Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 August 2018 doi:10.20944/preprints201808.0389.v1

Article

Measurement of Opportunity Cost of Travel Time for Predicting Future Residential Mobility based on the Smart Card Data of Public Transportation

Takashi Nicholas Maeda [1], Junichiro Mori [2], Masanao Ochi [1], Tetsuo Sakimoto [3] and Ichiro Sakata [1]

[1] Graduate School of Engineering, The University of Tokyo, 7-3-1, Hongo, Bunkyo, Tokyo 113-8656, Japan; maeda@ipr-ctr.t.u-tokyo.ac.jp (TN.M.), ochi@ipr-ctr.t.u-tokyo.ac.jp (M.O.), isakata@ipr-ctr.t.u-tokyo.ac.jp(I.S.)
[2] Mathematics and Informatics Center, The University of Tokyo, 7-3-1, Hongo, Bunkyo, Tokyo 113-8656, Japan; mori@mi.u-tokyo.ac.jp
[3] Policy Alternatives Research Institute, The University of Tokyo, 7-3-1, Hongo, Bunkyo, Tokyo 113-8656, Japan; sakimoto-t@ipr-ctr.t.u-tokyo.ac.jp
* Correspondence: maeda@ipr-ctr.t.u-tokyo.ac.jp; Tel.: +81-3-5841-1161

Abstract: This study attempts to investigate a method for creating an index from mobility data that not only correlates with the number of people who relocate to a place but also has causal influence on the number of such individuals. By creating an index based on human mobility data, it becomes possible to predict the influence of urban development on future residential movements. In this paper, we propose a method called the travel cost method for multiple places (TCM4MP) by extending the conventional travel cost method (TCM). We assume that the opportunity cost of travel time on non-working days reflects the convenience and amenities of a neighborhood. However, conventional TCM does not assume that the opportunity cost of travel time varies according to the departure place. In this paper, TCM4MP is proposed to estimate the opportunity cost of travel time with respect to the departure place. We consider such estimation as possible due to the use of massive mobility data. We assume that the opportunity cost of travel time on non-working days reflects the convenience and amenities of the neighborhood. Therefore, we consider that the opportunity cost of travel time has a causal influence on future residential mobility. In this paper, the validity of the proposed method is tested using the smart card data of public transportation in Western Japan. Our proposed method is beneficial for urban planners in estimating the effects of urban development and detecting the shrinkage and growth of a population.

Keywords: human mobility; residential mobility; smart card; public transportation; opportunity cost of travel time

1. Introduction

Urban planners are making efforts to make cities livable for residents; they are also making efforts to predict future population change. Therefore, it is highly important to detect the shrinkage or growth of a population and to estimate the effects of urban development on future populations. However, it is difficult immediately following urban development to estimate the effects of urban development on future residential mobility.

Previous studies have revealed that urban form influences both non-work travel behavior [1–4] and residential mobility [5,6]. The amenities of residential neighborhoods reduce the needs for non-work travel to distant places and increase the number in-migrants. For example, if shopping centers and parks are located in a neighborhood, residents do not have to go to shopping centers and parks located far away. Such convenience also contributes to an increase in residential movements to such a location.

© 2018 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 August 2018 doi:10.20944/preprints201808.0389.v1

Figure 1. Hypothesis on the causal relationship between urban form, non-work human mobility, and residential mobility. The objective of our paper is to create an index based on non-work travel behavior that not only correlates with the number of people who relocate to each place but also has causal influence on the number of residential movements.

We consider that the analysis of non-work travel contributes to the prediction of future residential mobility. Figure 1 shows our hypothesis on the causal relationship between urban form, non-work human mobility, and residential mobility. Changes in urban form (e.g., the amenities of a neighborhood) could be a cause of changes in non-work human mobility (e.g., average travel time and diversity of destinations), and changes in non-work human mobility could be a cause of changes in residential mobility (e.g., the number of people who relocate to each place). Therefore, we consider it possible to create an index based on mobility data to estimate the influence of urban development on residential mobility. Since human mobility changes immediately according to urban changes, such as the opening of a shopping center, we consider that mobility data enable us to immediately predict the outcome of urban development. Therefore, predicting the causal influence of changes in urban form on future residential movements does not take a long time. On the other hand, opportunities for analyzing massive real-time mobility data are increasing due to the automatic collection of mobility data such as smart card data of public transportation. Therefore, we consider that massive collected mobility data can contribute to the prediction of future residential mobility.

This study attempts to investigate a method for creating an index that not only correlates with future residential movements but also has causal influence on future residential movements. By creating an index based on human mobility data, it will be possible to predict the influence of urban development on future residential movements.

In this paper, we propose a method called *travel cost method for multiple places* (TCM4MP) by extending the conventional travel cost method (TCM). In previous studies, TCM was used for inferring the benefits of a recreational site [7,8] and the opportunity cost of travel time [9,10]. We consider that the opportunity cost of travel time on non-working days reflects the convenience and amenities of a neighborhood. However, conventional TCM does not assume that the opportunity cost of travel time varies according to the departure place. In this paper, TCM4MP is proposed to estimate the opportunity cost of travel time that varies according to the departure place. We consider such estimation as possible due to the use of massive mobility data. We assume that the opportunity cost of travel time on non-working days reflects the convenience and amenities of a neighborhood. Therefore, we consider that the opportunity cost of travel time has a causal influence on future residential movements.

The contributions of this paper are summarized as follows:

- We propose a method to infer the opportunity cost of travel time on non-working days that varies according to the departure place by extending the conventional travel cost method.
- We examine the extent to which the opportunity cost of travel time contributes to the prediction of future residential movements.

We compare the contribution of opportunity cost to predict the number of people who relocate to each place with other types of indices derived from smart card data of public transportation. It is insufficient to examine the contribution of the opportunity cost of travel time to the prediction of future residential movements only through the correlation to the number of people who relocate to each place. We infer a causal relation between the number of residential moves, the opportunity cost

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 August 2018 doi:10.20944/preprints201808.0389.v1

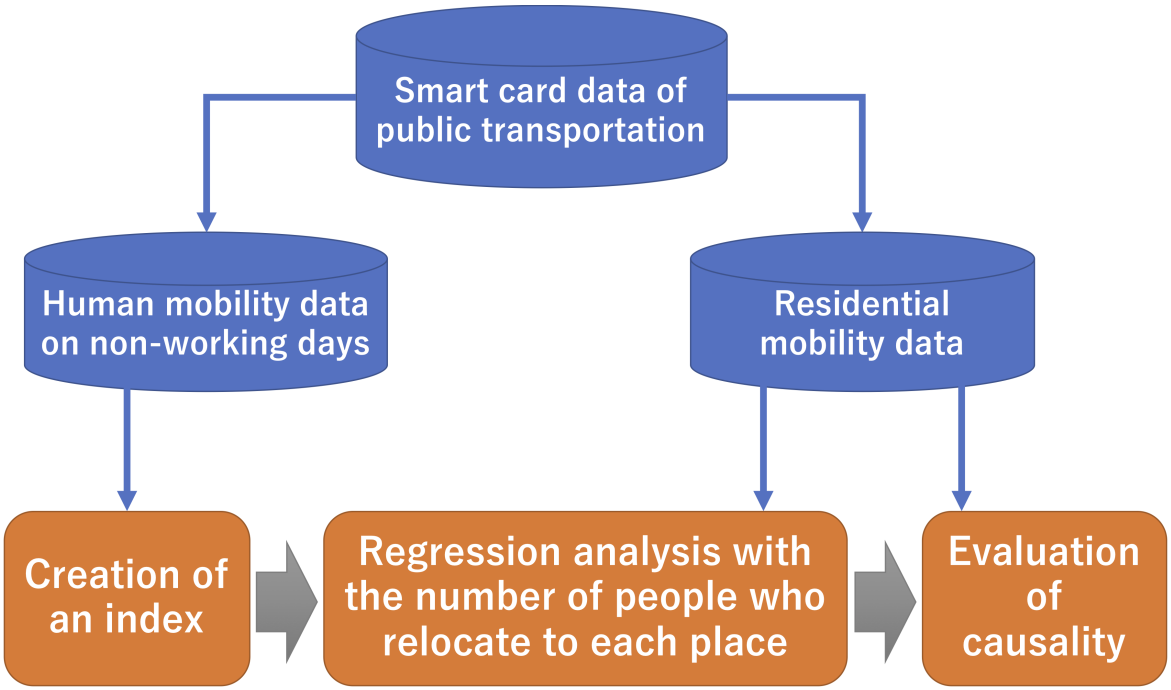


Figure 2. A flow-chart of the processes involved in this paper.

of travel time, and other indices that highly correlate with the number of residential moves. Thus, the research questions that we address in this paper are summarized as follows:

- **RQ1:** Does the opportunity cost of travel time calculated by an extended travel cost method contribute to the prediction of the number of people who relocate to each place compared to the current population and other indices?
- **RQ2:** Does the opportunity cost of travel time have a causal influence on the number of people who relocate to each place compared to the current population and other indices?

