CORRECTING ATTRITION BIAS IN STATED PREFERENCE MODE CHOICE MODELS*

Yoriyasu SUGIE, Professor
Graduate School for International Development and Cooperation
Hiroshima University

Akimasa FUJIWARA, Associate Professor
Graduate School for International Development and Cooperation
Hiroshima University
and
Hironori YAMANE
Fukken Chosa Sekkei Ltd

ABSTRACT

This paper aims at examining the extent of attrition biases specifically included in Stated Preference data and also demonstrates the effectiveness of a correcting method which exclude them in model building. SP panel data for the New Light Rail Transit System in Hiroshima, measured at two different points in time, was used for this purpose. As a result, the effect of attrition bias on mode choice model parameters based on SP data was empirically clarified. Attrition models which determine whether the respondents of the first wave participate in the second survey or not were developed using first wave data. In addition, a mode choice model was constructed based on the stayers (i.e. participants in the first wave who participated in the second wave) at the second wave. The attrition bias of this model was corrected by sequential steps on the assumption that the error in the mode choice model correlates with that of the attrition model. It was found that this type of correcting could effectively cancel out the biased share of each travel mode.

1. INTRODUCTION

Dynamic approach focusing on the time axis in behavioral mechanism has been taken up by many transport researchers as a main research subject in travel demand analysis since the Oxford Conference on Travel and Transportation in 1988. Repeated cross-sectional data, which is represented by traditional O-D travel survey data, is insufficient to understand the temporal change of travel behavior. However, panel data obtained by repeated surveys for the same group of individuals over time has been recognized as important for analyzing these types of travel behavior.

Panel data on transportation, including the Dutch Mobility Panel, has been collected and applied to various types of travel behavioral analysis in recent years. In the first U.S. Conference on Panels for Transportation Planning held in 1992, transport

* This is a revised edition of the paper presented at the 7th World Conference on Transport Research, Sydney, Australia, July 16-21, 1995.

panels collected in various areas were introduced and recent studies on survey methods and new methodologies of data analysis were presented. It was widely accepted by the participants in the Conference that the role of panel research would become greater in the field of transport studies for quite some time, since the demand for dynamic travel demand modeling has grown considerably in response to the needs of a rapidly changing society.

Panels are a powerful tool in analysis which researches the relationships between temporal change in behavior at the individual level and exogenous variables. They have another advantage in data collection in that they require a small number of samples, so they entail a lower level of data collection cost. However, they have several disadvantages in application to be overcome, one being that panel data is liable to have sample attrition biases caused by participants dropping out of subsequent surveys and fatigue biases caused by survey repetition. In addition, non-representation of data supplemented by newly recruited samples for the population sometimes becomes a serious problem in panel analysis.

This study focuses specifically on the attrition bias of panel data. The attrition bias does not always arise at the same time when dropping out of the panel survey occurs, but it does when dropping out has some correlation with travel behavior. This bias will become serious especially in dynamic modeling. It is known that respondents with less interest in the subject matter of a survey tend not to cooperate with the data collection effort (Brog et al, 1980). Bus users are more interested in the improvement of public transportation than car users and tend to participate in the second-wave survey. In this case, there is a correlation between the factors which determine the choice level for the new public transport policy and the number of surveys the respondents participated in. As a result, dynamic modeling of mode choice using this type of data will include some degree of bias.

In this study, a method of correcting the attrition bias included in the panels will be investigated by developing mode choice models using stated preference panel data in Hiroshima and some examples of correcting models will be presented.

2. REVIEW OF EXISTING LITERATURE

Attrition bias is similar to selectivity bias in the mechanism of its generation, in that both are characterized by the bias of sample features (Dubin et al, 1984). We often see that the correcting procedure for selectivity bias was applied to the correction for attrition bias. Selectivity bias becomes large when two different independent variables of discrete and continuous types are dependent on each other. In such discrete-continuous models, correcting the bias is made by assuming the existence of a correlation between the errors of these two models. It is applied, for instances, to vehicle choice/vehicle use (Mannering et al, 1987), household car ownership/mobility (Kitamura, 1987) and mode choice/parking location (Westin et al, 1978).

