Choice Models Using Matching Data

Sanko, Nobuhiro
Morikawa, Takayuki

(Citation)
The expanding sphere of travel behaviour research : selected papers from the 11th International Conference on Travel Behaviour Research:571-594

(Issue Date)
2009

(Resource Type)
book part

(Version)
Accepted Manuscript

(URL)
https://hdl.handle.net/20.500.14094/90001200

INSERT CHAPTER NUMBER

CHOICE MODELS USING MATCHING DATA

Nobuhiro Sanko, Kobe University, Japan

Takayuki Morikawa, Nagoya University, Japan

ABSTRACT

SP (Stated Preference) data are frequently used for travel behaviour analysis. In terms of the response formats of the data, *choice* has dominated due to the ease of answering and the development of an appropriate modelling technique called discrete choice modelling. Even when other response formats, such as *ranking* and *rating*, are employed, the data are often converted into choice data. Since choice offers an alternative with the highest priority, researchers use the priority information for SP analysis.

One of the tasks of this study is to examine the possibility of using preference indifference information (matching data) rather than priority information. Preference indifference is a condition in which more than one alternative has the same preference level (also considered a boundary where a behaviour changes from one alternative to another) and can contain wealth of information other than priority information.

The aims of this study are to: i) propose a methodology to utilise matching data (i.e., a response format for obtaining reliable matching data) and a corresponding model formulation in the framework of a discrete choice model; and ii) show that the proposed methodology for the SP and RP/SP models has higher estimation efficiency than models using choice data.

Matching data is obtained through a family of double-bounded (DB) response formats, which is relatively common in the CVM (Contingent Valuation Method). Data are collected in the *Keihanshin* (Kyoto – Osaka – Kobe) and *Chukyo* (Nagoya) metropolitan areas in Japan. In both sets of data, two commuting alternatives, auto and transit, are considered.

Estimation efficiency is evaluated based on *t*-statistics and standard errors of estimates. A family of DB formulations brought higher efficiency not only for both the SP and RP/SP models but also for both the *Keihanshin* and *Chukyo* data sets. A discussion on parameter equality between the RP and SP models revealed further insights and identified topics for future research.

1. INTRODUCTION

SP (Stated Preference) data of observations of respondents' preferences under hypothetical conditions have complementary relationships with RP (Revealed Preference) data of observations of their actual behaviours. SP data are widely used in the field of travel behaviour analysis. As the use of SP data spread, the world's researchers and practitioners improved their methodology, but a general consensus has yet to be achieved on either this or the SP response format for surveys, which is an element of SP experiment design. Researchers continue to use a process of trial and error.

While the *ranking* and *rating* response formats were used in the past, these days, the *choice* format is often used. The main reasons for this include: 1) the choice format allowing for simple answering, which ensures high data reliability, and 2) development of a discrete choice model that is appropriate for the analysis of choice data. Actually, the data obtained from response formats other than choice are often converted into choice data for analytical purposes. Since the alternative with the highest priority is selected in the choice format, researchers use the priority information in the SP analysis.

One task of this study is to examine the possibility of using preference indifference information rather than the priority information. Preference indifference is a condition in which more than one alternative has the same preference level (also considered a boundary where a behaviour changes from one alternative to another) and can contain wealth of information other than priority information. To obtain the preference indifference information, the *matching* format is sometimes used. In the matching format, a respondent faces two alternatives, the second of which is missing a value for one attribute. The respondent is asked to fill in the missing value so that the two alternatives are preference indifferent. Matching data is not often used for analyses, however, due to the lack of both data reliability and an appropriate modelling technique. The aims of this study are to: i) propose a methodology for utilising matching data (i.e., a response format for obtaining reliable matching data, or preference indifference information) and a corresponding model formulation in the framework of a discrete choice model; and ii) show that the proposed methodology for the SP and RP/SP models has higher estimation efficiency than models using choice data.

In section 2, SP response formats used in the transport research are revisited briefly. In addition, the formats used in the field of CVM (Contingent Valuation Method) are investigated to gain insights into obtaining matching data. In section 3, the response formats of CVM that were explained in section 2 are applied to transport research and a corresponding model is formulated. In section 4, the data used in this paper are explained, focusing on how to obtain matching information. In section 5, the estimation results are presented and a model using matching data is compared with a model using choice data. In section 6, the estimation results are further discussed from the viewpoints of parameter equality between the RP and SP models. Finally in section 7, the concluding remarks are presented.

2. SP DATA REVISITED

2.1. SP data in the field of transport research

SP data began to be used for transport research in the 1970s and have been widely used since then. The response formats used in transport research include ranking, choice, rating, and matching. A special case of matching is transfer price (TP). In TP response format, a respondent faces two alternatives in which the price of the second alternative has a missing value. The respondent is asked to fill in the price so that the two alternatives are preference indifferent. Response formats are explained in greater detail in some books and papers, including Pearmain *et al*. (1991), Payne *et al*. (1993), Hensher (1994) and Louviere *et al*. (2000).

Matching or TP formats might seem appropriate for obtaining matching data, but the data obtained with these formats are of low reliability. This is one reason why matching and TP formats are not used more frequently.

2.2. SP data in the CVM

The response formats used in the CVM are reviewed in order to find a way to more easily obtain highly reliable matching data. The CVM measures environmental resources in monetary terms. However, there is no market to trade environmental resources, and the only way to measure them is to ask the respondents. In the CVM, respondents are asked their willingness to pay (WTP) or willingness to accept compensation (WTA) when presented with hypothetical environmental improvement or hypothetical environmental deterioration. The monetary value of the environmental resources is then calculated based on their WTP or WTA.

The response formats for the CVM are summarised in Table 1. In the open-ended CVM, respondents are asked to fill in the price freely, and the open-ended CVM is similar to the matching and TP formats of SP. In the bidding game CVM, the price is determined through an auction sale. Respondents are asked to choose between paying and not paying the price written on a card, and the prices are presented from the lower price, for example. The bidding game CVM is similar to the repetition of the SP choice format. In the payment card CVM, researchers present several cards in order of prices (written on the cards) at the same time, and respondents are asked to choose one card from them. The payment card CVM is similar to the

SP ranking format, since the task can be similar to rank the cards shown and an additional virtual card written 'will not pay.' (In the payment card format, the card chosen is a next to the card of 'will not pay.') The dichotomous choice CVM is exactly the same as the SP choice format. The double-bounded dichotomous choice CVM (DB CVM) and one-and-a-half bound CVM (1.5B CVM) are special cases of the repetition of SP choice. In DB, a respondent is first asked if he/she is willing to pay price T. If he/she is willing to pay T, then he/she is asked if he/she is willing to pay price T^U ($>T$). If he/she is not willing to pay T, then he/she is asked if he/she is willing to pay price T^L ($<T$). In 1.5B, the second question is given either to those who are willing to pay price T or to those who are not willing to. Hanemann *et al.* (1991) compared the DB CVM with the dichotomous choice CVM, and concluded that the DB CVM was better from the viewpoints of parameter estimation efficiency, goodness of model fit, and efficiency of willingness to pay.

