Comparing travel demand forecasts between models with larger data from single time point and models with smaller data from two time points

Nobuhiro Sanko

Graduate School of Business Administration, Kobe University, Japan
2-1 Rokkodai-cho, Nada-ku, Kobe 657-8501, Japan
Phone & fax: +81-78-803-6987
E-mail: sanko@kobe-u.ac.jp
Abstract
Travel demand forecasting often is based on models estimated with only the most recent data even when data from multiple time points are available. This study aims to investigate a possibility of reducing the number of observations from each time point by using data from multiple time points jointly. Specifically, this study considers a case where data are collected from two points in time, estimates the following two models, and compares their forecasting performance. The first model utilises data only from the more recent time point, while the second model utilises data from two points in time, but the number of observations from each time point is smaller than that utilised in the first model. Models utilising data from two points in time assume that parameters are expressed as functions of gross domestic product per capita. Even when the number of observations from each of two points in time is reduced to 45–80% of that used in models using data only from the more recent time point, models using data from two points in time can produce statistically significantly better forecasts than models using only the more recent data.

Keywords
Travel demand forecast; Temporal transferability; Number of observations; Data collection time points; Mode choice model; Bootstrap
1. Introduction

Forecasts using disaggregate travel demand models often are based on data from the most recent time point, even when cross-sectional data is available from multiple time points. However, this is not a good use of the data. Sanko (2014) proposed a model that jointly utilises data from multiple time points and demonstrated that the proposed model outperformed models utilising only the most recent data in forecasting performance. The proposed model expressed parameters as functional forms, where parameter values vary over time.

Important question is which parameters are which functional forms (e.g., linear, square, square root, log, and exponential) of which variables (e.g., time, GDP per capita). Functions of time (year) (Sanko 2014) and those of gross domestic product (GDP) per capita (Sanko Forthcoming) were compared, and the latter produced better forecasts. However, the above two studies assumed that all the model parameters followed the same functional form of the same variable. Different parameters might be better expressed by different functional forms of different variables. In order to make a firm recommendation on the choice of functional form and the variable, deeper investigation into factors affecting changes in parameters are required. A series of studies by the present author have not reached clear recommendation since he utilised data from a limited number of time points. However, based on the following two reasons mentioned in Sanko (2016b), the author believes that assuming a GDP per capita in linear form for all the parameters is most appropriate among various functional forms. Firstly, a use of functions of time must be avoided due to a problem of overfitting. Secondly, Sanko (2016b) also considered functional forms of: female social participation (in linear form) and Nagoya City’s subway length (in linear form); however, the results differ little from the function of GDP per capita (in linear form). Functions of GDP per capita ascribe changes in parameters to economic conditions, which is easier to interpret among those examined by the present author.
Although above studies assumed that there are a lot of numbers of observations (and equal numbers of observations) from all time points, the author started investigating a case where the numbers of observations differ between time points. Sanko (2017) examined a case, where data are collected from two points in time and compared the forecasting performance of the following two models: (1) models estimated with data only from the more recent time point (hereinafter called ‘more recent data model’); and (2) models estimated with data from two points in time, where parameters are expressed as linear form of GDP per capita (hereinafter called ‘updating function model’, since the method assumes a function which updates parameters). He examined numerous combinations of data collection time points and the numbers of observations from each time point and demonstrated that the updating function models sometimes produced statistically significantly better forecasts than the more recent data models. This suggests that a use of data from two points in time contribute to improve forecasting performance.¹ Note that Sanko (2017) utilised the same number of observations from the more recent time point, which is utilised in both the more recent data models and the updating function models. Therefore, Sanko (2017) examined, in which case of time points and the numbers of observations, which method produced better forecasting performance. However, the updating function models, where the number of observations from each time point is smaller than that utilised for the more recent data models, might produce similar or better forecasts than the more recent data models. (For example, a use of 3000 observations from each of the two points in time (that is, 6000 observations in total) produces better forecasts than a use of 5000 observations from the more recent time point.) Therefore, the use of data from multiple time points might contribute to reduce the sample size and cost of the survey.

A research question of the present study is presented below in working-level.

Consider an older time point of \( y_1 \) and a more recent time point of \( y_2 \). Suppose the following

¹ The updating function model is a novel updating method, since none of model updating methods proposed so far (i.e., transfer scaling, Bayesian updating, combined transfer estimation, and joint context estimation) have produced statistically significantly better forecasts than the more recent data models (Sanko 2016a).
two models: (1) the more recent data models estimated with $n_L$ observations from $y_2$; and (2) updating function models (assuming linear form of GDP per capita) estimated with $n_S$ and $n_S$ observations from $y_1$ and $y_2$, respectively (that is, $n_S + n_S$ observations in total). Which of the above two models produced statistically significantly better forecasts? Note that $n_S \leq n_L$ must be satisfied, where the ‘$L$’ and ‘$S$’ stand for larger and smaller, respectively.