Figure 2 is a flow chart of this study. We extract human mobility data on non-working days and residential mobility data from smart card data of public transportation. Using human mobility data on non-working days, indices for the prediction of residential movements are created. We create a regression model for each index. Then, we perform the evaluation of causality between the number of people who relocate to each place and the created indices.

The remainder of this paper is organized as follows: Section 2 reviews previous studies. Section 3 describes the proposed method. Section 4 explains the baselines and evaluation method. Section 5 explains the data that we use. Section 6 describes the data preprocessing method. Section 7 reports on the results. Section 8 discusses the implications and limitations of our study. Section 9 draws our conclusions.

2. Related works

2.1. Travel cost method and opportunity cost of travel time

Travel is considered as a demand derived from the desire to engage in activities at destinations [11]. According to Becker's theory of the allocation of time [12], people allocate their limited time to activities to maximize utility. Since travel time is considered wasteful or unproductive, it is interpreted as a necessary evil to obtain benefits from activities at destinations. The travel cost method (TCM) is widely used for measuring the benefit of recreational sites [7,8] and for measuring the opportunity cost of travel time [9,10]. The method assumes that the ratio or the number of visits to a site and the cost

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 August 2018 doi:10.20944/preprints201808.0389.v1

of travel, including time and money, presents a demand curve. Regression analysis is performed for measuring the benefits of destinations and the opportunity cost of travel time. On the other hand, some studies [13–15] discuss that people spend time on productive or enjoyable activities while traveling using rapidly growing information and communications technologies (ICTs) such as smart phones, laptop computers, portable music players, and gaming devices. Therefore, the opportunity cost of travel time may differ based on such activities during traveling.

In our research, we measure the opportunity cost of travel time for multiple departure places using the smart card data of public transportation. Such measurement has never been performed in previous studies, and it requires an extension of the conventional TCM. In addition, we measure the opportunity cost of travel time for different demographic groups using gender and age data. The influence of the development of ICTs could be reflected in the difference in the opportunity cost of travel time among different demographic groups.

2.2. Interaction between urban form and travel behavior

In regard to the interaction between urban form and travel behavior, previous studies argue that urban form affects non-work travel behavior [1–4]. According to these studies, the ratio of people who walk on non-working days is strongly related to land use diversity, intersection density, and the number of destinations within walking distance. On the other hand, while other studies [16–20] also acknowledge that urban form influences non-work travel behavior, they argue that attitudes toward travel behavior and preferences toward residential location are more strongly associated with travel than are land use characteristics. These studies suggest that such preferences could affect residential choice. For example, people who prefer to drive may choose to live in suburban areas and travel by car, and people who prefer to walk may choose to live in areas with mixed land use and higher neighborhood accessibility. Mokhtarian and Cao [21] note the difficulty in finding causality between attitudinal factors, residential locations, and non-work travel behavior. They recommend the usage of longitudinal structural equations modeling for causal analysis. Krizek [22] performs analysis of the travel behavior of the same households in a longitudinal manner for causal analysis between urban form and travel behavior. The findings of the study suggest that households change travel behavior when exposed to differing urban forms. In particular, locating to areas with higher neighborhood accessibility decreases vehicle miles traveled (VMT) and person miles traveled (PMT). In addition, Handy et al. [23] acknowledge that residential self-selection plays an important role in influencing individuals' travel decisions, but they also suggest that mixed land uses tend to discourage auto travel.

Regarding our research, public transportation is the dominant measure for moving in the area where data that we use are collected. We assume that the opportunity cost of travel time increases when amenities and convenience in the neighborhood are high. In addition, we assume that convenience and amenities in the neighborhood contribute to an increase in the number of people who relocate to the place. Our study does not assume that preferences for residential locations and travel behavior differ among people. The limitation of our study is that it does not incorporate automobile travel, nor does it incorporate the difference in preferences for travel behavior and residential location. Therefore, it is necessary to further develop our method if applied to areas where people have varying preferences for travel behavior.

2.3. Influential factors on residential mobility

Influential factors on residential mobility and residential demand have long been studied [5,6]. By applying hedonic modeling, the influence of crime [24], consumption amenities [25–27], neighborhood parks [28], social capital [29], walkability [30], and air quality [31] have been discussed. By modeling individual or household choices regarding residential location, the relations between lifestyles, neighborhood characteristics, and the choice of residential location have been studied [32–34].

We consider that the opportunity cost of travel time reflects consumption amenities, neighborhood parks, and walkability. However, crime rate and air quality are not reflected in mobility data. We

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 August 2018 doi:10.20944/preprints201808.0389.v1

consider that the benefit of our method is that it only requires mobility data for predicting residential mobility and that it does not require additional secondary data such as information about crime rates. However, data on other factors, such as crime rate, can also be incorporated into our method to obtain better predictions of residential mobility.

2.4. Gap in the literature

In previous studies, data are collected through survey. It is costly and time consuming to collect survey data. The availability of massive mobility data collected from mobile phones and smart card of public transportation has been rapidly increasing.

Recent studies create indices for estimating urban characteristics using automatically collected mobility data. For example, Smith et al. [35] create an index for estimating community well-being by calculating the diversity of destinations people visit based on the smart card data of public transportation. Zhong et al. [36] create indices of the centrality of mobility for each station using a network science approach from the smart card data of public transportation. Yabe et al. [37] measure the fragility of people flows to appropriately plan future investments in infrastructure based on locational data collected from mobile phones. The benefit of such indices is as follows: First, we can detect urban changes without surveys that cost time and money. Second, we can evaluate the result of investments in urban development and predict future outcomes. Third, we can obtain knowledge about the nature of urban systems using those indices.

To the best of our knowledge, no method has been proposed to create an index based on mobility data to predict future residential movements. We propose a method called TCM4MP (travel cost method for multiple places) to infer the opportunity cost of travel time that varies according to the given place as an index for predicting the number of people who relocate to each place.

Our study is motivated by previous studies in the domain of transportation engineering and urban economics; however, the aim of our study is practical use. The final objective of our study is to contribute to the detection and prediction of changes in the trends of residential mobility. We consider that the current population of each place also correlates with the number of people who relocate to each place. However, this changes slowly compared to changes in travel behavior; therefore, it cannot contribute to the early detection and prediction of changes in the trends of residential mobility. We compare our proposed method with other indices derived from smart card data in terms of correlation and causality to the number of people who relocate to each place.

