Stated Preference (Conjoint) Market Research Data

Data for Estimation of Choice Models

Revealed Preferences (RP):

- observed or reported actual behavior
  - Travel diaries

Stated Preferences (SP):

- Response to hypothetical scenarios (or “experiments”)

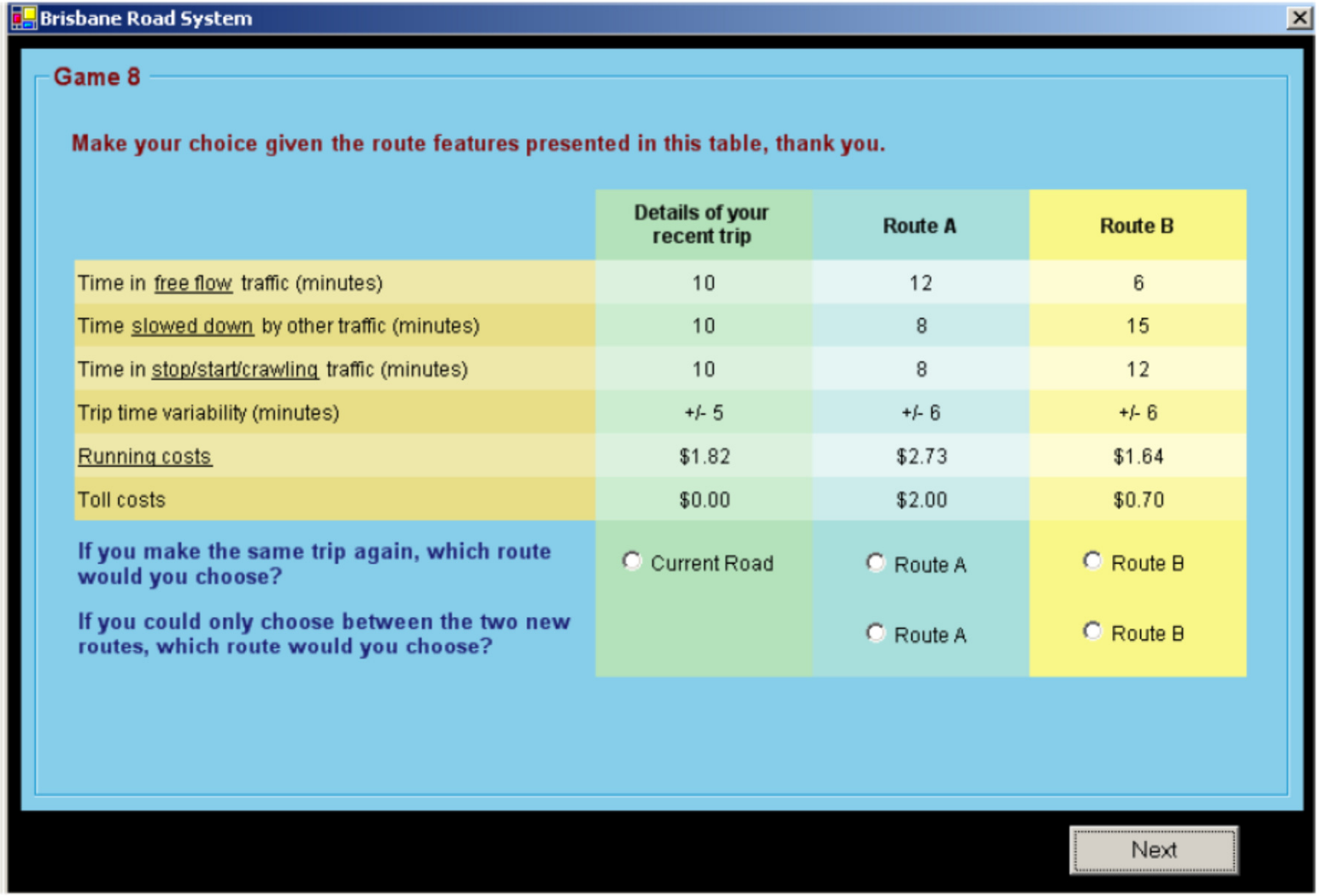


Figure 1: An example of a (unlabelled) stated choice situation

Stated Preferences: Motivation

Identification

- new alternatives
- new attributes
- attribute levels beyond range of RP data
- non-market applications