Table 1 CVM response formats

Response format	Explanation	Characteristics	Similar SP
Open-ended CVM	Fill in the price freely	Many non-responders; appearance of too high or too low prices	Matching, TP
Bidding game CVM	Determine price as in an auction sale	Time consuming; influenced by an initial bid	Repetition of choice
Payment card CVM	Choose a price from those set by a researcher	Influenced by a range of prices	Ranking
Dichotomous choice CVM	Answer 'yes' or 'no' to the price set by a researcher	Easier to answer; less bias	Choice
DB CVM†, 1.5B CVM‡	The 2nd price is set by a researcher based on the response to the 1st dichotomous choice	The price set in the 1st choice influences the 2nd response; inaccurate price set in the 1st choice is covered by the price set in the 2nd choice	Repetition of choice

†: double-bounded dichotomous choice CVM; ‡: one-and-a-half bound CVM
Source: modified from Kuriyama (1998) page 62

3. MATCHING DATA IN TRANSPORT RESEARCH: METHODOLOGY

3.1. Methodology for obtaining matching data from transport research

This subsection proposes a methodology for obtaining matching data from transport research. The CVM response formats considered in subsection 2.2 are the basis of the discussion here. However, the open-ended format is not discussed due to its low reliability. The dichotomous choice format is also not examined because of the difficulty of obtaining matching information without repeating dichotomous choices. The bidding game, payment card, DB, and 1.5B formats are examined. A binary situation comprising auto and transit is examined as an example. The questionnaire asks respondents their intentions to change their current transport mode. (It is possible to ask respondents their intentions to change their transport mode chosen under hypothetical conditions.)

3.1.1. Bidding game format

Those who currently use auto are asked their intentions when, for instance, the travel time by auto increases by 5 minutes, 10 minutes, 15 minutes. The questioning is continued until they choose transit. Matching information exists between the level of service at which they chose transit and the level of service at which they last chose auto. Other attributes can be changed and those who currently use transit can be respondents.

3.1.2. Payment card format

Those who currently use auto are shown some cards at the same time. For example, the respondents are shown five cards listing auto travel times increase ('by 5 minutes,' 'by 10 minutes,' 'by 15 minutes,' 'by 20 minutes,' and 'by 25 minutes'). They are asked to choose a minimum auto travel time change so that they choose transit. Matching information exists between the chosen level of service and the level of service next to the chosen card (in favour of the auto user). Other attributes can be changed and those who currently use transit can be respondents.

3.1.3. Double-bounded (DB) and 1.5 bound (1.5B) formats

The double-bounded format is discussed first. Those who currently use auto are asked to choose between auto and transit when one level of service (here, for example, auto travel time) is changed. a) They are asked to make choices when the travel time by auto becomes longer (1st bound, or 1st B). b) If they continue to choose auto in the 1st bound, they are asked to make choices when travel time by auto becomes much longer. c) If they change to transit in the 1st bound, they are asked to make choices when the travel time by auto is longer than that in the current situation but shorter than that in the 1st bound (b) and c) are called the 2nd bound, or 2nd B). Other attributes can be changed, and those who currently use transit can be respondents. In the 1.5 bound format, only ‘a) and b),’ or ‘a) and c)’ are used. DB or 1.5B are response formats that increase the chance of obtaining matching information (preference indifference information) by repeating reliable choice formats (responses in each bound can be considered choices).

Table 2 lists the response patterns for the DB and 1.5B formats. For the DB format, matching information lies between 1st and 2nd bounds in Nos. 1, 4, 5, and 8, and between the RP and the 2nd bound in Nos. 3 and 7. Matching information cannot be identified or bounded by two responses in Nos. 2 and 6. If matching information must be identified by SP responses, then it cannot be bounded by two responses in Nos. 3 and 7.

In 1.5B, the 2nd bounds of either Nos. 3, 4, 7, and 8 or Nos. 1, 2, 5, and 6 are omitted. Supposing that the 2nd bounds of Nos. 3, 4, 7, and 8 are not obtained, then matching information lies between the 1st and 2nd bounds in Nos. 1 and 5, and between the RP and the 1st bound in Nos. 3, 4, 7, and 8. Matching information cannot be identified or bounded by two responses in Nos. 2 and 6. If matching information must be identified by SP responses, then it cannot be bounded by two responses in Nos. 3, 4, 7, and 8.

In all response formats discussed in subsections 3.1.1–3.1.3, matching information can be identified by being bounded by two responses. Therefore, these types of response formats are generically called a family of double-bounded formats.

Table 2 Response patterns of DB and 1.5B formats

No.	RP	SP (1st Bound)	SP (2nd Bound)
1	Auto (i)	Auto (i)	Transit (j)
2	Auto (i)	Auto (i)	Auto (i)
3	Auto (i)	Transit (j)	Transit (j)
4	Auto (i)	Transit (j)	Auto (i)
5	Transit (j)	Transit (j)	Auto (i)
6	Transit (j)	Transit (j)	Transit (j)
7	Transit (j)	Auto (i)	Auto (i)
8	Transit (j)	Auto (i)	Transit (j)

Note: i and j in Table 2 will be explained in a later part of this paper.

3.2. Methodology of model formulation

In this subsection, data obtained through the response formats discussed in subsection 3.1 (bidding game, payment card, DB, and 1.5B formats) are modelled in the discrete choice modelling framework. The model is developed based on the DB format with two alternatives,

i and j. Data obtained through other formats can be modelled in the same manner, and these models are generically called a family of double-bounded models.

Let U_{in}^{B} be a utility of individual n choosing alternative i in bound B (B: 1*st* or 2*nd*) and assume that U_{in}^{B} can be divided into the systematic component, V_{in}^{B}, and the error component, ε_{in}^{B}.

$$U_{in}^{B} = V_{in}^{B} + \varepsilon_{in}^{B} \tag{1}$$

Suppose also that ε_{in}^{1st} and ε_{in}^{2nd} are the same as ε_{in}^{SP} (ε_{jn}^{1st} and ε_{jn}^{2nd} are the same as ε_{jn}^{SP}), and that $\varepsilon_{jn}^{SP} - \varepsilon_{in}^{SP}$ is standard normally distributed. The model is formulated as follows using data from the 1st and 2nd bounds: (Equations (2a) and (2b) are described in greater detail, and (2c) and (2d) are derived similarly.)

Nos. 1 and 8 in Table 2

$$\begin{aligned}
prob\left(d_n^{1st} = i, d_n^{2nd} = j\right) &= prob\left(U_{in}^{1st} > U_{jn}^{1st} \text{ and } U_{in}^{2nd} < U_{jn}^{2nd}\right) \\
&= prob\left(V_{in}^{1st} - V_{jn}^{1st} > \varepsilon_{jn}^{SP} - \varepsilon_{in}^{SP} \text{ and } V_{in}^{2nd} - V_{jn}^{2nd} < \varepsilon_{jn}^{SP} - \varepsilon_{in}^{SP}\right) \\
&= prob\left(V_{in}^{1st} - V_{jn}^{1st} > \varepsilon_{jn}^{SP} - \varepsilon_{in}^{SP} > V_{in}^{2nd} - V_{jn}^{2nd}\right) \\
&= \Phi\left(V_{in}^{1st} - V_{jn}^{1st}\right) - \Phi\left(V_{in}^{2nd} - V_{jn}^{2nd}\right)
\end{aligned} \tag{2a}$$

Nos. 2 and 7 in Table 2

$$\begin{aligned}
prob\left(d_n^{1st} = i, d_n^{2nd} = i\right) &= prob\left(U_{in}^{1st} > U_{jn}^{1st} \text{ and } U_{in}^{2nd} > U_{jn}^{2nd}\right) \\
&= prob\left(U_{in}^{1st} > U_{jn}^{1st} \middle| U_{in}^{2nd} > U_{jn}^{2nd}\right) prob\left(U_{in}^{2nd} > U_{jn}^{2nd}\right) \\
&= prob\left(U_{in}^{2nd} > U_{jn}^{2nd}\right) \text{ (since } prob\left(U_{in}^{1st} > U_{jn}^{1st} \middle| U_{in}^{2nd} > U_{jn}^{2nd}\right) = 1\text{)} \\
&= prob\left(V_{in}^{2nd} - V_{jn}^{2nd} > \varepsilon_{jn}^{SP} - \varepsilon_{in}^{SP}\right) \\
&= \Phi\left(V_{in}^{2nd} - V_{jn}^{2nd}\right)
\end{aligned} \tag{2b}$$

Nos. 3 and 6 in Table 2

$$prob\left(d_n^{1st} = j, d_n^{2nd} = j\right) = 1 - \Phi\left(V_{in}^{2nd} - V_{jn}^{2nd}\right) \tag{2c}$$

Nos. 4 and 5 in Table 2

$$prob\left(d_n^{1st} = j, d_n^{2nd} = i\right) = \Phi\left(V_{in}^{2nd} - V_{jn}^{2nd}\right) - \Phi\left(V_{in}^{1st} - V_{jn}^{1st}\right) \tag{2d}$$

where $\Phi(\bullet)$: standardised cumulative normal distribution, d_n^{B}: response in bound B of individual n.

