共同利用・共同研究拠点 Joint Usage/Research Center
東京大学 The University of Tokyo
空間情報科学研究センター
Center for Spatial Information Science

Discussion Paper Series

No. 167
CSIS Discussion Paper

August 2020

# Residential Agglomeration of the Homeless and Its Effects on Their Living Standards

**Mariko Nakagawa**
(Center for Spatial Information Science, University of Tokyo & Tohoku University)
**Kotaro Iizuka**
(Center for Spatial Information Science, University of Tokyo)

# Residential Agglomeration of the Homeless and Its Effects on Their Living Standards [*]

Mariko Nakagawa[†] and Kotaro Iizuka[‡]

July 31, 2020

**Abstract**

This research analyzes the benefits enjoyed by the homeless persons located on the riverbank of the Tama River in Tokyo when they reside in clusters. To conduct this research, we first detect house locations of homeless persons by using an Unmanned Aerial System—or a drone—and find the optimal number of clusters of houses based on single-linkage clustering, one of the hierarchical cluster-analysis methods. After detecting the clusters, we evaluate the effect of residing in a larger cluster on a homeless person's standard of living, which is measured by the temperature of each house in the winter. Our results show that the larger the cluster in which a house is located, the higher its temperature, indicating that living in a larger cluster makes their living standards higher. In addition, we find a negative bias in the uninstrumented estimates, which can be interpreted as an indication of negative self-selection of homeless persons to sort into a large community. That is, more disadvantaged homeless persons sort into larger clusters. By doing so, they benefit from living in a larger community via interaction with those living in the same cluster.

## 1 Introduction

There is a tendency for people to interact more with those located nearby according to various contexts. More active social interactions are associated with geographic proximity that brings about benefits

[*]The authors are grateful to Naoya Fujiwara, Vernon Henderson, Ryo Itoh, Keisuke Kawata, Tatsuhito Kono, Ryosuke Okamoto, Yasuhiro Sato, Dan Sasaki, Takatoshi Tabuchi, Takaaki Takahashi, Kohei Takeda, and Dao-Zhi Zeng for thoughtful comments and suggestions. We also thank seminar participants at the Urban Economics Workshop at the University of Tokyo and the Regional Science Workshop at Tohoku University. We appreciate outstanding research assistance by Toshikazu Fukuba. All remaining errors are our own.

[†]Center for Spatial Information Science, University of Tokyo and Graduate School of Information Sciences, Tohoku University. E-mail: mnakagawa@csis.u-tokyo.ac.jp.

[‡]Center for Spatial Information Science, University of Tokyo. E-mail: kiizuka@csis.u-tokyo.ac.jp.

from interaction with others without the costs that stem from long distances. Marmaros and Sacerdote (2006) study the likelihood of social interaction among students at Dartmouth College and find that geographic proximity matters when explaining the extent of interaction among students. By using data on adolescents in the United States, Kim et al. (2017) argue that a greater geographic dispersion lessens people's incentives to interact socially.

In the context of the labor market, Bayer et al. (2008) show that individuals residing in the same city block in the Boston Metropolitan Area are more likely to work together than those in nearby blocks. Under the assumption that job information is exchanged among agents residing in the same census block, they interpret the result as an outcome of a referral effect and conclude that the findings are evidence of the existence of significant social interactions at the block level. By using detailed geographic employer-employee data in the United States, Schmutte (2015) finds that workers are disproportionately more likely to become coworkers with their neighbors when changing jobs. Also, they are more likely to be employed in a high-wage firm when their neighbors are already employed in high-wage firms. Therefore, the author conclude that the role of social networks is important when connecting workers to jobs offered by firms paying high wages. From these findings, geographic proximity is considered to play a pivotal role in enhancing social interactions and benefits the relevant agents.

Moreover, disadvantaged workers may value the importance of social interactions more under geographic proximity. For instance, according to Hellerstein et al. (2011), residence-based labor market networks are more important for blacks than for whites and for workers with low education levels than for those with high education levels. Such networks are more crucial for Hispanic immigrants with poor language skills than for blacks.

These examples indicate how beneficial social networks are within a certain geographic range so that individuals can reach others, in friendship formation and job searching. These benefits are especially helpful for disadvantaged individuals. When we depart from these contexts, we can find another example in which social interaction works in a beneficial and effective way for a particular type of disadvantaged person in a society—the homeless. Homeless persons literally do not have secure residences. In many cases, they are also jobless or earn only a small amount of money. Due to a lack of money, they usually also suffer from food scarcity. They are at risk of violence from juvenile delinquents because of poor quality residences.

On the riverbank of the Tama River near Tokyo, Japan (the study area for this paper), homeless communities have been created. Homeless persons in that area usually earn a small amount of money.[1] They are engaged in illicit daily works that incorporate informal information (see Section 2). Because access to such informal information is basically possible via direct contacts and interactions with other homeless persons, social networks associated with geographic proximity are essential for their lives. Also, because their daily incomes tend to be unstable, mutual support through shared food stabilizes their lives when they cannot earn a lot. Moreover, other goods such as electromechanical devices, as well as food, may be shared.[2] Hence, mutual support within a homeless community should be one of the vital benefits of interaction.

As the homeless are one of the most disadvantaged groups in society, it is meaningful to examine how social interaction based on geographic proximity works in the homeless community and to see if it benefits a homeless person's living standard. In this paper, we focus on the benefits derived from agglomeration of homeless persons and examine whether living in a cluster improves their living standards in homeless communities created on the bank of the Tama River, which divides Tokyo Prefecture and Kanagawa Prefecture in Japan. The goal of this paper is to corroborate the existence of a positive agglomeration effect in a society of homeless persons in our study area.[3] Specifically, we investigate whether having interactions with more homeless persons leads to their higher living standards. To accomplish this goal, however, there are two main obstacles to overcome: (i) difficulty in detecting interaction among homeless persons associated with geographic proximity and (ii) difficulty in drawing out homeless persons' actual levels of living standards.

The first obstacle stems from the fact that we usually cannot observe who interacts with whom in the case of homeless persons, because they do not have fixed residential addresses. However, homeless persons in our study area are not nomadic but are settled in fixed places. They build cage-like houses to reside in on the riverbank. Moreover, the exteriors of the houses have special features. When seen from above, most of the houses are rectangular polygons and are covered by blue, green, or gray plastic sheets. Using these exterior characteristics, we can detect house-like objects in the aerial pictures.

[1]Homeless persons in this area rarely beg for money or food and earn money on their own (Division of Health & Welfare in Kawasaki City, 2019), so diffusion of job-related information is important for increasing their income levels.

[2]In the study area, not several homeless persons are known to own such electromechanical devices (Murata, 2015). For more details, see Section 2.

[3]Other mechanisms of agglomeration of homeless persons than the social interaction benefits can be considered. Homeless persons may gravitate to services such as shelters, meal programs, and medical clinics, which induce homeless agglomeration (Culhane, 2010; Corinth and Lucas, 2018; Lee and Price-Spratlen, 2004). Our study, however, does not shed light on this aspect, as such services are not occasionally provided on the riverbank in our study area. There are no facilities such as soup kitchens or shelters on the riverbank, and it is rare for food banks to provide free food services there. From our interview at the Division of the Medial Welfare Services of Kawasaki City, only three food distribution events were held per month during our study period, which implies that it is unlikely that concentration of homeless persons in the riverbank occurs due to this mechanism.

By conducting Unmanned Aerial Systems (usually called drones) flights, we obtain aerial images of our study area. With these, we create house-location data.[4] Based on the house-location data that we collected, we calculate the distance between houses. We obtain information about geographic proximity between homeless persons, which can be considered a proxy for social interactions among homeless persons.

The second obstacle is the difficulty in eliciting reliable information from homeless persons. Interviewing them is tough work even when they can be contacted. As is true elsewhere, in Japan there tends to be a high percentage of homeless persons with mental disorders, such as alcohol addiction. Also, some may have low IQ. Mental and intellectual problems make it difficult for the investigators and interviewers to draw out information about their actual levels of living standards.[5] To overcome this issue, we utilize house temperature data on winter nights as a proxy for the living standards of the resident in each house.

In our study area, it is known that several homeless persons have electronic generators that they can use to operate microwaves, rice cookers, television sets, and heaters if they have them. These electrical appliances generate heat when used, and the temperature of a homeless person who is well-off enough to own a lot of electrical devices should be higher than that of a homeless person with fewer means. Moreover, in the winter, they may use heaters to fend off the cold if they can afford it, which also raises house temperatures. Thus, we conjecture that house temperature indicates how affluent a resident is.

As we pinpointed the house locations, we obtain house temperature data by flying a drone loaded with a thermal sensor to get a thermal image. From the thermal image, we are able to extract temperature information of the areas overlapping the house polygons and to attach the extracted temperature values to each house.[6] The house temperature data sensed by the thermal camera capture the heat that leaked from inside the house.

We briefly summarize the analysis and results obtained in this paper. Based on the distance between houses, we conduct a hierarchical cluster analysis (single-linkage clustering) to group houses of homeless persons into clusters. Assuming that social interaction effectively works only within a cluster of houses, we adopt a cluster size (the number of houses in a cluster) as our agglomeration variable

[4]There is a demographic tendency for homeless persons in Japan to be single males, which leads to the assumption that each house is occupied by only one homeless person. Thus, house locations can be seen as the geographic location of a single homeless person. For details, see 2.

[5]For details, see Section 2.

[6]The exteriors of the houses of homeless persons, such as walls or roofs, are often thin and it is possible that heat inside a house can leak to the outside. For more explanation, see Section 2.

that expresses the magnitude of social interaction on the basis of geographic proximity. Because the variable cluster size is time invariant in our study, we stick with the random effects model.

As expected, the baseline estimation result confirms that the larger the cluster size (larger the social interaction network) he belongs to, the higher the living standard of a homeless person as captured by house temperature. Recognizing the existence of unobservable characteristics of each homeless person, such as the difference in ability, intellectual level, mental condition, and physical handicaps, we instrument the agglomeration variable, which may be correlated with a homeless individual's unobservable characteristics, because homeless persons with unobserved low abilities may sort into larger clusters, seeking more support via interaction with other homeless persons.

As an instrumental variable for the cluster size, we employ a dummy variable that equals one if a house is located in an inner curve of the meandering river and zero if it is located in an outer curve. This means that areas in the inner curve should be more suitable to build houses than those in the outer curve, implying that larger clusters of houses are more likely to be generated in the inner than the outer curve of the meandering river.

After instrumenting the cluster-size variable by the inner/outer curve dummy variable, the significantly positive coefficient on the cluster size variable remains unchanged, meaning that a larger cluster size leads to higher living standards of homeless persons in that cluster through a channel of a larger social interaction network. Comparison between the estimated coefficients of the cluster size in the instrumented and uninstrumented versions finds a downward bias of the uninstrumented one, with which we conclude that there is a negative sorting of homeless persons to enter a larger cluster. That is, less able or more disadvantaged homeless persons sort into larger clusters, and by doing so, they enjoy benefits of interaction with those in the same cluster and higher living standards.

The remainder of this paper is organized as follows. In Section 2, we first describe the background of our study area and why we used a drone for data collection. In Section 3, we describe how we collected the data. Based these collected data, we conduct a cluster analysis to divide observed houses into an appropriate number of clusters. Also, we propose measures for house agglomeration and we run regressions with them. In Section 4, we show the results based on the random effects model as our baseline model. After addressing the endogeneity issues of unobservable omitted variable bias, we show the instrumental variable results. Section 5 displays various robustness checks, which reveal that the results obtained in Section 4 remain unchanged. Section 6 proposes some policy possibilities and concludes the paper.

# 2 Background on the Homeless and Advantages of Utilizing a Drone for Data Collection

## 2.1 Benefits of Residential Agglomeration for Homeless Persons

As stated in Section 1, this paper tries to corroborate the positive effects of living in clusters for the well-being of homeless persons. A key factor in bringing about this positive impact of residential agglomeration is social interaction among homeless persons who live close. From the viewpoint of social interaction, there may be three types of benefits for homeless persons to live in a cluster: (i) gains from information diffusion, (ii) mutual aid, and (iii) mental well-being through friendships with other homeless persons.

Information diffusion is an important social interaction benefit to the homeless persons in our study area. Specifically, this is related to daily income and expenditures. One of the main income sources for homeless persons in this area is collecting aluminum cans or used metal found in garbage disposal spots set out in residential areas. Selling cans to scrap firms illegally is a way to earn some money.[7][8] The homeless must have knowledge of which scrap firms conducting illegal transactions. Without it, they can neither sell the cans they collected nor exchange them with money, because only a small number of scrap firms conduct illegal deals with homeless persons. Such informal knowledge is only available through interaction with homeless friends. This implies that information diffusion via interaction with other homeless persons is vital to their livelihoods.

On the expenditure side, knowing which supermarkets sell cheap food at discounted prices or food shops that let homeless persons have waste food is helpful. This type of knowledge helps homeless persons reduce their daily expenditures. Because food scarcity is one of the major obstacles for the homeless (Division of Health & Welfare in Kawasaki City, 2019), shared knowledge about access to cheap food providers alleviates the anxiety about food deficiency.

In terms of mutual aid, one of the most important benefits from living in a cluster is protection from the violence from passersby or juvenile delinquents. Harassment and violence from non-homeless people are ranked as the biggest anxieties for the homeless (Division of Health & Welfare in Kawasaki

[7] In most administrative divisions in Japan, the local ordinance forbids to collect or transport the resources disposed in the garbage spots, except for the business firms entrusted by the local government. This is why activities of homeless persons such as collecting cans and selling them to scrap firms are illegal in Japan.

[8] It is rare for homeless persons in Japan to beg for money or food from passersby. In the neighborhoods of our study area, about 70% earn money on their own. On average they earn approximately $ 350–400 US per month. Among homeless earners, 82% obtain income by collecting and selling cans to illegal scrap firms, and 11.5% work as daily construction workers (Division of Health & Welfare in Kawasaki City, 2019).

City, 2019; Sugita et al., 2010). Living near other homeless persons may keep them from being badly injured, compared to living alone. In addition, in daily life, sharing food when a person cannot earn enough money may be one instance that highlights the necessity of mutual aid among homeless neighbors. Because their sources of income are coming from collecting and selling cans and metal illicitly, the amount of money they earn changes day to day. An unreliable income may be relieved by mutual aid among homeless persons who live close to each other. Likewise, they can obtain hand-me-down electromechanical devices or useful goods from other homeless persons.

As for their mental well-being, daily contacts with other homeless persons in the same cluster may improve mental satisfaction. Loneliness is an essential factor of mental illness. Even simple daily communication will benefit homeless persons and improve their mental well-being.

These social interaction factors may improve quality of life for the homeless, so we expect that larger social networks (i.e., more homeless persons living close together) may raise the living standards of homeless persons in the larger cluster. To see the impact on homeless persons' material well-being provided by the amount of social interaction, we need to obtain two types of information about homeless persons in our study area: (i) interaction among homeless persons within the reach of their daily activity and (ii) their actual level of living standards. However, there is no straightforward way to extract such information from homeless persons. In Section 2.2, we discuss why it is difficult to obtain this kind of information and how we overcame or circumvent such obstacles in our study.

### 2.2 Observing Interactions among Homeless Persons Based on Geographic Proximity

The obstacle against obtaining the information about interaction among homeless persons within the reach of their daily activity, labeled (i) in the last paragraph in Section 2.1, can be recovered in the case of homeless persons in our study area. The homeless in our study area settle in fixed locations by building cage-like houses where they illegally occupy public land. Some unique features of the exteriors of their houses and their sociodemographic characteristics enable us to approximate the house locations, with which we can obtain information about their geographic proximity. We utilize the house-location information to group them in clusters based on the distance between houses. We implicitly assume that geographic proximity can be used as a proxy for social interactions,[9] and we

[9] Bayer et al. (2008), Hellerstein et al. (2011), and Schmutte (2015) use geographic proximity as a proxy for social interaction levels. This paper follows this method of identifying social interactions, and we assume that homeless persons in the same cluster interact with each other, while those in different clusters do not. Moreover, we additionally consider an effect of geographic distance between any two persons within the same cluster in Section 3.3.

use a cluster size (the number of houses in each cluster) as a proxy for the social network size.

We start with an assumption that each house is occupied by one resident so that the number of houses is considered to be the number of homeless persons in the network. We derive this by describing the sociodemographic characteristics of homeless persons in our study area. Unlike nomadic young homeless persons whose temporary locations are difficult to detect, some older homeless persons, mainly single males without families,[10] are settled in stations, city parks, or riverbanks in Japan (Division of Health & Welfare in Kawasaki City, 2019). In 2018, 49.7% of homeless men in Kawasaki Ward (where our study area is located) were settled in cage-like houses that they built themselves on the riverbank of the Tama River (Division of Health & Welfare in Kawasaki City, 2019). Sugita et al. (2010) find a similar settlement tendency for homeless persons in their field survey near the Tama River. The historical background of the residential style of homeless persons on the riverbank is attributed to an industrial structural shift that occurred at the end of Japan's high economic growth after WWII. Many of the current homeless persons in Japan served as construction workers or were engaged in the manufacturing sector during the era of high growth, and such workers lost their jobs after the industrial structural shift (Division of Health & Welfare in Kawasaki City, 2019; Iwata, 2004; Takano et al., 1999). Thus, most of the homeless persons in Japan are older single men (Ito et al., 2014; Okamura et al., 2014) and this situation is the same in our study area (Division of Health & Welfare in Kawasaki City, 2019).[11] These ex-construction workers may well have the ability to build simple-structured houses if they find a suitable location. The riverbank is an appropriate site for them to build cage-like houses to settle in, so they built their own houses there.

To obtain the location information of the cage-like houses of homeless persons, we took aerial pictures with latitude-longitude and height information by flying a drone and detecting house-like objects appearing in the high resolution aerial pictures. Typical exterior characteristics of the houses built by homeless persons in the study area enable us to detect what objects are the houses of homeless persons. The houses are usually rectangular or box-shaped polygons with the edge length longer than 1.5 meters when seen from above and covered by blue, green, or gray plastic sheets. We detect the houses by considering the objects' height information from a set of polygons that satisfy above

[10]The demographic characteristic that homeless persons tend to be single males in Japan are also common in other countries. In European counties, the homeless are predominantly men, and most of them are unemployed, unmarried men (Philippot et al., 2007). In the United States, the length of a homeless spell is longer for males, never-married persons (Allgood and Warren Jr, 2003). Similarly, in Australia, it is found that men are more likely to be sleeping on the streets (Cobb-Clark et al., 2016).