3. Method

3.1. Measurement of opportunity cost of travel time

Conventional studies using TCM do not assume that the opportunity cost of travel time varies according to the departure place. In addition, these studies usually calculate the benefit of one recreational site. Equation 1 is the equation of a demand curve defined in a conventional travel cost method.

$$v_i = C - Bt_i \quad (i \in \mathbf{D}) \tag{1}$$

In this equation, v_i denotes the visitation rate to a recreational site per person who lives in place i, C denotes the benefit one can obtain by visiting the recreational site, t_i denotes the time it takes to travel from place i to the recreational site. B denotes the opportunity cost of travel time per unit time, and $\mathbf{D}$ denotes the set of departure places.

We propose a method called TCM4MP by extending Equation 1 to measure the values of the opportunity cost that differ according to the departure place. Equation 2 is derived by assuming that

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 August 2018 doi:10.20944/preprints201808.0389.v1

the opportunity cost of travel time differs according to the departure place and by incorporating the values of the benefits of multiple arrival places.

$$v_{i,j} = C_j - B_i t_{i,j} \quad (i \in \mathbf{D},\ j \in \mathbf{A}) \tag{2}$$

In this equation, B_i is a parameter that represents the opportunity cost of people who live in departure place i. C_j is a parameter that represents the benefit that one can obtain by visiting arrival place j. $v_{i,j}$ denotes the average number of movements from departure place i to arrival place j per person, and $\mathbf{A}$ denotes the set of arrival places.

It is impossible to determine the values of B_i by simply applying linear regression analysis. We assume that the parameters B_i $(i \in \mathbf{D})$ and C_j $(j \in \mathbf{A})$ follow normal distributions as in Equations 3 and 4.

$$B_i \sim Normal(\mu_B, \sigma_B^2) \quad (i \in \mathbf{D}) \tag{3}$$

$$C_j \sim Normal(\mu_C, \sigma_C^2) \quad (j \in \mathbf{A}) \tag{4}$$

In these equations, μ_B, σ_B, μ_C, and σ_C are hyperparameters. All the parameters and hyperparameters are inferred by applying the hierarchical Bayesian inference method to Equations 2, 3, and 4 using mobility data on non-working days. In this study, we use statistical software, Stan [38], for Bayesian inference, which is based on No-U-Turn sampler (NUTS) [39], an extension of Hamiltonian Monte Carlo (HMC) [40]. Thus, the values of the opportunity cost of travel time B_i are obtained.

The opportunity cost B_i can be obtained separately according to people's demographic information. Our method divides mobility data according to demographic information about their gender and age, and we calculate B_i for each demographic group. We let k denote a demographic group, and $B_{i,k}$ denotes the opportunity cost of people living in place i in demographic group k.

3.2. Decision of prediction model

We assume that the opportunity cost of travel time reflects the convenience of the place. We also consider that the number of people who relocate to a place correlates with the convenience of the place. In addition, we consider that the convenience of a place varies depending on gender and age. Therefore, we use the opportunity cost of the travel time to a place for the prediction of the number of people who relocate to that place.

We create a power regression model φ_{O} for **opportunity cost (O)** to predict the amount of future residential mobility by Equation 5.

$$m_{i,k} = \varphi_{\mathrm{O}}(k, i) = \beta_k B_{i,k}^{\alpha_k} \tag{5}$$

Note that $m_{i,k}$ denotes the number of people in demographic group k who relocate to place i. The parameters α_k and β_k are unconditioned to place i. We consider that power regression is suitable because of the assumption that the number of residential movements are proportional to the product of the population of two areas (a gravity model [41]) and the assumption that the distribution of a population follows the power law (Zipf's law [42]). The parameters are calculated using the Levenberg-Marquardt algorithm [43].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 August 2018 doi:10.20944/preprints201808.0389.v1

4. Baselines and evaluation methods

4.1. Baselines

We evaluate how much the predictor using the opportunity cost (φ_{O}) contributes to the prediction of residential mobility by comparing it with other indices. We compare it with predictors using the following three indices:

- **Population (P)**: This index is simply the number of current residents in place i. We let $P_{i,k}$ denote the number of residents of demographic group k living in place i.
- **Average travel time (A)**: This index is a simplification of the opportunity cost of travel time. We let $T_{i,k}$ denote the multiplicative inverse of travel time per person.
- **Diversity (D)**: According to Smith et al. [35], the diversity of places people visit reflects the well-being of the community. The index is as follows:

$$diversity(u) = \frac{-\sum_{j \in \mathbf{S}_u} w_{u,j} \log(w_{u,j})}{|\mathbf{S}_u|} \tag{6}$$

$$H_i = \frac{1}{|\mathbf{M}_i|} \sum_{u \in \mathbf{M}_i} diversity(u) \tag{7}$$

In Equation 6, $\mathbf{S}_u$ is the set of places user u visited, and $w_{u,j}$ is the proportion of all u's visits to place j. The numerator in this equation is the Shannon entropy. In Equation 7, $\mathbf{M}_i$ is the set of users who live in place i. We let $H_{i,k}$ denote the diversity of places users in demographic group k living in place i visit.

As with Equation 5, prediction models φ_{P}, φ_{A}, and φ_{D} are determined by power regression.

4.2. Evaluation of correlation and causal relation between the number of relocations and the predictors

First, we evaluate the correlation between a predictor and the number of people who relocate to each place. If we found predictors that highly correlate with the number of people who relocate to each place, we perform causal inference between the predictors and the number of residential moves.

We apply linear non-Gaussian acyclic model (LiNGAM) [44] for the causal inference. By applying LiNGAM, causal relations between variables are obtained as described in Equation 8.

$$x_i = \sum b_{i,j} x_j + e_i \tag{8}$$

In this equation, x_i is an observed variable. The variable e_i is an exogenous variable (random variable) having a non-Gaussian distribution, and $b_{i,j}$ is the strength of the causal connection from x_j to x_i. The objective of causal inference is to determine $b_{i,j}$.

The key difference between LiNGAM and earlier works on causal inference is that LiNGAM assumes that exogenous variables are non-Gaussian. Under this assumption, it is possible to estimate a causal ordering of variables using passive observational data alone without any prior information on a causal ordering of the variables.

In our study, x_i denotes either the number of residential moves or the value of a predictor for predicting the number of residential moves. By applying LiNGAM, we obtain the causal networks between the number of residential moves and predictors that highly correlate with the number of residential moves. We use causal networks obtained using LiNGAM for the evaluation of the causal influence of each predictor on the number of residential moves.

5. Data

We use the smart card data of public transportation in the Kansai Area of Japan. The data are from March to April 2016 and from March to April 2017. There are two types of smart cards in this area. One

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 August 2018 doi:10.20944/preprints201808.0389.v1

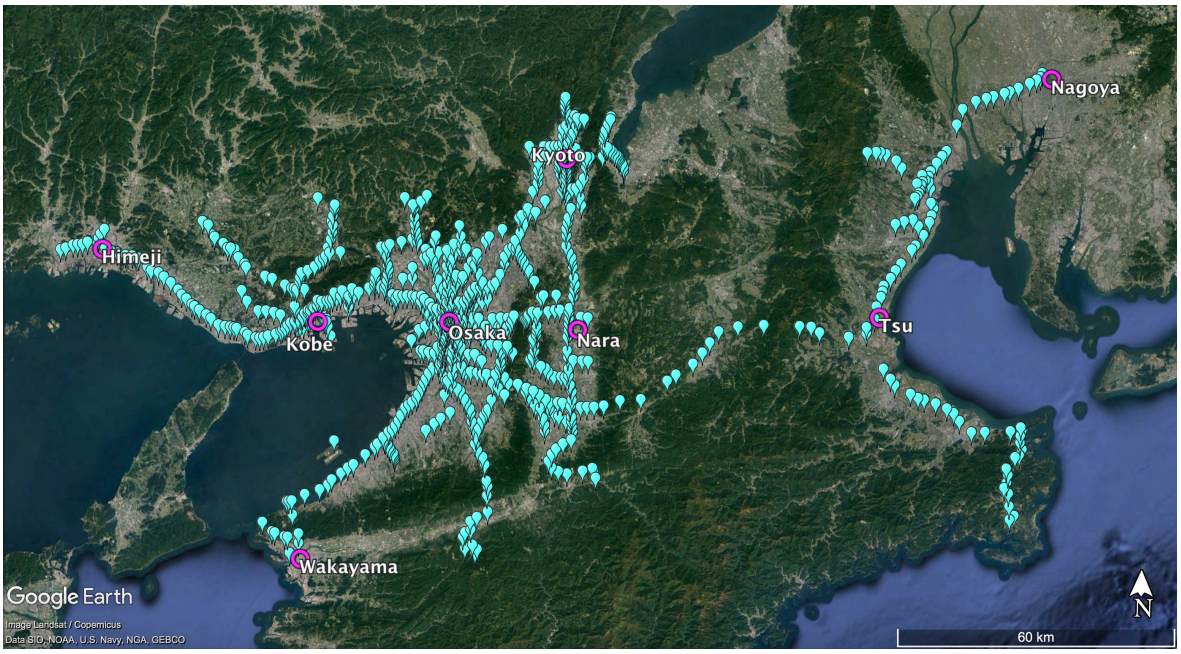


Figure 3. The locations of the 1024 railway stations used in this paper.