The author investigated this issue in the context of journey-to-work mode choice behaviours by utilising household travel survey data collected in Nagoya, Japan, in 1971, 1981, 1991, and 2001, where the first three time points were used for model estimation and the last time point was used for validation. Two time points (one for the older time point and the other for the more recent time point) are chosen from the 1971, 1981, and 1991, and different numbers of observations (ranging from 100 to 10000) are randomly selected from the chosen two points in time. Therefore, this study benefits from contrasting combinations of time points of the 1971/1981, 1971/1991 and 1981/1991, each of which has different progress of motorisation. In order to obtain insights with statistical meaning, the bootstrap technique is employed. All of the household travel survey data used for this study was implemented by the same governmental bodies, and the survey was conducted in a similar manner in each year. Therefore, it is reasonable to assume that the data is similar in quality across the years. Hence, the datasets used in the present study are appropriate for analysing the two time points, where the data are collected, and the numbers of observations collected from the two points in time.

Since this is the first attempt to compare forecasting performances between the more recent data models and the updating function models with special focus on the possibility of reducing sample size, the analysis is kept as simple as possible and utilises datasets available to the author. Data collection time points and the numbers of observations are the two dimensions of interest in the present study. Other dimensions which might affect temporal transferability but are not considered in the present study include: (i) the underlying theory of travel behaviour (e.g., utility maximisation vs. lexicographic; trip-based vs. tour-based), (ii) the
mathematical model structure (e.g., logit vs. nested logit), and (iii) the empirical specification (e.g., choice of explanatory variables, linear vs. non-linear formulation of explanatory variables, consideration of heterogeneity). This study assumes utility maximisation utilising linear-in-parameters multinomial logit models and uses a single model specification throughout the paper. The data used for this study comes from the 1971–2001 period. The 2001 data is 16 years old, but this is less of a concern.

This paper is organised as follows. Section 2 describes the datasets. Section 3 describes the more recent data models and the updating function models, bootstrapping procedure, and hypothesis testing. Section 4 reports the estimated parameters and compares and statistically tests the forecasting performance of the models. Section 5 presents the concluding remarks.

2. Data


---

2 The author benefited from Sikder’s (2013) explanations of the four levels of hierarchy that must be considered when analysing transferability. (Sikder analysed spatial transferability, but the present author believes that his explanations are still valid for temporal transferability.) The top three levels are (i)–(iii) in the main text. The main interest of this paper is (iv), model parameter estimates (e.g., transferability of coefficients of explanatory variables and other parameters such as elasticities and value of time measures). The two dimensions of data collection time points and the numbers of observations first impact the model parameter estimates and then the forecasting performance. Also note that the present study focuses more on survey than model estimation. (i)–(iii) are issues in the estimation stage and are controllable by researchers after the survey, while data collection time points and numbers of observations are issues in the data collection stage.

3 Model specifications reportedly affect temporal transferability, and most studies have shown that models with more explanatory variables are more temporally transferable (Badoe and Miller 1995b; Fox et al. 2014; Karasmäki and Pursula 1997; Train 1979). However, there is also a chance of overfitting (Badoe and Miller 1995a). The present study utilises the same model specification throughout the analysis, which means that sensitivity to different model specifications is less of a concern. Model specifications which produce better temporal transferability might exist. However, if these specifications were able to improve to the same extent both the more recent data models and the updating function models, then the impact of the model specifications would be cancelled out. If the specifications produced different degrees of improvement, then the model specifications would affect the results of the study. This is a topic for future study.
validation purposes. The household travel survey has been implemented in a similar manner by the same governmental bodies over the years. The modelled trips were journeys to work (commutes). Three alternative transportation modes were considered: rail, bus, and car. The datasets are fully described in Sanko (2014), but two points must be restated. First, this study does not consider travel cost, since most companies provide allowances for employees to purchase commuting passes or fuel. (Rules for allowances differ among companies, and some companies set an upper limit, which most employees do not exceed.) Second, the shares of travel modes have changed substantially between 1971 and 1981, but since then have changed less. After cleaning the data for estimation purposes, the shares of travel modes among rail, bus, and car for commuting purposes in 1971, 1981, 1991, and 2001 were 28%, 28%, 26%, and 25%, respectively, for rail, 21%, 9%, 5%, and 3%, respectively, for bus, and 51%, 63%, 68%, and 72%, respectively, for car. Therefore, this study benefits from contrasting older and recent time combinations of the 1971/1981 and 1971/1991 data, which include the 1971–1981 period that saw substantial share changes, with just the 1981/1991 combination, which does not include the 1971–1981 period.