When the correcting method for selectivity bias is applied to the correction of attrition bias included in the panel data, an attrition model describing the choice to stay in or drop out of the panel will be developed first, then travel behavioral models will be determined. This application can be seen in the models of income maintenance (Hauseman et al, 1979) and sub-compact car purchase (Winner, 1983) for economic modeling, and trip reporting (Kitamura et al, 1987) and vehicle use and possession (Hensher et al, 1992) for transportation planning models. All these cases employed probit models as a tool of analysis, and the effects of attrition bias on model estimates are measured by making use of parameter estimates for the correction term for attrition bias as an evaluation indicator.

It is possible to obtain consistent estimates by weighting $Q(i)/H(i)$ on the likelihood function (where, $H(i)$: share of alternative i in the population, $Q(i)$: share of alternative i in the samples) by assuming the panel as a kind of choice-based sampling data (Manski et al, 1977). An analogous method proposed by Nishii et al (1992) is that the log likelihood function of shopping destination choice be weighted by the reciprocal of attrition probability in the panel, which is given by the attrition model as a weighted factor.

Most of these studies have been done within the context of revealed preference panel data, whereas panel analysis based on SP data suitable for demand analysis of non-existing alternatives has started to make an appearance (Polak et al, 1994; Sugie et al, 1994). Our study is a three-time points mode choice analysis based on SP data, indicating that the stayers in the panel have a higher propensity to use the New Transit System (NTS), a commuter rail system proposed for Hiroshima more than the dropouts. It is expected that sample attrition biases would be generated in the panel data.

3. A CORRECTING METHOD FOR ATTRITION BIAS

Mode choice models will be developed together with attrition models, taking into account the correlation between the attrition in the panel and mode choice behavior. Its effectiveness will then be tested using SP data on mode choice. Both are discrete choice models consisting of binomial and multinomial logit type, respectively. A correcting equation for attrition bias will be derived in this chapter keeping in mind related ideas on correction methods for selectivity bias (Sano, 1990).

When some samples having a higher share for a specific transportation mode have a greater tendency to stay in subsequent surveys, factors in common are believed to be closely related to this phenomenon. Accordingly, the error terms based on panel data which include attrition biases, may not meet the assumption that they have zero means in the utility function. It is necessary to postulate a new error term with mean zero by eliminating the bias from the zero mean.

Latent variables A_{ns}^{t-1}, A_{nd}^{t-1}, associated with the choice to stay or drop out, at time $t-1$ of individual n can be expressed by defining their determinants and errors as V_{ns}^{t-1}, V_{nd}^{t-1} and ζ_{ns}^{t-1}, ζ_{nd}^{t-1} as follows:

$$A_{ns}^{t-1}=V_{ns}^{t-1}+\zeta_{ns}^{t-1} \qquad (1)$$

$$A_{nd}^{t-1}=V_{nd}^{t-1}+\zeta_{nd}^{t-1} \qquad (2)$$

On the assumption that the error is Gumbel-distributed, the attrition model will be formulated as a binary logit type. For that, the latent variable A_n^{t-1} is expressed as the difference of A_{ns}^{t-1} and A_{nd}^{t-1}.

$$A_n^{t-1}=V_n^{t-1}+\zeta_n^{t-1} \qquad (3)$$

Where,

$$A_n^{t-1}=A_{ns}^{t-1}-A_{nd}^{t-1} \qquad (4)$$

$$V_n^{t-1}=V_{ns}^{t-1}-V_{nd}^{t-1} \qquad (5)$$

$$\xi_n^{t-1}=\zeta_{ns}^{t-1}-\zeta_{nd}^{t-1} \qquad (6)$$

$$\begin{cases}\text{If } A_n^{t-1}\geq 0,\ \delta=1 \text{ (participation)}\\ \text{If } A_n^{t-1}<0,\ \delta=0 \text{ (non-participation)}\end{cases} \qquad (7)$$

Staying probability P_{ns}^{t-1} is given by the following equation.

$$P_{ns}^{t-1}=\frac{1}{1+\exp\{-(V_{ns}^{t-1}-V_{nd}^{t-1})\}} \qquad (8)$$

The error term of stayer n at time t for alternative i in the utility function ε_{in}^t is divided into a conditional expected value when participating in the panel at time $t-1$ $(\varepsilon_{in}^t \mid A_n^{t-1}\geq 0)$ and an error term with mean zero υ_{in}^t.