The formulation above uses matching information obtained through the responses of the 1st and 2nd bounds. If matching information is obtained not only through the responses of the 1st and 2nd bounds but also through the RP responses, then another assumption is required. The assumption is that $\varepsilon_{in}'^{RP}$ is the same as ε_{in}^{SP} ($\varepsilon_{jn}'^{RP}$ is the same as ε_{jn}^{SP}). ($\varepsilon_{\bullet n}'^{RP}$ is an error component of the SP model, where all attributes' levels are exactly the same as those of the RP model.) It may be difficult to accept this assumption that the RP and SP error components are the same. If, however, we assume that the error component of SP models, where all attributes' levels are exactly the same as those of RP models, is the same as the error component of 1st and 2nd bound SP's, then this assumption can be justified. Here $\varepsilon_{\bullet n}'^{RP}$ is used instead of $\varepsilon_{\bullet n}^{RP}$, since the authors consider not the error component of the RP model but that of the SP model, where all attributes' levels are the same as those of the RP model. In sections 5 and 6, the RP model uses the data in the RP column in Table 2. The SP model uses the data in the SP (1st Bound) and SP (2nd Bound) columns when $\varepsilon_{\bullet n}^{\prime RP}$ and $\varepsilon_{\bullet n}^{SP}$ are assumed to be different; the SP model uses the data in the RP, SP (1st Bound), and SP (2nd Bound) columns when $\varepsilon_{\bullet n}^{\prime RP}$ and $\varepsilon_{\bullet n}^{SP}$ are assumed to be the same.

4. DATA

4.1. *Keihanshin* (Kyoto – Osaka – Kobe) data

The data in the *Keihanshin* metropolitan area were from a supplementary survey of a person trip survey (household travel diary survey) taken in 2000. Besides the RP mode choice for commuting, the SP survey is based on the 1.5 bound format in a commuting situation in which both auto and transit are available. An example of a 1.5 bound question is shown in Figure 1. Only those who do not change their commuting mode in the 1st bound participate in the 2nd bound.

Question 1
If the cost of parking near your working place were to increase, would you commute by bus or rail? (For those who do not pay for parking, assume that you must pay the fee shown below.)
(I) If the cost of parking increased by 1,000 JPY per month,
 1. you would commute by bus or rail → go to **Question 2**
 2. you would commute by auto → go to **Question 1** (II)
(II) If the cost of parking increased by 3,000 JPY per month,
 1. you would commute by bus or rail
 2. you would commute by auto

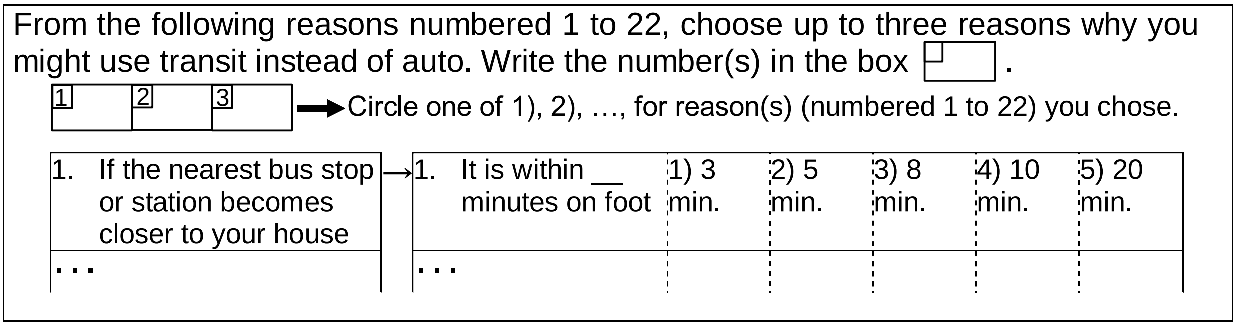
Note: This is a question for those who currently commute by auto.

Figure 1 An example of 1.5 bound SP in the *Keihanshin* survey

4.2. *Chukyo* (Nagoya) data

The data in the *Chukyo* metropolitan area were collected in cooperation with a small-scale person trip survey (household travel diary survey) in 1997. In addition to the RP mode choice for commuting, the SP survey is based on a payment card format for commuting situations in which auto and transit are available. An example of a payment card SP is shown in Figure 2.

Those who commute by auto but have an intention to use the bus or rail instead of auto are asked for their reasons (up to three reasons, for example, 'if the nearest bus stop or station becomes closer to your house') and the necessary change in the level of service (for example, choose one from within '3 min.,' '5 min.,' '8 min.,' '10 min.,' and '20 min.' on foot from your house) so that they choose the bus or rail. Note that those who currently commute by bus or rail do not have SP questions.



From the following reasons numbered 1 to 22, choose up to three reasons why you might use transit instead of auto. Write the number(s) in the box ☐ .

☐1 ☐2 ☐3 → Circle one of 1), 2), …, for reason(s) (numbered 1 to 22) you chose.

1. If the nearest bus stop or station becomes closer to your house	→	1. It is within __ minutes on foot	1) 3 min.	2) 5 min.	3) 8 min.	4) 10 min.	5) 20 min.
. . .		. . .					

Note: This is a question for those who commute by auto and who answered "yes" to the question, 'if the level of public transit service increased or the level of auto service decreased, would you stop commuting by auto and start commuting by public transit?'

Figure 2 An example of a payment card SP in the *Chukyo* survey

5. ESTIMATES[1]

5.1. *Keihanshin* data

Estimates from the *Keihanshin* data are shown in columns RP and i) – iii) of Table 3. For the RP model, the binary probit model is adopted. The details of the RP/SP model and the RP/SP model with serial correlation can be found in the appendices.

The estimates from the SP models are: i) Only 1st bound responses are modelled as choice data using binary probit (Table 3 column i)); ii) The 1st and 2nd bound responses are modelled as independent choice data using binary probit (Table 3 column ii)); and iii) The 1st and 2nd bound responses are modelled as 1.5B data using the formulation discussed in subsection 3.2 (Table 3 column iii)).

Accordingly, it is possible to compare the analyses of the 1st and 2nd bound SP's as 1.5B (Table 3 column iii)) and as independent choices (Table 3 column ii)). 1st bound SP model (Table 3 column i)) is also modelled for comparison purposes, since assuming that 1st and 2nd bound SP's are independent can be questionable.

---

[1] In the RP/SP model with serial correlation estimated in sections 5 and 6, a standard normal distribution is assumed for λ (see appendix A 2). A standard normal distribution is assumed for $v_{jn}^{SP} - v_{in}^{SP}$ (see appendix A 2).