3. Methodology

This study aims to compare and statistically test forecasting performances produced by the more recent data models and the updating function models. Since this is the first study to statistically test forecasting performances produced by these two models, simple multinomial logit models were employed. However, the methodology is applicable to other model formulations. This section presents multinomial logit models, followed by the more recent data models and the updating function models, bootstrapping procedure, and hypothesis testing utilising the bootstrap. In the following explanation, note that $t_1$ and $t_2$ represent the older and more recent time points, respectively.

3.1. Multinomial logit models
Random utility models are assumed and total utility is decomposed into a deterministic component and an error component. Under the assumptions of linear-in-parameters multinomial logit models, the deterministic component of individual $p$’s utility for alternative $i$, at $t$ ($t = t_1$ or $t_2$ in the following explanations), $V_{ip}^t$, is expressed as Eq. (1).

$$V_{ip}^t = \alpha_i^t + \sum_k \beta_{ik}^t x_{ikp}$$  \hspace{1cm} (1)

where $x_{ikp}$ denotes the $k$-th explanatory variable for individual $p$ for alternative $i$ at $t$, and $\beta_{ik}$ denotes its related parameter; $\alpha_i^t$ denotes an alternative-specific constant for alternative $i$ at $t$. However, a scale parameter, which is fixed to a unity value, is not explicitly formulated in Eq. (1), since the scale parameter and $\alpha$ and $\beta$ cannot be separately identified.

Assuming independent and identical Gumbel distributions for the error components, multinomial logit models are derived, where the probability of individual $p$’s choosing alternative $i$ among alternative $j$’s in his/her choice set at $t$, $P_{ip}^t$, is expressed as:

$$P_{ip}^t = \frac{\exp(V_{ip}^t)}{\sum_j \exp(V_{jp}^t)}$$  \hspace{1cm} (2)

The log-likelihood function, $L$, is defined by the sum of log-likelihood from each time point of $t$, $L^t$;

$$L = \sum_t L^t = \sum_t \sum_p \sum_j y_{jp}^t \ln(P_{jp}^t)$$  \hspace{1cm} (3)

where $y_{jp}^t$ denotes an indicator that takes a value of one if individual $p$ chose alternative $j$ at $t$ and zero otherwise. Parameters are estimated by maximising the log-likelihood function as
expressed in Eq. (3).

3.2. More recent data models and updating function models

3.2.1. More recent data models

Models are estimated as shown in Section 3.1. by utilising data only from \( t = t_2 \). The log-likelihood function as shown in Eq. (3) is expressed as \( L = L^2 \). A forecasting performance is evaluated by a log-likelihood on 2001 data \((L^{2001})\), which is calculated by utilising Eqs. (1)–(3), where \( \hat{\alpha}^{t_2} \) and \( \hat{\beta}^{t_2} \) are from the estimated models but \( x \) and \( y \) are from the 2001 dataset. Note that a hat (\(^\wedge\)) indicates an estimate.

3.2.2. Updating function models

In Eq. (1), the following formulations apply to \( \alpha_i \) and \( \beta_{ik} \).

\[
\begin{align*}
\alpha_i^t & = \alpha_i + \alpha_{i\alpha} gdpi \\
\beta_{ik}^t & = \beta_{ik} + \beta_{ik\alpha} gdpi
\end{align*}
\]

where, \( \alpha_i \) and \( \beta_{ik} \) denote parameters independent of time ('base parameters') and \( \alpha_{i\alpha} \) and \( \beta_{ik\alpha} \) denote historically changing factors for corresponding parameters ('historically changing parameters'). The \( \alpha_{i\alpha} \) and \( \beta_{ik\alpha} \) express parts changing according to the GDP per capita. The \( gdpi \) denotes GDP per capita (constant 2005 price) at \( t \) (units in 10 million JPY).

Models are formulated by applying \( t = t_1 \) and \( t_2 \) to Eqs. (1), (2), and (4). The log-likelihood function as shown in Eq. (3) is expressed as \( L = L^{t_1} + L^2 \). A forecasting performance is evaluated by a log-likelihood on 2001 data \((L^{2001})\), which is calculated by utilising Eqs. (1)–(4), where \( \hat{\alpha} \), \( \hat{\beta} \), \( \hat{\alpha}_d \), and \( \hat{\beta}_d \) are from the estimated models but \( x \) and \( y \)
are from the 2001 dataset, and the GDP per capita is from 2001 (constant 2005 price).4

3.3. Bootstrap

Bootstrapping, a method proposed by Efron and Tibshirani (1993), is applied to this study in the following way.