$$\begin{aligned}U_{in}^t&=V_{in}^t+\varepsilon_{in}^t\\ &=V_{in}^t+E(\varepsilon_{in}^t \mid A_n^{t-1}\geq 0)+\upsilon_{in}^t\end{aligned} \qquad (9)$$

Where,

$$E\,(\upsilon^t_{in})=0 \tag{10}$$

The conditional expected value of ε^t_{in}, given $A^{t-1}_n \geq 0$, can be given after obtaining the conditional expected values, ζ^{*t-1}_{ns} and ζ^{*t-1}_{nd} (Sano, 1990).

$$E\,(\varepsilon^t_{in}\,|\,A^{t-1}_n \geq 0) = E\,(\varepsilon^t_{in}\,|\,\zeta^{*t-1}_{ns},\ \zeta^{*t-1}_{nd}) \tag{11}$$

$$\zeta^{*t-1}_{ns} = E\,(\zeta^{t-1}_{ns}\,|\,A^{t-1}_n \geq 0) \tag{12}$$

$$\zeta^{*t-1}_{nd} = E\,(\zeta^{t-1}_{nd}\,|\,A^{t-1}_n \geq 0) \tag{13}$$

The following equations can be set up by transforming equations (12) and (13).

$$E\,(\zeta^{t-1}_{ns}\,|\,A^{t-1}_n \geq 0) = -\ln P^{t-1}_{ns} \cdot \theta^{-1} \tag{14}$$

$$E\,(\zeta^{t-1}_{nd}\,|\,A^{t-1}_n \geq 0) = \frac{P^{t-1}_{nd}}{1-P^{t-1}_{nd}} \ln P^{t-1}_{nd} \cdot \theta^{-1} \tag{15}$$

Where, P^{t-1}_{ns}: staying probability of individual n at time t (given by equation (8))
P^{t-1}_{nd}: dropout probability of individual n at time t
θ: scale parameter

The utility function (16) which corrects attrition biases is finally given by substituting equations (11), (14) and (15) for equation (9).

$$U^t_{in} = V^t_{in} + \gamma_{is}\left(\frac{P^{t-1}_{nd}}{1-P^{t-1}_{nd}} \ln P^{t-1}_{nd} + \ln P^{t-1}_{ns}\right) + \upsilon^t_{in} \tag{16}$$

Where, γ_{is}: parameter of attrition correction term for alternative i

4. ANALYSIS OF ATTRITION BIAS IN HIROSHIMA SP PANEL DATA

Of the five-wave panel surveys which were carried out in residential housing development in northwestern Hiroshima since 1987, the first two waves (1987 and 1988) were employed to investigate the biases generated between the two waves. The outline of the panel survey is presented in Table 1 and the sample attrition in the panel is shown in Figure 1 (Fujiwara et al, 1991).

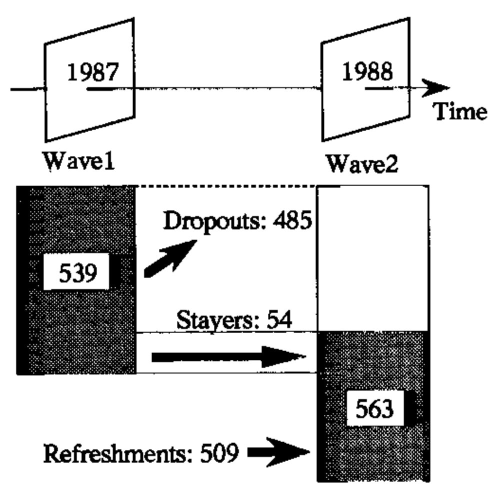
These are SP surveys with respect to the choice of the most preferred of three travel modes: the New Transit System (proposed for 1994), car and bus. Travel factors used in SP experiments are In-vehicle time, Travel cost, Wait time, and Access time. Twenty seven profiles combining three levels of these factors were set up based on the fractional factorial design. One of the profiles was randomly presented in the experiment which asked respondents to rank the three travel modes in order of preference. Three questions were made at the first wave, and five at the second wave.