Table 3 Estimates from the *Keihanshin* data

		i) 1st Bound Choice			ii) 1st and 2nd Bound Choice			iii) 1.5B			iv) 1.5B difference		
Variable name†	RP	SP	RP/SP	RP/SPsc‡	SP	RP/SP	RP/SPsc‡	SP	RP/SP	RP/SPsc‡	SP	RP/SP	RP/SPsc‡
RP constant (T)	1.04 (3.85)	--	0.984 (3.70)	1.42 (3.91)	--	0.914 (3.62)	1.23 (3.77)	--	0.305 (1.65)	0.755 (2.97)	--	0.836 (3.26)	1.21 (3.36)
SP constant (T)	--	-1.46 (-6.77)	-7.59 (-1.91)	-1.04 (-1.36)	-1.18 (-8.25)	-4.34 (-3.39)	-0.184 (-0.43)	-1.09 (-6.52)	-0.961 (-4.08)	0.624 (2.18)	-1.54 (-8.16)	-4.63 (-3.62)	-0.667 (-1.24)
Scale parameter μ	--	--	0.147 (2.29)	0.468 (4.03)	--	0.206 (4.34)	0.545 (5.58)	--	0.592 (8.21)	1.86 (7.16)	--	0.242 (4.58)	0.617 (5.63)
Inertia (T)	--	2.79 (15.30)	18.0 (2.14)	5.53 (3.59)	2.36 (20.44)	11.0 (4.03)	3.30 (4.40)	1.79 (12.98)	2.73 (6.80)	0.796 (2.99)	3.14 (16.53)	12.5 (4.10)	4.63 (4.44)
Travel time [hr]	-2.05 *0.304* (-6.75)	-0.323 *0.203* (-1.59)	-2.04 *0.299* (-6.82)	-2.82 *0.423* (-6.68)	-0.252 *0.135* (-1.87)	-1.91 *0.292* (-6.53)	-2.42 *0.381* (-6.37)	-0.757 *0.122* (-6.20)	-1.55 *0.221* (-7.02)	-2.11 *0.285* (-7.41)	-0.465 *0.166* (-2.80)	-2.04 *0.291* (-7.02)	-2.99 *0.413* (-7.24)
Cost (out of pocket) [1,000JPY]	-1.27 *0.240* (-5.30)	-0.180 *0.147* (-1.23)	-1.27 *0.235* (-5.40)	-1.84 *0.316* (-5.83)	-0.425 *0.0977* (-4.35)	-1.43 *0.222* (-6.43)	-2.19 *0.295* (-7.44)	-1.22 *0.0899* (-13.59)	-1.85 *0.201* (-9.22)	-2.84 *0.242* (-11.73)	-0.492 *0.113* (-4.35)	-1.45 *0.223* (-6.52)	-1.97 *0.313* (-6.30)
Civil servant (T)	0.652 (2.01)	0.304 (1.48)	0.718 (2.27)	0.862 (1.97)	0.178 (1.17)	0.676 (2.27)	0.769 (1.84)	0.232 (1.32)	0.492 (2.25)	0.579 (1.50)	0.105 (0.58)	0.607 (2.04)	0.731 (1.71)
Seat (T)*	0.354 (1.90)	0.147 (1.20)	0.382 (2.09)	0.515 (2.12)	0.0091 (0.11)	0.294 (1.73)	0.402 (1.96)	0.166 (2.17)	0.297 (2.70)	0.278 (2.39)	0.0710 (0.69)	0.344 (2.01)	0.539 (2.20)
Under the age of 30 (T)	0.482 (1.78)	0.391 (2.28)	0.589 (2.22)	0.746 (2.04)	0.163 (1.30)	0.526 (2.15)	0.621 (1.87)	0.324 (2.19)	0.550 (3.02)	0.692 (2.20)	0.151 (0.99)	0.507 (2.05)	0.608 (1.76)
Transfers more than twice (T)	-2.22 (-2.02)	-0.0083 (-0.02)	-2.08 (-1.98)	-2.28 (-1.65)	-0.382 (-1.07)	-2.18 (-2.37)	-2.36 (-2.19)	-0.450 (-1.12)	-1.31 (-2.31)	-2.08 (-2.21)	-0.715 (-1.65)	-2.55 (-2.72)	-3.44 (-2.70)
Car with free use (A)**	1.51 (6.92)	-0.170 (-0.95)	1.43 (6.46)	2.02 (6.59)	-0.0645 (-0.54)	1.30 (5.94)	1.73 (5.87)	0.0303 (0.21)	0.888 (5.01)	1.58 (5.95)	-0.125 (-0.86)	1.30 (5.93)	1.83 (6.10)
Free parking (A)***	0.347 (1.75)	0.233 (1.89)	0.410 (2.12)	0.640 (2.69)	0.220 (2.59)	0.464 (2.61)	0.739 (3.90)	0.0974 (0.96)	0.204 (1.62)	0.358 (3.59)	0.102 (0.86)	0.331 (1.81)	0.402 (1.54)
D Travel time [hr]	--	--	--	--	--	--	--	--	--	--	-1.97 (-6.68)	-8.39 (-4.06)	-5.22 (-5.18)
D Cost (out of pocket) [1,000JPY]	--	--	--	--	--	--	--	--	--	--	-4.83 (-14.36)	-19.4 (-4.57)	-11.4 (-6.27)
D Seat (T)*	--	--	--	--	--	--	--	--	--	--	0.284 (1.75)	1.15 (1.64)	0.668 (1.82)
D Free parking (A)***	--	--	--	--	--	--	--	--	--	--	0.488 (3.44)	1.87 (2.85)	0.936 (3.05)
Scale parameter θ (T)	--	--	--	2.72 (5.25)	--	--	1.93 (8.09)	--	--	1.77 (11.61)	--	--	2.29 (7.28)
N	326	960	1,286	326	1,737	2,063	326	948	1,274	326	948	1,274	326
Initial log-likelihood	-225.97	-665.42	-891.39	-891.39	-1,204.00	-1,429.96	-1,429.96	--	--	--	--	--	--
Final log-likelihood	-124.07	-268.29	-398.50	-376.65	-587.19	-721.37	-687.80	-684.24	-826.48	-711.61	-550.52	-682.43	-647.51
Adjusted ρ^2	0.411	0.582	0.539	0.563	0.504	0.487	0.510	--	--	--	--	--	--

Note: *T*-statistics are in parentheses. Standard errors are in italics. Standard errors other than travel time and cost are omitted. Column iv) is mentioned in section 6.
†: T and A in parentheses are alternative specific to transit and auto, respectively. Variable names without such indications are generic. “D” before a variable name indicates a difference term. sc‡: RP/SP models with serial correlation. *: Seat always or sometimes available. **: Having a car that is free to use. ***: Free parking near your working place.

Concerning iii), the 1.5B model, when all values of the explanatory variables are the same in the 1st and 2nd bounds (all values of explanatory variables may be the same even when the attributes' levels are not the same), the 1.5B model cannot be estimated due to the zero probability in equation (2). This is why some portions of the data used in models i) and ii) are not included in model iii). Additionally, in the 1.5B model, only matching information bounded by 1st and 2nd bound SP's is used, since the authors assume that the error component of the SP's, where all attributes' levels are exactly the same as those of the RP, and the error components of the 1st and 2nd bound SP's are not always the same.

An estimate from the RP model (binary probit model) is shown in the first column of Table 3. Travel time and cost estimates are significant.

Estimates from the SP models were also investigated. For comparison purposes, the same set of explanatory variables is employed in models i) – iii). Socio-economic characteristics such as "civil servant," "under the age of 30," and "car with free use," have the same values in both the RP and the SP conditions. One of the variables for level of service, "transfers more than twice," also has the same values in both the RP and the SP conditions. The investigation is focused more on travel time, cost, seat, and free parking, for which the attribute levels in the SP can differ from those in the RP.