[11]According to a report from the Division of Health & Welfare in Kawasaki City (2019), the average age of homeless persons in Kawasaki Ward was 62.3. 47.5% were in their sixties, and 24.9% were older than 70 in 2018.

features, i.e., we select polygons whose heights are taller than one meter. After detecting the houses of homeless persons, we calculate the distance between house centroids for each pair of houses. Based on this information, we group the detected houses into clusters within which homeless persons are assumed to interact with each other. As a majority of homeless persons in this area are single males, it is reasonable to assume that in most cases each house is occupied by a single man. This assumption allows the conversion of house locations into individual locations.[12]

### 2.3 Observing the Living Standards of Homeless Persons

The obstacle against an access to the information labeled (ii) in the last paragraph in Section 2.1 is a difficulty in eliciting homeless persons' actual levels of living standards stemming from mental illness, alcoholism, and intellectual impairment among the homeless. As in many other countries,[13] there is a strong tendency for homeless persons in Japan to have mental disorders or cognitive impairment such as hallucinations and delusions (Ito et al., 2014; Morikawa et al., 2011; Nishio et al., 2017; Okamura et al., 2014, 2015). Likewise, alcoholism is prevalent among homeless Japanese persons, and the morbidity rate of alcohol psychoses or alcohol-dependence syndrome is much higher than the Japanese average (Morikawa et al., 2011; Takano et al., 1999). As well as the problems of mental illness and alcoholism referred to above, there is a problematic prevalence of low IQ, cognitive disorders, and intellectual impairment among homeless persons in Japan (Nishio et al., 2017; Okamura et al., 2015).

These problems may induce severe difficulties in interviewing homeless persons because eliciting reliable information from them is extremely tough. In the survey of mental illness among homeless persons in Tokyo conducted by Morikawa et al. (2011), the authors were unable to interview homeless persons who were drunk or undergoing hallucinations. Due to the inaccessibility to these homeless persons, the authors conclude that the morbidity of mental illness among the homeless in their study area is underestimated. Alternatively, sample selection bias or measurement errors can arise when

[12]There still remains a possibility that some of the detected houses are vacant. To overcome this anxiety, we choose only house polygons showing variation of house temperatures in the final sample of houses. As will be explained below, we consider house temperature as a sign of human activity, house polygons without temperature variations, which exhibit high possibility of being abandoned, should be excluded from the sample.

[13]In Western Europe, the prevalence of mental disorder is high among the homeless (Philippot et al., 2007). In a field survey in the inner-city area of Mannheim, Germany, the prevalence of alcoholism or alcohol abuse are found to be high (Salize et al., 2002). Also, Philippot et al. (2007) report that the majority of homeless people in the Western European countries suffers from abuse of illegal drugs. Similarly, in the United States, a history of alcohol or illicit drug use increases the probability of being homeless (Early, 2005), and the homelessness spells tend to be longer for persons with such problems (Allgood and Warren Jr, 2003). In addition, significantly lower IQ scores and a stronger prevalence of intellectual disability among the homeless population than among the general population are found in the United Kingdom (Oakes and Davies, 2008) and Netherlands (Van Straaten et al., 2014). From these findings, the problems in conducting interviews targeting homeless persons are common in many other countries, given the similar tendency toward mental illness, substance abuse, and intellectual disorders.

the homeless interviewees have mental or intellectual issues, because the people who were interviewed were intentionally not drunk or hallucinating.

Another limitation of interviewing homeless persons is pointed out by Salize et al. (2002). Even when homeless persons do not suffer from intellectual disability, mental illness, or substance abuse, the interviewers can only easily reach accessible groups of homeless persons such as users of shelters or people who come for free meals. This, again, can induce a possible selection bias problem.

These problems would develop in our study area if we were to conduct a survey based on questionnaires or interviews of homeless persons. The report by the Division of Health & Welfare in Kawasaki City (2019) points out that homeless persons in our study area do suffer from addictions to alcohol or gambling. If the study were to be based on interviews, we expect it to be extremely difficult to draw out reliable answers from the homeless subjects. Therefore, we did not conduct an interview- or questionnaire-based approach to elicit the homeless persons' levels of material well-being. Instead, we collected house temperature data by using a thermal, infrared camera loaded on the drone. This approach was based on an assumption that house temperature in the winter can be used as a proxy for the standard of living of a resident in a house. The strength of using a drone loaded with a thermal sensor is that it can collect thermal data by sensing the various objects continuously and constantly from the sky. Such uniformity in data collection can circumvent sample selection problems and unreliable information from homeless persons with intellectual and mental disorders.

In a different context, sensing techniques using a thermal sensor and a drone have been utilized recently in environmental engineering. Akiyama et al. (2019) report the possibility of detecting vacant houses with this indirect surveying method. The presence of life activity producing waste heat was observed, while vacant houses showed no heat observation when the heat loss from house windows and walls was inspected. The authors emphasize that this observation is more valid in the winter and at night. In the winter, heat from solar radiation resolves faster during the day. Specular reflectance is also avoided at night, allowing the researchers to sense only the heat transfer from human activity. The advantage of utilizing the drone system enables the collection of high resolution information, with a flexible operation. Sensing from the sky up to 150 meters due to Japanese aviation laws allows observation of various objects or phenomena. It is evident that drones tend to observe things that are difficult to see on the ground.[14]

Despite the usefulness of the combination of infrared thermography and drones as above mentioned,

[14]Examples are animals (Oishi et al., 2018) or illegal miners (Iizuka et al., 2018).

one may cast doubt on the validity of using surface temperature of a house as a proxy for the living standard of its resident in our context. The following two questions about the validity may arise: (i) Does the night-time house temperature in the winter express the level of the living standard of the resident inside it? and (ii) If the answer is "yes," does the surface temperature of a house necessarily capture the inside temperature?

For the first question, we assume that the house temperature at night in the winter surely expresses the living standard of a homeless resident. As in Murata's (2015) report featuring homeless persons, not a few homeless persons in our study area have electronic generators and electromechanical devices. By using these electromechanical products generates heat, the house temperature of a homeless person who has a lot of electromechanical devices should be higher than that of a poor homeless person who does not have such heat-producing devices. Moreover, winter is a season when homeless persons suffer from the cold (Division of Health & Welfare in Kawasaki City, 2019). They would want to avoid being cold if they could manage to do so, and indeed, more well-off homeless persons can do so by using heat-generating equipment. By contrast, poor homeless persons may not be able to stay warm as they only have insufficient heat-generating products.[15]

Another important point about using this technique is to determine whether or not the residents are in their houses. We cannot give an exact answer to this question, but we believe that the homeless are in their houses at night, which was when we collected the thermal data. As addressed in Section 2.1, homeless persons earn money by collecting cans. In Japan, cans and garbage are disposed of in garbage sites in the morning, and thus, homeless persons can only collect cans early in the morning. This implies that the working hours of homeless persons are in the morning, and hence, they are more likely to be at home in the evening than in the morning.

In regard to the second question about external versus internal heat, we think that the outside surface temperatures of houses capture the inside temperatures to some extent. Balaras and Argiriou (2002) point out that thermal infrared images taken by thermography cameras are useful for non-contact detection of how energy leaks from a building's envelop. As the temperature of an object is correlated to the infrared radiation it emits, the infrared inspections of building envelops can be used to detect heat losses from walls or roofs. We can measure inside temperature that leaks outside though walls or roofs when the walls or roofs are thin. Indeed, the walls of houses of homeless persons tend to be very thin, and sometimes houses are simply made of frameworks covered by plastic sheets in

[15]In summer, people prefer to cool down the house, unlike in the winter, and it is not obvious if houses where more well-off homeless persons show high or low temperatures. This ambiguity does not occur in our data collection in winter.

order for houses to be well ventilated (Murata, 2015). Then, heat generated inside may leak outside, which can be observed as roof surface temperature. Hence, we presume that the house temperature measured from outside may express the inside temperature. In total, we conclude that these rationales underpin the usage of surface temperature of a house as a proxy for the living standard of its resident.

# 3 Data and the Empirical Model

## 3.1 Data Collected by a Drone

A flight campaign was conducted using the Matrice 210 drone with the thermal Zenmuse-XT2 camera (DJI, Shenzhen, China). The altitude of the drone was 140 meters above the ground and aerial images were continuously taken with the camera facing a base along the river edges. The aerial survey obtained thermal information and location data of houses of homeless persons on the Kanagawa Prefecture side along the Tama River, which divides Tokyo Prefecture and Kanagawa Prefecture. The flight permission was obtained regarding the aviation law of drones, and extreme care was taken with the flight altitude and the geographic position of the survey due to the presence of the nearby Haneda Airport. Despite the care taken to fly drones along the Tama River, limitations arise in collection of the data. One limitation is that we were unable to collect data of houses below bridges, over the river. This was to prevent accidents to avoid crashing into bridges. We will return to this point in Section 5.7 to check if the results do not change even if we exclude house observations near bridges from our sample.

The study area is known for homeless persons illegally squatting on the land in those cage-like, rectangular-shaped houses that are usually covered with blue, green, or gray plastic sheets.[16] Because these house features are observed mostly in the target area, we first detect house locations by viewing picture data collected by our drone in the daytime on February 17, 2019. We extracted house-like objects that have the following features: (i) a rectangular or box shape; (ii) covered by blue, green, or gray plastic sheets; (iii) enough size for a person to reside (e.g., with sides longer than about 1.5 meters and a height taller than about 1 meter). We obtained 63 house-like objects that satisfy these conditions in our study area. Figure 1a displays a sample of the detected house polygons highlighted by light blue lines.

[Figure 1 around here]

[16]Figure 7a in Appendix C shows a typical house of a homeless man in this area.

Figure 2a shows the entire study area, in which each house has been given a unique house ID from 1 to 63.

[Figure 2 around here]

Each point in Figure 2a is the centroid of a house, and the blue band running from the southeast to the northwest is the Tama River. Because houses are aligned in a narrow space along the river, which is nearly in a linear space, each house is attached with a unique ID whose numbering starts from the house located in the southeast end (house 1) and ends at the northwest end (house 63). The ID number is based on the number of houses up from house 1 along the river. The distance from house 1 to 63 along the river is about 7.1 kilometers.

To collect the thermal temperature data of each house,[17] we flew our drone three times in the winter of 2019. Our drone flights were conducted at 18:48–21:13 on February 17, 19:15–22:07 on March 14, and at 19:10–21:24 on March 15. The first reason for our collection of the temperature data in the nighttime is due to the disturbance of solar radiation during daytime. Heating material during the day causes limitations in observing the direct source of heat waste, and the specular reflectance of solar radiation causes observation errors.[18]

The second reason for nighttime collection relies on the behavior of homeless persons. Usually, homeless people earn money by collecting used cans and selling them to illegal scrap firms as mentioned in Section 2. In the morning, each Japanese household disposes of used cans at the garbage sites set by local administrations. Homeless persons collect these cans in the morning, which means that there is little possibility that homeless persons are at home in the daytime. Thus, it may not be appropriate to measure house temperatures during daytime, and we conjecture that the homeless are in residence in the evening. In comparison between data collection at daytime and at night, the latter should be preferable. That is why we collected the thermal data at night.

We match these two datasets, the house-location data and house-temperature data, by assuming that during our study period (from February 17 to March 15) there was no relocation of the houses

[17]Thermal temperature from an object can be calculated from the emitted electromagnetic energy based on the Stephan Bolzmann law. The Zenmuse-XT2 (DJI, Shenzhen, China) thermal infrared camera observes the spectral range of 7.5–13.5$\mu m$. The system will sense the thermal infrared waves and convert the electronic signals to the absolute temperature information.

[18]According to Balaras and Argiriou (2002), infrared measurements should be performed at night or during a cloudy day without high wind speeds. Drone flights cannot be performed on windy or rainy days, and our house temperature data (infrared thermal image data) were collected by a drone loading a thermal sensor at night, meaning that the temperature data were collected under appropriate conditions.

detected in a daytime picture taken on February 17.[19] Figure 1b displays a sample of thermal images matched with house polygons.[20] After separating out the part covered by edges and leaves in order not to capture temperature of vegetation covering rooftops, we calculate the average of temperature over the house polygon for each house and attach this average house temperature to each location of houses.[21] Figure 1c illustrates the average house temperature over each polygon, corresponding to the house polygons in Figures 1a and 1b. Hereafter, we simply call this average temperature over a house polygon the house temperature for each single house.

One thing to mention about the house temperature data is that, even though we have 63 house-location observations from the daytime flight, some of these 63 houses lack nighttime temperature values for the following two reasons. The first reason is the technical limitation. Some houses among our observations do not have temperature values due to a difference between resolution levels of the daytime picture data and the nighttime thermal images. The daytime pictures have a higher resolution than the nighttime thermal images. For observations with roofs covered by tiny leaves or thin edges, we cannot distinguish the roof part and the part covered by leaves or tree edges. As it is unable to separate out the uncovered and covered parts of the roof in the nighttime thermal images, the mixing of images occurred. This prevented us from calculating the average house temperature over the polygon by removing the part covered by vegetation. Houses 48, 49, and 53–55 are observations that have missing house temperature values for all dates.

The second reason for missing nighttime house temperature values is simply human error. The nighttime flights are less visible than the daytime flights, so that the former have more difficulty distinguishing objects than the latter. Then, it is more difficult to take thermal image pictures on the exact points, and some houses end up with missing temperature values. Therefore, houses 1, 12–17, and 47–52 lack house temperature data on February 17. On March 14 and 15, there are no missing house temperature values caused by this. By combining the first and second sources of missing house temperature values, finally we have 47 observations for date February 17, 58 observations for the other two dates, and 163 records in total.

[19]It is unlikely that there is house relocation in such a short period of time (about one month) because construction of houses may take sufficiently long even though the houses have simple structures. Thus, it is reasonable and natural to assume that relocations of houses are rare.

[20]The collected aerial imageries are in the format of Radiometric JPEG (R-JPEG). The temperature information is extracted from the R-JPEG data using the FLIR Tools software ver. 6.4 (FLIR, Wilsonville, U.S.). The "box tool" is used to cover the area of the house polygons, and the average and the maximum temperature values for each house polygon are extracted, based on the areas of house polygons covered by the boxes created by the tool.

[21]It is natural that a house polygon has different temperature values at different points. To obtain a representative variable expressing each house temperature, we calculate an average temperature over the polygon area for each house in our baseline model.

### 3.2 Cluster Analysis

As addressed in Section 1, the objective of this paper is to see if residing in clusters benefits homeless persons and the extent to which they are benefited. We first detect the clusters of houses of homeless persons that appear in Figure 2a, by using a hierarchical cluster analysis. Among the class of hierarchical cluster analyses, we adopt single-linkage clustering because of its chaining property. This property means that a cluster created in single-linkage clustering has the shape of a line (chain). As the houses of homeless persons are located in a narrow area along the river, they can be viewed as being aligned in this way. Thus, single-linkage clustering, which generates chain-like clusters, should be the appropriate method. In single-linkage clustering, a between-cluster distance (linkage function between two clusters, $C_1$ and $C_2$) is defined as

$$d_{single}(C_1, C_2) = \min_{i \in C_1, j \in C_2} d_{ij}, \tag{1}$$

where a distance between houses $i$ and $j$ is denoted as $d_{ij}$. In the single-linkage clustering, the linkage score, or between-cluster distance, in (1) is given by the distance between the closest pair of houses, one of which belongs to cluster $C_1$ and the other is in cluster $C_2$, which has no common elements (houses) with $C_1$. Starting from the singleton clusters, we repeat the assignment to merge a pair of clusters having shortest between-cluster distance (1), and the process ends when all houses are included in one identical cluster. The distance between houses, $d_{ij}$, should be defined in a way that expresses proximity between homeless persons. This is because the background motivation for this research is to see if a positive effect of interaction exists among homeless persons living close to each other. At this aim, $d_{ij}$ should represent how far it is for a homeless person $i$ to reach another homeless person $j$. As can be seen in Figure 2a, the river is meandering. This implies that it is inadequate to calculate the direct distance between two houses, because homeless persons do not walk over the river. Then, for $i < j$, where $i$ and $j$ indicate house IDs from 1 to 63, we define $d_{ij}$ as

$$d_{ij} = \sum_{k=i}^{j-1} d_{k,k+1},$$

which is obtained by summing up the direct length between the adjacent house centroids over house $i$ through house $j$ on the line connecting house centroids from house 1 to house 63.[22]

[22]By construction, the symmetry holds so that $d_{ij} = d_{ji}$.

Based on the dissimilarity matrix whose $(i, j)$-th entry is $d_{ij}$, we run a single-linkage clustering to obtain the dendrogram in Figure 3a.