Efficiency

- limited variability of attributes in RP data
- Co-linearity of attributes in RP data

Choice set definition

- choice set and attribute values are pre-specified

Data Collection Resources

- RP data may be too expensive and time-consuming to collect
- No measurement errors
- Easy to obtain multiple responses
- Various response types

Comparison between RP and SP

	RP data	SP data
Preference Information	● The result of the actual behaviour ● Consistent with the behaviour in the real market ● We can get "Choice" result	● Expression under the hypothetical situation ● Possibility of inconsistent with the behaviour in the real market ● We can get "Ranking", "Rating", "Choice", etc.
Alternatives	● Only existing alternatives	● Existing and non-existing alternatives
Attributes	● Measurement error ● Limited range of attributes' levels ● Possibility of collinearity among attributes	● No measurement error ● Extensibility of the range of attributes' levels ● Controllability of the collinearity among attributes
Choice Set	● Non-clear	● Clear
Number of Response(s)	● One response per respondent	● One or more response(s) per respondent

SP Approaches

Experimental setting:

- the context of the hypothetical scenarios
- alternatives or profiles are bundles of attributs
- respondents are presented with limited sets of alternatives

Expression of preferences:

- Rating
- Ranking
- Matching
- Choice

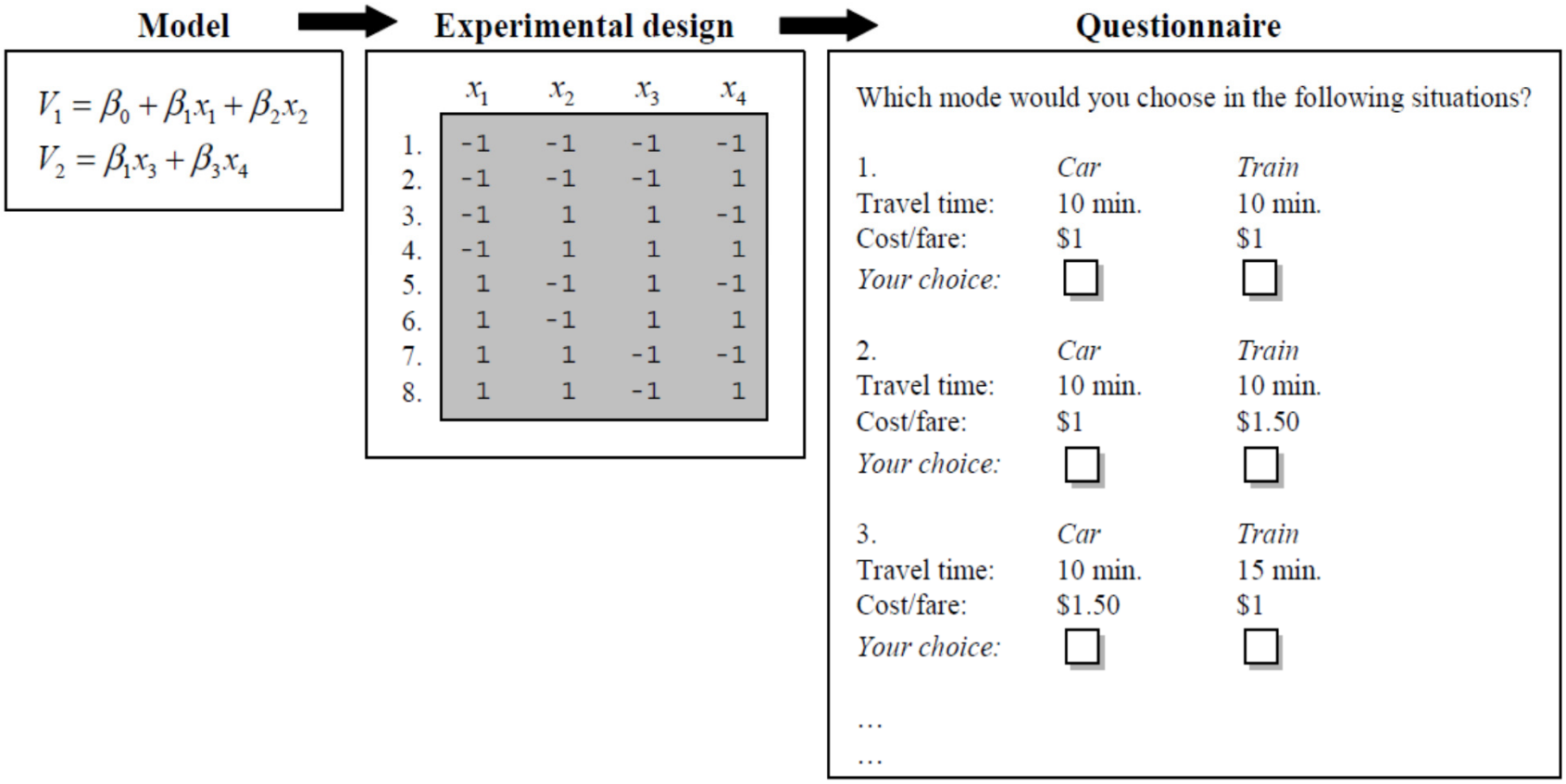
Potential Sources of Bias in SP Data

1. Indifference to the experimental task
2. Policy response bias (strategic misrepresentation)
3. Justification bias
4. Omission of situational constraints
5. Incomplete descriptions of alternatives
6. Context effects (anchoring, embedding)

Stated Preferences: Issues

- Validity -- SP response protocol vs. actual behavior
- Realism
- Complexity
- Difficulty
- Repetitions