In the 1st bound choice model (Table 3 column i)), the travel time and cost estimates are not significant. *T*-statistics indicate that seat dummy and free parking dummy estimates are less significant and a little bit more significant, respectively, compared to those for the RP model. In the 1st and 2nd bound choice model (Table 3 column ii)) the cost estimate is significant, but the travel time estimate is insignificant. *T*-statistics indicate that, compared to the RP model, the seat dummy estimate is less significant while the free parking dummy estimate is significant.

In the 1.5B SP model (Table 3 column iii)), the travel time and cost estimates are significant. *T*-statistics indicate that the seat dummy estimate is significant while the free parking dummy estimate is less significant compared to RP model. Comparing the three SP models, the *t*-statistics for travel time and cost are best in the 1.5B model (excluding the relatively less significant dummy variables). Standard errors for travel time and cost are smallest in the 1.5B model, suggesting the highest estimation efficiency. The constant and inertia terms are the closest to zero in the 1.5B model, suggesting that the other variables provide a better explanation.

Estimates from the RP/SP models were also examined. Four explanatory variables, the attribute levels of which can differ in the RP and SP conditions, are discussed here. In the 1st bound RP/SP model (Table 3 column i)), the *t*-statistics for the four estimates are slightly better than those in the RP model, but in reality, the *t*-statistics are almost the same as those in the RP model. In the 1st and 2nd bound RP/SP model (Table 3 column ii)), the *t*-statistics for cost and free parking dummy are better than they are in the RP model, but the *t*-statistics for travel time and seat dummy are worse.

In the 1.5B RP/SP model (Table 3 column iii)), the *t*-statistic for cost is greatly improved over that in the RP model while the *t*-statistics of travel time and seat dummy are improved. However, the *t*-statistic for free parking is worse than that in the RP model. Comparing the three models, the *t*-statistics for travel time and cost are best in the 1.5B model (excluding the relatively less significant dummy variables). Standard errors for travel time and cost are smallest in the 1.5B model, suggesting the highest estimation efficiency. The constant and inertia terms are closest to zero in the 1.5B model, suggesting that the other variables provide a better explanation. The scale parameter estimate is closest to unity in the 1.5B model, suggesting that the RP and 1.5B SP models differ little.

In addition, the RP/SP models with serial correlation were estimated (Table 3 columns i) – iii)). Some of the advantages of the RP/SP model with serial correlation over RP/SP model without serial correlation are summarised below:

- The SP constant and inertia terms are closer to zero, suggesting that the other variables provide a better explanation.
- The scale parameter is larger, suggesting that the variance of error component in the SP model is smaller. Particularly for the 1.5 bound model, the variance of error component is smaller in the SP than that in the RP, suggesting that the SP data are more reliable.

Parameter equality between the RP and SP models was also statistically tested. The results are summarised in Table 4. In the 1st bound choice model, parameter equality is not rejected at a 5% level of significance. On the other hand, both in the 1st and 2nd bound choice model and in the 1.5B model, parameter equality is rejected at a 1% level of significance. This means the assumption that the parameters for the RP and SP models are the same is not justified statistically.

Table 4 χ^2 test of the parameter equality (*Keihanshin*)

Model	χ^2 value
1st bound choice model (Table 3 column i))	12.27
1st and 2nd bound choice model (Table 3 column ii))	20.22
1.5 bound model (Table 3 column iii))	36.33

Note: 7 degrees of freedom, $\chi^2_7(.05)=14.07$, $\chi^2_7(.01)=18.48$.

5.2. *Chukyo* data

Estimates from the *Chukyo* data are shown in columns RP and i) – iii) of Table 5. For the RP model, the binary probit model is adopted. Details of the RP/SP model and the RP/SP model

Table 5 Estimates from the *Chukyo* data

		i) 1st Bound Choice			ii) 1st and 2nd Bound Choice			iii) Payment card			iv) Payment card difference		
Variable name†	RP	SP	RP/SP	RP/SPsc‡	SP	RP/SP	RP/SPsc‡	SP	RP/SP	RP/SPsc‡	SP	RP/SP	RP/SPsc‡
RP constant (A)	-0.403	--	--	--	--	-0.401	-0.564	--	-0.286	-0.406	--	-0.424	-0.601
	(-3.85)					(-3.85)	(-3.86)		(-3.11)	(-3.12)		(-4.16)	(-4.20)
SP constant (A)	--	--	-0.553	-0.826	-0.155	-1.03	-1.51	-1.39	-0.954	-1.41	0.643	1.35	1.21
			(-6.70)	(-6.82)	(-1.29)	(-2.93)	(-3.07)	(-13.01)	(-9.10)	(-9.41)	(3.65)	(2.08)	(2.05)
Scale parameter μ	--	--	--	--	--	0.184	0.137	--	1.45	3.20	--	0.513	0.436
						(2.30)	(2.35)		(9.29)	(7.05)		(3.97)	(5.09)
Travel time [hr]	-1.78	--	-1.27	-2.15	-0.393	-1.79	-2.52	-1.98	-1.44	-1.77	-0.740	-1.72	-2.45
	0.209		*0.144*	*0.208*	*0.179*	*0.207*	*0.290*	*0.135*	*0.158*	*0.215*	*0.204*	*0.221*	*0.287*
	(-8.49)		(-8.81)	(-10.33)	(-2.20)	(-8.63)	(-8.67)	(-14.67)	(-9.15)	(-8.20)	(-3.64)	(-7.76)	(-8.53)
Cost [1,000JPY]	-0.292	--	0.134	0.0417	0.134	-0.281	-0.398	-1.40	-0.754	-1.42	-0.240	-0.300	-0.432
	0.144		*0.128*	*0.187*	*0.259*	*0.143*	*0.201*	*0.160*	*0.0892*	*0.142*	*0.298*	*0.141*	*0.196*
	(-2.03)		(1.05)	(0.22)	(0.52)	(-1.97)	(-1.98)	(-8.74)	(-8.45)	(-10.01)	(-0.80)	(-2.12)	(-2.21)
Head [hr] (T)	-1.36	--	-0.903	-1.50	-0.268	-1.36	-1.90	-1.26	-0.975	-1.17	-0.979	-1.47	-2.06
	0.289		*0.198*	*0.290*	*0.271*	*0.284*	*0.397*	*0.194*	*0.154*	*0.188*	*0.310*	*0.261*	*0.376*
	(-4.72)		(-4.57)	(-5.17)	(-0.99)	(-4.78)	(-4.79)	(-6.49)	(-6.32)	(-6.21)	(-3.16)	(-5.64)	(-5.48)
D Travel time [hr]	--	--	--	--	--	--	--	--	--	--	-5.13	-10.0	-12.5
											(-16.71)	(-4.12)	(-5.81)
D Cost [1,000JPY]	--	--	--	--	--	--	--	--	--	--	-12.3	-24.0	-29.7
											(-11.93)	(-3.80)	(-5.15)
D Head [hr] (T)	--	--	--	--	--	--	--	--	--	--	-4.12	-7.78	-9.47
											(-9.02)	(-3.95)	(-5.01)
Scale parameter θ (A)	--	--	--	--	--	--	0.254	--	--	0.896	--	--	0.813
							(0.51)			(9.42)			(2.45)
N	679	--	887	679	416	1,095	679	208	887	679	208	887	679
Initial log-likelihood	-470.65	--	-614.82	-614.82	-288.35	-759.00	-759.00	--	--	--	--	--	--
Final log-likelihood	-335.79	--	-548.65	-557.04	-285.27	-621.34	-623.10	-684.33	-1,029.00	-936.03	-420.60	-756.95	-757.81
Adjusted ρ^2	0.278	--	0.101	0.0875	-0.00320	0.173	0.170	--	--	--	--	--	--

Note: *T*-statistics are in parentheses. Standard errors are in italics. Standard errors other than travel time, cost, and head are omitted. Column iv) is mentioned in section 6.