All samples at the first wave were classified into two

Table 1 Outline of SP Panel Surveys in Hiroshioma

	Wave 1	Wave 2
Time	1987	1988
Location	Northwestern residential area	
Respondents	Commuters/Students	
Collecting method	Home interviewing (self completion)	
Response format	Ranking of Bus, Car and NTS	
No. of questions	3	5
No. of respondents	539	563

Figure 1 Sample Attrition of SP Panel Data in Hiroshima



different data groups; stayers' and dropouts' according to whether the respondent chooses to stay in or drop out of the panel. Stayers comprised a small percentage of the total number of samples (i.e. 54) in the second wave, because questionnaire sheets were not distributed in the second wave to respondents who did not reply "Yes" to the question in the first wave which asked if he/she would be willingly to participate in supplementary interviewing or not. This will make the study of attrition issues more meaningful because the high attrition rate might yield a significant bias in the results of respondent preferred ranking of the three travel modes. In order to supplement the substantial decrease in samples due to dropouts, about 500 new respondents were recruited from the same residential area at the second wave.

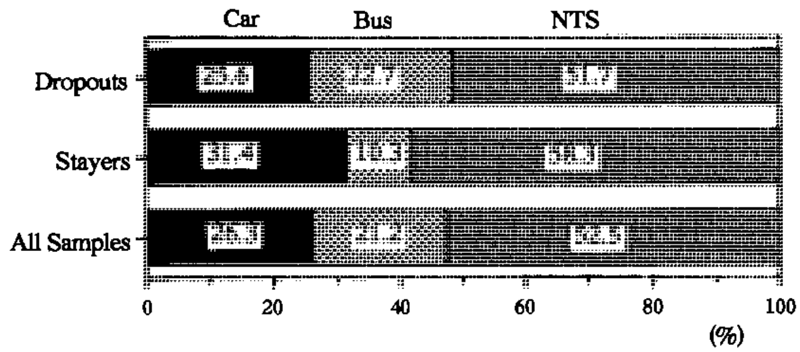
As the stayers' group seems to have a bias in the structure of socio-economic characteristics as compared with other respondents, Qualification Theory 2 (Hayashi, 1954) was employed to check the effect of socio-economic factors on sample attrition by setting the stayers and dropouts as an external criteria. Seven factors which have higher levels of partial correlation coefficients are listed in Table 2. It can be seen that the stayers' group belongs to several specific categories; age in the 20's or 50's, access to NTS stations by walking, owning one car, worker, and one commuter/student in the household.

Table 2 Causal Analysis Associated with Participation in Panel Surveys

Item	Category	Score	Partial correlation coefficient
Age	~19	−0.494	0.209
	20−29	2.077	
	30−39	−4.586	
	40−49	−0.092	
	50−59	7.820	
	60~	−4.830	
Access mode to NTS	Walk	1.772	0.121
	Bus	−3.327	
Car ownership	0	−5.059	0.093
	1	1.122	
	2~	−1.742	
Occupation	Worker	0.800	0.080
	Student	−4.673	
	Others	−1.438	
Response to TP questions	Complete	0.375	0.066
	Non-complete	−4.566	
No of commuters and students in household (15 years old and above)	1	0.811	0.065
	2	−0.462	
	3	−3.232	
	4	−7.477	
	5~	−1.638	
Stated best mode	Car	1.243	0.056
	Bus	−2.127	
	NTS	0.236	

External criteria: Stay+, Dropout− No of samples=1344 $\eta^2=0.101$

Figure 2 Shares of Ranked First Modes



"Response to TP questions" in Table 2 is an item which signifies whether questions on Transfer Price given in SP experiments are completely filled in or not. This was introduced in the model as a factor to indicate the level of magnitude for the cooperation in, and respondent interest in, the panel survey.

The results of mode choice questions, which are dependent variables of mode choice models were related to the two groups of respondents. As shown at the bottom of Table 2, the travel mode ranked first (i.e. stated best mode) in SP experiments is an important factor which determines whether the individual will remain in the panel or not.

Figure 2 presents the shares of the three travel modes which were ranked as the best mode at the first wave by stayers and dropouts. In the first wave, three responses were given by each respondent, so the number of samples is almost three times the total number of respondents. The share of bus for stayers is substantially lower than that for dropouts. It is, therefore, concluded that the stayers' data include a significant attrition bias concerning mode choice.

5. BUILDING SP MODELS CONCERNING MODE CHOICE

SP models on mode choice were developed as shown in Table 3 for all respondents, stayers and dropouts using the first wave data, then the difference of model parameters between stayers' and dropouts' models were statistically tested.