[Figure 3 around here]

Depending on a height at which the dendrogram is horizontally cut, the number of clusters differs. The next step is to determine the optimal number of clusters, accompanied by the determination of the grouping of the houses. To do so, we calculate the Calinski-Harabasz index ($CH$-index) under the possible numbers of clusters to find the number of clusters that returns the maximum $CH$-score. The $CH$-index is the ratio of the between-cluster variation, which measures how spread out the clusters are from each other, relative to the within-cluster variation, which is an inverse measure of how tightly the clusters are grouped.[23] Due to this definition, the number of clusters that gives the maximum $CH$-score should be the optimal number of clusters. Figure 3b shows the $CH$-scores under different numbers of clusters for $K = 2, \cdots, 30$. From Figure 3b, the $CH$-index calculated under the number of clusters $K = 4$ shows the local maximum value but not the global maximum. The $CH$-scores can be obtained for $K = 2, \cdots, 63$ although Figure 3b shows only $K = 2, \cdots, 30$. By calculating $CH$-scores exhaustively, it is revealed that the score basically increases for $K \geq 10$. In such a case, the clusters are not considered to be well separated because the $CH$-index does not have a global maximum. However, if an optimal number of clusters exits, the likeliest number is $K = 4$, according to the discussion and interpretation in Calínski and Harabasz (1974). Thus, we adopt $K = 4$ as our optimal number of clusters. From the dendrogram shown in Figure 3a, when the number of clusters

[23] Occasionally, for a given number of the clusters, $K$, the $CH$-index is defined as

$$CH(K) \equiv \frac{B(K)/(K-1)}{W(K)/(n-K)}, \tag{2}$$

where $n$ is the total number of houses. The within-cluster variation, $W(K)$, is given by $W(K) \equiv \sum_{k=1}^{K}\sum_{i=1}^{n_k}(x_i - \bar{x}_k)^2$, where $n_k$ is the number of houses in cluster $k$, $x_i$ is the position of house $i$, and $\bar{x}_k \equiv \frac{1}{n_k}\sum_{i=1}^{n_k} x_i$. The between-cluster variation, $B(K)$, is given by $B(K) \equiv \sum_{k=1}^{K} n_k(\bar{x}_k - \bar{x})^2$, where $\bar{x} \equiv \frac{1}{n}\sum_{i=1}^{n} x_i$. Because our cluster analysis is not directly based on the points of houses ($x_i$) but directly based on the distance between houses $i$ and $j$, $d_{ij}$, defined in (3.2), it is useful to rewrite the $CH$-index by using $d_{ij}$ and circumvent $x_i$, $\bar{x}$, and $\bar{x}_k$ appearing in the expression. Thus, we use the following version of the $CH$-index, which is equivalent with (2),

$$CH(K) \equiv \frac{[T - W(K)]/(K-1)}{W(K)/(n-K)},$$

where $T \equiv \frac{1}{n}\sum_{i=1}^{n}\sum_{j=i+1}^{n} d_{ij}^2$, and $W(K) = \sum_{k=1}^{K}\frac{1}{n_k}\sum_{i=1}^{n_k}\sum_{j=i+1}^{n_k} d_{ij}^2$. Calculation of this version of the $CH$-index is implemented by Halpin (2016).

is 4, the houses in the sample are grouped into the following 4 clusters: cluster 1 = [1–17], cluster 2 = [18–46], cluster 3 = [47–52], and cluster 4 = [53–63], where the numbers in the brackets indicate house IDs.[24] Figure 2b circles the four clusters of houses. Hereafter, clusters in our analysis are the four clusters depicted in Figure 2b.

### 3.3 Measures of House Agglomeration

We consider variables that measure the extent of agglomeration of homeless persons, namely, the cluster size. As mentioned in Section 2, homeless persons may improve their living standards by helping each other, so that the larger the number of homeless persons residing in a neighborhood, the more the person's life will improve. The number of houses in the same cluster, or the cluster size, may capture this aspect. As an immediate measure of the cluster size, we adopt $N(i)_{-i} \equiv N(i) - 1$, where $N(i)$ is the number of houses in cluster $C(i)$ that house $i$ belongs to.

Although $N(i)_{-i}$ is a direct variable, it does not consider the distance between houses within a cluster. Even within a cluster, distance may matter because people will more interact with those located closer than those who are farther away. Then, we modify $N(i)_{-i}$ to take into account this aspect by discounting the number of houses by distance to other houses in the same cluster. The simplest modification of $N(i)_{-i}$ is the distance-discounted number of houses in $C(i)$ except house $i$:

$$DN_{-i}(i) \equiv \sum_{j \in C(i) \setminus \{i\}} e^{-d_{ij}} = \sum_{j \in C(i)} e^{-d_{ij}} - 1 \equiv DN(i) - 1, \tag{3}$$

where $DN(i)$ is the distance-discounted total number of houses in cluster $C(i)$.[25]

### 3.4 Empirical Model

Our hypothesis is that homeless persons residing in larger clusters are better off, which is translated into the hypothesis that the temperature of a house located in a larger cluster is higher. Note here that $N(i)_{-i}$ and $DN(i)_{-i}$ are time-invariant variables in that they take the same values across all

[24] One may think that in our case, the houses are positioned in one-dimensional space, rather than a higher-dimensional one. With this interpretation of the space of our study area, some may prefer an analysis based on segmentation of the interval rather than clustering analysis. Then, we also conducted an analysis of segmentation to find a grouping of the houses that minimizes the sum of the within-group sums of squared deviations from the group means at a given number of groups $\hat{K}$, following the spirit of Fisher (1958). We used the user written Stata code "1dgroup" to implement the procedure. The results based on segmentation support the same classification of houses as in single-linkage clustering. For details, see Appendix A.

[25] $d_{ij}$ appearing in (3) are measured in kilometers. Then, for example in cluster 1, the magnitude of $e^{-d_{ij}}$ for house 1, the most peripheral house in the cluster, ranges from $e^{-d_{1,2}} \approx 0.797$ to $e^{-d_{1,17}} \approx 0.438$. Also, for house 9, the most central location in the cluster, it ranges from $e^{-d_{9,8}} \approx 0.974$ to $e^{-d_{9,1}} \approx 0.596$.

three dates, as they are only based on the house locations. Then, the estimation model is forced to be a random effects (RE) model:

$$\text{HouseTemperature}_{ictd} = \beta_0+\beta_1\text{Agglomeration}_{ic}+\beta_2\text{OutsideTemperature}_{itd}+\beta_3X_i+\theta_d+\nu_i+\epsilon_{ictd}, \quad (4)$$

where $\text{HouseTemperature}_{ictd}$ is the temperature of house $i$ in cluster $c$ at time $t$ (recording time of the temperature of house $i$) on date $d$.[26] $\theta_d$ is the date fixed effects, $\nu_i$ is house random effects assumed to satisfy $E(\nu_i) = 0$ and is uncorrelated with other explanatory variables of house $i$,[27] and $\epsilon_{ictd}$ is a stochastic disturbance. $\text{Agglomeration}_{ic}$ is either $N(i)_{-i}$ or $DN(i)_{-i}$ proposed in Section 3.3.

$\text{OutsideTemperature}_{itd}$ is the outside temperature in the neighborhood of house $i$ at time $t$ on date $d$. As mentioned in Section 3.1, the data collection time of house temperature for each house was obtained at the minute level. Ideally, we prefer to use the outside temperature data on exactly the same location of houses at the same time, which was not possible due to the limitation of data availability for outside temperatures. What we have is outside temperature data recorded every 10 minutes on two sites, Haneda and Fuchu, where the Automated Meteorological Data Acquisition System (AMeDAS) stations are located near our study area. The AMeDAS is a collection of Automatic Weather Stations run by the Japan Meteorological Agency for automatic observation of precipitation, wind direction/speed, temperature, and sunshine duration to support real-time monitoring of weather conditions with high temporal and spatial resolution.[28]

To fit temperature data recorded by the AMeDAS to our dataset, we construct the outside temperature data in the following procedure. First, from the set of reported times in every 10 minutes in the AMeDAS data, we choose the closest time to time $t$, at which house $i$'s temperature is collected by our drone. Then, we calculate the distance-weighted value of the outside temperature at the time

[26]Some may wonder why time fixed effects ($\mu_t$) do not appear in the estimation model (4). As will be described in the explanation of the variable of $\text{OutsideTemperature}_{itd}$, we have an access to outside temperature data recorded in every 10 minutes, but it is impossible to include time fixed effects at 10 minute intervals in our estimation model due to a small sample size (63 houses and 163 observations in total). Then, we prefer to make use of outside temperature data at 10 minute intervals instead of time fixed effects with time intervals longer than 10 minutes.

[27]This assumption of the RE model is *not* supported in most of the specifications, and we think that $\nu_i$ is correlated with $\text{Agglomeration}_{ic}$. This endogeneity problem is considered in Section 4.2.

[28]See Figure 8 in Appendix C the positional relation of the two AMeDAS stations, Fuchu AMeDAS station and Haneda AMeDAS station, and our study area.

chosen above, based on the distance from house $i$ to the two AMeDAS stations:

$$\begin{aligned}&\text{OutsideTemperature}_{itd}\\&=\frac{\text{OutsideTemperature}_{\text{Haneda},td} * \text{Distance}(\text{House}_i, \text{Fuchu})}{\text{Distance}(\text{House}_i, \text{Fuchu}) + \text{Distance}(\text{House}_i, \text{Haneda})}\\&+\frac{\text{OutsideTemperature}_{\text{Fuchu},td} * \text{Distance}(\text{House}_i, \text{Haneda})}{\text{Distance}(\text{House}_i, \text{Fuchu}) + \text{Distance}(\text{House}_i, \text{Haneda})}.\end{aligned}$$

$X_i$ is a set of time invariant variables of house $i$ other than Agglomeration$_{ic}$ to control for natural and geographic aspects and the amenity effect.[29] We include a forest-coverage variable, a grass-coverage variable, the distance to the river shore, and the distance to the nearest toilet/tap freely available in the riverside area.[30] Figure 5a is a map of a toilet/tap location layer superposed on the house-location layer.

[Figure 5 around here]

A sufficiently large number of toilets and taps are located at the riverside area, so homeless persons can have frequent and daily access to them. Hence, we expect that access to toilets and taps improves the living standards of homeless persons in this area. As for the forest and grass-coverage variables, we create a 5-meter buffer from house edges and calculate the percentage of land covered by either trees or grass.[31] The variable distance to the river shore is the shortest distance between the house edge and the river shore. These three variables are related to geographic and natural characteristics that may directly affect the house temperature. Land areas covered by grass or trees may be warmer than barren areas in the winter. Also, distance to the river shore may affect house temperature because the water temperature may be higher than the land temperature in the winter evenings due to the specific heat of water.[32] Unlike these natural characteristics variables, the distance to the nearest toilet/tap plays a different role in impacting the house temperature. Recall that the house temperature is a proxy for the living standard of a homeless person residing in the house, and our objective is to see how well-off he is. Distance to the nearest toilet/tap is considered to affect the living standards of homeless persons via a channel of a better access to the amenity, in contrast with variables such as outside temperature, vegetation coverage, and distance to the river shore, all of which directly affect

[29] Data sources are listed in Table 13 in Appendix B.

[30] Figure 7b in Appendix C is a picture of a toilet and a tap in the riverside area. As can be seen, the site where a toilet is located, tap water is also available.

[31] For the grass-coverage variable, we assume that grass covers the land hidden by trees in the aerial pictures.

[32] Indeed, by inspecting the thermal images we collected, water areas show higher temperatures than land areas.

house temperatures. We expect that a better access to the sanitary amenity improves a homeless persons' living standard, which leads to a higher house temperature.

### 3.5 Summary Statistics

Table 1 shows the summary statistics of the variables used in the estimation, which includes not only the variables used in the baseline model but also those used in the robustness checks that appear in Section 5.

[Table 1 around here]

One thing to notice about the statistics in 1 is the house temperature values. As can be seen, the house temperature variable shows values lower than the outside temperature, and they sometimes take negative values. This seemingly unnatural tendency of house temperature is in fact natural because the temperature of an artificial building measured from the outside is often lower than that in the vegetation area or the outside temperature, although the case can be different depending on various materials (Michell and Biggs, 1979).

Note that the fact that the artificial materials show lower temperatures than natural objects does not hinder our analysis, because we are measuring temperatures radiating from the inside to the outside of houses and conducting a comparison between the artificial objects (houses of homeless persons), not between artificial and natural objects. Also, as described in Section 2, because of the light structures of houses that have thin walls and roofs covered by plastic sheets to maintain sufficient permeability, we conjecture that it is possible to measure the heat radiating from inside.

Before going into the estimation results, we briefly look at the relationship between the house temperature and the cluster size for each date in Figure 4.

[Figure 4 around here]

In Figure 4, the horizontal axis is the number of houses in a cluster $N(i)$, and the vertical axis is the net house temperature of the outside temperature, defined as the house temperature minus the outside temperature. As explained in Section 3.1, there are houses with missing house temperature values, so that only three clusters appear in the panel for the date February 17. Panels of the other dates display four clusters; the largest cluster with $N(i) = 29$ is the cluster of houses 18–46 (cluster 2 in Figure 2b), the next largest cluster with $N(i) = 17$ is the cluster of houses 53–63 (cluster 1), the third largest cluster with $N(i) = 11$ is the cluster of houses 53–63 (cluster 4), and the smallest cluster with $N(i) = 6$ is the cluster of houses 47–52 (cluster 3). In all panels in Figure 4, the cluster size and the house temperature show a positive relationship, indicating a possibility that belonging to a larger cluster may improve a homeless person's living standard.

# 4 Results

## 4.1 Baseline Results

Because we aim to test whether or not there is a positive effect on the living standards of homeless persons from residing in a larger cluster, we expect a positive sign for $\beta_1$ in (4). Under the assumption that the geographic proximity (in our case, belonging to the same cluster) is a proxy of social interactions, $\beta_1 > 0$ can be interpreted as existence of benefits from social interactions within a cluster. Table 2 shows the results for the baseline model.

[Table 2 around here]

Columns (1) and (2) show the results of the pooled OLS (POLS) and the RE models, on a choice of $N(i)_{-i}$ as the agglomeration variable. Columns (4) and (5) are the corresponding results with a choice of $DN(i)_{-i}$. From the Breusch and Pagan test, the RE model is preferable than the POLS model in both specifications. On the other hand, the Sargan-Hansen statistic for the overidentification test against the orthogonality condition for the RE model does not support the choice of the RE model over the FE model. This implies that the RE model is inappropriate and shows a necessity of resolving the endogeneity issue. We tackle this problem in Section 4.2.

As expected, coefficients of the agglomeration variables ($N(i)_{-i}$ in column (2) and $DN(i)_{-i}$ in column (4)) are significantly positive at the 5% level. When the cluster size is larger, each house shows a higher temperature, meaning that residents with a larger within-cluster network may be better off. The finding that homeless persons located in a larger cluster tend to experience higher living standards does not come from the logic argued in Culhane (2010), Corinth and Lucas (2018), and Lee and Price-Spratlen (2004), which sheds lights on the gravitation of the homeless to services offered to them. In their research on homeless agglomeration, homeless persons in the area of the homeless concentration benefit by agglomeration of services that target them, such as shelters, free meals, and soup kitchens. However, this mechanism does not work in our case because there are no facilities that offer services to homeless persons in our study area. Also, from our interview with the Division of the Medical Welfare Services of Kawasaki City, we learned that there were only three free meal services offered there per month. Thus, it is unlikely that the logic proposed in Culhane (2010), Corinth and Lucas (2018), and Lee and Price-Spratlen (2004) will function in our case. Rather, our paper focuses on the aspects of interaction among homeless persons through channels of information diffusion and mutual aid that stem from interaction with other homeless persons living nearby.

Next, we look at the coefficients of other variables. The variable outside temperature is highly positively correlated with house temperature in all models. This result is natural because the house temperature should be affected by the outside temperature around the house.

Turning to the distance to toilets and taps, its significantly negative coefficient can be interpreted as a sign showing that better access to an amenity (clean water[33] improves a homeless persons' living standards. This result reflects the need for water to wash their bodies and clothes to maintain cleanness (Division of Health & Welfare in Kawasaki City, 2019), as well as for drinking. The finding that access to an amenity such as a toilet and water is valued by the residents is also observed in slums in developing countries. Brueckner (2013) find that residents in slum-like dwellings appreciate access to toilets and water sources in Indonesia. Similarly, having piped water connections to houses, toilets, and bathrooms are highly valued in slums in Mumbai, India (Takeuchi et al., 2008). In Lall et al. (2008), the authors find that households in squatter settlements in Pune, India, derive benefits from neighborhood amenities and the quality of the dwellings, especially from a connection to the sewer. Moreover, Feler and Henderson (2011) point out that, because water is essential, withdrawing water connections to a slum reduces growth rates of the slum population in Brazilian localities. It is natural to find the communality between the residents in slum and the homeless persons in our case because homeless persons in our study area occupy public lands and build houses illegally, similar to the activities of urban squatters.

The significance of the positive relationship of better access to an amenity and homeless persons' living standards can be explained by the unique attribute of the homeless, which is that they do not pay rent. By examining U.S. national data, Corinth and Lucas (2018) report that homeless persons are attracted to a mild climate, one of the important amenities. They ascribed the larger population of the homeless in places with warm climates to the fact that climate is not directly capitalized into the cost of living for homeless persons, who do not pay rent or mortgages. An analogous logic can be applied

[33]In Japan, tap water is good for drinking, so better access to taps means not only better access to water for washing but also for drinking, which is essential for living.

in our context. As a better access to toilets and water is not capitalized via rent, homeless persons living close to toilets/taps can benefit from better access to sanitation amenities without paying the pecuniary costs passed on in rent, which leads to those having toilets/taps in their neighborhood being better off.

As for the vegetation coverage variables (forest coverage and grass coverage), the forest coverage variable is positively correlated with house temperatures. When surrounded by vegetation, especially trees, the house temperature is considered stabilized. Such vegetation in their neighborhood may keep houses warmer in the winter and cooler in the summer. In addition, from the significantly negative coefficients of the distance to the river shore, house temperature is positively associated with closeness to the river shore, as expected. This may be due to the higher specific heat capacity of water than of land. Although we gave these interpretations of the coefficients of the vegetation variables and the distance to the river, another explanation may be possible. For homeless persons, being hidden from society at large may play a pivotal role. The closer to the river shore a house is, the farther it is from the area where the non-homeless people's activities are conducted. The denser the forest surrounding a house, the less likely it is to be seen by the non-homeless people who sometimes harm them through violence. The significantly positive sign of the forest coverage and the significantly negative sign of the distance to the river shore may be a reflection of this advantage. Isolated from the non-homeless society, homeless persons may live easier and enjoy better lives.

### 4.2 Identification

In Section 4.1, we confirmed that the house temperature, a proxy for how well-off each homeless person is, is positively associated with the cluster size. However, an endogeneity problem may be hidden behind this positive relationship: The cluster-size variable, our featured variable, may be endogenous.