†: T and A in parentheses are alternative specific to transit and auto, respectively. Variable names without such an indication are generic. “D” before a variable name indicates a difference term. sc‡: RP/SP models with serial correlation.

with serial correlation can be found in the appendices. In this section, for the purposes of explanation, the card chosen is called 1st bound and the next card chosen (in the auto user's favour) is called 2nd bound.

The estimates from the SP models are: i) Only 1st bound responses are modelled as choice data using binary probit (Table 5 column i)); ii) The responses of the 1st and 2nd bounds are modelled as independent choice data using binary probit (Table 5 column ii)); and iii) The responses of the 1st and 2nd bounds are modelled as payment card data using the formulation discussed in subsection 3.2 (Table 5 column iii)).

Accordingly, it is possible to compare the analyses of the 1st and 2nd bound SP's as payment card (Table 5 column iii)) and as independent choices (Table 5 column ii)). The 1st bound SP model (Table 5 column i)) is also modelled for comparison purposes, since assuming that the 1st and 2nd bound SP's are independent can be questionable.

Concerning iii), the payment card model, when all values of the explanatory variables are the same in the 1st and 2nd bounds, the payment card model cannot be used for estimations due to the zero probability in equation (2). This is why the data excluded in model iii) are also excluded in models i) and ii). Moreover, the payment card model uses matching information bounded not only by 1st and 2nd bound SP's but also by the RP and the SP. For example, in Figure 2, if a respondent chooses reason number 1 and '5) 20 min.,' then the matching information, which lies between '20 min.' and the current access time (usually greater than 20 min.), is used. If it is equal or greater than the current access time, or if this causes zero probability, then the data are excluded from the analysis. In the *Chukyo* payment card data, levels of services are listed in rows, and the authors assume that the respondents have a current level of service explicitly in their mind when choosing one of the listed level of services. In other words, the authors assume that the error component of the SP, where all attributes' levels are exactly the same as those of the RP, and the error components of the 1st and 2nd bound SP's are the same.

An estimate from the RP model (binary probit model) is shown in the first column of Table 5. Estimates for travel time, cost, and head are significant.

Estimates from the SP models were also investigated. For comparison purposes, the same set of explanatory variables is employed in models i) – iii). In the 1st bound choice model (Table 5 column i)), the SP model is not estimated since, in *Chukyo* survey, only current auto users were respondents. That is, all SP responses in the 1st bound are transit. In the 1st and 2nd bound choice model (Table 5 column ii)), the *t*-statistics for three of the level of service variables are less significant than those in the RP model. The cost estimate has a positive sign.

In the payment card SP model (Table 5 column iii)), the *t*-statistics of three of the level of service variables are better than those in the RP model. Compared with those of the 1st and 2nd bound choice model (Table 5 column ii)), the *t*-statistics of three of the estimates are better in the payment card model (Table 5 column iii)). Standard errors for three of the estimates are smaller in the payment card model, suggesting higher estimation efficiency. The inertia term is not included in the model because all of the respondents currently use auto.

Estimates for the RP/SP models were also examined. In the 1st bound RP/SP model (Table 5 column i)), the *t*-statistics for travel time and head are almost the same as those in the RP model. However, the cost estimate has a positive sign. Note that the scale parameter cannot be estimated and only one constant (the same constant term in the RP and SP models) is estimated[2]. In the 1st and 2nd bound RP/SP model (Table 5 column ii)), the *t*-statistics of the three of the level of service estimates are almost the same as those in the RP model. The scale parameter indicates that the variance of error component is larger in the SP model.

In the payment card RP/SP model (Table 5 column iii)), the *t*-statistics of three of the level of service variables are better than those in the RP model. The scale parameter indicates a smaller variance of error component in the SP model. Comparing the three models, the *t*-statistics of three of the level of service variables are best in the payment card model. The standard errors for these three variables are smaller than in the 1st and 2nd bound choice model, in which the sign for cost is correctly estimated, suggesting better estimation efficiency in the payment card model.

The RP/SP models with serial correlation were also estimated (Table 5 columns i) – iii)). The scale parameter in the payment card model is larger, suggesting that the SP is more reliable.

Parameter equality between the RP and SP models was tested statistically. The results are summarised in Table 6. In the 1st and 2nd bound choice model, parameter equality is not

Table 6 χ^2 test of the parameter equality (*Chukyo*)

Model	χ^2 value
1st bound choice model (Table 5 column i))	--
1st and 2nd bound choice model (Table 5 column ii))	0.57
Payment card model (Table 5 column iii))	17.76

Note: 2 degrees of freedom, $\chi^2_2(.01)=9.21$, $\chi^2_2(.50)=1.39$.

---

[2] 1) Since the SP constant includes an SP-specific bias, different constants are estimated in the RP and SP models. 2) The scale parameter μ is introduced in order to share parameters between the RP and SP models after adjusting the variances of error components in the two models. 3) The scale parameter θ is introduced, since there is no guarantee that the scale of the systematic component of error terms is the same in both the RP and SP models (discussed in the RP/SP models with serial correlation). However, none of the above is considered in this model. Since in the *Chukyo* data, the 1st bound choice result is transit only, and the SP-specific constant and scale parameters are not reasonably estimated. For comparison purposes only, a simplified model is estimated here.

rejected at a 50% level of significance. In the payment card model, however, parameter equality is rejected at a 1% level of significance. This means the assumption that parameters of RP and SP models are the same is not justified statistically.

6. PARAMETER EQUALITY BETWEEN THE RP AND SP MODELS

6.1. Parameter equality between the RP and SP models, and a proposal

In the previous section, a DB model family offered the highest estimation efficiency not only for both the SP and RP/SP models but also for both the *Keihanshin* and *Chukyo* data. However, in the DB model family, parameter equality between the RP and SP models is not justified statistically. Two interpretations of this are explored in this subsection.

In the first interpretation, each individual is assumed to follow the same behavioural norm in both the RP and SP models, and a rejection of the parameter equality can be caused by SP response bias. Since only one level of service has changed from the RP condition, exaggerated responses can be observed and can include a bias.

In the second and a little bit more interesting interpretation, each individual is assumed to have a different behavioural norm in the RP and SP models. Parameters obtained from the RP model describe the behaviour of the whole sample and express average relationships among the variables. However, a specific individual has a chance to change his/her behavioural norm in accordance with a change in the level of service, especially when only one level of service is changed. In other words, when an individual faces a situation where only one level of service is changed, he/she can be forced to restructure his/her preference, leading to a change in preference. In marketing science, Mizuno and Katahira (2003) pointed out that a consumer can be forced to reconstruct his/her preference, which can be a cause of a preference change, when the product space expands thanks to a technological innovation.

Parameter equality is evaluated by statistical testing and the equality is rejected in a double-bounded model family. However, many studies justify the equality when more than one level of service is changed from the current situation. In our data, where only one level of service is changed from the current situation, the equality is justified in the 1st bound choice model in the *Keihanshin* data and in the 1st and 2nd bound choice model in the *Chukyo* data. The justification for parameter equality can depend on the response formats and/or model formulation.

In any case, parameter equality is rejected in the double-bounded model family and the problem remains. To solve this problem, the authors present the following proposal.