Explanatory variables considered in the model consist of travel factors set up in SP experiments. In addition, Car ownership in the household, which is expected to greatly influence the stated preference for car use, was also included in the explanatory variables, even though it was not presented in the experiments. As three SP responses can be obtained under the different travel conditions in the first wave, the total number of samples is given by subtracting the number of unusable responses from the number of all respondents times three (i.e. number of questions per respondent).

It can be seen that the stayers' model yields the highest degree of fit (i.e. Rho-bar squared), even though parameter estimates for the three models resemble each other. This means that travel factors set up in SP experiments can explain to some degree the stated preference, especially for the stayers' group. As the dropouts account for 85 percent of all respondents, it is natural that the estimated results for all respondents' and dropouts' models are quite similar.

Based on these results, the t test for the difference of parameters between the stayers' and dropouts' models was carried out using equation (17) in order to

Table 3 Mode Choice Models for the First Wave

Variable	All respondents	Stayers	Dropouts
Car ownership	0.440	0.420	0.465
	(4.33)	(1.87)	(4.39)
Access time	−0.080	−0.085	−0.079
	(−4.81)	(−1.47)	(−4.54)
Wait time	−0.086	−0.381	−0.060
	(−4.22)	(−4.25)	(−2.86)
In-vehicle time	−0.035	−0.040	−0.035
	(−6.30)	(−2.21)	(−6.00)
Travel cost	−0.003	−0.002	−0.004
	(−8.53)	(−1.58)	(−8.56)
Car specific constant	−1.936	−2.053	−1.976
	(−6.34)	(−2.21)	(−6.07)
Bus specific constant	−0.455	0.980	−0.574
	(−1.75)	(1.05)	(−2.10)
L(0)	−1475.2	−171.4	−1303.8
L($\hat{\beta}$)	−1255.7	−122.5	−1118.8
Rho-bar squared	0.147	0.269	0.139
% correct	56.7	58.3	55.8
No. of sampels	1343	156	1187

Numbers in parentheses are t-statistics

check the difference in weights for each variable.

$$t=\frac{|\hat{\beta}_s-\hat{\beta}_d|}{\sqrt{\frac{(n_s-1)\,n_s S_s^2+(n_d-1)\,n_d S_d^2}{n_s+n_d-2}}\sqrt{\frac{1}{n_s}+\frac{1}{n_d}}} \tag{17}$$

Where,, $\hat{\beta}_s$, $\hat{\beta}_d$: parameter estimates for stayers' and dropouts' models

S_s, S_d: standard deviations of estimated parameters for both models

n_s, n_d: number of samples for both models

As can be seen in Table 4, the significant difference for the Wait time parameter was statistically accepted. Table 3 demonstrates that the absolute Wait time parameter value for the stayers' model is considerably bigger than that for the dropouts' model, so we can see that the stayers' group places great important on Wait time. It can therefore be said that the SP data for stayers' group includes significant

Table 4 Difference of Parameters between Stayers and Dropouts

Variable	t-statistics
Car ownership	0.148
Access time	0.114
Wait time	4.801*
In-vehicle time	0.281
Travel cost	1.603
Car specific constant	0.081
Bus specific constant	1.884

*Significant at 1%

attrition biases for all samples and this greatly affects the parameter estimate for Wait time.

The likelihood ratio statistics χ^2 (7 degrees of freedom) given by equation (18) were employed in order to identify more firmly the overall difference between the stayers' and dropouts' models. The null hypothesis to be tested is that parameter vectors of the two models are equivalent.

Table 5 Difference of Parameter Vector between Stayers and Dropouts

χ^2 Statistic	Degree of freedom	Critical value at 5%
28.8	7	14.1

$$\chi^2 = -2\{L(\widehat{\beta}_{s+d}) - L(\widehat{\beta}_s) - L(\widehat{\beta}_d)\} \quad (18)$$

Where, $L(\widehat{\beta}_{s+d})$: maximum log likelihood for all respondents' model

$L(\widehat{\beta}_s)$: maximum log likelihood for stayers' model

$L(\widehat{\beta}_d)$: maximum log likelihood for dropouts' model

As a consequence of the χ^2 test, the null hypothesis was rejected, because the χ^2 statistic is greater than its critical value (see Table 5). This indicates that the parameter vectors in both models are significantly different. It became more apparent that the stayers' model is affected by attrition bias.

