How to Reinforce the Stated Preference Methods Using the Potential of Computer Based Questionnaires ?

Cedric Taverne

PhD Committee : Lambert Ph., Govaert B., Mouchart M., Hubert M.

Centre de recherche en statistique, biostatistique et sciences actuarielles (ISBA)
Institut de recherche multidisciplinaire pour la modélisation et l'analyse quantitative (IMMAQ)
Université catholique de Louvain (UCLouvain)

cedric.taverne@uclouvain.be

Young Researchers Day - February 5, 2010

Outline

If you are interested in :

Studying how people choose
a particular alternative in a given situation

If you are looking for :

A quantitative answer

Then you would like to use :

The stated preference methods

...or the stated choice, the conjoint analysis, etc.
(synonymous)

What are the Stated Preference Methods ?

- Stated preference are methods to study individual choices

  Ex : A transport mode, a product, a service...

- Instead of submitting a long list of focused questions

  Ex : Is this side effect would change your choice of medicine?
  What price would you pay?
  Would you accept two connections on your way home?

- The idea is to put the respondents in front of complex **scenarios**

  where all the **attributes** of interest are gathered

  and are changing systematically

Ex 1 : Choice of transport mode in home-work trips

Scénario 2 : Choisissez le mode de transport qui vous convient le mieux pour vous rendre sur votre lieu de travail :

Ne répondez pas nécessairement en fonction de ce que vous faites actuellement mais bien en fonction des éléments décrits, toutes choses égales par ailleurs. Evaluez ce que vous choisiriez dans la situation qui vous est présentée.

Option 1 : En voiture 	Option 2 : En transport public 
Votre trajet durera entre **25** minutes et **45** minutes selon le trafic fluctuant de jour en jour (embouteillages, feux, etc.)	Votre trajet durera entre **25** minutes et **30** minutes en fonction de votre temps d'attente aux arrêts et de l'état du trafic. Vous n'aurez **pas** à effectuer de correspondance durant ce trajet.
Dans ce scénario, les prix des carburants auraient **augmenté de 50%** par rapport à aujourd'hui; soit un prix à la pompe de **1.425 €** pour le diesel, **2.025 €** pour l'Eurosuper 98, **1.845 €** pour la Superplus 95.	Dans ce scénario, votre trajet vous coûtera **2.4 €** (prix par trajet pour un abonnement mensuel)
Sur votre trajet, **un supermarché de proximité avec une boulangerie** est accessible facilement.	Il n'y a **pas de distributeur de billet, ni de supérette** accessible dans les points d'arrêts de votre parcours.

Quel serait votre choix en pareille situation :

○ Automobile ○ Transport public ou multimodal

suivant >

Ex 2 : Choice of a medication to lose weight

	Drug 1	**Drug 2**	**Drug 3**	
Medicine :	Original on prescription	Generic on prescription	Original on prescription	
Price per month :	55 € per m.	35 € per m.	75 € per m.	
Regular side effects :	Headaches	Diarrhea	None	
Services :	None	A recipe book	1 hour with a personal coach for free	
Your preference :	☐ Drug 1	☐ Drug 2	☐ Drug 3	☐ No Choice

The Steps of a Stated Preference Study (Louviere *et al.*, 2000, p.255)

1. Define study objectives
   - A step centred on the client's needs
   - Selection of the variables of interest
2. Conduct supporting qualitative study
   - A step centred on the respondent's needs
   - Selection of the variables necessary
3. Develop and pilot the data collection instrument
   - Design of the set(s) of scenarios
   - Writing (and programming) of the questionnaire
   - Pilot testing and validation of the questionnaire
4. Define sample characteristics
5. Perform data collection
6. Conduct model estimation
7. Conduct policy analysis

SWOT of the Stated Preference Methods

Strengths

- The context is under control
- The complexity is optimally designed for your model

Weaknesses

- The design is a prior which can be wrong !

Opportunities

- *The world as it could be*

  (new products, changes in personal and/or market constrains, etc.)

  $\longrightarrow$ New markets

  $\longrightarrow$ Trade-off analysis

Threats

- *All other things being equal*
- Declarative $\neq$ Observed

  Stated vs **Revealed** preferences

Common Hypotheses to Model the Choice

We have to impose some restrictions to define the scope of our model :

Hyp. 1 : The choice is supposed to be open and realistic

Hyp. 2 : The choice is made in context

Hyp. 3 : The alternative chosen in a particular set
is the one that maximizes its utility

Hyp. 4 : "Utility is derived from properties of things (. . .)
rather than the goods *per se*"

(Louviere *et al.*, 2000, p.2)

Utility Formula and Discrete Choice

$$U_{in} = V_{in} + \varepsilon_{in} = \sum_{k=1}^{K} \beta_{ik} X_{ikn} + \sum_{l=1}^{L} \gamma_{ip} Z_{iln} + \epsilon_{in} \qquad \forall i \in \mathbf{C}_n$$

$$\text{Alternative } i \text{ is chosen} \leftrightarrow U_{in} = \max_{j \in \mathbf{C}_n} U_{jn}$$

With
- U_{in} = The utility of the alternative i for the observation n
- β_{ik} = The utility parameter associated to the attribute k for the alternative i
- X_{ikn} = The attribute k for the alternative i and observation n
- γ_{il} = The random utility parameter associated to the design attribute l for the alternative i
- Z_{iln} = The design attribute l at the alternative i and observation n
- ϵ_{in} = The residual random component of the utility U_{in}
- $\mathbf{C}_n$ = The set of alternatives available for the observation n

