Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2018) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

International Symposium of Transport Simulation (ISTS'18) and the International Workshop on Traffic Data Collection and its Standardization (IWTDCS'18)

Identifying passengers who are at risk of reducing public transport use: A survival time analysis using smart card data

Hiroaki Nishiuchi[a]*, Makoto Chikaraishi[b]

[a]Kochi University of Technology, 185 Tosayamada-cho Miyanokuchi, Kami City, Kochi, 782-8502, Japan
[b] Hiroshima University, 1-3-2 Kagamiyama, Higashi-Hiroshima City, Hiroshima, 739-8511, Japan

Abstract

This study investigates what types of passengers in the rural city of Kochi, Japan are most likely to reduce their use of buses and trams. In the analysis, two main efforts are made. First, a Cox proportional hazard model is fitted to data from public transport smart cards to identify the "at risk" users, where the risk is identified based on a consecutive reduction of public transport use. This allows us to distinguish changes in the frequency (i.e., lasting reductions of public transport use) from variations of the frequency (i.e., temporal fluctuation of public transport use) in a coarse but simple manner. Second, we develop a set of indicators to characterize the passengers' travel patterns, allowing us to make policy implications for promoting public transport use. In an empirical analysis, the impact of a reduction in tram frequency by public transport authorities on usage of transport services is initially examined using the model and is found not to have any significant short-term effects. Factors affecting public transport use are then identified and discussed. Based on the analysis, suggestions are made to encourage use of public transport in rural cities by selected demographics and travel patterns.

Keywords: Public transport; Smart card data; Cox proportional hazard model; Tarvel pattern

* Corresponding author. Tel.: +81-887-57-2406; fax: +81-887-57-2420.
E-mail address: nishiuchi.hiroaki@kochi-tech.ac.jp

1. INTRODUCTION

In recent years, an aging society, reduction in population, and migration of young people to big cities are common issues in rural cities of Japan. In addition, a decrease in the number of customers and the resultant loss of income for public transport companies are serious problems for rural cities in Japan. However, the role of public transport remains an important one as it is most suitable for an aging society experiencing a reduction in population. Since Japan's public transport has generally been operated by private sectors under an independent accounting system, public transport authorities and related organizations need to attract a certain amount of users in order to sustain public transport services. If the amount is not sufficiently high, the level of service is often reduced. In practice, such strategies are discussed based mainly on information related to income and cost, without consideration to the nature of trips made by public transport passengers. Interviews with employees of public transport authorities in Kochi City (conducted by the authors in January, 2014) revealed that service cuts had been made to the number of routes and trip frequencies of trams and buses. Public transport authorities also say that it is difficult to win back passengers once they stop using public transport following a cut in services. To discuss the above issue, understanding the behavior of public transport passengers and the characteristics of public transport usage is the first step in gaining fundamental information vital for relevant policy debates.

Monitoring devices have been installed in public transportation in many cities, such as sensors that automatically count the number of boarding and alighting passengers (Tamin, 1997; Kita and Tsukioka, 2005; Zhang and Teng, 2013). While data from these devices can be used for analyzing changes in and variations of the number of passengers, it is not directly possible to analyze the Origin-Destination (OD) of passengers. Due to this limitation, there has recently been a growing body of literature using the data from smart card fare-collection systems to directly measure OD between boarding and alighting stops. In particular, many researchers have analyzed the trip patterns and trip regularity of passengers by using smart card data so as to understand passenger behavior (Chapleau and Chu, 2007; Morency et al., 2007; Pelletier et al., 2011; Devillaine et al., 2012; Asakura et al., 2012; Nishiuchi et al., 2013; Ma et al., 2013; Kieu et al., 2014; Kusakabe and Asakura, 2014). Smart card data have also been analyzed to provide details of passenger behavior for marketing purposes (Seaborn et al., 2009; Páez et al., 2011; Asakura et al., 2012). Such information is fundamental to understanding passenger usage characteristics of public transport from a marketing strategy point of view and thus, important for the revitalization of public transport in rural cities in Japan. Most marketing strategies are seeking ways to increase demand but measures to stop the reduction of public transport use have steadily become more important within a declining population.