6. DEVELOPING CORRECTED MODELS AND THEIR EFFECTIVENESS

6.1 Attrition Models

Specification of attrition models based on equation (8) is shown in Table 6. Positive values for certain parameters in the model imply a positive contribution of the variable to the probability of participation in the second-wave survey, while negative values suggest non-participation. The explanatory variable at the bottom, "Supplementary survey", in Table 6 is a dummy variable which indicates whether the respondent agreed to participate in the supplementary interview survey or not. This question was posed at the end of questionnaire sheet at the first wave. Other variables were determined by consulting the result of causal analysis done in Table 2. Some variables were displaced by other related ones because of the generation of multicolinearity among independent variables.

Table 6 Attrition Models Using First Wave Data

Variable	Model A	Model B
Const	−2.907	−2.140
	(−31.2)	(−41.7)
Sex	0.250	0.305
(1: male, 0: female)	(1.69)	(2.23)
Age	0.001	0.005
	(0.16)	(1.58)
Car license	0.380	0.420
(1: holder, 0: non-holder)	(2.65)	(3.16)
Car owner in the future	0.133	0.530
(1: yes, 0: no)	(0.81)	(3.47)
Working/studying place	−0.185	−0.124
(1: CBD, 0: others)	(−1.65)	(−1.24)
Access mode to NTS	0.369	0.481
(1: walk, 0: others)	(3.02)	(4.39)
Car user at present	0.196	−0.094
(1: yes, 0: no)	(1.79)	(−0.99)
No of respondents in household	0.027	0.022
	(0.54)	(0.47)
Supplementary survey	1.645	
(1: yes, 0: no)	(11.4)	
L(0)	−473.6	−473.6
$L(\hat{\beta})$	−328.7	−439.7
Rho-bar squared	0.300	0.065
% correct	88.4	88.0
No. of samples	1289	1289

Numbers in parentheses are t-statistics

It can be seen that car license holders and walkers to the nearest NTS station tend to remain in the panel. The "Supplementary survey" value in model A has a higher explanatory power and improved greatly the goodness-of-fit for the attrition model. Since the effectiveness of correcting attrition biases greatly depends on the estimation accuracy of the attrition model, the attrition model A will be used for the correction of attrition bias in the remainder of this paper.

6.2 Correction of Attrition Biases

The mode choice model for the second wave which was corrected for attrition bias using equation (16) is specified in Table 7. Another model, which was not corrected, is listed in the same Table for comparison.

The utility functions of individual n for the three choices (i.e. car, bus and NTS) at time t are expressed as:

$$U^t_{car,n} = V^t_{car,n} + \gamma_{car,s}\,\Omega + \upsilon^t_{car,n} \tag{19}$$

$$U^t_{bus,n} = V^t_{bus,n} + \gamma_{bus,s}\,\Omega + \upsilon^t_{bus,n} \tag{20}$$

$$U^t_{NTS,n} = V^t_{NTS,n} \qquad + \upsilon^t_{NTS,n} \tag{21}$$

Where,

$$\Omega = \left(\frac{P^{t-1}_{nd}}{1-P^{t-1}_{nd}} \ln P^{t-1}_{nd} + \ln P^{t-1}_{ns} \right) \tag{22}$$

As the attrition correction term, γ_{is} (see equation (16)) can only be set up for two alternatives (i.e. number of alternatives-1), utility functions only for car and bus incorporated the correction terms by setting the correction for NTS to zero.

The number of responses given by each respondent is three for the first wave, while it is five for the second wave, so the response matching for two time points becomes 15 for each respondent. Supposing that responses given by each respondent are independent of each other for the two time points, each response pair can be regarded as one independent sample. Therefore, the number of samples used here is equivalent to that the number of stayers (i.e. 54) times the number of response pairs (i.e. 15), minus the number of unusable responses.

Model estimation was made by two sequential steps; P^{t-1}_{nd} and P^{t-1}_{ns} of equation (22) were first calculated based on attrition model A shown in Table 6, using the data at time $t-1$ (first wave), then equations (19)–(21) were specified using the data at time t (second wave).

