel times are greatest, indicating that agglomeration benefits are obtained with the higher travel costs in these cities.

Table 6. Share of commercial facilities by type of agglomerated areas

Group	Type of agglomerated areas				
	(1) Station	**(2) CBD**	**(3) Residence**	**(4) Roadside**	**(5) Others**
Group 1	56.74%	9.69%	19.29%	11.25%	3.02%
Group 2	56.87%	11.30%	26.04%	5.00%	0.79%
Group 3	12.08%	35.96%	28.19%	19.16%	4.61%
Group 4	22.81%	31.78%	34.02%	11.23%	0.16%
Group 5	24.01%	28.37%	33.08%	9.68%	4.87%
Group 6	41.83%	35.10%	15.23%	6.32%	1.52%
All groups	35.72%	24.54%	26.81%	10.44%	2.49%

Note: Shaded cells indicate the highest share in each group.

Table 7. City transport characteristics

Group	Avg. travel distance	Average travel time	Average percentage of transport mode usage (%)						Average population size (people)	Average area size (km²)	Average density (population/ km²)
			Train	Bus	Car	Motor-cycle	Bicycle	Walk			
1	11.89	29.26	15.53	1.18	54.43	2.28	9.20	17.39	179,010	35.92	4,806
2	12.09	27.77	11.59	0.87	61.10	2.24	10.13	14.07	268,511	123.87	2,241
3	12.17	22.78	1.27	0.84	77.27	1.68	8.48	10.46	123,615	467.56	300
4	11.76	26.91	4.56	2.61	67.79	1.73	8.60	14.72	490,789	483.60	1,083
5	11.77	24.57	3.50	2.16	69.35	2.07	8.38	14.54	274,971	524.87	628
6	12.80	30.59	19.31	3.57	41.61	1.94	12.07	21.51	1,707,435	503.36	4,502
Average	12.08	27.10	9.50	1.83	61.73	2.00	9.55	15.39	503,527	342.11	2,257

Considering the types of agglomeration forces, we found that: (1) cities with larger areas and higher train shares experience agglomeration, presumably through *matching* and/or *trip chaining*, while cities with smaller areas have less agglomeration despite high train share; and (2) car-dependent cities enjoy agglomerations presumably through *sharing*, particularly by agglomerating their residential and roadside areas. These results indicate that transport systems may moderate agglomeration forces, i.e., the dominant agglomeration forces vary across cities depending on their transport systems. These results also indicate the importance of managing how transport systems moderate the agglomeration economy in the development of theoretical and empirical models, rather than simply identifying the degree to which the level of accessibility or density (which can be seen as simplified transport system performance measures) leads to an agglomeration economy.

The above findings are crucial in shaping relevant policies, particularly in transport investment appraisals. Recently, evaluating the broader economic impact of transport infrastructure investment in an agglomeration economy has gained popularity (Chatman & Noland 2011, 2014; Graham, 2007; Graham & Gibbons, 2019; Kidokoro, 2015). Graham (2007) estimates the additional benefits from the agglomeration would be around 25%, while Horcher et al. (2020) show that agglomeration benefits strongly affect optimal public transport policies, indicating that optimizing transport services without considering agglomeration effects would not be optimal in the long run.

Thus, the literature clearly indicates the importance of considering agglomeration effects in decision making on transport investment and management strategies. However, there are several important limitations in the literature. Chatman and Noland (2011) state that:

> *the challenges are numerous in conducting research to determine whether and when public transport improvements increase agglomeration economies. The possible agglomeration mechanisms at work imply a dizzying array of possible measures and methods. Tracing the links between public transport and agglomeration is an important step that has not been explored yet. (p. 740).*

Chatman and Noland (2014) also conduct empirical studies using data on US metropolitan areas and identify the link between public transport and benefits of agglomeration. Our study improved our understanding of the association between agglomeration and city characteristics, including transport characteristics.