First, 10000 commuting trips were randomly selected from each time point of 1971, 1981, and 1991. In the following analysis, a smaller number of observations was chosen from these 10000 observations. The same number of observations was chosen from each time point to avoid any impact on forecasting performance that might occur should different numbers of observations from each time point be used. 10000 commuting trips also were selected randomly from the 2001 dataset that was used for evaluating forecasting performance.

Three notations—\(y\), \(n\), and \(b\)—are defined below.

- \(y\) denotes the year when the data was collected (1971, 1981, and 1991).
- \(n\) denotes the number of observations. The author examined 23 values for \(n\) (100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 6000, 7000, 8000, 9000, and 10000).
- \(b\) denotes a bootstrap repetition. Bootstrapping was repeated 200 times (\(b = 1, 2, \ldots, 200\)).

From each \(y\), \(n\) observations were randomly drawn 200 times, with replacement from 10000 commuting trips already selected from each year. (Note that for each of the \(b\)-th draw from the same \(y\), large \(n\) observations contain all the records included in the small \(n\).

---

4 The updating function model is identical to estimating the main effects and interactions between explanatory variables and the GDP per capita. The location parameters and scale parameters of the Gumbel distributions might differ between time points; the author assumes that alternative-specific constants with functional forms account for the differences in the location parameters and that the scale parameters are constant over time. While conventional models assume that neither the scale parameters nor the parameters of explanatory variables change over time (see Fox and Hess (2010) for a review), the updating function model allows the parameters related to explanatory variables to change, which is more flexible.
observations.) In total, 3 \( y \)'s \times 23 n \('s \times 200 b \)'s = 13800 datasets were generated.

This study examines combinations of data collection time points and the numbers of observations, further notations—\( y_1, y_2, n_L, \) and \( n_S \)—are defined in Section 1 but restated below.

- \( y_1 \) and \( y_2 \) represent an older and more recent time points, respectively (\( y_1 < y_2 \)).
- \( n_L \) represents the number of observations from \( y_2 \) when the more recent data model is utilised, while \( n_S \) represents the number of observations from each of \( y_1 \) and \( y_2 \) when the updating function model is utilised. (\( n_L \geq n_S \)).

A working procedure is summarised as follows:

- Estimate the more recent data models and forecast behaviours for the 2001 dataset 9200 times, which is \( 2 y_2 \)'s \times 23 \( n_L \)'s \times 200 b \('s \).
- Estimate the updating function models and forecast behaviours for the 2001 dataset 13800 times, which is 3 combinations of \( y_1 \) and \( y_2 \times 23 n_S \)'s \times 200 b \('s \).

Note that realisable combinations of \( (y_1, y_2) \) are (1971, 1981), (1971, 1991), and (1981, 1991); however, realisable \( y_2 \) in the above combinations is 1981 and 1991. This is why the number of estimations required is different between the above two models.

A forecasting performance is evaluated by a log-likelihood on the 2001 dataset. Let \( L_1 (\cdot, y_2, \cdot, n_L, b) \) and \( L_2 (y_1, y_2, n_S, n_S, b) \) represent log-likelihoods on the 2001 dataset by the more recent data model and updating function model utilising the \( b \)-th repetition for the \( y_1, y_2, n_S, \) and \( n_L \), respectively. Note that the numbering of \( L \)'s corresponds to the time points, where data are collected for model estimation; that is, \( L_1 \) and \( L_2 \) were calculated by models utilising data from one and two points in time, respectively.

3.4. Hypothesis testing

This section proposes tests to compare the forecasting performances of the above two models. Suppose that the more recent data model is estimated with \( n_L \) observations from \( y_2 \) and that the updating function model is estimated with \( n_S \) and \( n_S \) observations from \( y_1 \) and \( y_2 \),
respectively. With a use of $L1$ and $L2$ defined in Section 3.3., the variable $x_b$ is defined as:

$$x_b = L2 (y_1, y_2, n_S, s_L, b) - L1 (\cdot, y_2, \cdot, n_L, b)$$

(5)

Note that $x_b$ is defined only when both $L$’s are calculated. The calculation of $x_b$’s is unsuccessful for some $b$’s, which is likely to happen when $n_L$ and/or $n_S$ are small. If the updating function model produces better forecasts, then $x_b$ is more positive.