With the aim of contributing to the revitalization of public transport, Nishiuchi et al. (2014) analyzed variations in time series data of OD between tram stops (tram OD) in Kochi City using the data from "DESUCA," Kochi's public transportation smart card. Using a state-space model, the OD volume was divided into trend components, daily variation components, and autoregressive components. The results show the temporal rhythms of tram passengers, such as fewer passengers on Mondays and more passengers on Fridays. In addition, Nishiuchi et al. (2015) analyzed the impact of the reduction in tram services in Kochi, implemented on 1 November, 2012. They developed a model to estimate the probability of change in the number of passengers after 1 November, 2012 by applying a Cox hazard model using smart card data. The results showed that OD pairs from a suburban area to a city center could maintain the number of tram users, even after a reduction in the level of service. However, the study was conducted based on an aggregated number of OD passengers. Therefore, it does not show what types of passengers may change their frequency of public transport use, while analyzing smart card data at the disaggregated level can help public transport authorities identify suitable targets for revitalization measures. In particular, it is necessary to identify the types of passengers who are at a higher risk of reducing public transport use in order to avoid a loss of passengers.

This paper employs a Cox hazard model to identify factors affecting the risk of reducing public transport use by utilizing smart card data. There are two objectives, i.e., (1) identifying the impact of the reduction in tram services in Kochi implemented on 1 November, 2012, and (2) analyzing the trip characteristics of passengers who might reduce their use of public transport.

This paper is organized as follows. In the next section, the study area and data are briefly introduced. We then describe a model to estimate the probability of passengers reducing their public transport use during the data collection period, followed by an empirical examination of the characteristics of passengers who are at risk of reducing their use of public transport. The last section summarizes findings and remaining future tasks.

2. Study area and Smart Card DESUCA

DESUCA is a smart card that became available on January 25, 2009, and can be used on trams and buses in Kochi Prefecture. Kochi Prefecture is located on Shikoku Island in the southern part of Japan. The population in residential areas within 500m of a tram stop has been estimated at approximately 154,000 (Kochi Prefecture, 2011). The tram network in Kochi City consists of an east-west line and a north-south line and cross at Harimayabashi Tram Stop where passengers can transfer between the two lines (see Fig. 1). The smart card payment system is accepted by Tosa Electric Railway Co., Ltd., Tosaden Dream Service K.K., Kochikenkotsu, Inc. rail services, and buses operating over a wide area in Kochi Prefecture (ie., Kenkohokubukotsu Co., Ltd., and Kochikenkotsu Inc. in Susaki, Suginokawa, and Yusuhara) . These cards serve as commuter passes and can be issued anonymously with personalized data recorded on them so they can be returned to the owner if lost. Five types of cards are issued, including a card for children up to elementary school and “Nice Age” cards for users 65 years and older. Cards available for adults include personalized and non-personalized types, as well as a special card for disabled passengers.

The data used in this study are acquired from DESUCA records collected over 18 weeks from September 1 to December 31, 2012. However, to not consider the difference in behavioral characteristics during the year-end period, the period chosen for data analysis was from September 1 to December 21, 2012 (16 weeks). In addition, the data collected in the period shown in Table 1 did not include information on the type of card, as mentioned above. Therefore, this paper matched card IDs with those appearing in the DESUCA database collected for use by Nishiuchi et al. (2013) which included the passenger card type. As a result of card ID matching, the number of analyzed card IDs in this paper is 11,631. Table 1 shows a summary of the data used in the analysis.

Fig. 1. Tram network and frequency reduction rate by region(map courtesy Tosa Electric Railway Co., Ltd.).

3. Cox Proportional Hazard Model for Survival Time Analysis

3.1. Outline of the model

Survival time analysis is a method of analyzing the relationship between the occurrence times of two events, such as the occurrence of a disease and the death of a patient in the medical field, or the failure of a product in the engineering field. The Cox proportional hazard model is classified as a semi-nonparametric model that can introduce a survival function by using the Kaplan-Meier method without assumption of a probability distribution. The Cox proportional hazard model can be defined as follows.

$$\lambda_k(t) = \lambda_0(t)\exp\left(\sum_{i=1}^{m}\beta_i z_i\right) \tag{1}$$

$\lambda_k(t)$: probability of frequency reduction of passenger k at week t

$\lambda_0(t)$: baseline hazard function

β_i: parameter for covariates i

z_i : covariates

Table 1. Summary of DESUCA data used in this study.