We did not find any clear difference in Table 7 in terms of the goodness-of-fit results (e.g. Rho-bar squared and Percent correct) between the two models, one of which was corrected for bias and the other which was not corrected. It can be seen that

Table 7 Models Correcting Attrition Biases for Stayers in the Second Wave

Variable	Non-corrected	Corrected
Car ownership	−0.569	−0.571
	(−3.08)	(−3.06)
Access time	−0.118	−0.125
	(−3.02)	(−3.17)
Wait time	−0.068	−0.070
	(−2.12)	(−2.15)
In-vehicle time	−0.035	−0.036
	(−7.19)	(−7.41)
Travel cost	−0.003	−0.003
	(−4.16)	(−4.15)
Car specific constant	−1.313	−0.655
	(−2.51)	(−1.11)
Bus specific constnat	−1.407	−1.226
	(−4.36)	(−3.09)
Attrition correction for car		0.332
γ *car, s*		(2.28)
Attrition correction for bus		0.092
γ *bus, s*		(0.08)
L(0)	−729.4	−729.4
$L(\hat{\beta})$	−466.5	−463.5
Rho-bar squared	0.357	0.361
% correct	74.4	73.5
No. of samples	661	661

Numbers in parentheses are t-statistics

the utilities for car and bus were biased in such a way as to be smaller than they actually were before the correction, since the parameter estimates for the attrition correction terms are positive for both modes. This

indicates that the share of NTS will be overestimated without the correction. It appears that the correction of biases was effectively done, because the t statistics for the correction variable for car preference, $\gamma_{car,s}$ is statistically significant.

All samples for the second wave seem likely to represent the population more than the stayers' data, so the effectiveness of the corrected model was evaluated by comparing the goodness-of-fit results for the two transferred models to all samples for the second wave with and without correction of attrition biases.

All samples for the second wave are defined as the sum of stayers in the panel and newly recruited samples who for the second wave were used to replace the dropouts from the panel as shown in Figure 1. The goodness-of-fit indicators for the transferred models are shown in Table 8. We can see that the correction of attrition bias improved slightly the accuracy of transferred model judging from its higher Rho-bar squared value, but this does not completely verify the effectiveness of the correction.

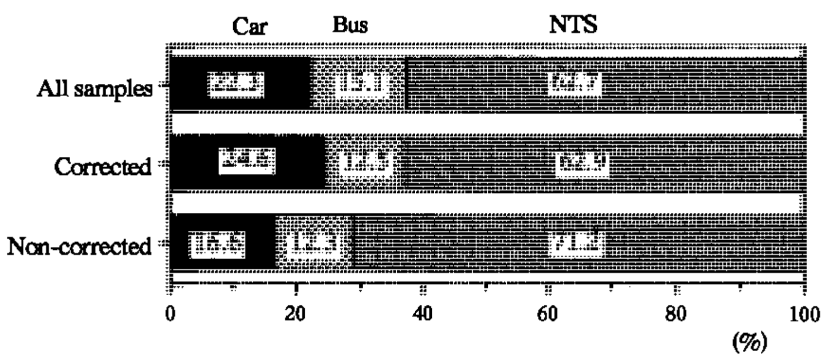
The shares of the three travel modes for all samples at the second wave were estimated with and without correction using the sample enumeration technique which is one of the sample aggregation methods (Koppelman, 1976).

Table 8 Transfer of Stayers' Models to All Respondents

Goodness-of-fit indicators	Non-corrected	Corrected
L(0)	−1859.6	−1859.6
$L(\hat{\beta})$	−1431.0	−1401.9
Rho-bar squared	0.228	0.245
% correct	64.7	64.3
No. of smaples	1693	1693

The estimated shares for car and NTS as shown in Figure 3 account for 24.6% and 62.9% for corrected model. These became closer to the shares of both modes for all samples; 22.3% and 62.7%. Quite notable is that the discrepancy was reduced to 0.2% by the correction of biases, even though the non-corrected mode share of NTS was overestimated by 8.5% (i.e. 71.2−62.7) as compared with that for all samples. This will lead to the conclusion that the correction of attrition biases was effectively done.

Figure 3 Estimated Modal Shares



7. CONCLUSIONS

It was attempted in this study to eliminate sample attrition biases included in the panel data by making use of the correlation between error terms of discrete attrition models and mode choice models. The correction of attrition biases is important in analyzing panel data, so these results are expected to contribute to the further development of dynamic approaches using SP panel data. The results of this study are summarized as follows:

It is demonstrated at the outset that the panel data was biased by sample attrition by comparing the modal shares of all samples with those of stayers' from the first wave and also by comparing mode choice models based on these data sets.