The formulation of the traditional RP/SP model is shown in equations (3a) – (3c).

RP model

$$U_{in}^{RP} = \boldsymbol{\beta}'\mathbf{x}_{in}^{RP} + \boldsymbol{\alpha}'\mathbf{w}_{in}^{RP} + \varepsilon_{in}^{RP} \tag{3a}$$

SP model

$$U_{in}^{SP} = \boldsymbol{\beta}'\mathbf{x}_{in}^{SP} + \boldsymbol{\gamma}'\mathbf{z}_{in}^{SP} + \varepsilon_{in}^{SP} \tag{3b}$$

Ratio of variances between the RP and SP error components

$$Var\left(\varepsilon_{in}^{RP}\right) = \mu^2 Var\left(\varepsilon_{in}^{SP}\right), \qquad \forall i, n \tag{3c}$$

where

U_{in}^{M} : total utility of individual *n* choosing alternative *i* in *M* model
ε_{in}^{M} : error component of total utility U_{in}^{M}
$\mathbf{x}_{in}^{M}$, $\mathbf{w}_{in}^{M}$, $\mathbf{z}_{in}^{M}$: explanatory variable vectors of deterministic utility of individual *n* choosing alternative *i* in *M* model
$\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$: unknown parameter vectors to be estimated
M : type of model
μ : scale parameter explaining the differences of variances between the RP and SP error components

By introducing scale parameter, parameter $\boldsymbol{\beta}$ can be shared between the RP and SP models. On the other hand, the model proposed here separates the explanatory variables for the SP into an "RP part" and a "difference from the RP part." Here, equation (3b) is replaced by (3d).[3]

SP model

$$U_{in}^{SP} = \boldsymbol{\beta}'\mathbf{x}_{in}^{RP} + \tilde{\boldsymbol{\beta}}'\left(\mathbf{x}_{in}^{SP} - \mathbf{x}_{in}^{RP}\right) + \boldsymbol{\gamma}'\mathbf{z}_{in}^{SP} + \varepsilon_{in}^{SP} \tag{3d}$$

In the SP model, the SP attributes are divided into an "RP part" ($\mathbf{x}_{in}^{RP}$) and a "difference from the RP part" ($\mathbf{x}_{in}^{SP} - \mathbf{x}_{in}^{RP}$), and parameters $\boldsymbol{\beta}$ and $\tilde{\boldsymbol{\beta}}$ are set, respectively. The parameter vector

---

[3] Instead of (3d), the following formulation is available. The same estimate of $\boldsymbol{\beta}$ as in equation (3d) is obtained, but the different estimate of $\tilde{\boldsymbol{\beta}}$ from equation (3d) is obtained.

$$U_{in}^{SP} = \boldsymbol{\beta}'\mathbf{x}_{in}^{SP} + \tilde{\boldsymbol{\beta}}'\left(\mathbf{x}_{in}^{SP} - \mathbf{x}_{in}^{RP}\right) + \boldsymbol{\gamma}'\mathbf{z}_{in}^{SP} + \varepsilon_{in}^{SP}$$

for the “RP part,” that is, $\boldsymbol{\beta}$, is shared between the two models. The parameter vector for the “difference part,” ($\mathbf{x}_{in}^{SP} - \mathbf{x}_{in}^{RP}$), that is, $\tilde{\boldsymbol{\beta}}$, is interpreted as a respondent’s contingent preference when facing the SP question in the first interpretation, and as a respondent’s preference change when facing the SP question in the second interpretation. The parameter vector for the “RP part” ($\mathbf{x}_{in}^{RP}$), that is, $\boldsymbol{\beta}$, is interpreted as a respondent’s core preference in the first interpretation and as an average preference of the whole sample in the second interpretation.

When the proposed model is used for forecasting, parameter vector $\boldsymbol{\beta}$, which explains the core preference, must be used in the first interpretation. In the second interpretation, only parameter vector $\boldsymbol{\beta}$, which explains the average preference of the whole sample, can be used. Preference change can occur when a level of service is changed, but not all respondents can make this preference change. However, when all respondents make a preference change, parameter vectors $\boldsymbol{\beta}$ and $\tilde{\boldsymbol{\beta}}$ must be used. This must be concluded through empirical analysis using, for example, panel data.

6.2. *Keihanshin* data

The estimates from the *Keihanshin* data are shown in Table 3 column iv). The chi-squared value is 15.70 and the parameter equality is not rejected at a 2.5% level of significance, suggesting that parameter equality is justified statistically.

An estimate for the SP model was also examined. The estimated parameters for the “difference from the RP part” are generally larger (about 4 to 10 times the size of the “RP part”) and more significant than the “RP part.” The *t*-statistics of the “RP part’s” cost and time parameters are better compared to the 1st bound choice model (Table 3 column i)) and 1st and 2nd bound choice model (Table 3 column ii)). Compared to the 1st bound choice model (Table 3 column i)), in which the parameter equality is justified, the standard errors of “RP part” are smaller in the time and cost variables, suggesting higher estimation efficiency in the proposed model.

An estimate for the RP/SP model was also examined. Compared with Table 3 columns i) and ii), the *t*-statistics for the time and cost estimates in the “RP part” are improved. Compared with the 1st bound RP/SP model (Table 3 column i)), in which the parameter equality is justified, standard errors of the “RP part” are smaller in the time and cost estimates, suggesting that the proposed model has higher estimation efficiency.

An estimate of the RP/SP model with serial correlation was also examined. Compared with the RP/SP model in Table 3 column iv), the SP constant and inertia terms are again approaching zero. The scale parameter is approaching unity, suggesting reliable SP data. The usefulness of the models with serial correlation is also suggested.

6.3. *Chukyo* data

Estimates for the *Chukyo* data are shown in Table 5 column iv). The chi-squared value is 1.13 and parameter equality is not rejected at a 50% level of significance, suggesting that parameter equality is statistically justified.

An estimate for the SP model was also examined. Parameters for the “difference from the RP part” are generally larger (about 4 to 50 times the size of the “RP part”) and more significant than the “RP part.” The *t*-statistics for the “RP part” are better than those for the 1st and 2nd bound choice model (Table 5 column ii)) and the signs are as expected.

An estimate for the RP/SP model was also examined. The *t*-statistics are generally improved (or at least remain at the same level) and the signs are as expected, as compared to the 1st bound choice model and the 1st and 2nd bound choice model (Table 5 columns i) and ii)). Compared to model ii), where the parameter signs are reasonable, standard errors for three variables generally remain almost at the same level or are improved, suggesting higher or at least the same level of estimation efficiency.

An estimate for the RP/SP model with serial correlation was also examined. Compared with the RP/SP model in Table 5 column iv), the SP constant is smaller, suggesting that the other variables provide a better explanation. Generally speaking, the merits of models with serial correlation are limited in the *Chukyo* data as compared to the *Keihanshin* data. A possible reason for this is that only those who commute by auto have the maximum three SP data, leading to a relatively smaller number of SP responses per respondent.

6.4. Summary of this section

In the previous section, parameter equality is not justified in a double-bounded family formulation. A model dividing the SP variable into the “RP part” and the “difference from the RP part” is developed, and parameter equality in the “RP part” is justified. The formulation described in this section indicates at least the same or a higher level of estimation efficiency compared to the traditional choice model, suggesting the usefulness of the proposed formulation. In subsection 6.1, this formulation is interpreted from two points of view: SP bias and preference change.

In the transport behaviour model demonstrated in this paper, the validity of parameter equality can be statistically tested. In some cases of CVM, however, RP data is impossible to obtain, and validation cannot be performed. This is why detailed guidelines for CVM surveys have been developed (for example, Arrow *et al.*, 1993). For transport behaviour modelling, however, few comprehensive guidelines for SP surveys exist. The required level of SP survey design again must be discussed.