Steps in designing a stated choice experiment



Model

- Each stated choice experiment is specifically created for estimating a specific model
- Therefore, one need to specify the model and the parameters to be estimated:
  - Number of alternatives
  - Attribute related
  - Attribute levels
  - Model type/structure

Degree of Freedom

- The experiment degree of freedom is the total number of parameters (excluding the constants) to be estimated plus one (accounting for all constants)
- Number of parameters:
  - Generic vs. alternative specific
  - Interaction effect
  - Nonlinear effects
- The number of choice situations must be equal or greater than the degree of freedom

Experimental Design for Three Attributes with Two Levels Each

Options	Attributes		
	Fare	Travel Time	Frequency
1	Low	Fast	Infrequent
2	Low	Fast	Frequent
3	Low	Slow	Infrequent
4	Low	Slow	Frequent
5	High	Fast	Infrequent
6	High	Fast	Frequent
7	High	Slow	Infrequent
8	High	Slow	Frequent

Numeric representation:

Options	Attributes (-1 = poor; 1 = good)		
	1	2	3
1	1	1	-1
2	1	1	1
3	1	-1	-1
4	1	-1	1
5	-1	1	-1
6	-1	1	1
7	-1	-1	-1
8	-1	-1	1

Presentation of Public Transport Options

Public Transport Service		(Option 1)
Fare = $0.30	Travel time = 15 mins	Frequency = every 30 minutes

Public Transport Service		(Option 2)
Fare = $0.30	Travel time = 15 mins	Frequency = every 15 minutes

Public Transport Service		(Option 3)
Fare = $0.30	Travel time = 25 mins	Frequency = every 30 minutes

Public Transport Service		(Option 4)
Fare = $0.30	Travel time = 25 mins	Frequency = every 15 minutes

Public Transport Service		(Option 5)
Fare = $0.50	Travel time = 15 mins	Frequency = every 30 minutes

Public Transport Service		(Option 6)
Fare = $0.50	Travel time = 15 mins	Frequency = every 15 minutes

Public Transport Service		(Option 7)
Fare = $0.50	Travel time = 25 mins	Frequency = every 30 minutes