type of smart card requires applicants to submit their personal information when applying to obtain the card. We use data collected from that type of smart card. The data include information about the user's ID, user's age, user's gender, the postal code of the location of user's residence, boarding station, boarding time, alighting station, and alighting time. The data were completely anonymized before the data are provided to us by the railway companies.

The data include 1,024 stations held by 14 railway companies and 3 agencies of city governments. The locations of the stations are shown in Figure 3.

The demographic composition of the card holders is given in Table 1. The number of male card holders and that of female card holders exceed one million in both 2016 and 2017.

We compare the demographic composition of the smart card data and that of the data published by the government of Osaka City. We categorize the smart card data by gender and ages in groups of 10 years. Osaka City has 24 wards. We therefore further divide the smart card data according to the ward in which each card holder's residence is located. We compare the number of residents by ward for each demographic group. Table 2 shows the correlation coefficients between the number of each demographic group of the smart card data and that of the governmental data. We assume that smart card data reflect real demographic compositions.

In this paper, we use smart card data collected from March to May 2016 for obtaining predictors for the amount of future residential mobility. We recognize residential moves between 2016 and 2017 by the changes in stations at which card holders most frequently board for the first time on a day.

We do not use data on individuals younger than 20 or older than 79 because we assume that most of these people do not relocate to other areas by their own will. In the following, we divide card holders into demographic groups by gender and in age groups of 10 years.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 August 2018 doi:10.20944/preprints201808.0389.v1

Table 1. Demographic composition of card holders

	2016		2017	
Age	**Male**	**Female**	**Male**	**Female**
0-9	11,331	12,274	12,252	13,308
10-19	127,968	141,913	133,691	147,673
20-29	146,365	215,936	148,487	217,400
30-39	198,082	251,226	193,508	248,822
40-49	275,129	310,716	279,966	322,486
50-59	219,646	225,320	230,026	242,502
60-69	143,823	151,027	157,696	169,407
70-79	133,427	182,992	130,555	180,812
80-89	40,429	53,125	43,489	58,318
90-99	1,557	1,840	1,786	2,083
Total	1,297,757	1,546,369	1,331,456	1,602,811

Table 2. Correlation coefficients of the comparison of 24 wards in Osaka City between the population of smart card holders and the information about the demographic composition published by the government of Osaka City.

Age	2016	2017
0-9	0.787010	0.823162
10-19	0.937348	0.948269
20-29	0.958899	0.969244
30-39	0.935792	0.946171
40-49	0.925023	0.942998
50-59	0.927487	0.941806
60-69	0.897694	0.927388
70-79	0.950722	0.950101
80-89	0.948302	0.941291
90-99	0.882572	0.904075

6. Data pre-processing

This section explains the data pre-processing procedure for the smart card data that we used. Some thresholds used in this section are arbitrarily set, and we do not investigate a method to automatically find suitable such thresholds in this paper. Further research is needed to investigate a method to automatically find suitable thresholds.

6.1. Decision of travel time between two stations

We determine the time it takes to travel from departure station d_1 to arrival station a_1 by $t(d_1, a_1)$ defined in Equation 9.

$$t(d_1, a_1) = \mathrm{Med}(\mathbf{H}(d_1, a_1)) \tag{9}$$

In this equation, $\mathbf{H}(d_1, a_1)$ denotes a set of all the values of durations for traveling from departure station d_1 to arrival station a_1. $\mathrm{Med}(\mathbf{H}(d_1, a_1))$ denotes the median of $\mathbf{H}(d_1, a_1)$.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 August 2018 doi:10.20944/preprints201808.0389.v1

6.2. Grouping departure stations, arrival stations, and smart card holders

We group departure stations by the physical proximity between stations and similarity of boardings by smart card holders. Departure station d_1 and departure station d_2 are grouped together when Equation 10 or Equation 11 is satisfied.

$$l(d_1, d_2) < 300 \tag{10}$$

$$\Big(|\mathbf{V}(d_1) \cap \mathbf{V}(d_2)| > 0.1 \cdot |\mathbf{V}(d_1)|\Big) \wedge \Big(|\mathbf{V}(d_1) \cap \mathbf{V}(d_2)| > 0.1 \cdot |\mathbf{V}(d_2)|\Big) \tag{11}$$

In Equation 10, $l(d_1, d_2)$ denotes the distance between departure station d_1 and departure station d_2 in meters. In Equation 11, $\mathbf{V}(d_1)$ denotes a set of smart card holders who had boarded at departure station d_1 for the first time on a day during the period from March to May 2016 (we assume that a station that a user boards at for the first time on a day is located near the user's residential location). Equation 10 means that the distance between departure station d_1 and departure station d_2 is less than 300 meters. Equation 11 means that the number of smart card users who have boarded at both departure station d_1 and departure station d_2 is greater than 10% of the number of smart card holders who had boarded at departure station d_1 and that of smart card users who had boarded at departure station d_2.

We group arrival stations by the physical proximity and walkability between stations. Arrival station a_1 and arrival station a_2 are grouped together when Equation 12 or Equation 13 is satisfied.

$$l(a_1, a_2) < 300 \tag{12}$$

$$\min \mathbf{W}(a_1, a_2) < 30 \tag{13}$$

In Equation 13, $\mathbf{W}(a_1, a_2)$ denotes the set of all the values of the duration between a user alighting at station a_1 and the time they consequently board at station a_2. Equation 13 means that the minimum period between when users alight at station a_1 and when the users consequently board at station a_2 is less than 30 minutes. We assume that this means that people can walk from station a_1 to station a_2 within 30 minutes.

Finally, 599 groups of departure stations and 541 groups of arrival stations are obtained. In the following, the group of departure stations is simply referred to as the departure place, and the group of arrival stations is simply referred to as the arrival place.

We determine the time it takes to travel from departure place i to arrival place place j by Equation 14.

$$T(i, j) = \min_{d \in \mathbf{F}_{\mathrm{D}}(i),\ a \in \mathbf{F}_{\mathrm{A}}(j)} t(d, a) \tag{14}$$

In this equation, $\mathbf{F}_{\mathrm{D}}(i)$ denotes the set of all stations in departure place i, and $\mathbf{F}_{\mathrm{A}}(j)$ denotes the set of all stations in arrival place j.

We also group smart card holders. Smart card holders are grouped by the stations at which they most frequently board for the first time on a day.

$$\operatorname*{argmax}_{k}\ n(u, k) = i \ \Leftrightarrow\ u \in \mathbf{U}(i) \tag{15}$$

In this equation, $n(u, k)$ denotes the number of days user u boards at place k. This equation means that user u is grouped into set $\mathbf{U}(i)$ when the number of days user u boards at departure place i is the highest among all the departure places.