7. CONCLUSIONS

The results of this study can be summarised as follows.

- The value of obtaining matching data is discussed.
- Matching data is effectively obtained in a double-bounded family format and formulated within the framework of discrete choice modelling.
- Estimates using the *Keihanshin* and *Chukyo* data show increased estimation efficiency in a double-bounded model family formulation.
- In a double-bounded family formulation, however, parameter equality between the RP and SP models is not justified statistically. A model formulation in which the SP level of service is divided into the "RP part" and the "difference from the RP part" is proposed. The estimates justify parameter equality in the "RP part" in the proposed model. The proposed model has at least the same or a higher level of estimation efficiency compared to traditional choice formulation. The proposed model is interpreted from two aspects: SP bias and preference change. (Another specification, noted in footnote 3, is available.)

Topics for further research:

- When applying a double-bounded family format to actual demand forecasting, compare the results with the results of the traditional choice format.
- Matching information can be obtained through the payment card, bidding game, DB, and 1.5B formats, and a better response format for obtaining matching information must be discussed.
- The validity of dividing the SP level of service into the "RP part" and the "difference from the RP part" must be verified through more case studies.

ACKNOWLEDGEMENTS

The authors wish to thank the anonymous reviewers for their comments. The authors also wish to thank Dr. Toshiyuki Yamamoto (Associate Professor at Nagoya University, Japan) for his valuable comments. This study is supported by Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology (Japan).

REFERENCES

Arrow, K., R. Solow, P. R. Portney, E. E. Leamer, R. Radner and H. Schuman (1993). *Report of NOAA Panel on Contingent Valuation*. 58 Federal Register 4601 (January 15).

Ben-Akiva, M. and T. Morikawa (1990). Estimation of Travel Demand Models from Multiple Data Sources. In: *Transportation and Traffic Theory* (Proc. of the 11th Int'l Symp. on Transportation and Traffic Theory) (M. Koshi, ed.), pp. 461-476, Elsevier, New York.

Hanemann, M., J. Loomis and B. Kanninen (1991). Statistical Efficiency of Double-Bounded Dichotomous Choice Contingent Valuation. *American Journal of Agricultural Economics*, **73 (4)**, 1255-1263.

Hensher, D. A. (1994). Stated Preference Analysis of Travel Choices: the State of Practice. *Transportation*, **21**, 107-133.

Kuriyama, K. (1998). *Environmental Value and Valuation Method*. Hokkaido University Press, Sapporo. (in Japanese)

Louviere, J. J., D. A. Hensher and J. D. Swait (2000). *Stated Choice Methods: Analysis and Application*. Cambridge University Press, Cambridge.

Mizuno, M. and H. Katahira (2003). Expanding Product Space and the Formation of Preferential Decision Rules: An Evolutionary Process of Products and Consumer Preferences. *Journal of Marketing Science*, **11(1-2)**, 1-21. (in Japanese)

Morikawa, T. (1989). *Incorporating Stated Preference Data in Travel Demand Analysis*. Ph. D. dissertation, Department of Civil Engineering, MIT.

Morikawa, T. (1994). Correcting State Dependence and Serial Correlation in the RP/SP Combined Estimation Method. *Transportation*, **21**, 153-165.

Payne, J., J. Bettman and E. Johnson (1993). *The Adaptive Decision Maker*. Cambridge University Press, Cambridge.

Pearmain, D., J. Swanson, E. Kroes and M. Bradley (1991). *Stated Preference Technique: A Guide to Practice* (2nd Ed.). Steer Davies Gleave and Hague Consulting Group, Richmond.

APPENDIX A: RP/SP COMBINED ESTIMATION METHOD

A 1: RP/SP combined estimation method (Morikawa, 1989)

The RP/SP combined estimation method is briefly explained. More details are available from Morikawa (1989), Ben-Akiva and Morikawa (1990), and others. The RP and SP models and the ratio of variances between the RP and SP error components are formulated in equations (A1) – (A3).

RP model

$$U_{in}^{RP} = \boldsymbol{\beta}'\mathbf{x}_{in}^{RP} + \boldsymbol{\alpha}'\mathbf{w}_{in}^{RP} + \varepsilon_{in}^{RP}, \; i = 1, \cdots, I_n^{RP} \text{ and } n = 1, \cdots, N^{RP} \qquad \text{(A1)}$$

SP model

$$U_{in}^{SP} = \boldsymbol{\beta}'\mathbf{x}_{in}^{SP} + \boldsymbol{\gamma}'\mathbf{z}_{in}^{SP} + \varepsilon_{in}^{SP}, \; i = 1, \cdots, I_n^{SP} \text{ and } n = 1, \cdots, N^{SP} \qquad \text{(A2)}$$

Ratio of variances between the RP and SP error components

$$Var\left(\varepsilon_{in}^{RP}\right) = \mu^2 Var\left(\varepsilon_{in}^{SP}\right), \qquad \forall i, n \qquad \text{(A3)}$$

where

U_{in}^{M} : total utility of individual *n* choosing alternative *i* in *M* model
ε_{in}^{M} : error component of total utility U_{in}^{M}
$\mathbf{x}_{in}^{M}$, $\mathbf{w}_{in}^{M}$, $\mathbf{z}_{in}^{M}$: explanatory variable vectors of deterministic component of individual *n* choosing alternative *i* in *M* model
$\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$: unknown parameter vectors to be estimated
I_{n}^{M} : number of alternatives included in the choice set of individual *n* in *M* model
N^{M} : number of observations in *M* model
M : type of model
μ : scale parameter explaining the ratio of variances between the RP and SP error components

A 2: RP/SP combined estimation method with serial correlation (Morikawa, 1994)

The RP/SP combined estimation method with serial correlation is briefly explained. More details are available from Morikawa (1994) and others. Unlike subsection A 1, the error components are divided into λ (which is common to each individual and alternative) and ν (which is truly random for researchers (white noise)), and λ is shared between the two models.

$$\varepsilon_{in}^{RP} = \lambda_{in} + \nu_{in}^{RP} \tag{A4}$$
$$\varepsilon_{in}^{SP} = \theta_{i}\lambda_{in} + \nu_{in}^{SP} \tag{A5}$$

The RP and SP models and the ratio of variances between the RP and SP error components are then formulated using equations (A6) – (A8).

RP model

$$U_{in}^{RP} = \boldsymbol{\beta}'\mathbf{x}_{in}^{RP} + \boldsymbol{\alpha}'\mathbf{w}_{in}^{RP} + \lambda_{in} + \nu_{in}^{RP} \text{ , } i = 1,\cdots,I_{n}^{RP} \text{ and } n = 1,\cdots,N^{RP} \tag{A6}$$

SP model

$$U_{in}^{SP} = \boldsymbol{\beta}'\mathbf{x}_{in}^{SP} + \boldsymbol{\gamma}'\mathbf{z}_{in}^{SP} + \theta_{i}\lambda_{in} + \nu_{in}^{SP} \text{ , } i = 1,\cdots,I_{n}^{SP} \text{ and } n = 1,\cdots,N^{SP} \tag{A7}$$

Ratio of variances between the RP and SP error components

$$Var\left(\nu_{in}^{RP}\right) = \mu^{2} Var\left(\nu_{in}^{SP}\right), \qquad \forall i, n \tag{A8}$$

where

λ_{in} : systematic component of error component of individual *n* choosing alternative *i*
ν_{in}^{M} : white noise of error component
θ_{i} : unknown parameter to be estimated