As mentioned in Section 4.1, the estimation result based on the RE model suffers from correlation between the error term and the independent variables. We are concerned over the endogeneity issue of $\text{Agglomeration}_{ic}$. Homeless persons who are less able or more disadvantaged may prefer to reside in a larger cluster to seek more support from other homeless persons. Those who do not have an ability to build quality houses due to physical disadvantages may want some surrounding homeless persons to help them. Those who cannot consistently earn enough money may want to stabilize their lives by sharing food or daily necessaries with others. In contrast, those with sufficiently high abilities or who are not as disadvantaged do not need others to rely on, and may prefer to live in isolation. In short, homeless persons unobservably more disadvantaged or less able may sort into larger clusters, while those with unobservable high abilities may sort into smaller clusters.[34] To deal with this omitted variable problem, we construct an instrumental variable that may satisfy exogeneity and relevance.

An appropriate instrumental variable should affect the living standard of a homeless person only though the channel of house distribution. One possibility is a variable that expresses suitability for house construction in a particular area on the riverbank, as it is more likely that a larger cluster of houses is generated in areas suitable to build houses. We would use a variable that indicates whether an area is located in an inner curve or an outer curve as an instrumental variable that would capture this aspect.

To start with, we inspected house distributions on both sides (the Kanagawa and Tokyo sides) along the Tama River in 2020, as shown in Figure 5c.[35] From a casual observation of the distribution of houses on both sides, we find a tendency for there to be almost no houses on the right-opposite side of the shore along which a house cluster is generated. In addition, more houses are located along the inner curves than the outer curves. Usually, a set of two shores of a river are shaped as a pair of nearly parallel lines, implying that the right-opposite side of an inner curve is an outer curve, and vice versa. From this observation, the following conjecture about the distribution of houses may arise: there are more houses located in an inner curve than an outer curve. The tendency that riverside areas located in the inner curve accommodate more houses than those located in the outer curve may be due to the flatness of the land in the inner curve compared to that in the other curve. Usually, flatter spaces are generated in the inner curve because of sedimentation. By contrast, steeper and narrower spaces are generated in the outer curve because of erosion. Areas located in the inner curve should be more suitable for building houses than those located in the outer curve, which leads to larger clusters of houses in the inner than in the outer curve.

[34]Unobservable heterogeneity among homeless persons does not largely come from demographic difference such as sex, age, and marital status, as most of the homeless persons are single older males in our study area. Thus, we focus on heterogeneity of ability or how disadvantaged a homeless person is as unobservable characteristics that cause an omitted variable problem.

[35]Although we collected house distribution data in 2020, they are not suitable for our analysis as the house distribution is unstable and changed in a short period in the winter of 2020. In October 2020, a typhoon hit the study area and most of the houses were washed away. After a few months, we collected aerial pictures of houses built after the typhoon in January 2020, which is depicted in Figure 5c. However, when collecting the house temperature data in February 2020, the house distribution seems to have changed from that collected in January 2020, which may be because this area was still in recovery. Because of the difference between the house location detected in aerial pictures in January 2020 and the thermal images taken in February 2020, the latter of which lack daytime pictures of houses, we were unable to match house locations between these two data collection dates. Due to this instability of the house location distribution and impossibility of matching aerial pictures and thermal images, we utilized only the data from 2019.

Although it is impossible to prove that the inner/outer curve variable affects the house temperature only through the channel of the cluster size, we address the concern that we suspect on the violation of the exogeneity condition. The inner/outer curve variable may be correlated with the distance from a house to the river shore. As the area in an inner curve tends to be flat and roomy, houses can be built inland farther away from the river shore. This indicates the possibility of houses being located in an inner curve is correlated with the distance from houses to the river shore. However, this may not be a crucial concern in the case of our study area. In the riverbank area of the Tama River, even in an inner curve, only the areas very close to the river shore are available for homeless persons, similar to an outer curve. The inner curve area with roomy flat land is used as playgrounds for citizens, which implies that whether a house location is in an inner or outer curve is not correlated with distance between the shore and a house. Thus, the inner/outer curve variable does not affect house temperatures through the channel of distance to the river shore.

To construct the instrumental variable of inner/outer curves, we take the following five steps.[36] In the first step, we locate points every five meters along the river shore line[37] in our study area. In the second step, we draw straight lines by connecting some of the points generated in the first step. In doing so, we admit that newly drawn straight lines can depart from the actual river shore line by 100 meters.[38] In the third step, we identify whether the points generated in the first step are located in an inner curve or an outer curve: (i) if the straight line connecting two points drawn in the second step runs in the land area, we determine that the points between these two points (two ends of the straight line) are located in an inner curve and (ii) if the straight line connecting two points generated in the first step runs in the water area, the points between the two ends of the straight line are considered to be in an outer curve. In the fourth step, we find the nearest point among those generated in the first step for each house, and attach the nearest point's characteristic identified in the third step (that is, whether the point is located in an inner curve or an outer curve) to each house. Then, we obtain a dummy variable which is 1 if a house is located in an inner curve and 0 if it is located in an outer curve as the instrumental variable to implement the cluster size.

[36]An illustrative image of constructing the instrumental variable is depicted in Figure 9 in Appendix C.

[37]Instead of the line data of the river shore provided by the National Land Numerical Information Download Service, we use line data of the river shore that we constructed by delineating a boundary between the water area and the land area based on the aerial photographs provided by the Geospatial Information Authority of Japan (GIAJ). This is because the GIAJ line data is more convenient due to the line smoothness.

[38]We also constructed an instrumental variable based on 50-meter and 10-meter divergences from the river line. However, instrumental variables based on such smaller divergence did not show sufficient relevance, i.e., the first stage F-statistics are small. In addition, simultaneously including all instrumental variables is not supported by the overidentification test. This is why we used an instrumental variable based on a 100-meter divergence from the river shore.

Columns (3) and (6) in Table 2 exhibit the RE-2SLS results for specifications based on the choice of $N(i)_{-i}$ and $DN(i)_{-i}$, respectively. In the first stage estimation, the F-statistic is larger than 10, and the instrumental variable, the inner/outer curve dummy variable, shows strong significance with the expected sign in all specifications, which supports the relevance of the instrumental variable. Turning to the second stage estimation, the coefficients of the agglomeration variables, $N(i)_{-i}$ and $DN(i)_{-i}$ exhibit the strong positive significance of the cluster size, with which we can verify that living with more homeless persons may improve their living standards.

Next, to consider the direction of the bias, we compare the RE and RE-2SLS results. By comparing the coefficients of agglomeration variables in columns (2) and (3) for the $N(i)_{-i}$ specification and columns (5) and (6) for the $DN(i)_{-i}$ specification, we see that the RE estimates are negatively biased in both specifications. We interpret this downward bias as a sign of more disadvantaged or less able homeless persons sorting into a larger cluster. By combining the positive effect of larger clusters on the living standards of homeless persons and this downward bias, we conclude that more disadvantaged homeless persons tend to rely on help from other homeless persons in their neighborhood. By cooperating with others residing in the same cluster, they can decrease difficulty in their lives.[39]

# 5 Robustness Checks

## 5.1 House Size Effect

This section offers robustness checks. First, we considered the effects of the floor area of houses. There may be two opposite effects of the floor area on the house temperature. On one hand, it may be more difficult to heat the whole house if the house floor area is large. Hence, a larger floor area may lower house temperature. On the other hand, more well-off homeless persons may live in larger houses, which can lead to a positive sign of the house floor area. Table 3 is the results when controlling for the house floor size.

[Table 3 around here]

In all specifications, the variable of house floor area is insignificant and indistinguishable from 0,

[39] In the context of immigrant segregation, similar tendency is corroborated. Cutler et al. (2008) study how ethnic segregation influence socioeconomic outcomes of immigrants in metropolitan statistical areas in the United States. The authors find that immigrants who possess unfavorable unobserved characteristics sort themselves into ethnic enclaves more, and ethnic clusters or ethnic concentration gives significantly positive impacts on socioeconomic outcomes. They interpret the downward bias of simple estimates of the relationship between segregation and socioeconomic outcomes of immigrants as a reflection of selection bias.

which may be because the opposite effects of the house floor area cancel each other out. As for the agglomeration variables, both cluster-size variables ($N(i)_{-i}$ in column (2), and $DN(i)_{-i}$ in column (4)) show positively significant impacts on house temperature at the 1% level. Moreover, in a comparison of the RE results and the RE-2SLS results (columns (1) and (2) for the specification of $N(i)_{-i}$; columns (3) and (4) for that of $DN(i)_{-i}$), the negative bias of the uninstrumented simple estimation is verified in both specifications.

### 5.2 Number of Supermarkets

The next robustness check accounts for the effect of the number of supermarkets near the houses of homeless persons. As mentioned in Section 2, homeless persons in our study area rarely beg for food or money. Instead, they earn money and buy food on their own. Also, it is rare that free meals are available in the riverbank of our study area. The Division of Health & Welfare in Kawasaki City (2019) reported that one of the worries for homeless persons is food deficiency. Better access to supermarkets (especially those selling food at discounted prices) in their neighborhood improves their living standards by decreasing expenditures for food.[40] We consider the aspect of access to (cheap) food by adding a variable of the number of supermarkets in a 1- and 2-kilometer radius buffers from the house centroids.

[Table 4 around here]

In Table 4, columns (1)–(4) are for the results of the 1-kilometer buffers and columns (5)–(8) are for the 2-kilometer buffers. In all specifications, the number of supermarkets near houses does not improve the living standards of homeless persons. This may be because our target area is highly urbanized, and there are plenty of food shops, which leads to an insufficient variation in the number of food shops. Also, homeless persons may buy food in supermarkets that sell them at discounted prices. Looking only at the number of supermarkets does not capture what supermarkets homeless persons choose to buy food from at low prices. For the agglomeration variables, the positive coefficients of $N(i)_{-i}$ (columns (2) and (6)) or $DN(i)_{-i}$ (columns (4) and (8)) and the negative biases (comparison between columns (1) and (2) or (5) and (6) for $N(i)_{-i}$, and that between columns (3) and (4) or (7) and (8) for $DN(i)_{-i}$) are verified when the number of supermarkets is added to the set of control variables.

[40] Indeed, Suzuki (2008) interpret a higher density of supermarkets as a better access to food for homeless persons in the research on homeless persons in Osaka, Japan.

### 5.3 Number of Scrap Firms

On the income side, we include the number of scrap firms within 2-kilometer buffers from house centroids. As mentioned in Section 2, homeless persons' incomes mainly come from illegally collecting cans and metal scraps and conducting illicit transactions with the scrap firms. We expect that the number of scrap firms near houses of homeless persons have a positive impact on their living standards. Table 5 displays the results with an additional control variable, the number of scrap firms within 2-kilometer buffers from each house centroid.

[Table 5 around here]

The table reveals that the number of scrap firms has almost no impact on the living standards of homeless persons although the sign of the scrap firm variable is positive, as expected. This insignificance may be because of the closed transactions of used cans and metal scraps. Even though homeless persons earn money by selling the cans and metals they illegally collect to scrap firms, such illicit transactions may be conducted in only a small number of scrap firms. In addition, information about which firms conduct illegal deals with homeless persons is not open. Therefore, we cannot detect the effect of access to such illicit scrap buyers. As for the agglomeration effect, even after controlling for the number of the scrap firms, the significantly positive effect of larger clusters on house temperature still exists (column (2) for $N(i)_{-i}$ and column (4) for $DN(i)_{-i}$) as well as the downward bias of the RE result compared to the 2SLS result (by comparison between columns (1) and (2) for $N(i)_{-i}$ and between columns (3) and (4) for $DN(i)_{-i}$).

### 5.4 Distance to Nisshin Town

We consider another possible type of earnings. The majority of the homeless persons rely on earnings from can or metal scrap collection, but some of them earn money by temporarily working in the construction sector. To take into account this possibility, we include the distance to Nisshin Town in the model. Nisshin Town is known as the site that provides informal job opportunities for daily construction workers and is about seven kilometers away from the southeast end of our study area. We expect a negative sign for the variable distance to Nisshin Town, because the better access (shorter distance) to Nisshin Town is considered to lead to more job opportunities for homeless persons engaged in the construction sector. Table 6 displays the results of this specification.

[Table 6 around here]

The coefficients of the distance to Nisshin Town are positive against our expectation, although they are insignificant in the 2SLS results. The positive sign of the coefficient means that the farther from Nisshin Town homeless persons' houses are, the better off they are. There are two explanations for this unexpected sign. First, homeless persons in the study area may not find daily jobs in Nisshin Town. Cheap hotels for homeless daily construction workers are located in the neighborhood of Nisshin Town. Daily jobs offered around Nisshin Town are considered to be supplied to homeless persons staying in such hotels, not to those living in our study area. Thus, homeless persons in our study area may not find job opportunities in Nisshin Town; instead they earn money by collecting cans or metals at garbage places. In this case, better access to Nisshin Town does not have a significant impact on the homeless person's earnings.

Second, there are gambling facilities located around Kawasaki Station, which is adjacent to Nisshin Town. Some of the homeless persons suffer from gambling addiction in Japan as described in Section 2.3, and being close to an area with a lot of gambling facilities may worsen homeless persons' situations because they waste money in an uneconomical way. Then, the variable of a shorter distance to Nisshin Town may negatively affect house temperatures. That is, residents in houses located close to Nisshin Town may lose more money by gambling, which gives a negative effect to the variable of better access to Nisshin Town on the living standards of homeless persons. Our featured variables, $N(i)_{-i}$ in column (2) and $DN(i)_{-i}$ in column (4) are revealed to have a significant positive effect on house temperatures in this robustness check. Also, a comparison between the RE and RE-2SLS results robustly exhibits the negative bias of the agglomeration effect in both specifications of $N(i)_{-i}$ (columns (1) and (2)) and $DN(i)_{-i}$ (columns (3) and (4)).

### 5.5 Vegetation Coverage

In the baseline model, we choose a 5-meter buffer range for the variables of vegetation coverage. In this robustness check, we changed the buffer range to 10 meters and 15 meters from the house edges. Table 7 shows the results of different vegetation buffers (10-meter buffers for columns (1)–(4) and 15-meter buffers for columns (5)–(8)).

[Table 7 around here]

Unlike in the case of 5 meter buffer from the house perimeters, the forest coverage variable in 15-meter buffer ranges is insignificant, but the forest coverage of 10-meter buffer is significant as the 5-meter buffer. As for the agglomeration variables, the positive impact of $N(i)_{-i}$ and $DN(i)_{-i}$ (columns (2) and (4), respectively) and the negative bias of the RE result remain unchanged effectively (columns (1) and (2) for $N(i)_{-i}$; columns (3) and (4) for $DN(i)_{-i}$).

### 5.6 Wind Speed Effect

We consider the effect of wind speed at the time of the collection of house temperature values. Similar to the calculation of the outside temperature, we calculate the wind speed variable as follows:

$$
\begin{aligned}
&\text{WindSpeed}_{itd} \\
&= \frac{\text{WindSpeed}_{\text{Haneda},td} * \text{Distance}(\text{House}_i, \text{Fuchu})}{\text{Distance}(\text{House}_i, \text{Fuchu}) + \text{Distance}(\text{House}_i, \text{Haneda})} \\
&+ \frac{\text{WindSpeed}_{\text{Fuchu},td} * \text{Distance}(\text{House}_i, \text{Haneda})}{\text{Distance}(\text{House}_i, \text{Fuchu}) + \text{Distance}(\text{House}_i, \text{Haneda})}.
\end{aligned}
$$

We expect a negative effect of the wind speed on house temperatures. Table 8 displays the results of estimations with the baseline variables and the wind speed variable as well.

[Table 8 around here]

As expected, the wind speed variable is negative, but it is insignificant in both specifications. This result is reasonable under the drone flight conditions. As drone flights are prohibited on windy days, and the flights to collect house temperature data were conducted under a light wind, it is natural that the wind speed variable only affects house temperatures insignificantly. Turning to the coefficients on the agglomeration effects within a house cluster, the result remains unchanged: larger cluster sizes ($N(i)_{-i}$ in column (2) and $DN(i)_{-i}$ in column (4)) raise house temperatures, which can be translated into the statement that agglomeration of homeless persons benefits residents in a large cluster. In addition, by comparing the coefficient estimates of the RE and 2SLS-RE (columns (1) and (2) for $N(i)_{-i}$; columns (3) and (4) for $DN(i)_{-i}$), the downward bias is observed as in the baseline specification.

### 5.7 Effect of Bridges

As mentioned in Section 3.1, we cannot collect house location and temperature data for houses located below bridges because we cannot fly drones below them.[41] However, houses of homeless persons may

[41] Figure 5b shows bridges (both pedestrian and non-pedestrian bridges) in the study area.

also be located below bridges, which are invisible in the aerial pictures. In order to address this point, we drop houses sufficiently close to bridges and redo the estimation. In addition, we conduct an analogous estimation by dropping a cluster as a whole, which contains houses sufficiently close to bridges.

The distance between adjacent houses differs across pairs of houses and the average distance between adjacent houses within a cluster also differs across clusters. Then, it is not adequate to set the same distance for all houses and clusters, for instance, at 100 meters or 200 meters, from bridges to create bridge buffers, based on whether we decide a house or a cluster should be dropped from the sample. Instead, we prefer to employ the following distance measure to create buffers from bridges:

$$d_{C(i)}^{\max} \equiv \max_{k \in C(i) \setminus \{\bar{k}_{C(i)}\}} d_{k,k+1}, \tag{5}$$

where $\bar{k}_{C(i)}$ is the maximum house ID in cluster $C(i)$, e.g., $\bar{k}_1 = 17$ for cluster 1, $\bar{k}_2 = 46$ for cluster 2, and so on. In (5), $d_{C(i)}^{\max}$ expresses the maximum distance among distances between two adjacent houses belonging to the same cluster $C(i)$. The assumption behind the adoption of (5) for buffers from bridges is as follows: for each cluster, there are no invisible houses that are farther than the maximum distance between the two visible houses in that cluster. Based on this assumption, if a distance between a bridge and a house closest to the bridge among houses in cluster $C(i)$ is longer than $d_{C(i)}^{\max}$, there should be no houses below the bridge. By contrast, if a distance between a bridge and a house closest to the bridge among houses in cluster $C(i)$ is shorter than $d_{C(i)}^{\max}$, there is possibility that there are houses below the bridge. Such clusters may contain a larger number of houses than originally observed because of the hidden houses below the bridge. Thus, we drop such cluster(s) and repeat the same analysis as the baseline model. More specifically, we run regressions by dropping cluster $C(i)$ such that

$$\min_{k \in C(i)} d_{k,\text{bridge}} < d_{C(i)}^{\max}, \tag{6}$$

where $\min\limits_{k \in C(i)} d_{k,\text{bridge}}$ is the distance between a bridge and the house closest to the bridge among houses in cluster $C(i)$.