Public Transport Service		(Option 8)
Fare = $0.50	Travel time = 25 mins	Frequency = every 15 minutes

Note:

Experiment attributes are:

Fare: Low=$0.30; High=$0.50

Time: Fast=15min; Slow=25min

Frequency:

Frequent=every 15min;
Infrequent=every 30min

Examples of a Fractional Factorial Design Derived from a Full Factorial Design

Full Factorial Design

	Attributes		
	1	2	3
Options:			
1	+1	+1	-1
2	+1	+1	+1
3	+1	-1	-1
4	+1	-1	+1
5	-1	+1	-1
6	-1	+1	+1
7	-1	-1	-1
8	-1	-1	+1

Fractional Factorial Design:

2	+1	+1	+1
3	+1	-1	-1
5	-1	+1	-1
8	-1	-1	+1

A Definition of Attribute Levels Dependent on the Characteristics of a Respondent's Actual Trip

	Cost	Travel Time	Frequency of Service
Respondent's Actual Trip	$1.00	20 mins	Bus every 20 mins

Definitions of Attribute Levels

Stated Preference Alternatives	Cost	Travel Time	Frequency of Service
(As absolute changes)			
1	+30¢	+10 mins	-10 mins
2	-20¢	-5 mins	+20 mins
(As proportional changes)			
1	+30%	+50%	-50%
2	-20%	-25%	+100%

Presentation of Choices

	Recent Trip			Alternative Trip		
	Cost	Time	Frequency	Cost	Time	Frequency
1	$1.00	20 mins	1 every 20 mins	$1.30	30 mins	1 every 10 mins
2	$1.00	20 mins	1 every 20 mins	$0.80	15 mins	1 every 40 mins

Experimental Design

- Alternatives
  - Label: car, train, bus….
  - Unlabel: Route A, Route B….
- Attributs z_k, k = 1,...,K .
- Levels $z_{k\ell}$, ℓ = 1,...,L_k.
- Profiles per choice set i = {z_k(i), k = 1,...,K} .
- Full factorial: $\prod_{k=1}^{K} L_k$ po possible profiles.
- Fractional factorial design: an “optimal” subset of profiles.
- Number of choice set (scenarios) per survey
- Number of different survey forms

Experiment Design – Factorial experiments

- Analyze joint effect of several attributes
  - e.g. travel time, travel cost
- Each attribute has multiple levels
  - e.g. low, medium, high cost
- Full factorial experiment
  - All combinations of attribute levels
  - Orthogonal design: attributes should vary independently from one another (correlations are zero)
  - Balance design: combination of the different attribute levels are distributed equally

Choice set creation

- N alternatives
- M attributes
- L levels
- Full factorial design produces L^{MN} games
- N=2, M=3, L=2 -> 64 games
- Avoid:
  - Dominant games
  - Transitivity + dominance
  - Contextual constraints (combination that don't make sense)

Non-Orthogoanl Design

- Ratio of parameters (value of time)
- Most useful information is obtained when respondents are in the borderline between choices
- Choose values to limit the difference in utility
- Knowledge of the parameter values can help to design it this way.

Efficient Design

- Efficiency mean minimizing the variances of parameters' estimation.
- Therefore, an efficient design depends on the model to be estimated from the data and expectations about the parameters.
- For linear models with no prior information, the more efficient design is usually the more it will converge toward **balance and orthogonality**
- This should minimize the resulting standard errors when estimating the model.
- In non-linear models (like discrete choice) efficiency do not necessarily implicate orthogonality
- Exclusion of dominant or unrealistic choice alternatives violate strict orthogonality and balance

Efficient Choice Design

Specify the utility specification for the likely model to be estimated from the data

1. Generic vs. alternative specific parameters
2. Dummy vs. other variables
3. Main effects or also interaction effects
4. The value of the parameters likely to be obtained
5. The precise econometric model

Points 1-3 influence directly the design matrix

Point 4 and 5 affect the SE

Experimental design

- Define utility functions
- Define number of rows
- Define number of blocks