4 Conclusions

In this study, we conducted an exploratory analysis of the association between the urban agglomeration of commercial facilities and the city-level characteristics of 69 Japanese cities with a particular focus on transport systems. We developed a simple agglomeration index inspired by Duranton and Overman (2005) that we could use to compare between cities. The major findings from our empirical analysis are that: (1) cities with larger areas and higher train shares experience agglomeration, presumably through *matching* and/or *trip chaining* (Koster et al., 2019; Takahashi, 2013), while cities with smaller areas have less agglomeration despite high train shares; (2) car-dependent cities experience agglomeration, presumably through sharing (Pashigian & Gould, 1998), particularly by agglomerating in their residential and roadside areas. These findings also lead to two academic implications. First, our results highlight the importance of managing how transport systems moderate the agglomeration economy in the development of theoretical and empirical models, rather than simply identifying the degree to which the level of accessibility or density leads to an agglomeration economy. Second, in empirical analysis, it is crucial to use travel times or network distance rather than Euclidian distance, as the results could differ substantially.

We should also note that this study involves a number of limitations. First, while we simply use a city's administrative boundary to define the unit of analysis, this may be inappropriate. One simple solution would be to deal with the whole of Japan as a unit and compute the pairwise distance between all facilities. Unfortunately, this would be computationally expensive. For example, there are 52,394 commercial facilities in Osaka, and thus around 2.7 billion pairs (52,394 x 52,393) exist in just one city. As the numbers of pairs increase exponentially, there is a need for an alternative solution to this issue. Second, the approach we took in this paper was more statistical than economic; therefore, it is not possible to connect our analysis directly with a discussion on transport investment appraisal, which would require a solid microeconomic foundation that could differ depending on the agglomeration forces. Third, we assume that agglomeration does not exist for public facilities, but we should conduct further analyses to confirm whether this assumption is correct. Fourth, in the analysis, we did not distinguish agglomeration in a building (such as a shopping mall) from agglomeration on a street, but the difference between these two would be important for urban planners. Related to this, the design of streets, including size, speed limits, width, and zoning constraints would also affect the emergence of agglomeration.

Extending the method to consider these aspects is an important challenge. Finally, Safira and Chikaraishi (2022) argue that on-demand transport services change the dominant type of agglomeration force. More empirical studies on how these emerging forms of mobility change agglomeration forces would also be needed to discuss their wider impacts.

Appendix

Appendix available as a supplemental file at www.jtlu.org/index.php/jtlu/rt/suppFiles/1968/0.

References

Anas, A., Arnott, R., & Small, K. A. (1998). Urban spatial structure. *Journal of Economic Literature, 36*(3), 1426–1464.

Anderson, S. P., Goeree, J. K., & Ramer, R. (1997). Location, location, location. *Journal of Economic Theory, 77*(1), 102–127.

Baum-Snow, N., & Ferreira, F. (2015). Causal inference in urban and regional economics. *Handbook of Regional and Urban Economics, 5*, 3–38.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees.* Boca Raton, FL: CRC Press.

Chatman, D. G., & Noland, R. B. (2011). Do public transport improvements increase agglomeration economies? A review of literature and an agenda for research. *Transport Reviews, 31*(6), 725–742.

Chatman, D. G., & Noland, R. B. (2014). Transit service, physical agglomeration and productivity in US metropolitan areas. *Urban Studies, 51*(5), 917–937.

Combes, P., Mayer, T., & Thisse, J.-F. (2008). *Economic geography.* Princeton, NJ: Princeton University Press.

Combes, P. P., & Gobillon, L. (2015). The empirics of agglomeration economies. *Handbook of Regional and Urban, 5*, 247–348.

Duranton, G., & Overman, H. G. (2005). Testing for localization using micro-geographic data. *The Review of Economic Studies, 72*(4), 1077–1106.

Duranton, G., & Puga, D. (2004). Micro-foundations of urban agglomeration economies. *Handbook of Regional and Urban Economics, 4,* 2063–2117.