We also conduct a less straightforward but similar analysis to the one stated above. Instead of dropping all houses belonging to the cluster satisfying (6), we drop houses that are sufficiently close to the bridge. Namely, we drop house $i$, a member of cluster $C(i)$, such that

$$d_{i,\text{bridge}} < d_{C(i)}^{\max}. \tag{7}$$

Columns (1)–(4) in Table 9 show the results when houses satisfying (7) are dropped, and (5)–(8) are the results when a cluster satisfying (6) is dropped.[42]

[Table 9 around here]

The results are effectively the same for both ways of dropping observations near bridges. That is, there is a positive effect of the cluster size on the living standards of homeless persons, and there is a negative bias that suggests the possibility that more disadvantaged homeless persons tend to enter a larger cluster, and by doing so, they enjoy the benefits of agglomeration.

### 5.8 Effect of Houses on the Opposite Shore

As displayed in Figure 5c, although this map is based on the 2020 house distribution, houses of homeless persons are similarly located on the opposite shore of the river, implying that interaction with homeless persons located on the opposite side may be possible. Because meeting with those located on the other shore is possible only by crossing pedestrian bridges, we exclude houses sufficiently close to pedestrian bridges in the study area. As in Figure 5b, there are three pedestrian bridges in the study area, so we drop samples within 500-meter buffers from each pedestrian bridge, which left 137 observations.[43] Table 10 reports the results.

[Table 10 around here]

Although, in the specification with $N(i)_{-i}$, the first stage F-statistic is not sufficiently high in column (2), we find a positive effect of the cluster size on the house temperature. The result based

[42] In our dataset, only cluster 4 satisfies (6), so that the number of clusters dropped from the analysis is 1. Other clusters than this dropped cluster belong to the same administrative division, so that the administrative fixed effect is not included in the model in Table 9.

[43] In this robustness check, not only the house distribution on the shore of our study area (Kanagawa Prefecture) but also on the other side of the river (Tokyo side) matters. Unfortunately, however, we did not collect house location data on the other shore in 2019. In order to construct buffer lengths from pedestrian bridges as we did in Section 5.7 to deal with the possibility of interaction with homeless persons on the other shore, house location data on the other side of the river, which we do not have, is indispensable. Then, we are forced to set a fixed value of the buffer length common to all houses. As the river width in our study area is about 300–400 meters, we set the buffer length at 500 meters in this robustness check. Unlike in Section 5.7, we only conduct a robustness check by dropping houses that are located near pedestrian bridges, but we cannot implement a robustness check by dropping clusters that contain at least one house included in 500-meter buffers from the pedestrian bridges. This is because the latter robustness check based on dropping clusters as a whole leaves us with only cluster 3, which contains six houses.

on $DN(i)_{-i}$ appearing in column (4) also shows a significantly positive impact. In addition, from a comparison between columns (1) and (2) in the case of $N(i)_{-i}$ and between columns (3) and (4) in the case of $DN(i)_{-i}$, we can confirm that the simple estimates of the relationship between the cluster size and the levels of living standards are biased downwards. Hence, excluding observations that may have interaction with homeless persons on the other side of the river does not change the result.

### 5.9 Different Cluster-Size Variable (Sum of House Temperatures)

We consider a variable expressing the aggregate living standards for each cluster, not just the cluster size itself as we have done throughout the paper. The idea behind the adoption of the aggregate living standards is that a homeless person may be better off if homeless persons in his neighborhood also live relatively well. We basically calculate the sum of house temperatures in cluster $C(i)$ exclusive of house $i$'s contribution. Assuming that house temperatures in winter may express how well-off homeless persons are, the sum of temperatures of houses in a cluster may represent the extent of the total living standards in that cluster. However, the sum of house temperatures in each cluster should be corrected due to the missing values of house temperatures for some houses, as explained in Section 3.1. Then, we propose an adjusted version of the sum of temperatures of houses in cluster $C(i)$, on the basis of the idea that the missing temperature values are replaced with the average of the observed house temperatures in cluster $C(i)$. The variable should be as follows:

$$ST_{-i}(i) \equiv \sum_{j \in C'(i) \setminus \{i\}} T_j + \sum_{j \in C(i) \setminus C'(i)} \bar{T}_{C'(i)} = \sum_{j \in C'(i)} T_j + \sum_{j \in C(i) \setminus C'(i)} \bar{T}_{C'(i)} - T_i = ST(i) - T_i,$$

where $T_j$ is the temperature of house $j$, $C'(i)$ is the subset of houses in cluster $C(i)$ whose house temperatures are observed, $N'(i)$ is the number of houses in the set of houses in $C'(i)$,

$$\bar{T}_{C'(i)} \equiv \frac{1}{N'(i)} \sum_{j \in C'(i)} T_j,$$

which is the average house temperature of houses whose temperatures are observed, and

$$ST(i) \equiv \frac{N(i)}{N'(i)} \sum_{j \in C'(i)} T_j,$$

which expresses the total affluence of cluster $C(i)$. This index is computed by assuming that the unobserved house temperatures are on average equal to observed house temperatures in the same cluster. Then, the missing values of the house temperature are replaced with the average house temperatures of non-missing values in the same cluster. $N(i) - N'(i)$ is consequently the number of houses missing their temperature values in the observation.

As in the case of $N(i)_{-i}$ and $DN(i)_{-i}$, the distance-discounted version of $ST(i)_{-i}$ can be considered as well:

$$DST_{-i}(i) \equiv \sum_{j \in C'(i) \setminus \{i\}} e^{-d_{ij}} T_j + \sum_{j \in C(i) \setminus C'(i)} e^{-d_{ij}} \bar{T}_{C'(i)} = DST(i) - T_i,$$

where

$$DST(i) \equiv \sum_{j \in C'(i)} e^{-d_{ij}} T_j + \sum_{j \in C(i) \setminus C'(i)} e^{-d_{ij}} \bar{T}_{C'(i)}.$$

Turning to the results in Table 11, which are based on the choice of $ST(i)_{-i}$ and $DST(i)_{-i}$ with a replacement of $N(i)_{-i}$ and $DN(i)_{-i}$, the coefficient on $ST(i)_{-i}$ shows significantly positive values in the RE model in column (2).

[Table 11 around here]

On the other hand, the coefficient of $DST(i)_{-i}$ in the RE model is insignificant in column (3), although the sign is positive as expected. As for the 2SLS results, the first stage F-statistic is very low, indicating that the instrumental variable may not explain the endogenous variable, and the relevance does not seem to be satisfied. Despite this weak correlation in the first stage, we look at the coefficients of $ST(i)_{-i}$ and $DST(i)_{-i}$, simply for reference. $ST(i)_{-i}$ is significantly positive at the 5%-level in column (2), while $DST(i)_{-i}$ is only significant at the 10%-level in column (4). This somewhat noisy result comes from the choice of summation of house temperatures over a cluster, which may be partly due to the replacement of lacking house temperatures $T_{j \in C(i) \setminus C'(i)}$ with $\bar{T}_{C'(i)}$. In addition, the house temperature values themselves capture aspects that are not only related to the levels of the living standards but also to the natural conditions at the time of data collection, such as outside temperatures and river water temperatures. Hence, simply summing up house temperatures without partially removing factors not related to the living standards may not express the total living standards for each cluster, which leads to the noisy results.

Similar to the case of $N(i)_{-i}$ or $DN(i)_{-i}$, a downward bias can be found when comparing the results of the RE and the RE-2SLS (columns (1) and (2) for $ST(i)_{-i}$; columns (3) and (4) for $DST(i)_{-i}$). This means that disadvantaged homeless persons sort themselves into more well-off clusters. By locating in the well-off cluster, they benefit from the aggregate well-off status of the homeless persons in the same cluster.

### 5.10 Different Choice of the House Temperature Variable

Despite its instability as a measure of house temperature compared to the average value of the temperatures within house polygons,[44] the maximum value of the house temperature for each house polygon may well represent the inside temperature, because in the thin part of the roof, more heat may radiate from inside. To see if the result remains unchanged, we adopt the maximum value of house temperature instead of the average value as our dependent variable. Accordingly, we recalculate $ST(i)_{-i}$ and $DST(i)_{-i}$ based on the maximum value of house temperature. Table 12 shows the results.

[Table 12 around here]

Columns (1) and (2) are for $N(i)_{-i}$, columns (3) and (4) are for $DN(i)_{-i}$, columns (5) and (6) are for $ST(i)_{-i}$, and columns (7) and (8) are for $DST(i)_{-i}$. The RE results based on the cluster size (column (1) for $N(i)_{-i}$ and column (3) for $DN(i)_{-i}$) are insignificant, although as expected they are positive. However, looking at the results of the RE-2SLS, both specifications show significantly positive coefficients on the agglomeration variable at the 5%-level (column (2) for $N(i)_{-i}$ and column (4) for $DN(i)_{-i}$). As in Section 5.9, in the specifications based on $ST(i)_{-i}$ or $DST(i)_{-i}$, the 2SLS results are significantly positive but only at the 10%-level (column (6) for $ST(i)_{-i}$ and column (8) for $DST(i)_{-i}$). By comparing the RE and RE-2SLS results (comparison between column (1) and (2) for $N(i)_{-i}$, (3) and (4) for $DN(i)_{-i}$, (5) and (6) for $ST(i)_{-i}$, and (7) and (8) for $DST(i)_{-i}$), we can verify the expected downward biases for the uninstrumented coefficients, with which we can conclude that the negative self-selection story may hold even when the variable choice of the living standards of the homeless persons is based on the maximum value of house temperature within each house polygon.

## 6 Concluding Remarks and Policy Implications

In this paper, we investigated how homeless persons benefit from living in clusters via a channel of social interactions with those located in the same cluster. We observed that cage-like houses built by homeless persons are distributed in clusters on the riverbank along the Tama River near Tokyo.

[44] For instance, if heat-producing objects such as birds and small animals are on the roof, the maximum house temperature measured from outside may catch these heat origins that do not express inside temperature. Then, the maximum temperature value of each house polygon may not represent the inside house temperature. To circumvent such mistakes in collecting house temperature data, we prefer to use the average temperature values rather than extreme values such as the highest temperature within house polygons.

Through this observation, we predicted that the homeless enjoyed benefits by living close to other homeless persons. To examine our hypothesis, first we detected locations of houses of homeless persons based on aerial pictures taken by flying a drone. The detection of house locations was possible due to the characteristics of homeless persons on the riverbank—they are not nomadic but settled in fixed locations. By investigating aerial pictures, we detected house-like objects characterized by exterior features of blue, gray, or green rectangular polygons whose edges are longer than 1.5 meters and the heights are taller than 1 meter. Also, we calculated the distance between the detected houses. Using the house location and distance data we created, we found the optimal number of clusters of houses based on single-linkage clustering, a hierarchical cluster-analysis method. After detecting the clusters, we estimated the impact of being located in a larger cluster on the living standards of homeless persons, which were measured by the house temperatures in the winter. This data was collected by using a thermal camera loaded on a drone. Our estimation results showed that if a house is located in a larger cluster, its house temperature is higher, indicating that a homeless person living in a larger cluster is better off.

The uninstrumented estimates may suffer from an endogeneity problem due to the cluster-size variable, which is our featured variable that measures the extent of social interactions among homeless persons. It is likely that more disadvantaged homeless persons tend to sort themselves into larger clusters to seek more support from or cooperation with other homeless persons. Such unobservable characteristics of homeless persons can bring about an omitted variable bias. To instrument the cluster-size variable, we used a dummy variable indicating whether a house is located in an inner or outer curve along the river. This instrumental variable works, because flatter spaces are generated in the inner curve due to sedimentation. Thus, areas located in the inner curve should be more suitable for building houses than those located in the outer curve, which leads to larger clusters of houses in the inner than the outer curve.

By comparing the uninstrumented and instrumented results, we found a negative bias with the uninstrumented estimates, which can be interpreted as an indication of a negative sorting of homeless persons into a large community. In other words, disadvantaged homeless persons sort into a larger cluster. By doing so, they benefit from living in a larger community via interaction with those living in the same cluster. Aside from cluster size effects, we also found that better access to amenities such as toilets and taps improves the living standard of homeless persons.

Based on these findings, we address some comments on the possibilities of policies. If we stand on the viewpoint that we should encourage homeless persons to escape from homelessness, it is important to consider (i) how to decrease the utility they obtain when living under homelessness and (ii) how to increase the utility they gain when living in an apartment or a dormitory as employed workers. Because some of homeless persons choose to be homeless because they are satisfied with the current situation,[45] it is necessary that the utility gained when living in an apartment or a dormitory exceeds that gained when living in homelessness, in order for such autonomous homeless persons to decide not to be homeless.

One way to decrease the utility level gained under homelessness is to clamp down on the illegal scrap firms that conduct transactions with homeless persons. By doing so, the information diffusion about which scrap firms make deals in cans or scrap metals with homeless persons is less meaningful. This decreases the benefit from social interactions with other homeless persons located in the same cluster and makes staying homeless less attractive. From the viewpoint of an amenity, withholding water connections to the riverbank may reduce the homeless population in our study area, although the idea does not involve interaction among homeless persons.[46] More specifically, if the taps do not turn on and toilets are unavailable at night, the homeless lose free access to the public service of sanitation facilities in their neighborhood.[47] Because better access to the sanitation amenity increases the well-being of the homeless, shown in Section 4.1, cutting access to such an amenity may decrease the utility gained when living in homelessness on the riverbank.

As for increasing the utility gained when living as non-homeless citizens, we focus on the mental health benefit that social interaction bears, which was not well considered by the house temperature as a proxy of the level of living standards. Living in an apartment or a dormitory as employed persons means that they are no longer homeless and jobless. Then, the benefits from information diffusion (to earn money by collecting and selling cans to illicit scrap firms) and from mutual aid (to avoid violence from passersby or to help each other when they cannot earn enough money) may be less important than they are now. The most important remaining benefit from having social interactions is the mental well-being from friendships with other ex-homeless friends. One possible way to maintain the mental benefits from social interaction under non-homelessness is to let ex-homeless persons who

[45]Division of Health & Welfare in Kawasaki City (2019) report that 37% of homeless persons in Kawasaki Ward answered that they are satisfied with the current life. By contrast, 11% answered that they want to live in an apartment or dormitory and find a regular job.

[46]This discussion comes from Feler and Henderson (2011), who argue that withholding water connections to slums reduces the population growth in slums, especially for low-education households, as water is an essential and indispensable good.

[47]Basically, taps and toilets set in the riverbank are for playground users. To maintain convenience for the non-homeless users, the amenities on the riverbank should be available only in the daytime.

lived close to each other work in the same place or live in the same apartment building.[48] In this way, ex-homeless persons can enjoy mental benefits by maintaining ties with neighboring homeless persons even after they become non-homeless, which raises the utility level after becoming non-homeless. A combination of these policy implementations can push homeless persons out from homelessness.

To close the paper, we address the limitation of our research. In the dataset that we collected, we have no information about pairwise interactions among homeless persons. What we can observe in our dataset is geographic proximity among homeless persons. Ideally, we would prefer to use the homeless persons' pairwise contact information and the intensity of their interactions. This, however, is not possible due to the data collection difficulties when targeting homeless persons, as mentioned in Section 2.3, because interviewing those with alcoholism or mental and intellectual disorders may cause unreliable answers and a sample selection bias. Thus, the lack of rich information of pairwise interactions is a limitation of the dataset. Despite this limitation, we believe that this paper is innovative in that, by utilizing the recent remote sensing technique (a data collection method that makes use of a thermal sensor loaded on a drone), we were able to analyze the benefits from social interaction among homeless persons, whose actual conditions are difficult to grasp, based on the idea that house temperatures in the winter can be interpreted as a proxy for the living standards of the residents.

[48]This policy is not adequate when keeping contacts among homeless persons induces crime or illicit transactions such as drug dealing. In our case, however, this may not be a big concern, as crimes committed by homeless persons are rarely reported in our study area. Homeless persons themselves do not commit crimes even though they sometimes become victims of violence from strangers.

## References


Akiyama, Y., K. Iizuka, O. Yachida, and S. Sugita (2019): "Fundamental Research on Estimating the Vacant Houses using Thermal Infrared and Visible Imagery Observed from Unmanned Aerial Vehicle (UAV)," *Proceedings of the 28th Annual Meeting on Geographic Information Systems Association of Japan (in Japanese)*, F-2-3.

Allgood, S. and R. S. Warren Jr (2003): "The Duration of Homelessness: Evidence from a National Survey," *Journal of Housing ESconomics*, 12, 273–290.

Balaras, C. A. and A. Argiriou (2002): "Infrared Thermography for Building Diagnostics," *Energy and Buildings*, 34, 171–183.

Bayer, P., S. L. Ross, and G. Topa (2008): "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes," *Journal of Political Economy*, 116, 1150–1196.

Brueckner, J. K. (2013): "Slums in Developing Countries: New Evidence for Indonesia," *Journal of Housing Economics*, 22, 278–290.

Calínski, T. and J. Harabasz (1974): "A Dendrite Method for Cluster Analysis," *Communications in Statistics-theory and Methods*, 3, 1–27.

Cobb-Clark, D. A., N. Herault, R. Scutella, and Y.-P. Tseng (2016): "A Journey Home: What Drives How Long People Are Homeless?" *Journal of Urban Economics*, 91, 57–72.

Corinth, K. and D. S. Lucas (2018): "When Warm and Cold Don't Mix: The Implications of Climate for the Determinants of Homelessness," *Journal of Housing Economics*, 41, 45–56.

Culhane, D. P. (2010): "Tackling Homelessness in Los Angeles' Skid Row: The Role of Policing Strategies and the Spatial Deconcentration of Homelessness," *Criminology & Public Policy*, 9, 851–857.

Cutler, D. M., E. L. Glaeser, and J. L. Vigdor (2008): "When Are Ghettos Bad? Lessons from Immigrant Segregation in the United States," *Journal of Urban Economics*, 63, 759–774.