Null and alternative hypotheses, represented as $H_0$ and $H_1$, respectively, are defined below.

$$H_0: x_b = 0$$

$$H_1: x_b \neq 0$$

The statistic, $z$, is defined as Eq. (6).

$$z = \frac{x_b}{s(x_b)}$$

(6)

where, $\bar{x}_b$ and $s(x_b)$ denote mean and standard deviation of $x_b$, respectively.

If $x_b$ is assumed to follow a normal distribution, then the null hypotheses are rejected at the five percent level of significance when $z \geq 1.96$, indicating that the updating function models produce better forecasting performance than the more recent data models.

4. Results and discussion

This section presents estimates utilising all the 10000 observations chosen from each year, followed by the results of hypothesis testing.
4.1. Estimates

The following dummy variables are defined: male (1 for male, 0 for female), 20 years old or older (1 if 20 years old or older, 0 if younger), 65 years old or older (1 if 65 years old or older, 0 if younger), and Nagoya (1 if origin and/or destination of the trip are in Nagoya City, 0 if not). Descriptive statistics for the variables included in the mode choice models are fully interpreted in Sanko (2014), but one point must be restated. Extremely large and small shares of 20 years old or older and 65 years old or older, respectively, together with smaller share of bus users presented in Section 2 result in poor estimates when the number of observations is small. (In 1971, 1981, 1991, and 2001, percentages of 20 years old or older are 94.1%, 96.2%, 97.1%, and 98.8%, respectively; percentages of 65 years old or older are 1.5%, 1.6%, 2.1%, and 3.1%, respectively.)

Table 1 reproduced the journey-to-work multinomial logit mode choice model estimates using data from each time point independently (Sanko 2014). The author examined numerous combinations of variables and reported the best results. The author did not include car ownership as an explanatory variable, since it is highly related to mode choice (or car choice) and the two are regarded as being endogenous to some extent. The model specification presented here is used throughout the present paper. Note that travel cost is not included in the models for the reason mentioned in Section 2.\(^5\) Models are fully interpreted in Sanko (2014).

***** Table 1 *****

Table 2 reproduced the journey-to-work multinomial logit mode choice model estimates by the updating function models (Sanko 2017). The estimates of the base parameters and historically changing parameters are shown at the top and bottom of the

\(^5\) Sanko et al. (2013) estimated commuting mode choice models between car and public transportation for the Nagoya metropolitan area and found that the estimate for the car cost parameter was not significant. Moreover, the public transportation cost parameter was not included because it had the wrong sign. This empirically justifies the author’s approach of not considering the travel cost.
table, respectively. The base parameters must be interpreted with special care, however, since they express parameters where the GDP per capita = 0, which is highly unlikely. For the comparison to be fair, the parameters for 1971, 1981, and 1991 must be calculated using the estimates in Table 2 and Eq. (4). (For example, the travel time parameter for 1991 in 1981/1991 model is \(-2.27 + 1.92 \times 0.354 (= \text{GDP per capita in 1991}) = -1.59\). Readers might refer to Sanko (Forthcoming) for more detailed interpretation.) The choice of explanatory variables is the same as that for the more recent data models shown in Table 1, and the model specification presented here is used throughout the present paper.

***** Table 2 *****

Forecasting performances of models shown in Tables 1 and 2 for the 2001 dataset are compared (see the rows labelled ‘Log-likelihood on 2001 data’). The 1971/1991 model produced the highest forecasting performance, followed by the 1981/1991, 1991, 1971/1981, and 1981. This means that if the more recent data comes from 1991, additional use of data from older time point in updating function models contributes to improve the forecasting performance. The same applies to a case where the more recent data comes from 1981.

4.2. Results of hypothesis testing

Figure 1 shows the results of tests to determine in which case the updating function models produce better forecasts than the more recent data models. Also tested were combinations of older and more recent time points: 1971 and 1981 in panel (a), 1971 and 1991 in panel (b), and 1981 and 1991 in panel (c).

***** Fig. 1 *****

In each panel, the horizontal and vertical axes represent the \(n_s\) (the number of
observations from each of the two points in time for the updating function models) and \( n_L \) (the number of observations from the more recent time point for the more recent data models), respectively. Parts shaded in light grey in the lower right of the panel do not satisfy \( n_S \leq n_L \) and are excluded from the analysis. Therefore, the present study focuses on the upper left of the panel. The cells are shaded in black and dark grey if \( z \geq 1.96 \) and \( 1.96 > z > 0.0 \), respectively, indicating that the updating function models produced statistically significantly better forecasts at five percent level of significance and produced better forecasts without five percent level of significance, respectively. Note that Kolmogorov–Smirnov tests (not presented in this paper) did not reject the hypothesis that the \( x_b \) is normally distributed in all combinations of \( y_1 \) and \( y_2 \) when \( n_S = n_L = 10000 \). Quantile–Quantile plots (not presented in this paper) also suggested that the \( x_b \) is normally distributed. This justifies the author’s proposed tests, which assume that \( x_b \) is normally distributed.