Elison, G., Glaeser, E., & Kerr, W. (2010). What causes industry agglomeration? Evidence from coagglomeration patterns. *American Economic Review, 100*(3), 1195–1213.

Fujita, M., Krugman, P., & Venables, A. J. (1999). *The spatial economies: Cities, regions, and international trade.* Cambridge, MA: MIT Press.

Giuliano, G. (2004). Land-use impacts of transportation investments. *The Geography of Urban Transportation, 3,* 237–273.

Glaeser, E. L., & Kahn, M. E. (2004). Sprawl and urban growth. *Handbook of Regional and Urban Economics, 4*, 2481–2527.

Graham, D. J. (2007). Agglomeration, productivity and transport investment. *Journal of Transport Economics and Policy, 41*(3), 317–343.

Graham, D. J., & Gibbons, S. (2019). Quantifying wider economic impacts of agglomeration for transport appraisal: Existing evidence and future directions. *Economics of Transportation, 19,* 100121.

Greenwald, M. J., & McNally, M. G. (2008). Land-use influences on trip-chaining in Portland, Oregon. *Proceedings of the Institution of Civil Engineers-Urban Design and Planning, 161*(2), 61–73.

Holmes, T. J., & Sieg, H. (2015). Structural estimation in urban economics. *Handbook of Regional and Urban Economics, 5*, 3–68.

Hörcher, D., De Borger, B., Seifu, W., & Graham, D. J. (2020). Public transport provision under agglomeration economies. *Regional Science and Urban Economics, 81,* 103503.

Imabari City. (2022). *Website of Imabari City.* Retrieved from https://www.city.imabari.ehime.jp/profile/keikaku.html

Kidokoro, Y. (2015). Cost–benefit analysis for transport projects in an agglomeration economy. *Journal of Transport Economics and Policy, 49*(3), 454–474.

Kolko, J. (2007). *Agglomeration and co-agglomeration of services industries.* Retrieved from http://dx.doi.org/10.2139/ssrn.985711

Koster, H. R., Pasidis, I., & van Ommeren, J. (2019). Shopping externalities and retail concentration: Evidence from Dutch shopping streets. *Journal of Urban Economics, 114,* 103–194.

Krugman, P. R. (1991). *Geography and trade.* Cambridge, MA: MIT Press.

Malmberg, A., Malmberg, B., & Lundequist, P. (2000). Agglomeration and firm performance: Economies of scale, localization, and urbanization among Swedish export firms. *Environment and Planning A, 32*(2), 305–321.

Malmberg, A., & Maskell, P. (2002). The elusive concept of localization economies: Toward a knowledge-based theory of spatial clustering. *Environment and Planning A: Economy and Space, 34*(3), 429–449.

Marshall, A. (1920). *Principles of economics.* London: MacMillan.

Nara City. (2022). *Website of Nara City.* Retrieved from https://www.city.nara.lg.jp/ site/ kankou/9732.html

Pashigian, B. P., & Gould, E. D. (1998). Internalizing externalities: The pricing of space in shopping malls. *The Journal of Law and Economics, 41*(1), 115–142.

Safira, M., & Chikaraishi, M. (2022). Exploring the impacts of online food delivery service on facility distribution: A case of Jakarta, Indonesia. *Asian Transport Studies, 8,* 100078.

Tabuchi, T. (1998). Urban agglomeration and dispersion: A synthesis of Alonso and Krugman. *Journal of Urban Economics, 44*(3), 333–351.

Takahashi, T. (2013). Agglomeration in a city with choosy consumers under imperfect information. *Journal of Urban Economics, 76,* 28–42.

Takayama, Y., Ikeda, K., & Thisse, J. F. (2020). Stability and sustainability of urban systems under commuting and transportation costs. *Regional Science and Urban Economics, 84*, 103553.

Therneau T., Atkinson B., Ripley B., & Ripley, M. B. (2019). *Package 'rpart'.* Retrieved from https://cran.pau.edu.tr/web/packages/rpart/rpart.pdf