Data Collection Term (Analysis period)	1 September - 31 December 2012
Data Collection Hours	05:00 to 23:00
Variables measured/recorded	Date: Year, Month, Day Card ID: Provided for all DESUCA cards (always the same once given) Type of Card: With/Without Registration, Handicapped Passenger, Child Passenger, Elderly Passenger Time (Boarding): Hour and min when boarding tram Time (Alighting): Hour and min when exiting tram Tram Stops: Name and code of each stop Distance between Tram Stops
Bus/Tram Companies	Tram: Tosa Denki-Testudo Bus: Tosa Denki-Testudo Tosaden Service Kochi-ken Kotsu Kenko Hokubu Kotsu

In this study, "a reduction in frequency of use" is defined based on the following procedure:

1. calculate passenger k's average number of trips per week during the first 4 observed weeks in September (called the criteria value of reduction)；
2. identify whether or not passenger k's average number of trips at week t (from October to December) is smaller than the criteria value of reduction, and;
3. judge that "a reduction in frequency of use" occurs when passenger k's average number of trips at week t is lower than the criteria value for successive n weeks.

The key idea in taking these steps is to distinguish changes in the frequency from variations of frequency in a simple manner. If an appropriate n could be provided, changes in the frequency could be explored with less noise caused by the fluctuation of the frequency.

To find the appropriate value of n, first, different values of n are applied and then multiple observations are made for each value of n where the multiple observations indicate that the reduction may not mean change, but fluctuation. When n =3, multiple observations occurred for approximately 25% of passengers who were at risk of reductions, while the frequency of multiple observations dramatically reduced to around 10% when n =4. Thus, the value of n is defined as 4 in this paper (see Fig. 2). It should be noted that there is some room for improvement of the above definition as "a reduction in frequency of use." For example, strictly speaking, the appropriate value of n would actually vary across passengers due to an existing heterogeneity in the rhythm of public transport use. Though this would add some burden in applying the model to a practical context, these would be worth exploring in the future.

Since the first 4 weeks of data are used to obtain the initial baseline value and the probability of a reduction in frequency would be zero during the initial 4 weeks (since n =4), the remaining 12 weeks of data are used in the empirical analysis.

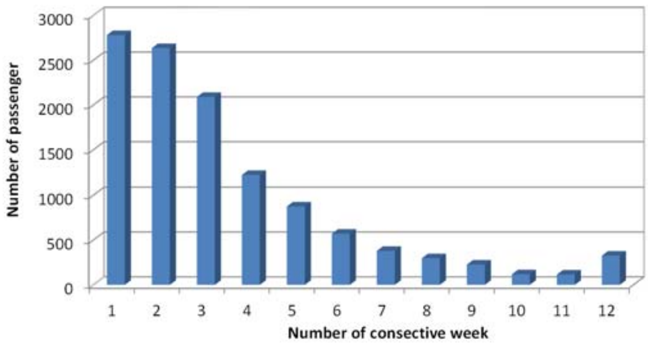

Fig. 2. Histogram of the maximum number of consecutive weeks for which passengers' use of public transport was below the criteria value.

3.2. Definition of baseline hazard function

The baseline hazard function is assumed to be a value changing by time period, but the trend is not dependent on the difference in the values of the covariate vector *z*. This means that the function is not dependent on each individual subject of analysis. In this study, it is set as one function for all passengers as describing a common trend of probability that the frequency of passenger use will decrease.

3.3. Data formulation for covariate vectors

The covariate vector z in Equation (1) is formulated based on individual information of passengers collected from smart card data. Table 2 shows the prepared variables for this study. The combination of the variables tested and parameters are estimated. As the smart card data used in this paper contain information that is a combination of each boarding and alighting in a day, the proposed model includes variables related to the travel patterns. The "ratio of number of OD pairs in a weekday or on another day (weekend or national holiday) describes how passengers are moving to and from different places by public transport in one day (weekday or another day (weekend or national holiday)). If the passenger is simply traveling from home to the workplace every day, the value of the parameter for "Ratio of number of OD pairs on a weekday or on another day (weekend or national holiday)" becomes lower. "Spatial diversity index of public transport use" was proposed in Nishiuchi et al. (2013), and this was defined to clarify whether a passenger is traveling by the same route (same combination of boarding and alighting bus or tram stops) every day or not. This index is calculated as the proportion of the number of trip days with the most travelled combination of boarding and alighting bus or tram stops pair in a day against a traveller's total number of trip days. Therefore, if the value of "Spatial diversity index of public transport use" is high (close to 1.0), the passenger is traveling using exactly the same daily combination of boarding and alighting bus or tram stop pairs every day.

Table 2. Prepared covariates vectors.