In all three panels, the updating function models sometimes produced statistically significantly better forecasts than the more recent data models. Results are interpreted when the \( n_L = 10000 \). In panel (a), the updating function models with at least \( n_S = 6000 \) from each of 1971 and 1981 (that is, 12000 observations in total) produced statistically significantly better forecasts than the more recent data model with \( n_L = 10000 \) from 1981. The corresponding numbers of observations are \( n_S = 4500 \) for 1971 and 1991 (that is, 9000 observations in total) in panel (b) and \( n_S = 8000 \) for 1981 and 1991 (that is, 16000 observations in total) in panel (c). The interpretation is extended to a case where \( n_L < 10000 \) is included, and at least \( n_S = 6000 \), \( n_S = 1500 \), and \( n_S = 3000 \) in panels (a), (b), and (c), respectively contributed to produce statistically significantly better forecasts than the more recent data models.

Next, the updating function models produced better forecasts than the more recent data models (without five percent level of statistical significance). Results are interpreted when the \( n_L = 10000 \). In panel (a), the updating function models with at least \( n_S = 1500 \) from each of 1971 and 1981 (that is, 3000 observations in total) produced better forecasts than the more recent data model with \( n_L = 10000 \) from 1981. The corresponding numbers of
observations are $n_S = 1000$ for 1971 and 1991 (that is, 2000 observations in total) in panel (b) and $n_S = 2000$ for 1981 and 1991 (that is, 4000 observations in total) in panel (c). The interpretation is extended to a case where $n_L < 10000$ is included, and at least $n_S = 1500$, $n_S = 400$, and $n_S = 700$ in panels (a), (b), and (c), respectively contributed to produce better forecasting performances than the more recent data models.

On the other hand, in any combinations of data collection time points and the numbers of observations, the more recent data models never produced statistically significantly better forecasts than the updating function models (i.e., $z \leq -1.96$ was not reported). For example, the more recent data model with 10000 observations did not produce statistically better forecast than updating function models with 100 observations from each of the two points in time (that is, 200 observations in total). (In panel (c), note that statistical test has not been performed due to problems in calculating $x_0$ when $n_S = 100$.)

This suggests a possibility of reducing sample sizes from each time point. In order to obtain statistically significantly better forecasts than the more recent data models, the number of observations from each time point in updating function models can be reduced to 45% (see panel (b) of Fig. 1) to 80% (see panel (c) of Fig. 1) when the number of observations used for the more recent models is 10000. In order to obtain better (without five percent level of statistical significance) forecasts than the more recent data models, the number of observations from each time point in updating function models can be reduced to 10% (see panel (b) of Fig. 1) to 20% (see panel (c) of Fig. 1) when the number of observations used for the more recent models is 10000. Of course, the above discussion is based on an assumption that the future GDP per capita is correctly forecast. Therefore, a sensitivity analysis with respect to the future GDP per capita is required. Sanko (Forthcoming) conducted a sensitivity analysis with respect to the future GDP per capita, when 10000 observations are utilised from each of 1971, 1981, and 1991. A similar approach is applicable to the present study.

Figure 1 panels (b) and (c), where the more recent data comes from 1991, are
compared. The two time points span 20 and 10 years in panels (b) and (c), respectively. The panel (b) has more cells shaded in black and dark grey, implying that a use of data from wider range of time points contributes to improve the forecasting performance. Comparing Fig. 1 panels (a) and (c), both of which have a 10-year-interval, the differences were not substantial.

5. Conclusions

This study questioned a common practice of data usage, that is, a use of data only from the most recent time point even when data from multiple time points are available. This study examined possibilities of reducing the number of observations from each time point of repeated cross-sectional survey, by utilising data from multiple time points jointly. Models with a use of data from multiple time points were formulated based on Sanko (2014, Forthcoming), where parameters are expressed as functions of GDP per capita in linear form. The multiple time points examined in the present study were two points in time, where an older and more recent time points are represented by $y_1$ and $y_2$, respectively. Specifically, the following two models were estimated and compared with respect to their forecasting performance: (1) the more recent data model estimated with $n_L$ observations from $y_2$; and (2) the updating function model estimated with $n_S$ observations from each of $y_1$ and $y_2$ (that is, $n_S + n_S$ observations in total). Note that $n_S \leq n_L$ must be satisfied. The author examined numerous combinations of $y_1$, $y_2$, $n_S$, and $n_L$ and utilised bootstrapping methods to find results with statistical meaning. An empirical case studied is commuting mode choice behaviours in Nagoya, Japan by utilising repeated cross-sectional data.