Division of Health & Welfare in Kawasaki City (2019): "Project for Supporting the Homeless to be Independent (2019–2023) (in Japanese)," .

EARLY, D. W. (2005): "An Empirical Investigation of the Determinants of Street Homelessness," *Journal of Housing Economics*, 14, 27–47.

FELER, L. AND J. V. HENDERSON (2011): "Exclusionary Policies in Urban Development: Underservicing Migrant Households in Brazilian Cities," *Journal of Urban Economics*, 69, 253–272.

FISHER, W. D. (1958): "On Grouping for Maximum Homogeneity," *Journal of the American Statistical Association*, 53, 789–798.

HALPIN, B. (2016): "Cluster Analysis Stopping Rules in Stata," *University of Limerick Working Paper*, WP2016-01.

HELLERSTEIN, J. K., M. MCINERNEY, AND D. NEUMARK (2011): "Neighbors and Coworkers: The Importance of Residential Labor Market Networks," *Journal of Labor Economics*, 29, 659–695.

IIZUKA, K., M. ITOH, S. SHIODERA, T. MATSUBARA, M. DOHAR, AND K. WATANABE (2018): "Advantages of Unmanned Aerial Vehicle (UAV) Photogrammetry for Landscape Analysis Compared with Satellite Data: A Case Study of Postmining Sites in Indonesia," *Cogent Geoscience*, 4, 1498180.

ITO, K., S. MORIKAWA, T. OKAMURA, K. SHIMOKADO, AND S. AWATA (2014): "Factors Associated with Mental Well-being of Homeless People in Japan," *Psychiatry and Clinical Neurosciences*, 68, 145–153.

IWATA, M. (2004): "Who Becomes Homeless? (in Japanese)," *Japan Labor Issues*, 49–58.

KIM, J. S., E. PATACCHINI, P. M. PICARD, AND Y. ZENOU (2017): "Urban Interactions," *IFN Working Paper*, 1192.

LALL, S. V., M. K. LUNDBERG, AND Z. SHALIZI (2008): "Implications of Alternate Policies on Welfare of Slum Dwellers: Evidence from Pune, India," *Journal of Urban Economics*, 63, 56–73.

LEE, B. A. AND T. PRICE-SPRATLEN (2004): "The Geography of Homelessness in American Communities: Concentration or Dispersion?" *City & Community*, 3, 3–27.

MARMAROS, D. AND B. SACERDOTE (2006): "How Do Friendships Form?" *The Quarterly Journal of Economics*, 121, 79–119.

Michell, D. and K. Biggs (1979): "Radiation Cooling of Buildings at Night," *Applied Energy*, 5, 263–275.

Morikawa, S., R. Uehara, K. Okuda, Y. Shimizu, and Y. Nakamura (2011): "Morbidity of Mental Illness among Homeless People in One Area of Tokyo (in Japanese)," *Japanese Journal of Public Health*, 58, 331–339.

Murata, R., ed. (2015): *Homeless Superstar Legends (in Japanese)*, Loft Books.

Nishio, A., R. Horita, T. Sado, S. Mizutani, T. Watanabe, R. Uehara, and M. Yamamoto (2017): "Causes of Homelessness Prevalence: Relationship Between Homelessness and Disability," *Psychiatry and Clinical Neurosciences*, 71, 180–188.

Oakes, P. M. and R. C. Davies (2008): "Intellectual Disability in Homeless Adults: A Prevalence Study," *Journal of Intellectual Disabilities*, 12, 325–334.

Oishi, Y., H. Oguma, A. Tamura, R. Nakamura, and T. Matsunaga (2018): "Animal Detection Using Thermal Images and Its Required Observation Conditions," *Remote Sensing*, 10, 1050.

Okamura, T., K. Ito, S. Morikawa, and S. Awata (2014): "Suicidal Behavior Among Homeless People in Japan," *Social Psychiatry and Psychiatric Epidemiology*, 49, 573–582.

Okamura, T., T. Takeshima, H. Tachimori, K. Takiwaki, Y. Matoba, and S. Awata (2015): "Characteristics of Individuals with Mental Illness in Tokyo Homeless Shelters," *Psychiatric Services*, 66, 1290–1295.

Philippot, P., C. Lecocq, F. Sempoux, H. Nachtergael, and B. Galand (2007): "Psychological Research on Homelessness in Western Europe: A Review from 1970 to 2001," *Journal of Social Issues*, 63, 483–503.

Salize, H. J., C. Dillmann-Lange, G. Stern, B. Kentner-Figura, K. Stamm, W. Rössler, and F. Henn (2002): "Alcoholism and Somatic Comorbidity Among Homeless People in Mannheim, Germany," *Addiction*, 97, 1593–1600.

Schmutte, I. M. (2015): "Job Referral Networks and the Determination of Earnings in Local Labor Markets," *Journal of Labor Economics*, 33, 1–32.

Sugita, S., N. Kobayashi, and M. Dohi (2010): "The Movement of the Sleeping Places of Homeless People and Exclusion of those People (in Japanese)," *Journal of the City Planning Institute of Japan*, 45, 751–756.

Suzuki, W. (2008): "What Determines the Spatial Distribution of Homeless People in Japan?" *Applied Economics Letters*, 15, 1023–1026.

Takano, T., K. Nakamura, S. Takeuchi, and M. Watanabe (1999): "Disease Patterns of the Homeless in Tokyo," *Journal of Urban Health*, 76, 73–84.

Takeuchi, A., M. Cropper, and A. Bento (2008): "Measuring the Welfare Effects of Slum Improvement Programs: The Case of Mumbai," *Journal of Urban Economics*, 64, 65–84.

Van Straaten, B., C. T. Schrijvers, J. Van der Laan, S. N. Boersma, G. Rodenburg, J. R. Wolf, and D. Van de Mheen (2014): "Intellectual Disability Among Dutch Homeless People: Prevalence and Related Psychosocial Problems," *PLoS One*, 9.

# 7 Tables and Figures

## 7.1 Tables

Table 1: Summary Statistics

| Variable | Obs | Mean | Std.Dev | Min | Max |
|---|---|---|---|---|---|
| House Temperature (average) (deg. C) | 163 | 0.28 | 4.30 | -12.8 | 7.5 |
| House Temperature (max) (deg. C) | 163 | 2.17 | 3.86 | -11.1 | 9.3 |
| $N(i)_{-i}$ | 163 | 20.98 | 7.98 | 5 | 28 |
| $DN(i)_i$ | 163 | 16.01 | 6.17 | 4.23 | 22.57 |
| $ST(i)_{-i}$ (average) | 163 | 9.51 | 70.80 | -131.59 | 125.40 |
| $DST(i)_{-i}$ (average) | 163 | 21.10 | 103.76 | -140.53 | 206.02 |
| $ST(i)_{-i}$ (max) | 163 | 46.49 | 67.28 | -93.70 | 159.40 |
| $DST(i)_{-i}$ (max) | 163 | 75.55 | 100.91 | -100.47 | 261.00 |
| Outside temperature (deg. C) | 163 | 9.02 | 2.22 | 5.44 | 11.80 |
| Distance to toilet/tap ($m$) | 163 | 310.69 | 152.19 | 58.33 | 632.96 |
| Forest coverage (5$m$, %) | 163 | 22.78 | 23.74 | 0 | 100 |
| Forest coverage (10$m$, %) | 163 | 21.34 | 18.68 | 0 | 84.75 |
| Forest coverage (15$m$, %) | 163 | 17.05 | 14.84 | 0 | 63.72 |
| Grass coverage (5$m$, %) | 163 | 62.08 | 22.87 | 0 | 100 |
| Grass coverage (10$m$, %) | 163 | 64.25 | 19.35 | 9.28 | 97.84 |
| Grass coverage (15$m$, %) | 163 | 64.44 | 18.82 | 8.73 | 97.83 |
| Distance to the shore ($m$) | 163 | 6.96 | 1.74 | 4.19 | 10.64 |
| House floor area ($m^2$) | 163 | 11.78 | 7.47 | 2.44 | 35.64 |
| Distance to Nisshin Town ($km$) | 163 | 6.96 | 1.74 | 4.19 | 10.64 |
| Wind speed ($m/s$) | 163 | 3.97 | 0.75 | 2.27 | 5.12 |
| Number of supermarkets (1$km$) | 163 | 10.47 | 2.59 | 4 | 15 |
| Number of supermarkets (2$km$) | 163 | 65.60 | 17.14 | 44 | 114 |
| Number of scrap firms (2$km$) | 163 | 14.72 | 2.12 | 11 | 20 |
| Inner/outer curve (=1 if inner, 0 if outer) | 163 | 0.50 | 0.50 | 0 | 1 |

Table 2: Result: Baseline Model

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Model | POLS | RE | RE-2SLS | POLS | RE | RE-2SLS |
| Dependent variable | House Temperature (deg. C) | | | | | |
| $N(i)_{-i}$ | 0.099** | 0.092** | 0.209*** | | | |
| | (0.04) | (0.04) | (0.07) | | | |
| $DN(i)_{-i}$ | | | | 0.134** | 0.121** | 0.286*** |
| | | | | (0.05) | (0.05) | (0.10) |
| Outside temperature (deg. C) | 3.363*** | 2.908*** | 2.812*** | 3.451*** | 2.945*** | 2.879*** |
| | (0.65) | (0.52) | (0.50) | (0.68) | (0.53) | (0.52) |
| Distance to toilet/tap water ($m$) | -0.005** | -0.005*** | -0.007*** | -0.005** | -0.004** | -0.006*** |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| Forest coverage ($5m$, %) | 0.074*** | 0.074*** | 0.070*** | 0.073*** | 0.074*** | 0.069*** |
| | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.01) |
| Grass coverage ($5m$, %) | 0.027 | 0.028 | 0.025 | 0.028 | 0.028* | 0.026 |
| | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) |
| Distance to the river shore ($m$) | -0.061* | -0.058** | -0.099** | -0.066** | -0.062** | -0.111** |
| | (0.03) | (0.03) | (0.04) | (0.03) | (0.03) | (0.05) |
| Date FE | Y | Y | Y | Y | Y | Y |
| Administrative FE | Y | Y | Y | Y | Y | Y |
| Observations | 163 | 163 | 163 | 163 | 163 | 163 |
| Adjusted $R^2$ | 0.71 | | | 0.71 | | |
| Within $R^2$ | | 0.82 | | | 0.82 | |
| Between $R^2$ | | 0.64 | | | 0.64 | |
| Breusch and Pagan LM $\chi^2$ [$p$-value] | | 26.12[0.00] | | | 25.95[0.00] | |
| Sargan-Hansen stat. $\chi^2$ [$p$-value] | | 5.376[0.07] | | | 6.30[0.04] | |
| First stage | | | | | | |
| Inner/outer curve (=1 if inner, 0 if outer) | | | 9.368*** | | | 6.831*** |
| | | | (1.77) | | | (1.40) |
| First stage F-stat. | | | 28.09 | | | 23.92 |

Standard errors clustered by houses are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 3: Result: House Floor Area

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Model | RE | RE-2SLS | RE | RE-2SLS |
| Dependent variable | House Temperature (deg. C) | | | |
| $N(i)_{-i}$ | 0.088** | 0.226*** | | |
| | (0.04) | (0.08) | | |
| $DN(i)_{-i}$ | | | 0.115* | 0.310*** |
| | | | (0.06) | (0.11) |
| House floor area ($m^2$) | 0.007 | -0.027 | 0.008 | -0.027 |
| | (0.03) | (0.03) | (0.03) | (0.03) |
| Outside temperature (deg. C) | 2.898*** | 2.817*** | 2.932*** | 2.889*** |
| | (0.52) | (0.50) | (0.53) | (0.52) |
| Distance to toilet/tap water ($m$) | -0.005** | -0.008*** | -0.004** | -0.007*** |
| | (0.00) | (0.00) | (0.00) | (0.00) |
| Forest coverage ($5m$, %) | 0.074*** | 0.070*** | 0.074*** | 0.068*** |
| | (0.02) | (0.02) | (0.02) | (0.01) |
| Grass coverage ($5m$, %) | 0.028 | 0.025 | 0.028 | 0.026 |
| | (0.02) | (0.02) | (0.02) | (0.02) |
| Distance to the river shore ($m$) | -0.057* | -0.107** | -0.060** | -0.119** |
| | (0.03) | (0.04) | (0.03) | (0.05) |
| Date FE | Y | Y | Y | Y |
| Administrative FE | Y | Y | Y | Y |
| Observations | 163 | 163 | 163 | 163 |
| Within $R^2$ | 0.82 | | 0.82 | |
| Between $R^2$ | 0.65 | | 0.64 | |
| Breusch and Pagan LM $\chi^2$ [$p$-value] | 26.15[0.00] | | 26.00[0.00] | |
| Sargan-Hansen stat. $\chi^2$ [$p$-value] | 5.20[0.07] | | 6.09[0.05] | |
| First stage | | | | |
| Inner/outer curve (=1 if inner, 0 if outer) | | 8.679*** | | 6.329*** |
| | | (1.80) | | (1.44) |
| First stage F-stat. | | 23.17 | | 19.41 |

Standard errors clustered by houses are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Result: Number of Supermarkets

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| Model | RE | RE-2SLS | RE | RE-2SLS | RE | RE-2SLS | RE | RE-2SLS |
| Dependent variable | House Temperature (deg. C) | | | | | | | |
| $N(i)_{-i}$ | 0.092** | 0.200*** | | | 0.065 | 0.325** | | |
| | (0.04) | (0.06) | | | (0.05) | (0.14) | | |
| $DN(i)_{-i}$ | | | 0.122** | 0.275*** | | | 0.081 | 0.429** |
| | | | (0.05) | (0.09) | | | (0.06) | (0.19) |
| Number of supermarkets ($1km$) | -0.120 | -0.127 | -0.119 | -0.127 | | | | |
| | (0.12) | (0.13) | (0.12) | (0.13) | | | | |
| Number of supermarkets ($2km$) | | | | | -0.027 | 0.066 | -0.030 | 0.059 |
| | | | | | (0.03) | (0.06) | (0.03) | (0.05) |
| Outside temperature (deg. C) | 2.731*** | 2.662*** | 2.768*** | 2.724*** | 2.825*** | 2.783*** | 2.843*** | 2.825*** |
| | (0.53) | (0.50) | (0.53) | (0.51) | (0.51) | (0.50) | (0.51) | (0.50) |
| Distance to toilet/tap water ($m$) | -0.005*** | -0.007*** | -0.004** | -0.006*** | -0.004* | -0.011*** | -0.003 | -0.009*** |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| Forest coverage ($5m$, %) | 0.072*** | 0.068*** | 0.071*** | 0.066*** | 0.072*** | 0.074*** | 0.071*** | 0.071*** |
| | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) |
| Grass coverage ($5m$, %) | 0.026 | 0.023 | 0.026 | 0.024 | 0.024 | 0.034* | 0.024 | 0.034 |
| | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) |
| Distance to the river shore ($m$) | -0.065** | -0.104*** | -0.069** | -0.115*** | -0.064** | -0.104** | -0.066** | -0.121** |
| | (0.03) | (0.04) | (0.03) | (0.04) | (0.03) | (0.05) | (0.03) | (0.06) |
| Date FE | Y | Y | Y | Y | Y | Y | Y | Y |
| Administrative FE | Y | Y | Y | Y | Y | Y | Y | Y |
| Observations | 163 | 163 | 163 | 163 | 163 | 163 | 163 | 163 |
| Within $R^2$ | 0.82 | | 0.82 | | 0.82 | | 0.82 | |
| Between $R^2$ | 0.65 | | 0.65 | | 0.65 | | 0.65 | |
| Breusch and Pagan LM $\chi^2$ [$p$-value] | 26.79[0.00] | | 26.76[0.00] | | 26.51[0.00] | | 26.47[0.00] | |
| Sargan-Hansen stat. $\chi^2$ [$p$-value] | 3.57[0.17] | | 6.04[0.05] | | 4.82[0.09] | | 6.47[0.04] | |
| First stage | | | | | | | | |
| Inner/outer curve (=1 if inner, 0 if outer) | | 9.485*** | | 6.915*** | | 7.025*** | | 5.330*** |
| | | (1.77) | | (1.40) | | (2.16) | | (1.74) |
| First stage F-stat. | | 28.6 | | 24.32 | | 10.6 | | 9.41 |

Standard errors clustered by houses are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Result: Number of Scrap Firms

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Model | RE | RE-2SLS | RE | RE-2SLS |
| Dependent variable | House Temperature (deg. C) | | | |
| $N(i)_{-i}$ | 0.102** | 0.254*** | | |
| | (0.05) | (0.08) | | |
| $DN(i)_{-i}$ | | | 0.127* | 0.332*** |
| | | | (0.07) | (0.11) |
| Number of scrap firms ($2km$) | 0.070 | 0.297 | 0.032 | 0.219 |
| | (0.18) | (0.24) | (0.18) | (0.23) |
| Outside temperature (deg. C) | 2.857*** | 2.624*** | 2.921*** | 2.737*** |
| | (0.53) | (0.49) | (0.55) | (0.50) |
| Distance to toilet/tap water ($m$) | -0.005** | -0.010** | -0.005* | -0.009** |
| | (0.00) | (0.00) | (0.00) | (0.00) |
| Forest coverage ($5m$, %) | 0.075*** | 0.071*** | 0.074*** | 0.069*** |
| | (0.02) | (0.01) | (0.02) | (0.01) |
| Grass coverage ($5m$, %) | 0.028* | 0.026 | 0.029* | 0.026 |
| | (0.02) | (0.02) | (0.02) | (0.02) |
| Distance to the river shore ($m$) | -0.060** | -0.108*** | -0.063** | -0.119*** |
| | (0.03) | (0.04) | (0.03) | (0.05) |
| Date FE | Y | Y | Y | Y |
| Administrative FE | Y | Y | Y | Y |
| Observations | 163 | 163 | 163 | 163 |
| Within $R^2$ | 0.82 | | 0.82 | |
| Between $R^2$ | 0.64 | | 0.64 | |
| Breusch and Pagan LM $\chi^2$ [$p$-value] | 26.19[0.00] | | 26.00[0.00] | |
| Sargan-Hansen stat. $\chi^2$ [$p$-value] | 5.64[0.06] | | 6.85[0.03] | |
| First stage | | | | |
| Inner/outer curve (=1 if inner, 0 if outer) | | 8.085*** | | 6.189*** |
| | | (1.47) | | (1.13) |
| First stage F-stat. | | 30.07 | | 29.76 |