In addition, we exclude the data of user $u \in \mathbf{U}(i)$ if place i is far from the location of his/her residence estimated by his/her postal code. In this study, we set the threshold to 5,000 meters.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 August 2018 doi:10.20944/preprints201808.0389.v1

6.3. Removing records of arrival stations each user frequently visits

We apply the travel cost method to the data of non-work travel. Therefore, we apply the method to data collected on holidays and weekends. However, some people commute to work on holidays or weekends. To remove such records, we count the number of days each user has alighted at each station. We ignore the records of the stations at which a user frequently alighted. The records of arrival place j of user u that satisfy Equation 16 are ignored.

$$\frac{g(u,j)}{s} \geq \frac{2}{7} \tag{16}$$

In this equation, $g(u,j)$ denotes the number of days (including weekdays) user u has alighted at arrival station j. s denotes the number of all the days in the period of data.

Table 3 shows the number of valid users obtained after data preprocessing.

Table 3. Valid card holders

Age	Male	Female
20-29	115,926	176,803
30-39	153,922	206,359
40-49	217,804	266,191
50-59	171,413	192,379
60-69	117,933	130,967
70-79	114,981	160,849

7. Results

7.1. Results of regression of the predictors to the number of residential moves.

Table 4 shows the results of the regression analyses between the number of residential movements $m_{i,k}$ and predictors φ_{O}, φ_{P}, φ_{A}, and φ_{D}. The predictor using opportunity cost (O) is superior to other baselines for the data of male individuals of all ages and female individuals less than 60 years old. On the other hand, the predictor using population (P) scores the highest for the data of female individuals who are 60 or older.

The coefficients of determination of the predictors using opportunity cost (O) and population (P) are higher than 0.6, but the other predictors are lower than 0.6.

7.2. Causal influences on the number of residential moves

We measure causal influences from the two predictors using opportunity cost (O) and population (P) on the number of residential moves using linear non-Gaussian acyclic model (LiNGAM) [44]. LiNGAM assumes that exogenous variables are non-Gaussian. We test whether the predictors and the number of residential moves satisfy this assumption. It is impossible to test the distributions of exogenous variables. Therefore, we test the distributions of observed variables instead.

The Shapiro-Wilk test [45] is used to test normality. The Shapiro-Wilk test tests the null hypothesis that a sample came from a normal distribution. The null hypothesis of this test is that the population of the data is normally distributed. If the p-value is less than the chosen alpha level, then the null hypothesis is rejected, and there is evidence that the data tested are not from a normally distributed population. Table 5 shows the results of the Shapiro-Wilk test. Every p-value listed in Table 5 is small enough to conclude that the predictors O and P and the number of residential moves to each place (M) do not follow a normal distribution.

In addition to non-Gaussianity, LiNGAM assumes that relations between variables are linear. The linearity between the number of residential moves ($m_{i,k}$) and the predictor O (φ_{O}) and the linearity

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 August 2018 doi:10.20944/preprints201808.0389.v1

between the number of residential moves ($m_{i,k}$) and the predictor P (φ_{P}) have already been confirmed, as shown in Table 4. In regard to the relation between the two predictors (φ_{O} and φ_{P}), the correlation coefficient is shown in Table 6. It is high enough to assume that the relation between the two predictors is linear. Therefore, we consider that the values satisfy the assumption of LiNGAM.

Figure 4 shows the results of the application of LiNGAM. The directions of the obtained causal networks are the same for every demographic group. The number of people who relocate to each place (M) is affected by both the predictor using the opportunity cost of travel time (O) and the predictor using the current population (P). The strength of causality from O to M is greater from P to M except for the demographic group of female individuals between 60 and 69 years of age. In addition, the strength of causality from P to O is greater from P to M except for the demographic group of female individuals between 60 and 69 years of age.

We measure the reproducibility of the results. First, we examine the reproducibility of the directions of causal networks. We repeat the experiment by making 1000 causal networks for each demographic group by randomly taking 400 departure places out of 599 departure places and making causal networks using the data of 400 selected departure places. The results are shown in Table 7. The first type of network has the same causal directions as the network that we obtain using all the departure places. The number of individuals in the first type in network is 1000 for the demographic groups of male individuals less than 70 years of age and female individuals less than 60 years of age. On the other hand, the number of individuals in the first type of network for the groups of male individuals in their 70s is 772. Therefore, there might be missing causal factors for the demographic group; thus, it is necessary to seek other causal factors to incorporate.

Next, we examine the reproducibility of the order of the strength between the causalities to the number of residential moves from the opportunity cost of travel time and current population. We conduct the same random sampling until we obtain 1000 networks whose directions of causality are the same as the network that we obtain using all the departure places. Figure 5 shows the ranges of the strengths of causality. Comparing the medians of the strengths of causality to the number of residential moves, the orders are consistent with the results obtained using all the departure places, as shown in Figure 4. The ranges of the causalities (O→M and P→M) overlap only in regard to the demographic group of female individuals 60 years or older. Therefore, we consider that the reproducibility is unstable in regard to the demographic groups, and it is stable in regard to the other demographic groups.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 August 2018

doi:10.20944/preprints201808.0389.v1

Table 4. Coefficients of determination (R^2) between the number of people who relocate to each place ($m_{i,k}$) and predicted numbers (φ_O, φ_P, φ_A, φ_D).

	Male						Female					
Predictor	**20-29**	**30-39**	**40-49**	**50-59**	**60-69**	**70-79**	**20-29**	**30-39**	**40-49**	**50-59**	**60-69**	**70-79**
O (φ_O)	**0.677131**	**0.682127**	**0.702986**	**0.686997**	**0.695490**	**0.677117**	**0.641250**	**0.685919**	**0.685707**	**0.669508**	0.646802	0.621696
P (φ_P)	0.637419	0.625829	0.638509	0.627181	0.632797	0.662053	0.631902	0.654962	0.683051	0.662978	**0.680595**	**0.693258**
A (φ_A)	0.476285	0.462466	0.501236	0.484837	0.420105	0.397013	0.517185	0.459414	0.506208	0.459616	0.381638	0.390667
D (φ_D)	0.344690	0.337078	0.398216	0.410906	0.397417	0.230339	0.325484	0.344409	0.389053	0.372604	0.345309	0.227351

Table 5. The result of the Shapiro-Wilk test.

	Male						Female					
Valuable	**20-29**	**30-39**	**40-49**	**50-59**	**60-69**	**70-79**	**20-29**	**30-39**	**40-49**	**50-59**	**60-69**	**70-79**
O (φ_O)	1.18E-25	9.34E-28	5.78E-25	9.99E-23	1.59E-24	9.59E-32	1.04E-25	1.45E-27	9.53E-24	5.21E-24	9.02E-25	2.44E-30
P (φ_P)	7.73E-27	5.09E-28	1.13E-26	6.38E-26	9.00E-27	1.46E-31	5.26E-28	1.93E-28	9.89E-28	5.59E-28	3.06E-29	4.13E-32
M ($m_{i,k}$)	2.02E-35	2.11E-36	5.82E-34	1.75E-33	9.16E-34	3.88E-37	1.57E-36	4.94E-36	4.03E-34	1.18E-34	4.11E-35	2.78E-37