Name of variables	**Explanation**
Fraction of days travelled on weekdays	= total number of travelled days on weekdays / total number of weekdays
Fraction of days travelled on weekends and holidays	= total number of travelled days on weekends and holidays / total number of weekends and holidays
Proportion of number of trips on weekdays	= total number of trips on weekdays / number of all days
Fraction of days travelled in the morning on weekdays	= total number of travelled days in the morning on weekdays / number of days travelled on weekdays
Number of OD pairs on weekdays	= total number of OD pairs on weekdays / number of days travelled on weekdays [OD pairs/ travelled weekdays]
Number of OD pairs on weekends and holidays	= total number of OD pairs on wekends and holidays / number of days travelled on weekends and holidays [OD pair/ travelled weekends and holidays]
Tram use rate	= the number of days traveled by tram / the number of days traveled by any public transport
Spatial diversity index of public transport use	= number of days traveled on the most frequently observed combination of OD pairs (i.e., a sequence of tram stations and bus stops used in a day) / total number of traveled days
Average travel time on weekday	= total travel time on weekdays / number of days travelled on weekdays [minutes/day]
Average travel time on weekends and holidays	= total travel time on holidays / number of days travelled on holidays [minutes/day]
Registration card type dummy	1: if the card type is registered as personalized card 0 :others
Children card type dummy	1: if the card type is registered as children passenger 0 :others
Handicap card type dummy	1: if the card type is registered as handicap passenger 0 :others
Elderly card type dummy	1: if the card type is registered as elderly passenger 0 :others

4. Development of Cox proportional hazard model

4.1. Estimation of parameter of baseline hazard function

Fig. 3 shows the estimation results of the baseline hazard function. A multi-term approximation is applied to decide the parameter of the function. The identified function is as follows:

$$\lambda_0(t) = -0.0024t^2 + 0.0081t + 0.9969 \tag{2}$$

From Fig. 3 and Equation (2), the probability of maintaining the same number of passengers decreases with the passage of time. This means that the total number of passengers for the entire tram network in Kochi is steadily decreasing under the definition of $\lambda_k(t)$. The values from weeks 1 through 3 shows 1.0. This means frequency of use of all passengers will not decrease during weeks 1 through 3 as the definition of probability in this paper follows that the number of trips per week during October through December for passenger k is smaller than the initial baseline value for continually longer than 4 weeks based on Fig. 2. Therefore, the frequency of use of an individual is not considered during weeks 1 through 3.

Also, from Fig. 3, note that public transport authorities start to reduce frequency of number of trams from November 1, 2012. Public transport authorities and other organizations expected that a certain amount of passengers would stop using public transportation. However, such a sudden reduction in passenger usage frequency cannot be confirmed from the figure. Therefore, it can be said that there was no immediate impact caused by the reduction in tram frequency. This insight was also confirmed in the previous study by Nishiuchi et. al. (2015). One possible reason behind it is that in a rural area, public transport tends to be used when there is no alternative mode of travel, indicating that they may not be able to shift to other modes even after the frequency reduction.

On the other hand, there is a possibility that the demand is slowly decreasing due to the frequency reduction, but to explore such long-term impacts of the reduction of frequency would require an analysis using smart card data recorded over a longer period of time.

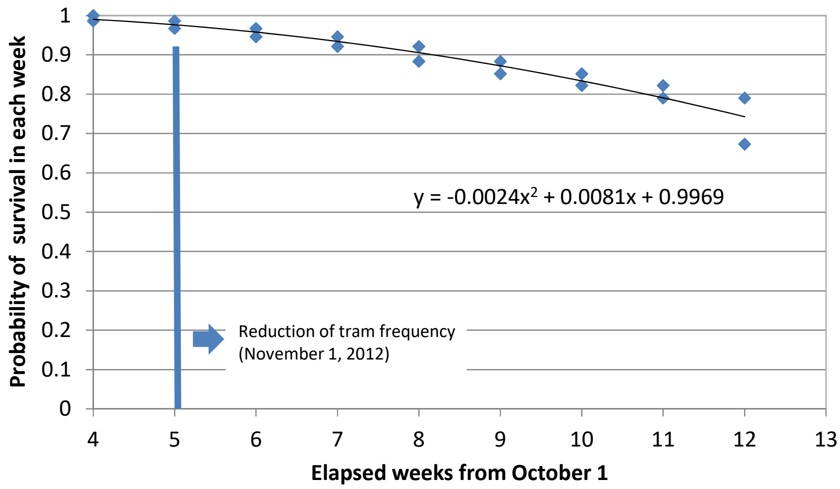

Fig. 3. Estimated approximation curve of baseline hazard function.