Standard errors clustered by houses are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Result: Distance to Nisshin Town

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Model | RE | RE-2SLS | RE | RE-2SLS |
| Dependent variable | House Temperature (deg. C) | | | |
| $N(i)_{-i}$ | 0.083** | 0.190*** | | |
| | (0.03) | (0.07) | | |
| $DN(i)_{-i}$ | | | 0.115** | 0.256*** |
| | | | (0.05) | (0.09) |
| Distance to Nisshin Town ($m$) | 0.000** | 0.000 | 0.000** | 0.000 |
| | (0.00) | (0.00) | (0.00) | (0.00) |
| Outside temperature (deg. C) | 2.606*** | 2.613*** | 2.612*** | 2.624*** |
| | (0.47) | (0.47) | (0.47) | (0.47) |
| Distance to toilet/tap water ($m$) | -0.005** | -0.007*** | -0.005** | -0.006*** |
| | (0.00) | (0.00) | (0.00) | (0.00) |
| Forest coverage (5$m$, %) | 0.074*** | 0.071*** | 0.073*** | 0.069*** |
| | (0.02) | (0.01) | (0.02) | (0.01) |
| Grass coverage (5$m$, %) | 0.029* | 0.026 | 0.029* | 0.027 |
| | (0.02) | (0.02) | (0.02) | (0.02) |
| Distance to the river shore ($m$) | -0.075** | -0.110*** | -0.082** | -0.123*** |
| | (0.03) | (0.04) | (0.03) | (0.04) |
| Date FE | Y | Y | Y | Y |
| Administrative FE | Y | Y | Y | Y |
| Observations | 163 | 163 | 163 | 163 |
| Within $R^2$ | 0.82 | | 0.82 | |
| Between $R^2$ | 0.65 | | 0.65 | |
| Breusch and Pagan LM $\chi^2$ [$p$-value] | 26.35[0.00] | | 25.93[0.00] | |
| Sargan-Hansen stat. $\chi^2$ [$p$-value] | 2.56[0.28] | | 3.70[0.16] | |
| First stage | | | | |
| Inner/outer curve (=1 if inner, 0 if outer) | | 9.445*** | | 7.026*** |
| | | (1.94) | | (1.51) |
| First stage F-stat. | | 23.67 | | 21.71 |

Standard errors clustered by houses are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 7: Result: Vegetation Coverage

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| Model | RE | RE-2SLS | RE | RE-2SLS | RE | RE-2SLS | RE | RE-2SLS |
| Dependent variable | House Temperature (deg. C) | | | | | | | |
| $N(i)_{-i}$ | 0.095** | 0.205*** | | | 0.111** | 0.226** | | |
| | (0.04) | (0.07) | | | (0.06) | (0.09) | | |
| $DN(i)_{-i}$ | | | 0.132** | 0.280*** | | | 0.157** | 0.304** |
| | | | (0.06) | (0.10) | | | (0.07) | (0.13) |
| Forest coverage ($10m$, %) | 0.074*** | 0.063*** | 0.073*** | 0.062*** | | | | |
| | (0.02) | (0.02) | (0.02) | (0.02) | | | | |
| Grass coverage ($10m$, %) | 0.032 | 0.020 | 0.033 | 0.023 | | | | |
| | (0.02) | (0.03) | (0.02) | (0.03) | | | | |
| Forest coverage ($15m$, %) | | | | | 0.035 | 0.015 | 0.035 | 0.017 |
| | | | | | (0.03) | (0.03) | (0.03) | (0.03) |
| Grass coverage ($15m$, %) | | | | | 0.003 | -0.017 | 0.004 | -0.013 |
| | | | | | (0.03) | (0.03) | (0.02) | (0.03) |
| Outside temperature (deg. C) | 2.950*** | 2.865*** | 2.991*** | 2.933*** | 2.909*** | 2.851*** | 2.956*** | 2.924*** |
| | (0.56) | (0.53) | (0.58) | (0.56) | (0.57) | (0.54) | (0.59) | (0.57) |
| Distance to toilet/tap water ($m$) | -0.006*** | -0.008*** | -0.005*** | -0.007*** | -0.006*** | -0.009*** | -0.006*** | -0.008*** |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| Distance to the river shore ($m$) | -0.065** | -0.100** | -0.070** | -0.112** | -0.057* | -0.078** | -0.065** | -0.094** |
| | (0.03) | (0.04) | (0.03) | (0.05) | (0.03) | (0.04) | (0.03) | (0.04) |
| Date FE | Y | Y | Y | Y | Y | Y | Y | Y |
| Administrative FE | Y | Y | Y | Y | Y | Y | Y | Y |
| Observations | 163 | 163 | 163 | 163 | 163 | 163 | 163 | 163 |
| Within $R^2$ | 0.82 | | 0.82 | | 0.82 | | 0.82 | |
| Between $R^2$ | 0.54 | | 0.55 | | 0.48 | | 0.49 | |
| Breusch and Pagan LM $\chi^2$ [$p$-value] | 37.54[0.00] | | 36.61[0.00] | | 44.19[0.00] | | 42.66[0.00] | |
| Sargan-Hansen stat. $\chi^2$ [$p$-value] | 8.81[0.01] | | 11.21[0.00] | | 10.01[0.01] | | 13.69[0.00] | |
| First stage | | | | | | | | |
| Inner/outer curve (=1 if inner, 0 if outer) | | 8.891*** | | 6.516*** | | 8.088*** | | 6.000*** |
| | | (1.80) | | (1.46) | | (1.72) | | (1.45) |
| First stage F-stat. | | 24.43 | | 20.02 | | 21.99 | | 17.10 |

Standard errors clustered by houses are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 8: Result: Wind Speed

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Model | RE | RE-2SLS | RE | RE-2SLS |
| Dependent variable | House Temperature (deg. C) | | | |
| $N(i)_{-i}$ | 0.093** | 0.217*** | | |
| | (0.04) | (0.07) | | |
| $DN(i)_{-i}$ | | | 0.122** | 0.297*** |
| | | | (0.05) | (0.10) |
| Wind speed ($m/s$) | -0.397 | -0.500 | -0.373 | -0.459 |
| | (0.69) | (0.76) | (0.68) | (0.74) |
| Outside temperature (deg. C) | 2.672*** | 2.527*** | 2.723*** | 2.619*** |
| | (0.69) | (0.74) | (0.69) | (0.74) |
| Distance to toilet/tap water ($m$) | -0.005*** | -0.007*** | -0.004** | -0.007*** |
| | (0.00) | (0.00) | (0.00) | (0.00) |
| Forest coverage ($5m$, %) | 0.075*** | 0.070*** | 0.074*** | 0.069*** |
| | (0.02) | (0.02) | (0.02) | (0.02) |
| Grass coverage ($5m$, %) | 0.028 | 0.025 | 0.029 | 0.026 |
| | (0.02) | (0.02) | (0.02) | (0.02) |
| Distance to the river shore ($m$) | -0.057** | -0.101** | -0.061** | -0.113** |
| | (0.03) | (0.04) | (0.03) | (0.05) |
| Date FE | Y | Y | Y | Y |
| Administrative FE | Y | Y | Y | Y |
| Observations | 163 | 163 | 163 | 163 |
| Within $R^2$ | 0.82 | | 0.82 | |
| Between $R^2$ | 0.64 | | 0.64 | |
| Breusch and Pagan LM $\chi^2$ [$p$-value] | 26.21[0.00] | | 25.98[0.00] | |
| Sargan-Hansen stat. $\chi^2$ [$p$-value] | 7.40[0.06] | | 23.19[0.00] | |
| First stage | | | | |
| Inner/outer curve (=1 if inner, 0 if outer) | | 9.465*** | | 6.924*** |
| | | (1.79) | | (1.41) |
| First stage F-stat. | | 28.09 | | 24.11 |

Standard errors clustered by houses are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 9: Result: Bridge Effects

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| Sample | Dropping houses near bridges | | | | Dropping a cluster near bridges | | | |
| Model | RE | RE-2SLS | RE | RE-2SLS | RE | RE-2SLS | RE | RE-2SLS |
| Dependent variable | House Temperature (deg. C) | | | | | | | |
| $N(i)_{-i}$ | 0.088** | 0.186*** | | | 0.076** | 0.146** | | |
| | (0.04) | (0.07) | | | (0.04) | (0.06) | | |
| $DN(i)_{-i}$ | | | 0.119** | 0.255*** | | | 0.109** | 0.197** |
| | | | (0.05) | (0.09) | | | (0.05) | (0.08) |
| Outside temperature (deg. C) | 3.966*** | 3.733*** | 4.068*** | 3.906*** | 6.605*** | 6.587*** | 6.754*** | 6.846*** |
| | (1.10) | (1.06) | (1.14) | (1.11) | (1.50) | (1.54) | (1.54) | (1.55) |
| Distance to toilet/tap water ($m$) | -0.005*** | -0.007*** | -0.005*** | -0.007*** | -0.005*** | -0.007*** | -0.005*** | -0.006*** |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| Forest coverage ($5m$, %) | 0.078*** | 0.075*** | 0.077*** | 0.074*** | 0.075*** | 0.072*** | 0.074*** | 0.072*** |
| | (0.02) | (0.02) | (0.02) | (0.01) | (0.02) | (0.02) | (0.02) | (0.02) |
| Grass coverage ($5m$, %) | 0.035* | 0.032* | 0.036* | 0.034* | 0.032* | 0.030 | 0.033* | 0.031* |
| | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) |
| Distance to the river shore ($m$) | -0.064** | -0.097*** | -0.069** | -0.109*** | -0.079** | -0.104*** | -0.086*** | -0.114*** |
| | (0.03) | (0.04) | (0.03) | (0.04) | (0.03) | (0.03) | (0.03) | (0.03) |
| Date FE | Y | Y | Y | Y | Y | Y | Y | Y |
| Administrative FE | Y | Y | Y | Y | - | - | - | - |
| Observations | 145 | 145 | 145 | 145 | 139 | 139 | 139 | 139 |
| Within $R^2$ | 0.82 | | 0.82 | | 0.83 | | 0.83 | |
| Between $R^2$ | 0.67 | | 0.67 | | 0.65 | | 0.65 | |
| Breusch and Pagan LM $\chi^2$ [$p$-value] | 18.33[0.00] | | 17.90[0.00] | | 20.29[0.00] | | 19.94[0.00] | |
| Sargan-Hansen stat. $\chi^2$ [$p$-value] | 3.41[0.18] | | 4.79[0.09] | | 3.97[0.14] | | 4.91[0.09] | |
| First stage | | | | | | | | |
| Inner/outer curve (=1 if inner, 0 if outer) | | 9.508*** | | 6.929*** | | 9.687*** | | 7.188*** |
| | | (1.80) | | (1.46) | | (1.91) | | (1.53) |
| First stage F-stat. | | 27.82 | | 22.52 | | 25.65 | | 22.15 |

Standard errors clustered by houses are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 10: Result: Effect of Houses on the Opposite Shore (Pedestrian Bridges)

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Model | RE | RE-2SLS | RE | RE-2SLS |
| Dependent variable | House Temperature (deg. C) | | | |
| $N(i)_{-i}$ | 0.120** | 0.346** | | |
| | (0.05) | (0.16) | | |
| $DN(i)_{-i}$ | | | 0.165** | 0.353** |
| | | | (0.06) | (0.14) |
| Outside temperature (deg. C) | 3.312*** | 3.532*** | 3.363*** | 3.546*** |
| | (0.60) | (0.66) | (0.60) | (0.66) |
| Distance to toilet/tap water ($m$) | -0.006** | -0.013*** | -0.006*** | -0.010*** |
| | (0.00) | (0.00) | (0.00) | (0.00) |
| Forest coverage ($5m$, %) | 0.073*** | 0.055*** | 0.073*** | 0.062*** |
| | (0.02) | (0.02) | (0.02) | (0.02) |
| Grass coverage ($5m$, %) | 0.021 | 0.009 | 0.020 | 0.013 |
| | (0.02) | (0.02) | (0.02) | (0.02) |
| Distance to the river shore ($m$) | -0.054* | -0.137* | -0.059** | -0.115** |
| | (0.03) | (0.08) | (0.03) | (0.06) |
| Date FE | Y | Y | Y | Y |
| Administrative FE | Y | Y | Y | Y |
| Observations | 137 | 137 | 137 | 137 |
| Within $R^2$ | 0.82 | | 0.82 | |
| Between $R^2$ | 0.70 | | 0.71 | |
| Breusch and Pagan LM $\chi^2$ [$p$-value] | 15.70[0.00] | | 14.76[0.00] | |
| Sargan-Hansen stat. $\chi^2$ [$p$-value] | 9.61[0.01] | | 10.42[0.01] | |
| First stage | | | | |
| Inner/outer curve (=1 if inner, 0 if outer) | | 5.915** | | 5.819*** |
| | | (2.32) | | (1.79) |
| First stage F-stat. | | 6.47 | | 10.52 |

Standard errors clustered by houses are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 11: Result: Choice of $ST(i)_{-i}$ and $DST(i)_{-i}$

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Model | RE | RE-2SLS | RE | RE-2SLS |
| Dependent variable | House Temperature (deg. C) | | | |
| $ST(i)_{-i}$ | 0.016*** | 0.111** | | |
| | (0.00) | (0.05) | | |
| $DST(i)_{-i}$ | | | 0.002 | 0.082* |
| | | | (0.00) | (0.04) |
| Outside temperature (deg. C) | 1.643*** | -6.126 | 2.652*** | -7.313 |
| | (0.50) | (4.50) | (0.81) | (5.82) |
| Distance to toilet/tap water ($m$) | -0.004** | -0.008*** | -0.003* | -0.007*** |
| | (0.00) | (0.00) | (0.00) | (0.00) |
| Forest coverage ($5m$, %) | 0.077*** | 0.074*** | 0.077*** | 0.076*** |
| | (0.02) | (0.02) | (0.02) | (0.02) |
| Grass coverage ($5m$, %) | 0.030* | 0.027 | 0.030* | 0.027 |
| | (0.02) | (0.02) | (0.02) | (0.02) |
| Distance to the river shore ($m$) | -0.036 | -0.100 | -0.027 | -0.087 |
| | (0.03) | (0.06) | (0.03) | (0.07) |
| Date FE | Y | Y | Y | Y |
| Administrative FE | Y | Y | Y | Y |
| Observations | 163 | 163 | 163 | 163 |
| Within $R^2$ | 0.82 | | 0.82 | |
| Between $R^2$ | 0.62 | | 0.62 | |
| Breusch and Pagan LM $\chi^2$ ($p$-value) | 31.11[0.00] | | 29.64[0.00] | |
| Sargan-Hansen stat. $\chi^2$ ($p$-value) | 9.14[0.03] | | 14.77[ 0.01] | |
| First stage | | | | |
| Inner/outer curve (=1 if inner, 0 if outer) | | 17.589*** | | 23.960** |
| | | (6.59) | | (10.45) |
| First stage F-stat. | | 7.13 | | 5.25 |

Standard errors clustered by houses are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

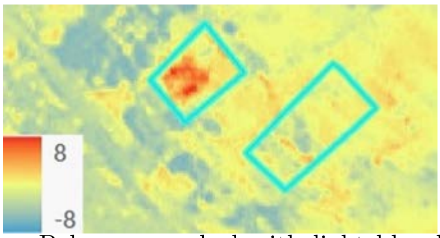
Table 12: Result: Maximum House Temperature within a House Polygon

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| Model | RE | RE-2SLS | RE | RE-2SLS | RE | RE-2SLS | RE | RE-2SLS |
| Dependent variable | House Temperature (maximum) (deg. C) | | | | | | | |
| $N(i)_{-i}$ | 0.064 | 0.156** | | | | | | |
| | (0.04) | (0.07) | | | | | | |
| $DN(i)_{-i}$ | | | 0.067 | 0.213** | | | | |
| | | | (0.06) | (0.10) | | | | |
| $ST(i)_{-i}$ | | | | | 0.015*** | 0.046* | | |
| | | | | | (0.00) | (0.02) | | |
| $DST(i)_{-i}$ | | | | | | | 0.007* | 0.032* |
| | | | | | | | (0.00) | (0.02) |
| Outside temperature (deg. C) | 3.707*** | 3.641*** | 3.725*** | 3.681*** | 2.474*** | -0 .247 | 2.895*** | -0.403 |
| | (0.54) | (0.53) | (0.55) | (0.54) | (0.54) | (2.10) | (0.65) | (2.33) |
| Distance to toilet/tap water ($m$) | -0.004 | -0.005** | -0.003 | -0.005** | -0.003* | -0.006** | -0.003 | -0.005** |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| Forest coverage ($5m$, %) | 0.059*** | 0.056*** | 0.059*** | 0.055*** | 0.060*** | 0.057*** | 0.060*** | 0.056*** |
| | (0.02) | (0.01) | (0.02) | (0.01) | (0.02) | (0.02) | (0.02) | (0.02) |
| Grass coverage ($5m$, %) | 0.014 | 0.012 | 0.015 | 0.012 | 0.014 | 0.012 | 0.015 | 0.012 |
| | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) |
| Distance to the river shore ($m$) | -0.051** | -0.083** | -0.048* | -0.092** | -0.045* | -0.081* | -0.039 | -0.082 |
| | (0.03) | (0.04) | (0.03) | (0.05) | (0.02) | (0.05) | (0.02) | (0.05) |
| Date FE | Y | Y | Y | Y | Y | Y | Y | Y |
| Administrative FE | Y | Y | Y | Y | Y | Y | Y | Y |
| Observations | 163 | 163 | 163 | 163 | 163 | 163 | 163 | 163 |
| Within $R^2$ | 0.79 | | 0.79 | | 0.80 | | 0.80 | |
| Between $R^2$ | 0.59 | | 0.58 | | 0.57 | | 0.58 | |
| Breusch and Pagan LM $\chi^2$ ($p$-value) | 21.89[0.00] | | 22.56[0.00] | | 25.31[0.00] | | 24.44[0.00] | |
| Sargan-Hansen stat. $\chi^2$ ($p$-value) | 3.81[ 0.15] | | 4.38[0.11] | | 23.26[0.00] | | 5.29[0.15] | |
| First stage | | | | | | | | |
| Inner/outer curve (=1 if inner, 0 if outer) | | 9.365*** | | 6.834*** | | 31.478*** | | 46.051*** |
| | | (1.77) | | (1.40) | | (9.60) | | (15.27) |
| First stage F-stat. | | 28.10 | | 23.92 | | 10.75 | | 9.09 |

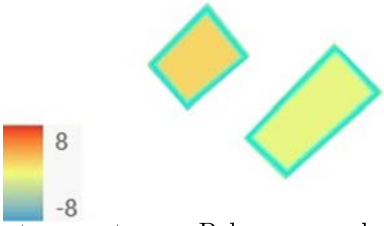
Standard errors clustered by houses are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

## 7.2 Figures

(a) Sample of houses (orthophoto): Polygons marked with light blue lines are detected houses (on February 17th).