Findings of the analysis are summarised below.

- The updating function models sometimes produced statistically significantly better forecasts than the more recent data models. If recommendations are made based on the five percent level of significance, a use of the updating function models can reduce the number of observations from each time point to 45–80% of that utilised in the more recent data models when $n_L = 10000$. 
The updating function models sometimes produced better forecasts than the more recent data models without statistical significance. If recommendations are made based on average forecasting performance without any statistical level of significance, a use of the updating function models can reduce the number of observations from each time point to 10–20% of that utilised in the more recent data models when $n_L = 10000$.

The more recent data models never produced statistically significantly better forecasts than the updating function models.

Topics for future study are summarised as follows:

- The present study examined a case where data are available from two points in time. Including a case where data from three points in time are available will produce more findings.
- The findings of the present study are based on an accurate GDP per capita in the target year of forecast. A sensitivity analysis with respect to the future GDP per capita is required to check robustness of the present study.

Budget constraints for transport surveys have been discussed in many places, so analyses such as the present study are required to determine a survey interval and the number of observations.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Numbers 25380564 and 16K03931. The author acknowledges the use of data provided by the Chubu Regional Bureau, Japan’s Ministry of Land, Infrastructure, Transport and Tourism, and the NUTREND (Nagoya University TRTransport and ENvironment Dynamics) Research Group.

References


disaggregate logit mode choice models. Transportation Research Record 1493,
90–100.


Fox, J., Daly, A., Hess, S., Miller, E., 2014. Temporal transferability of models of
mode-destination choice for the Greater Toronto and Hamilton Area. The Journal of
Transport and Land Use 7 (2), 41–62.

models. Transportation Research Record 2175, 74–83.

area travel forecasting models. Transportation Research Record 1607, 38–44.

Sanko, N., 2014. Travel demand forecasts improved by using cross-sectional data from
multiple time points. Transportation 41 (4), 673–695.

Sanko, N., 2016a. Criteria for selecting model updating methods for better temporal
transferability. Compendium of Papers of the 95th Annual Meeting of the
Transportation Research Board, Washington D.C., U.S.A.

Sanko, N., 2016b. Factors affecting temporal changes in mode choice model parameters.
Transportation Planning and Technology, 39 (7), 641–652.

domestic product per capita. Compendium of Papers of the 96th Annual Meeting of
the Transportation Research Board, Washington D.C., U.S.A.

Sanko, N., Forthcoming. Travel demand forecasts improved by using cross-sectional data
from multiple time points: enhancing their quality by linkage to gross domestic product.
Transportation.

Sanko, N., Morikawa, T., Kurauchi, S., 2013. Mode choice models’ ability to express intention
to change travel behaviour considering non-compensatory rules and latent variables.

Note: Parts shaded in light grey in the lower right of the panel do not satisfy $n_S \leq n_L$ and are excluded from the analysis. The cells are shaded in black and dark grey if $z \geq 1.96$ and $1.96 > z > 0.0$, respectively. In panel (c), note that statistical test has not been performed due to problems in calculating $x_b$ when $n_S = 100$.

**Figure 1** Statistical tests with respect to forecasting performance.
Table 1 Estimates of more recent data models

<table>
<thead>
<tr>
<th>Variables</th>
<th>1981</th>
<th>1991</th>
<th>2001a</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>t-stat.</td>
<td>Est.</td>
</tr>
<tr>
<td>Constant (B)</td>
<td>-0.392</td>
<td>-6.21</td>
<td>-0.638</td>
</tr>
<tr>
<td>Constant (C)</td>
<td>-0.645</td>
<td>-4.65</td>
<td>0.301</td>
</tr>
<tr>
<td>Travel time [hr]</td>
<td>-1.81</td>
<td>-16.47</td>
<td>-1.59</td>
</tr>
<tr>
<td>Male dummy (R)</td>
<td>0.787</td>
<td>8.70</td>
<td>0.812</td>
</tr>
<tr>
<td>Male dummy (C)</td>
<td>2.17</td>
<td>25.22</td>
<td>1.78</td>
</tr>
<tr>
<td>20 years old or older dummy (C)</td>
<td>0.764</td>
<td>5.78</td>
<td>0.776</td>
</tr>
<tr>
<td>65 years old or older dummy (B)</td>
<td>1.37</td>
<td>5.73</td>
<td>1.33</td>
</tr>
<tr>
<td>Nagoya dummy (C)</td>
<td>-1.77</td>
<td>-33.21</td>
<td>-2.18</td>
</tr>
<tr>
<td>N</td>
<td>10000</td>
<td>10000</td>
<td>10000</td>
</tr>
<tr>
<td>$L(\beta)$</td>
<td>-5985.02</td>
<td>-5300.58</td>
<td>-4716.28</td>
</tr>
<tr>
<td>$L(0)$</td>
<td>-8593.88</td>
<td>-8398.85</td>
<td>-8159.63</td>
</tr>
<tr>
<td>Adj rho-squared</td>
<td>0.303</td>
<td>0.368</td>
<td>0.421</td>
</tr>
<tr>
<td>Log-likelihood on 2001 data</td>
<td>-5225.15</td>
<td>-4801.79</td>
<td>Not applicable</td>
</tr>
</tbody>
</table>