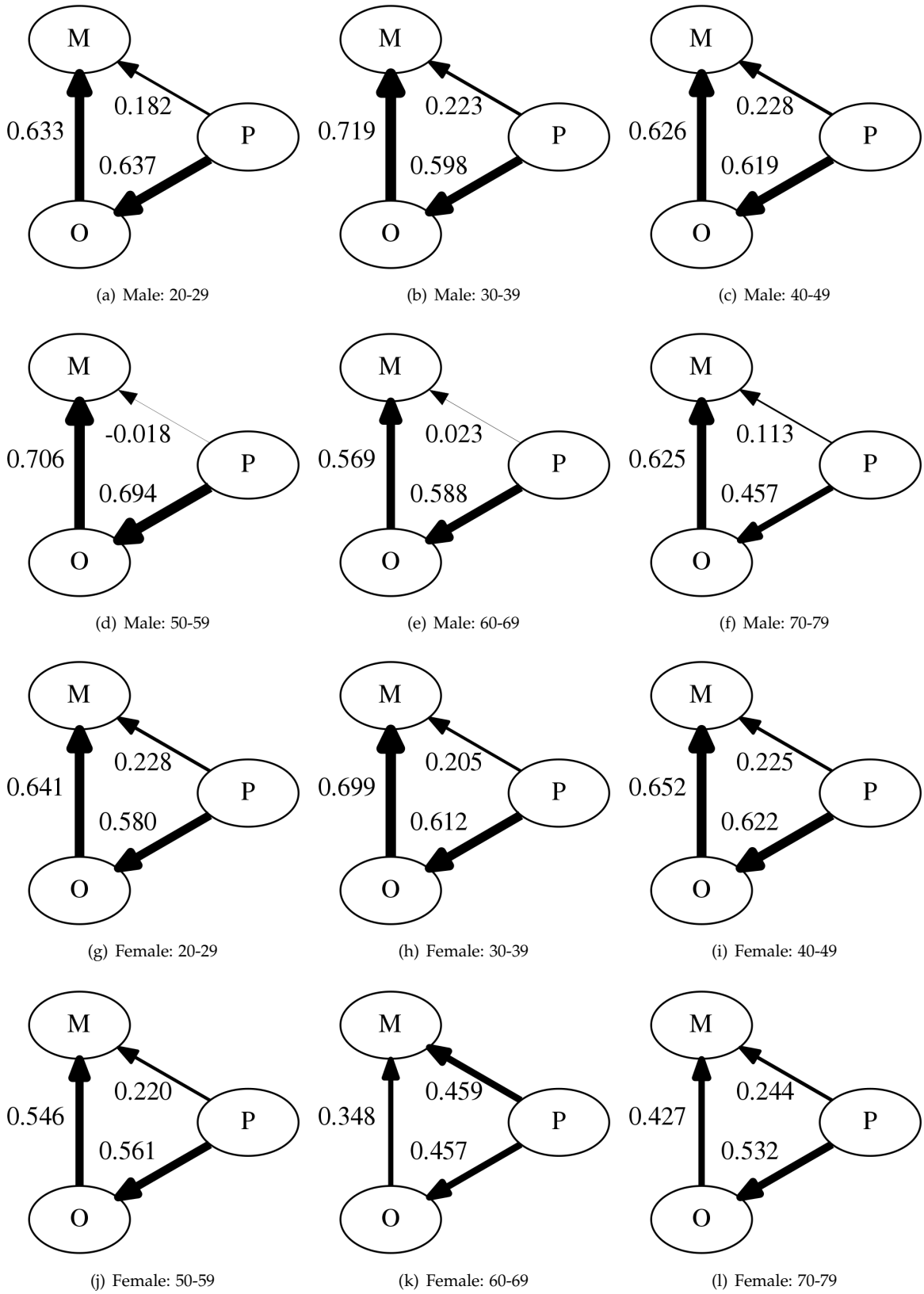
Table 6. Correlation coefficient (R) between the predictor using opportunity cost (φ_O) and the predictor using current population (φ_P).

Male						Female					
20-29	**30-39**	**40-49**	**50-59**	**60-69**	**70-79**	**20-29**	**30-39**	**40-49**	**50-59**	**60-69**	**70-79**
0.827688	0.794022	0.783471	0.797551	0.805344	0.849521	0.823154	0.835350	0.816704	0.806756	0.813640	0.847267

Table 7. The reproducibility of causal networks obtained using LiNGAM: We repeat the experiment by making 1000 causal networks for each demographic group by randomly taking 400 departure places out of 599 departure places and making causal networks using the data of 400 departure places.

	Male						Female					
Direction of causality	**20-29**	**30-39**	**40-49**	**50-59**	**60-69**	**70-79**	**20-29**	**30-39**	**40-49**	**50-59**	**60-69**	**70-79**
1: O to M, P to O, P to M	1000	1000	1000	1000	1000	772	1000	1000	1000	1000	923	975
2: O to M, O to P, P to M	0	0	0	0	0	228	0	0	0	0	77	19
3: O to P, M to O, M to P	0	0	0	0	0	0	0	0	0	0	0	6

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 August 2018 doi:10.20944/preprints201808.0389.v1



(a) Male: 20-29 (b) Male: 30-39 (c) Male: 40-49

(d) Male: 50-59 (e) Male: 60-69 (f) Male: 70-79

(g) Female: 20-29 (h) Female: 30-39 (i) Female: 40-49

(j) Female: 50-59 (k) Female: 60-69 (l) Female: 70-79

Figure 4. Causal networks between the predictor using the opportunity cost of travel time (O), the predictor using population (P), and the number of residential moves to each place (M).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 August 2018 doi:10.20944/preprints201808.0389.v1