4.2. Estimation of parameters of covariates

Table 3 shows the estimation results of the parameters for the covariates. Covariates are chosen by the stepwise method by which the combinations of covariates, which are the estimated minimum AIC values, are sequentially decided. From Table 3, the sign of each parameter can be understood, and a statistical test shows the most of explanatory variables are significant at the 5% level at least, except “Proportion of number of trips on weekdays”, "Tram use rate" and "Children card type dummy". Therefore, the defined covariates are factors explaining the probability of the reduction in frequency by each passenger type.

From the results, values of "Fraction of days travelled on weekdays", "Fraction of days travelled on weekends and holidays", "Number of OD pairs in a weekend and holiday", "Average travel time on weekends and holidays," "Elderly card type dummy" and "Children card type dummy" are significant and have a negative sign. A negative sign means that the probability of a reduction in frequency of use by a passenger becomes lower, which indicates that a reduction in frequency of use by a passenger does not tend to happen when those values become higher. On the other hand, parameters "Diversity index of public transport use", "Registration card type dummy" and "Tram use rate" show positive values, indicating that a reduction in the frequency of use by passengers tends to increase with an increase in the values for those covariates. In summary, the results indicate that (1) passengers who frequently use public transport both on weekdays and weekends are at lower risk of reducing public transport use, i.e., high frequent users may continue to use public transport, even after a decrease in tram frequency; (2) passengers who use public transport on weekdays rather than weekends are at lower risk of the reduction; (3) passengers who travel different OD pairs by public transport (i.e., those who use it for a variety of purposes) are at lower risk of reduction, while those who use public transport only on certain routes (and using a commuter pass (=registration card) by tram) are at a higher risk of reducing public transport use; (4) passengers who travel longer (ie., longer distances) on weekends are at a lower risk of reduction; and (5) Children, and Elderly passengers are at a lower risk of reduction. The most interesting finding may be (3) in terms of promoting public transport use. It indicates that public transport authorities and related organizations should be encouraged to introduce measures to boost public transport use such as getting commuter pass holders to visit the city centre. For example, currently, the commuter pass can only be used for a fixed OD pair in the study area. Introducing zone-based commuter passes could be an interesting policy as this would allow passengers to travel different OD pairs with one commuter pass. In addition, boosting demand for passengers with elderly and children's cards should be considered by public transport authorities because these passengers are at a lower risk of reducing public transport use and hence may contribute to maintaining a certain amount of travel demand.

Table 3. Estimation of parameters for covariates.

Covariates	Unit	Coefficient	z-value	p-value	Significance
Fraction of days travelled on weekdays	trip weekdays / weekday	-1.7605	-16.97	0.0000	**
Fraction of days travelled on weekends and holidays	trip holidays / holiday	-0.4135	-3.08	0.0021	**
Proportion of number of trip on weekdays	-	-0.2056	-1.75	0.0803	*
Number of OD pairs on weekdays	OD pair / trip weekday	-0.0552	-4.31	0.0000	**
Number of OD pairs on weekends and holidays	OD pair / trip holiday	-0.0032	-2.16	0.0310	**
Tram use rate	-	0.0600	1.60	0.1091	
Spatial diversity index of public transport use	-	0.2492	3.42	0.0006	**
Average travel time on weekends and holidays	minutes / holiday	-0.0018	-2.00	0.0451	**
Registration card type dummy	-	0.1607	4.20	0.0000	**
Children card type dummy	-	-0.2693	-1.95	0.0515	*
Elderly card type dummy	-	-0.2441	-6.19	0.0000	**
Number of samples		11,631			
AIC		74,376			

**: 5% level of significance, *: 10% level of significance

5. Conclusions

This paper proposed a Cox proportional hazard model using public transport smart card data to understand what passenger types might tend to reduce their frequency of public transport use in rural cities in Japan. We believe that this is the first study identifying who are at higher/lower risk of reducing public transport use. We have developed a simple method to identify the risk, where changes in the frequency (i.e., lasting reductions of public transport use) are coarsely but simply distinguished from variations of the frequency (i.e., temporal fluctuation of public transport use) based on information on a consecutive reduction of public transport use.