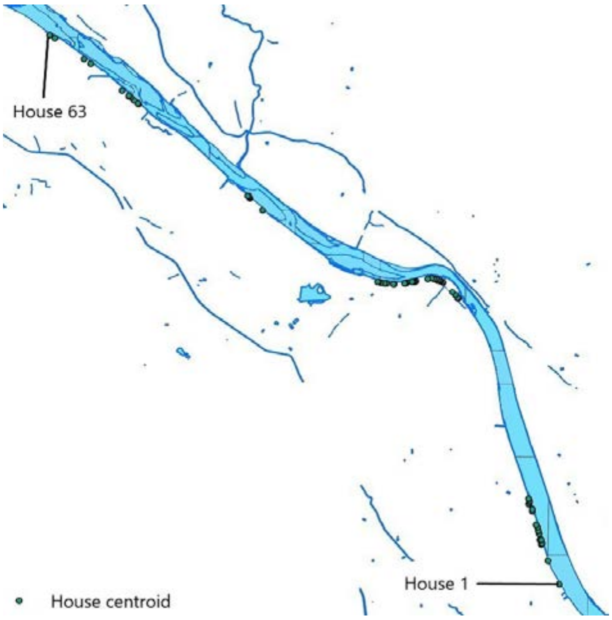



(b) Sample of thermal images: Polygons marked with light blue lines correspond to the detected houses in (1a). Reddish (bluish, respectively) area indicates higher (lower, respectively) temperatures. Temperature is measured in deg. C. Data collected on February 17th.

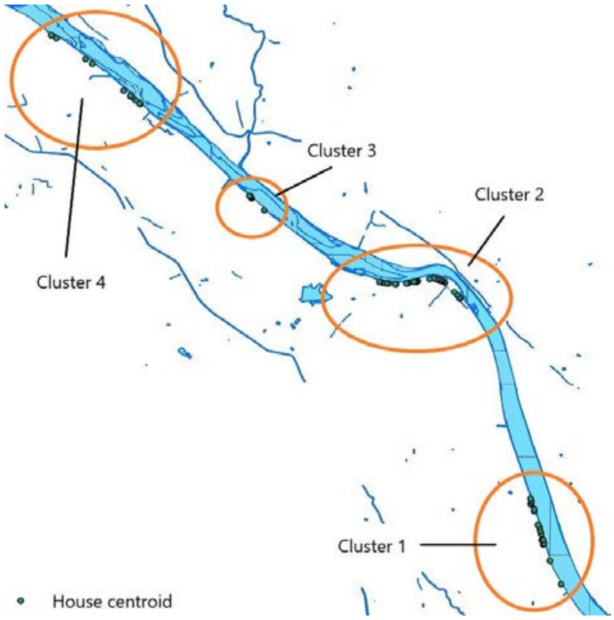



(c) Sample of average house temperatures: Polygons marked with light blue correspond to the detected houses in (1a). Each polygon is colored by the average temperature within each polygon area. Data collected on February 17th.

Figure 1: House polygons and temperatures

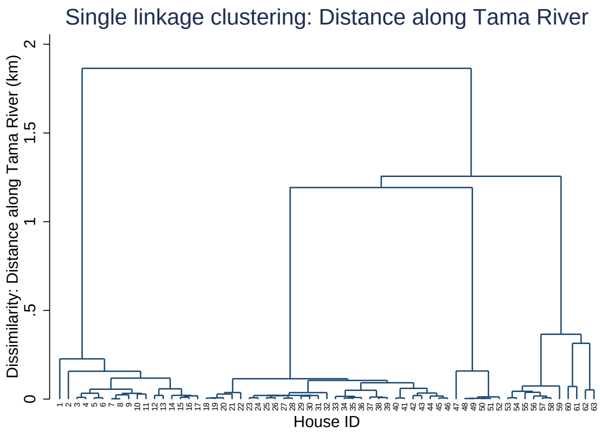



(a) Location of house centroids in the whole study area

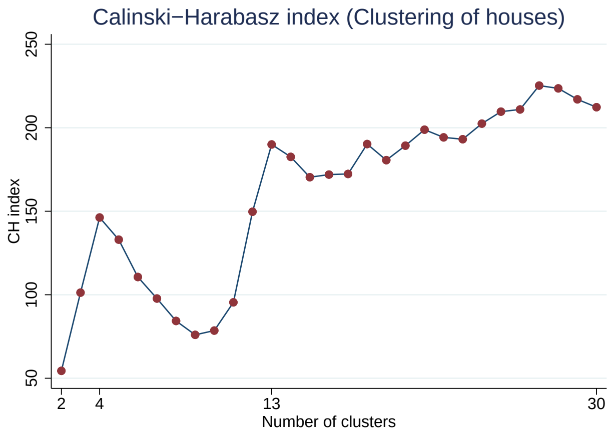



(b) House centroids and clusters

Figure 2: House locations and clusters




(a) Dendrogram based on single linkage cluster




(b) $CH$-index in single linkage clustering

Figure 3: Single linkage clustering and optimal number of clusters

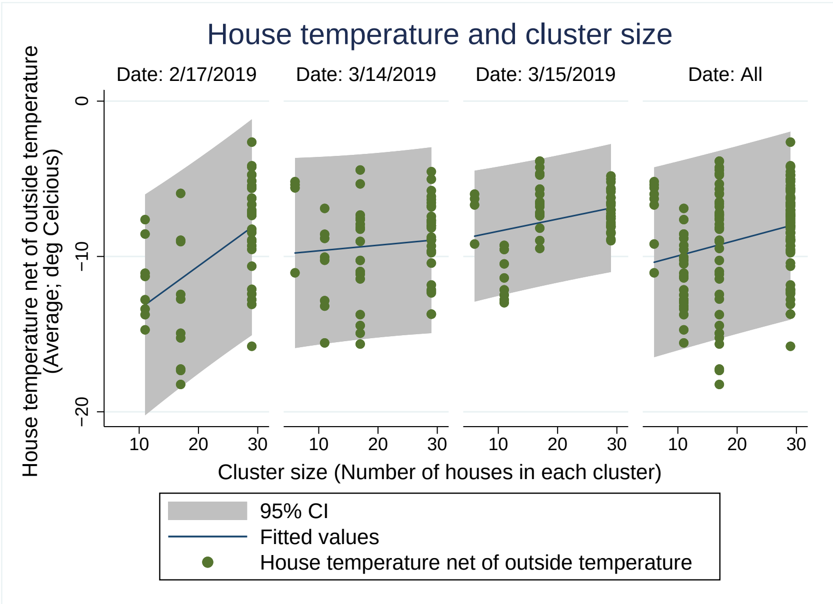
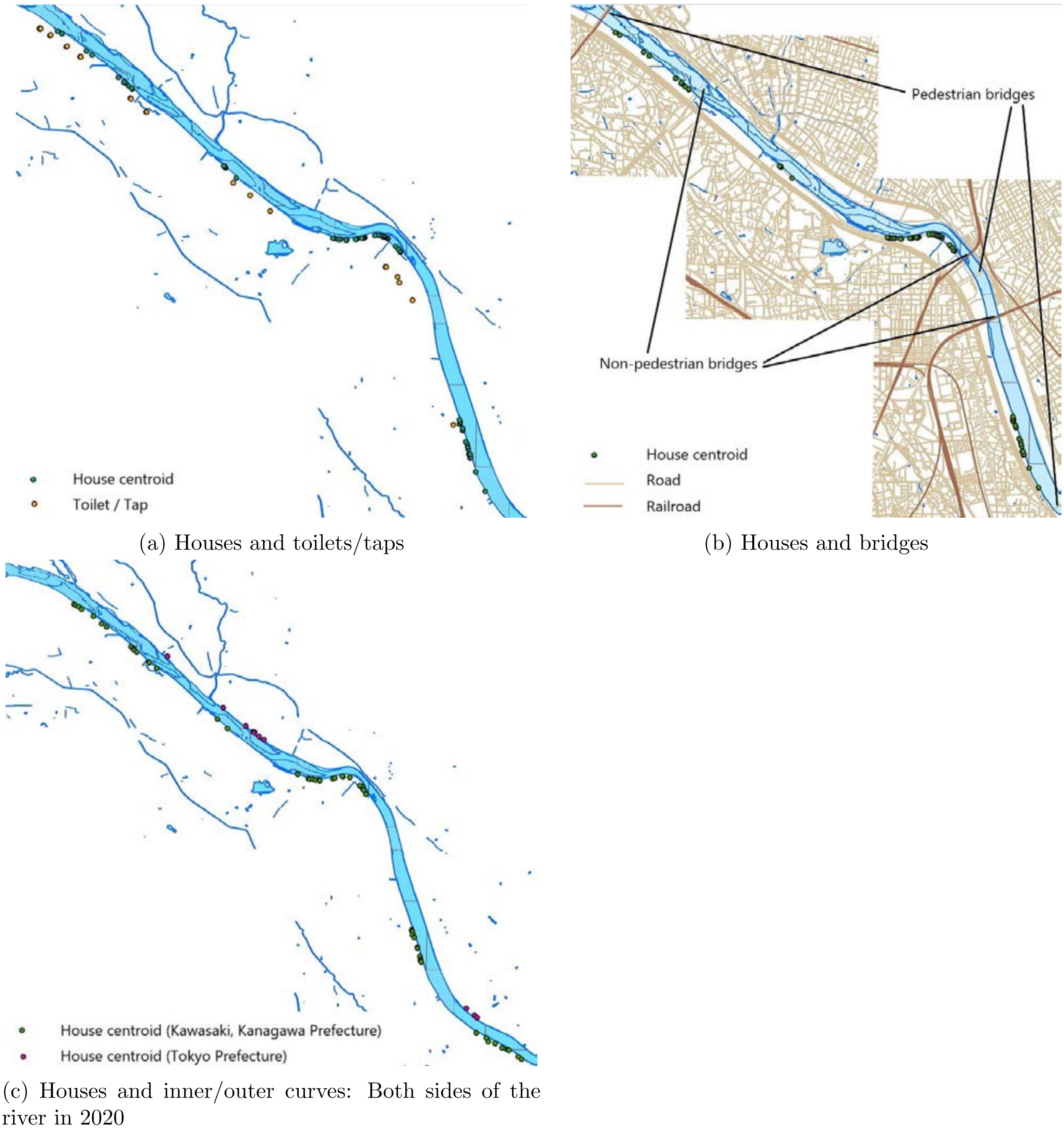



Figure 4: Cluster sizes and house temperature net of the outside temperature




(a) Houses and toilets/taps

(b) Houses and bridges

(c) Houses and inner/outer curves: Both sides of the river in 2020

Figure 5: Maps of the study area

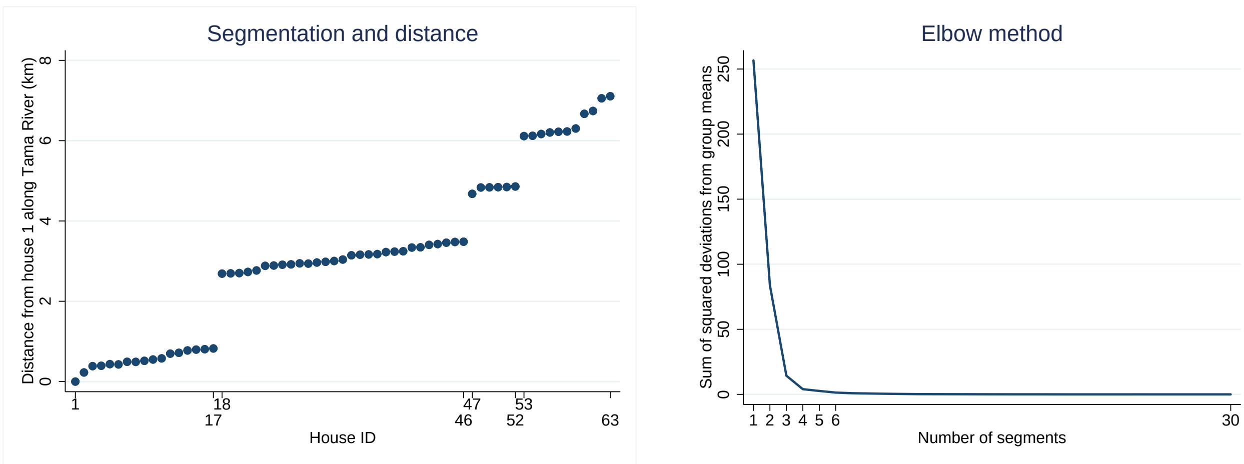
# Appendix A Segmentation in a One-Dimensional Space

This appendix shows the results of grouping houses based on segmentation. In segmentation, each house is on a line that has a measurement. Figure 6a shows $d_{1i}, i \in \{2, \cdots, 63\}$ on the vertical axis and the house ID on the horizontal axis.

[Figure 6 around here]

We consider a line starting from house 1. All houses are on the line and ordered by house IDs. Put differently, they are ordered in terms of $d_{1i}$ from small to large. By choosing $\hat{K}-1$ boundaries from 62 intervals to minimize the sum of the within-group sums of squared deviations from group means, houses are optimally grouped into $\hat{K}$ groups.[49]

Next, we casually investigate the optimal number of groups, using the elbow method. The vertical axis in Figure 6b is the sum of the within-group sums of squared deviations from group means at a given number of clusters (segments) $\hat{K}$. From Figure 6b, $\hat{K}=3$ or 4 seems to be the elbow. That is, $\hat{K}=3$ and 4 are candidates of the optimal number of segments. By a casual inspection of Figure 6a, the houses may be grouped into four groups rather than three, the first of which consists of houses 1 to 17, the second consists of houses 18 to 46, the third consists of 47 to 53, and the fourth consists of 53 to 63. This is the same classification as in the single-linkage clustering.




(a) Distance from house 1

(b) Elbow method

Figure 6: Segmentation and optimal number of segments

[49]This procedure is implemented by a Stata user written command "group1d."

# Appendix B Data source

Table 13: Data source

| Variable | Source | |
|---|---|---|
| House Temperature | Own collection and calculation | |
| House location | Own collection | |
| Outside temperature | | |
| Past temperature | Japan Meteorological Agency (JMA) | http://www.data.jma.go.jp/obd/stats/etrn/index.php |
| Distance from the AMeDAS to houses | Own calculation | |
| $N(i), N(i)_{-i}$ | Own calculation from the house location data | |
| $DN(i), DN(i)_i$ | Own calculation from the house location data | |
| $ST(i), ST(i)_{-i}$ | Own calculation from the house location data and the JMA temperature data | |
| $DST(i), DST(i)_{-i}$ | Own calculation from the house location data and the JMA temperature data | |
| Inner/outer curve (=1 if inner, 0 if outer) | Own calculation from Geospatial Information Authority of Japan (GIAJ) | https://www.gsi.go.jp/ENGLISH/index.html |
| Distance to toilet/tap ($m$) | Own calculation from the house location data and the GIAJ data | |
| Forest coverage | Own collection and calculation | |
| Grass coverage | Own collection and calculation | |
| Distance to the river shore ($m$) | Own calculation from the house location data and the line data of the shore of Tama River provided by the National Land Numerical Information Download Service | |
| Distance to Nisshin Town ($km$) | Own calculation from the house location data and the GIAJ data | |
| House floor area | Own collection and calculation | |
| Past wind speed | Japan Meteorological Agency | |
| Past wind speed | Japan Meteorological Agency (JMA) | |
| Distance from the AMeDAS to houses | Own calculation | |
| Number of supermarkets ($1km$) | Own data transformation and calculation from the data of NTT Hello Page | |
| Data transformation | Address matching service provided by Center for Spatial Information Science (CSIS), University of Tokyo | http://newspat.csis.u-tokyo.ac.jp/geocode/ |
| Number of supermarkets ($2km$) | Own data transformation and calculation from the data of NTT Hello Page | |
| Data transformation | Address matching service provided by CSIS | |
| Number of scrap firms ($2km$) | Own data transformation and calculation from the data of NTT Hello Page | |
| Data transformation | Address matching service provided by CSIS | |
| Distance from bridges | Own calculation from the line data of the rail roads and roads provided by the National Land Numerical Information Download Service | http://nlftp.mlit.go.jp/ksj-e/index.html |

## Appendix C Maps and pictures

(a) A typical house of a homeless man in this area

(b) Toilet and tap in the bank area

Figure 7: Pictures


Figure 8: Two AMeDAS stations and the study area

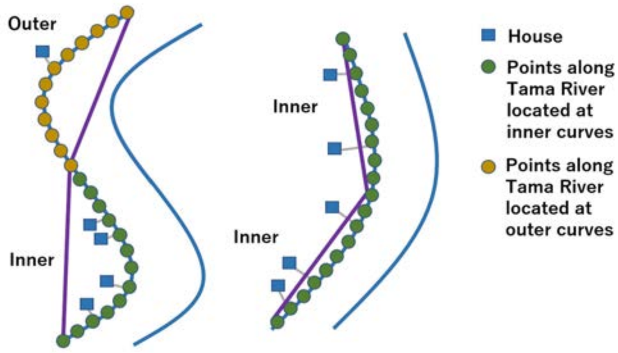



Figure 9: Image of constructing the instrumental variable