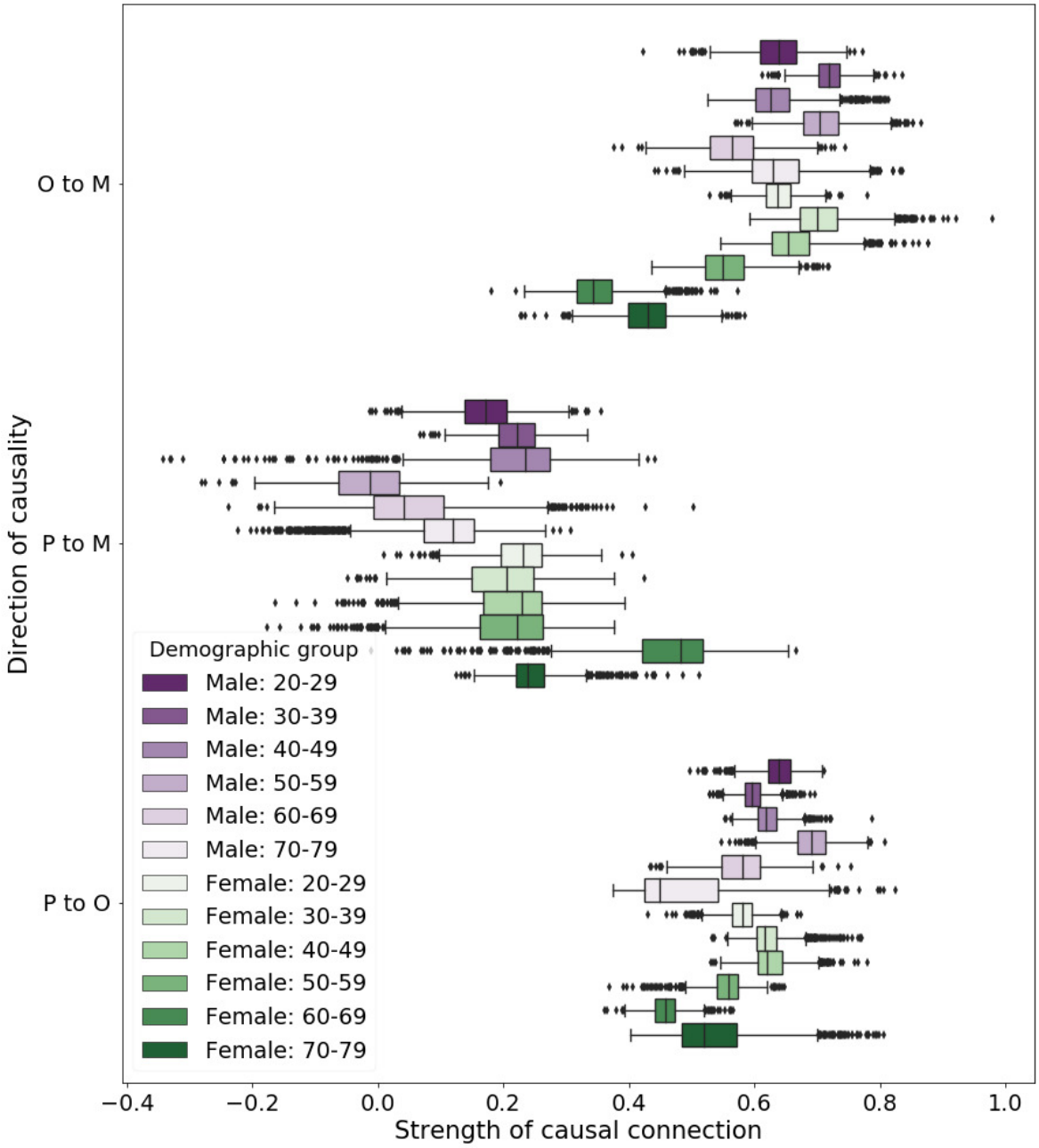


Figure 5. The reproducibility of the strength of causalities: The ranges of the causalities (O→M and P→M) overlap only in regard to the demographic group of female individuals 60 years or older. Therefore, we consider that the reproducibility is unstable in regard to the demographic group, and it is stable in regard to the other demographic groups.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 August 2018 doi:10.20944/preprints201808.0389.v1

8. Discussion

We have observed that the opportunity cost of travel time has strong causality to the number of people who relocate to each place. On the other hand, the index is influenced by the current population in each place. We assume that this is because a more heavily populated area will be more convenient because commercial developers and residential developers focus on such areas. This leads to the conclusion that most of the influence from the current population on residential mobility is mediated indirectly through the opportunity cost of travel time. The opportunity cost of travel time is calculated from non-work travel behavior in a short period of time (3 months), and it directly reflects the effects of urban development such as the construction of parks and the development of commercial facilities. Therefore, the opportunity cost of travel time is an efficient and effective index for predicting future residential mobility.

We can list many other factors that are not investigated in our paper. Figure 6 includes some factors related to residential mobility that were not investigated in this paper. The factors in green are not investigated in our paper.

Human mobility on non-working days is influenced by urban form such as amenities of the neighborhood and land use diversity. Urban form is influenced by investment from both industries and governments. Industries tend to invest in areas with high population densities. Therefore, investment from industries is influenced by the current population. However, governments can change the trend of residential movements by investing in urban development.

There is also causality from the current population to the number of people who relocate to each place that is not mediated through non-work human mobility. Residential movement is influenced by land prices and the social capital of the local community. Those factors are influenced by the current population. We assume that these causalities could be why the causality from O to M is not always greater than that from P to M in regard to the demographic group of female individuals 60 years or older.

There are also factors influencing residential mobility not being mediated by human mobility on non-working days and the current population. For example, closeness to business areas and air quality could have strong causality to residential mobility.

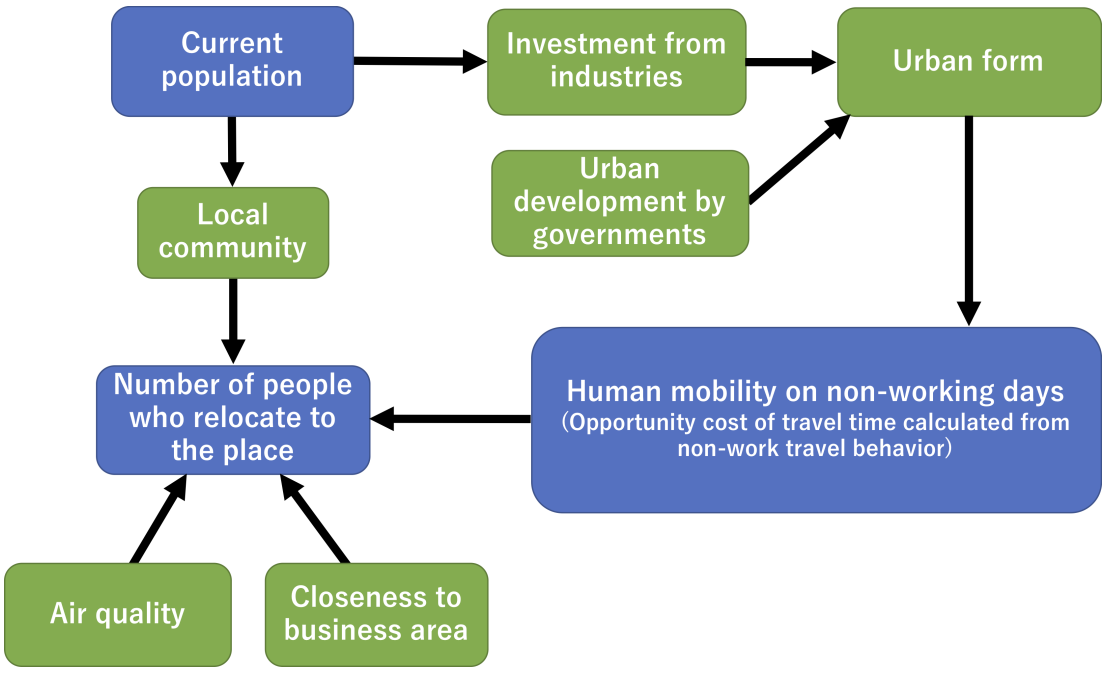


Figure 6. Detailed hypothesis on the causalities: Factors in green are not investigated in this paper, and factors in blue are investigated.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 August 2018 doi:10.20944/preprints201808.0389.v1

9. Conclusion

In this work, we have investigated a method for creating an index based on mobility data to estimate the effects of urban development on residential movements. The contributions of this work are summarized as follows:

- We have extended the conventional travel cost method to estimate the opportunity cost of travel time as a function of the departure place.
- We have confirmed that both the current population and the opportunity cost of travel time contribute to the prediction of the number of relocations.
- We confirmed that most of the causal influence from the current population to residential mobility is mediated indirectly through the opportunity cost of travel time. Therefore, the opportunity cost of travel time is more effective at estimating changes in residential mobility caused by urban development.

We note some limitations of our work. First, the opportunity cost of travel time is influenced by not only urban form but also other factors such as people's disposable time and activities using ICTs while traveling (e.g., mobile phones and potable gaming devices). Second, residential mobility is determined by other factors such as closeness to business areas and air quality. Third, the causal networks are not theoretically derived; therefore, it is necessary to analyze the causalities from a theoretical perspective.

Our future work will investigate which factors of urban form influence the opportunity cost of travel time. In addition, further research is needed to extend the method to include other types of transportation such as automobiles.

Despite the above limitations, our proposed method is beneficial to urban planners for estimating the effects of urban development and detecting the shrinkage and growth of populations.

Acknowledgments: This study is partly supported by the Leading Graduates Schools Program "Global Leader Program for Social Design and Management (GSDM)" run by the Ministry of Education, Culture, Sports, Science and Technology, Japan.