Attrition models which were used to correct the attrition bias were developed by introducing individuals' socio-economic attributes as main explanatory variables. The result of model goodness-of-fit test proved satisfactory,

In addition, a mode choice model was constructed based on the stayers at the second wave. The attrition bias of this model was corrected by sequential steps on the assumption that the error in the mode choice model correlates with that of the attrition model. It was found that this type of correcting could effectively cancel out the biased share of each travel mode.

However, a significant factor, "Supplementary survey", is not in general suitable for inclusion in the explanatory variables. Consequently, there remains some things to be reconsidered concerning the set of variables used to explain the sample attrition more accurately.

There were only 54 stayers in the panel which was used in the analysis. It is important to ameliorate the survey method to avoid such a high rate of sample attrition. It is generally possible to improve the attrition rate up to the level of 30% or less which has been reported in the RP panel surveys. It actually decreased to 31% from the fourth wave (1993) to the fifth wave (1994), both of which were carried out in the same way as the first and second waves in Hiroshima.

REFERENCES

1) Brog, W. and Meyburg, A. "Nonresponse problem in travel surveys: an empirical investigation", *Transportation Research Record* 775, pp. 34-38, 1980.
2) Dubin, J.A. and MacFadden, D.L. "An economic analysis of residential electric appliance holdings and consumption", *Econometrica*, Vol. 52, pp. 345- 362, 1984.
3) Fujiwara, A. and Sugie, Y. "Temporal stability of stated preference data", Presented for the 6th IATBR Conference, Quebec, 1991.
4) Hausman, J. and Wise, D. "Attrition bias in experimental and panel data: The Gary income maintenance experiment", *Econometrica*, Vol. 47, No. 2, pp. 55-473, 1979.
5) Hayashi, C. "Multidimentional quantification with the applications to analysis of social phenomena", *Annals of the Institute of Statistical Mathematics*, Vol. 5, No. 2, pp. 121-143, 1954.
6) Hensher, D., Smith, N., Milthorpe, F. and Barnard, P., *Dimensions of Automobile Demand*-A longitudinal study of household automobile ownership and use, North-Holland, 1992.
7) Kitamura, R. "A panel analysis of household car ownership and mobility", *Proceedings of JSCE*, No. 383/4-7, pp. 13-27, 1987.
8) Kitamura, R. "Panel analysis in transportation planning: An overview", *Transportation Research*, Vol. 24A, No. 6, pp. 401-415, 1990.
9) Kitamura, R. and Bovy, P. "Analysis of attrition biases and trip reporting errors for panel data", *Transportation Research*, Vol. 21A, No. 4/5, pp. 287-302, 1987.
10) Koppelman, F. "Guidelines for aggregate travel prediction using disaggregate choice models", *Transportation Research Record* 610, pp. 19-24, 1976.
11) Mannering, F. and Hensher, D. "Discrete/continuous econometric models and their application to transport analysis", *Transport Reviews*, Vol. 7, No. 3, pp. 227-244, 1987.
12) Manski, C. and Lerman, S. "The estimation of choice probabilities from choice based samples", *Econometrica*, Vol. 45, No. 8, pp. 1977-1988, 1977.
13) Nishii, K. and Kondo, K. "Panel analysis of shopping destination choice behavior in Japan", Paper presented at the First U.S. Conference on Panels for Transportation Planning, Lake Arrowhead, California, 1992.
14) Polak, J. and Meland, S. "Evidence on the temporal stability and inter-temporal properties of stated preference data", *Proceedings of Seminar G*, The 22nd European Transport Forum, pp. 315-325, 1994.
15) Sano, S., *Discrete Choice Analysis-Theory and Application*, Mitsubishi Economics Institute (Japanese), 1990.
16) Sugie, Y. and Fujiwara, A. "Dynamic analysis of stated preference for travel modes using panel data", Presented at 7th IATBR Conference, Valle Navade, Santiago, 1994.

17) Westin, R. and Gillen, D. "Parking location and transit demand", *Journal of Econometrics*, Vol. 8, pp. 75–101, 1978.

18) Winner, R. "Attrition bias in econometric models estimated with panel data", *Journal of Marketing Research*, Vol. 20, pp. 177–186, 1983.