Note: (R), (B), and (C) notations refer to alternative-specific variables for rail, bus, and car, respectively. Variables without notations are generic.

a 2001 is the target year of forecast, and a model from 2001 is not required but is presented for a comparison purpose.
Table 2 Estimates of updating function models

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.  t-stat.</td>
<td>Est.  t-stat.</td>
<td>Est.  t-stat.</td>
</tr>
<tr>
<td><strong>Base parameters</strong> ($\alpha_i, \beta_{ik}$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant (B)</td>
<td>1.45</td>
<td>6.74</td>
<td>0.856</td>
</tr>
<tr>
<td>Constant (C)</td>
<td>-2.51</td>
<td>-4.54</td>
<td>-2.54</td>
</tr>
<tr>
<td>Travel time [hr]</td>
<td>2.53</td>
<td>5.94</td>
<td>0.334</td>
</tr>
<tr>
<td>Male dummy (R)</td>
<td>-0.0156</td>
<td>-0.06</td>
<td>0.353</td>
</tr>
<tr>
<td>Male dummy (C)</td>
<td>1.43</td>
<td>4.86</td>
<td>2.15</td>
</tr>
<tr>
<td>20 years old or older dummy (C)</td>
<td>1.28</td>
<td>2.45</td>
<td>1.02</td>
</tr>
<tr>
<td>65 years old or older dummy (B)</td>
<td>3.27</td>
<td>3.29</td>
<td>2.47</td>
</tr>
<tr>
<td>Nagoya dummy (C)</td>
<td>0.572</td>
<td>2.62</td>
<td>-0.118</td>
</tr>
<tr>
<td><strong>Historically changing parameters</strong> ($\alpha_{di}, \beta_{dik}$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant (B)</td>
<td>-7.71</td>
<td>-7.20</td>
<td>-4.22</td>
</tr>
<tr>
<td>Constant (C)</td>
<td>7.85</td>
<td>2.87</td>
<td>8.02</td>
</tr>
<tr>
<td>Travel time [hr]</td>
<td>-18.2</td>
<td>-8.56</td>
<td>-5.44</td>
</tr>
<tr>
<td>Male dummy (R)</td>
<td>3.40</td>
<td>2.74</td>
<td>1.29</td>
</tr>
<tr>
<td>Male dummy (C)</td>
<td>3.11</td>
<td>2.12</td>
<td>1.04</td>
</tr>
<tr>
<td>20 years old or older dummy (C)</td>
<td>-2.18</td>
<td>-0.84</td>
<td>-0.677</td>
</tr>
<tr>
<td>65 years old or older dummy (B)</td>
<td>-7.89</td>
<td>-1.63</td>
<td>-3.24</td>
</tr>
<tr>
<td>Nagoya dummy (C)</td>
<td>-9.82</td>
<td>-9.16</td>
<td>-5.82</td>
</tr>
<tr>
<td>N</td>
<td>20000</td>
<td>20000</td>
<td>20000</td>
</tr>
<tr>
<td>$L(\beta)$</td>
<td>-13761.89</td>
<td>-13077.44</td>
<td>-11285.60</td>
</tr>
<tr>
<td>$L(0)$</td>
<td>-17542.13</td>
<td>-17347.11</td>
<td>-16992.73</td>
</tr>
<tr>
<td>Adj rho-squared</td>
<td>0.215</td>
<td>0.245</td>
<td>0.335</td>
</tr>
<tr>
<td>Log-likelihood on 2001 data</td>
<td>-4996.66</td>
<td>-4764.18</td>
<td>-4779.85</td>
</tr>
</tbody>
</table>

Note: (R), (B), and (C) notations refer to alternative-specific variables for rail, bus, and car, respectively. Variables without notations are generic.