Author Contributions: All authors discussed and designed the experiments as well as contributing to the writing of the paper. T. N. Maeda defined the research agenda, implemented the experiments and wrote the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflicts of interest. The funding source had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

A	average travel time
D	diversity
ICTs	information and communications technologies
LiNGAM	linear non-Gaussian acyclic model
O	opportunity cost of travel time
P	current population
PMT	person miles traveled
TCM	travel cost method
TCM4MP	travel cost method for multiple places
VMT	vehicle miles traveled

References

1. Cervero, R.; Kockelman, K. Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment* **1997**, *2*, 199 – 219.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 August 2018 doi:10.20944/preprints201808.0389.v1

2. Rajamani, J.; Bhat, C.; Handy, S.; Knaap, G.; Song, Y. Assessing impact of urban form measures on nonwork trip mode choice after controlling for demographic and level-of-service effects. *Transportation Research Record: Journal of the Transportation Research Board* **2003**, pp. 158–165.
3. Zhang, M. Exploring the relationship between urban form and nonwork travel through time use analysis. *Landscape and Urban Planning* **2005**, *73*, 244 – 261. Research on the Built and Virtual Environments.
4. Ewing, R.; Cervero, R. Travel and the Built Environment. *Journal of the American Planning Association* **2010**, *76*, 265–294.
5. Quigley, J.M.; Weinberg, D.H. Intra- Urban Residential Mobility: A Review and Synthesis. *International Regional Science Review* **1977**, *2*, 41–66.
6. Dieleman, F.M. Modelling residential mobility; a review of recent trends in research. *Journal of Housing and the Built Environment* **2001**, *16*, 249–265.
7. Trice, A.H.; Wood, S.E. Measurement of recreation benefits. *Land economics* **1958**, *34*, 195–207.
8. McConnell, K.E. Chapter 15 The economics of outdoor recreation. In *Handbook of Natural Resource and Energy Economics*; Elsevier, 1985; Vol. 2, *Handbook of Natural Resource and Energy Economics*, pp. 677 – 722.
9. Walsh, R.G.; Sanders, L.D.; Mckean, J.R. The Consumptive Value Of Travel Time On Recreation Trips. *Journal of Travel Research* **1990**, *29*, 17–24.
10. Bockstael, N.E.; Strand, I.E.; Hanemann, W.M. Time and the Recreational Demand Model. *American Journal of Agricultural Economics* **1987**, *69*, 293–302.
11. Jones, P.M. *New approaches to understanding travel behavior: the human activity approach*; Oxford University, 1977.
12. Becker, G.S. A Theory of the Allocation of Time. *The economic journal* **1965**, pp. 493–517.
13. Lyons, G.; Urry, J. Travel time use in the information age. *Transportation Research Part A: Policy and Practice* **2005**, *39*, 257 – 276. Positive Utility of Travel.
14. Lyons, G.; Jain, J.; Holley, D. The use of travel time by rail passengers in Great Britain. *Transportation Research Part A: Policy and Practice* **2007**, *41*, 107 – 120.
15. Ohmori, N.; Harata, N. How different are activities while commuting by train? A case in Tokyo. *Tijdschrift voor economische en sociale geografie* **2008**, *99*, 547–561.
16. Kitamura, R.; Mokhtarian, P.L.; Laidet, L. A micro-analysis of land use and travel in five neighborhoods in the San Francisco Bay Area. *Transportation* **1997**, *24*, 125–158.
17. Handy, S.L.; Clifton, K.J. Local shopping as a strategy for reducing automobile travel. *Transportation* **2001**, *28*, 317–346.
18. Bagley, M.N.; Mokhtarian, P.L. The impact of residential neighborhood type on travel behavior: A structural equations modeling approach. *The Annals of Regional Science* **2002**, *36*, 279–297.
19. Bhat, C.R.; Guo, J.Y. A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. *Transportation Research Part B: Methodological* **2007**, *41*, 506 – 526.
20. Cao, X.J.; Mokhtarian, P.L.; Handy, S.L. The relationship between the built environment and nonwork travel: A case study of Northern California. *Transportation Research Part A: Policy and Practice* **2009**, *43*, 548 – 559.
21. Mokhtarian, P.L.; Cao, X. Examining the impacts of residential self-selection on travel behavior: A focus on methodologies. *Transportation Research Part B: Methodological* **2008**, *42*, 204 – 228. A Tribute to the Career of Frank Koppelman.
22. Krizek, K.J. Residential Relocation and Changes in Urban Travel: Does Neighborhood-Scale Urban Form Matter? *Journal of the American Planning Association* **2003**, *69*, 265–281.
23. Handy, S.; Cao, X.; Mokhtarian, P. Correlation or causality between the built environment and travel behavior? Evidence from Northern California. *Transportation Research Part D: Transport and Environment* **2005**, *10*, 427 – 444.
24. Lynch, A.K.; Rasmussen, D.W. Measuring the impact of crime on house prices. *Applied Economics* **2001**, *33*, 1981–1989.
25. Kuang, C. Does quality matter in local consumption amenities? An empirical investigation with Yelp. *Journal of Urban Economics* **2017**, *100*, 1 – 18.
26. Rappaport, J. Consumption amenities and city population density. *Regional Science and Urban Economics* **2008**, *38*, 533 – 552.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 August 2018 doi:10.20944/preprints201808.0389.v1

27. Pope, D.G.; Pope, J.C. When Walmart comes to town: Always low housing prices? Always? *Journal of Urban Economics* **2015**, *87*, 1 – 13.
28. Jim, C.; Chen, W.Y. External effects of neighbourhood parks and landscape elements on high-rise residential value. *Land Use Policy* **2010**, *27*, 662 – 670. Forest transitions Wind power planning, landscapes and publics.
29. Kan, K. Residential mobility and social capital. *Journal of Urban Economics* **2007**, *61*, 436 – 457.
30. Gilderbloom, J.I.; Riggs, W.W.; Meares, W.L. Does walkability matter? An examination of walkability's impact on housing values, foreclosures and crime. *Cities* **2015**, *42*, 13–24.
31. Harrison, D.; Rubinfeld, D.L. Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management* **1978**, *5*, 81 – 102.
32. Haase, D.; Lautenbach, S.; Seppelt, R. Modeling and simulating residential mobility in a shrinking city using an agent-based approach. *Environmental Modelling & Software* **2010**, *25*, 1225–1240.
33. Krizek, K. Lifestyles, residential location decisions, and pedestrian and transit activity. *Transportation Research Record: Journal of the Transportation Research Board* **2006**, pp. 171–178.
34. Ben-Akiva, M.; Bowman, J.L. Integration of an Activity-based Model System and a Residential Location Model. *Urban Studies* **1998**, *35*, 1131–1153.
35. Smith, C.; Quercia, D.; Capra, L. Finger on the Pulse: Identifying Deprivation Using Transit Flow Analysis. Proceedings of the 2013 Conference on Computer Supported Cooperative Work; ACM: New York, NY, USA, 2013; CSCW '13, pp. 683–692.
36. Zhong, C.; Schläpfer, M.; Arisona, S.M.; Batty, M.; Ratti, C.; Schmitt, G. Revealing centrality in the spatial structure of cities from human activity patterns. *Urban Studies* **2017**, *54*, 437–455.
37. Yabe, T.; Tsubouchi, K.; Sekimoto, Y. CityFlowFragility: Measuring the Fragility of People Flow in Cities to Disasters Using GPS Data Collected from Smartphones. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2017**, *1*, 117:1–117:17.
38. Carpenter, B.; Gelman, A.; Hoffman, M.D.; Lee, D.; Goodrich, B.; Betancourt, M.; Brubaker, M.; Guo, J.; Li, P.; Riddell, A. Stan: A probabilistic programming language. *Journal of statistical software* **2017**, *76*.
39. Hoffman, M.D.; Gelman, A. The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research* **2014**, *15*, 1593–1623.
40. Neal, R.M. MCMC Using Hamiltonian Dynamics. *Handbook of Markov Chain Monte Carlo* **2010**, *54*, 113–162.
41. Zipf, G.K. The P1P2/D Hypothesis: On the Intercity Movement of Persons. *American Sociological Review* **1946**.
42. Gabaix, X. Zipf's Law for Cities: An Explanation. *The Quarterly Journal of Economics* **1999**, *114*, 739–767.
43. Moré, J.J. The Levenberg-Marquardt algorithm: Implementation and theory. Numerical Analysis; Watson, G.A., Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, 1978; pp. 105–116.
44. Shimizu, S.; Hoyer, P.O.; Hyvärinen, A.; Kerminen, A. A linear non-Gaussian acyclic model for causal discovery. *Journal of Machine Learning Research* **2006**, *7*, 2003–2030.
45. Shapiro, S.S.; Wilk, M.B. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika* **1965**, *52*, 591–611.