The empirical results show that the reduction of tram frequency by public transport authorities had no immediate effect on the reduction of passengers, though additional analysis using smart card data for a longer period is needed to monitor long-term changes in public transport use and to obtain a firm conclusion. The parameter estimation results indicate that public transport authorities should introduce measures to boost demand by passengers who are traveling by exactly the same route every day, such as commuters. One possible measure is to introduce zone-based public transport passes. This could encourage, commuters to travel not only to their workplaces but also to visit other destinations on their way to/from the workplace. The identification of “at risk” users could provide an alternative view on marketing strategies for increasing the use of public transport as most marketing strategies are seeking ways in which to increase demand. The proposed framework focuses on measures to stop reducing public transport demand, which will become more and more necessary under a declining population.

Though smart card data can identify dynamics in public transport use at the individual level, allowing us to explore “at risk” users as we did, it will become necessary in future studies to extend the period of data acquisition and include other factors in the model such as the location of the point of origin and destination of passengers, conditions of land use at origin and destination, weather conditions, and length of time spent in the city. In addition to that, we should also consider the timing of starting the services of each passenger, especially focusing on new passengers during the data period.

Acknowledgements

We would like to thank Mr. Kenichi Uchiyama of DESUCA Co. Ltd. for providing the data set used in this research. We hereby extend our gratitude. In addition to that this work was supported by JSPS KAKENHI Grant Number 17K06605.

References

Asakura, Y., Iryo, T., Nakajima, Y., and Kusakabe, T. , 2012. Estimation of behavioural change of railway passengers using smart card data. Public Transport 4.1, 1-16.

Chapleau, R., and Chu, K.K. A. ,2007. Modeling transit travel patterns from location-stamped smart card data using a disaggregate approach. Proceedings of World Conference on Transportation Research 11. CD-ROM.

Devillaine, F., Munizaga, M., and Trépanier, M., 2012. Detection of activities of public transport users by analyzing smart card data. Transportation Research Record: Journal of the Transportation Research Board 2276, 48-55.

Kieu, L. M., Bhaskar, A., and Chung, E., 2014. Transit passenger segmentation using travel regularity mined from smart card transactions data. Proceedings of the 93rd Annual Meeting of the Transportation Research Board, USB.

Kita, H., and Tsukioka, S., 2005. OD-Trips estimation of bus passengers based on getting-on-and-off volume data. Infrastructure Planning Review, 22, 145-152 (*in Japanese*).

Kusakabe, T., and Asakura, Y., 2014. Behavioral data mining of transit smart card data: A data fusion approach. Transportation Research Part C: Emerging Technologies 46, 179-191.

Ma, X., Wu, Y. J., Wang, Y., Chen. F., and Liu, J., 2013. Mining smart card data for transit riders’ travel patterns. Transportation Research Part C 36, 1-12.

Morency, C., Trépanier, M., and Agard, B., 2007. Measuring transit use variability with smart-card data. Transport Policy 14, 193-203.

Nishiuchi, H., King, J., and Todoroki, T., 2013. Spatial-temporal daily frequent trip pattern of public transport passengers using smart card data. International Journal of ITS Research 11.1, 1-10.

Nishiuchi, H., Kobayashi, Y., Todoroki, T., and Kawasaki T., 2014. A study of variability of the public transportation users using state space model: A case study of tosa electric railway. Proceedings of the Japan Society of Traffic Engineers 34, 497-500 (*in Japanese*).

Nishiuchi, H., Todoroki, T., and Kawasaki T., 2015. A Study on Change of Number of Tram Passenger by Survival Time Analysis –Case Study on Tosa-Electronic Railway Company-. Journal of the Japan Society of Transportation Economics 58, 113-120 (*in Japanese*).

Páez, A., Trépanier, M., and Morency C.,2011. Geodemographic analysis and the identification of potential business partnerships enabled by transit smart cards. Transportation Research Part A 45, 640-652.

Pelletier, M. P., Trépanier, M., and Morency, C., 2011. Smart card data use in public transit: A literature review. Transportation Research Part C 19, 557-568.

Public Transport Division in Kochi Prefectural Government, 2011. Business report of measures of verification, investigation and cost-benefit analysis related to public transportation. Kochi Prefectural Government (*in Japanese*).

Seaborn, C., Attanucci, J., and Wilson, N., 2009. Analyzing multimodal public transport journeys in London with smart card fare payment data. Transportation Research Record: Journal of the Transportation Research Board 2121, 119–126.

Tamin, O. Z., 1997. Application of transport demand models for inter-regional vehicle movements in West-Java (Indonesia). Journal of the Eastern Asia Society for Transportation Studies, 2. 3, 903-916.

Zhang, C., and Teng, J., 2013. Bus dwell time estimation and prediction: A study case in Shanghai-China. Procedia - Social and Behavioral Sciences 96, 1329-1340.