Debatable Hypotheses to Simplify the Model

IIA : ***Independence from irrelevant alternatives***

Impose that the ratio between two alternative's utilities would be same even if we add or delete an irrelevant alternative

IID : ***Independently and identically distributed***

The random components of the utility (ε_{in} or ϵ_{in}) are independently and identically distributed amongst alternatives

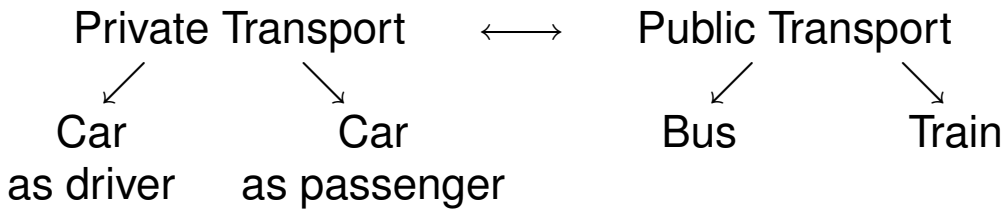
The Multinomial Logit (MNL) model

	Car option	**Public Transport Option**	**Bicycle Option**
Time and its variability :	Between 25 and 45 minutes according to the traffic flow	Between 25 and 30 minutes according to the waiting time	Around 45 minutes (speed : 15 km/h)
Connections :		No connection needed	
Price :	Fuel prices 50% higher Diesel : 1.659 € / l Euro 95 : 2.114 € / l	2.4 € / trip with a monthly pass	Cyclist indemnity 1 € / trip
Amenities on your way :	Supermarket with bakery	(none)	Supermarket with bakery
Your preference :	☐ Car Option	☐ Public Transport Option	☐ Bicycle Option

If IIA and IID hypotheses hold $\longrightarrow$ **Multinomial Logit** (MNL)

Relaxing the IID hypothesis

- **Idea 1 :** One choice can be split into successive choices



$\Rightarrow$ **Nested Logit** (NL), **Generalized Extreme Value** (GEV)

(Ben-Akiva, 1973 ; McFadden, 1978, 1979)

- Relax the independence (I) within the nests
- Relax the identical distribution (ID) between the nests

$$U_{in} = \left(V_{gn} + \mu_{gn}\right) + \left(V_{in|g} + \varepsilon_{in|g}\right) \qquad \begin{matrix} i \in \mathbf{C}_{gn} \\ \mathbf{C}_{gn} \subset \mathbf{C}_n \end{matrix}$$

With
U_{in} = The utility of the alternative i for the observation n
V_{gn} = The deterministic part of U_{in} at the 1st level
$\mu_{in|g}$ = The random component of U_{in} at the 1st level,
$V_{in|g}$ = The deterministic part of U_{in} at the 2nd level
$\varepsilon_{in|g}$ = The random component of U_{in} at the 2ndt level

$\mathbf{C}_{gn}$ = The subset of alternatives available for the observation n if g is chosen at the first level
$\mathbf{C}_n$ = The set of alternatives available for the observation n

Relaxing the IID hypothesis

- **Idea 2 :** Relaxing the identical distribution (ID) hypothesis

  $\Rightarrow$ **Heteroscedastic Extreme Value** (HEV) (Bhat, 1995)

  - Useful to test the ID hypothesis itself
  - Useful to test which alternatives gathering in a NL model
  - Comparing the estimated variance of the error terms

- **Idea 3 :** Relaxing the identical distribution (ID) hypothesis
  or / and modeling the dependence (I) of the error terms

  $\Rightarrow$ **Covariance Heterogeneity HEV, Covariance pattern MNL**
  (Swait and Adamowicz, 1996 ; Swait and Stacey, 1996)

Relaxing the IIA hypothesis

- **Idea 1 :** The utility parameters may depend of the set of alternatives

  $\Rightarrow$ **Mixed Logit** (ML), **Random Parameter Logit** (RPL)

  (McFadden and Train, 1996 ; Ben-Akiva *et al.*, 1993)

$$P_{in|\mathbf{C}_n} = \frac{\exp(\sum_{k=1}^{K} \beta_{ik|\mathbf{C}_n} X_{ikn} + \varepsilon_{in})}{\sum_{j \in \mathbf{C}_n} \exp(\sum_{k=1}^{K} \beta_{jk|\mathbf{C}_n} X_{jkn} + \varepsilon_{jn})}, \qquad \begin{matrix} i \in \mathbf{C}_n \\ \varepsilon_{in} \sim IID \text{ GEV} \end{matrix}$$

- **Idea 2 :** Asking for a ranking instead of a discrete choice

  (Beuthe *et al.*, 2008)

  - More information on the choice structure
  - Return to the common hypothesis 3 ($\max_{i \in \mathbf{C}_n} U_{in}$) to focus on $\mathbf{C}_n$

Relaxing the IIA hypothesis

- **Idea 3 :** The choice can be split into the choice of a set of considered alternatives and the choice itself

  $\Rightarrow$ **Independent Availability Logit** (IAL)

  (Swait, 1984 ; Erdem and Swait, 1998)

$$P_{in} = \sum_{\mathbf{C}_m \in \Gamma_n} P_{in|\mathbf{C}_m} P_{\mathbf{C}_m n}, \quad \mathbf{C}_m \subseteq \mathbf{C}_n$$

With
P_{in} = The probability that the respondent n chooses the alternative i
$P_{in|\mathbf{C}_m}$ = The probability that the respondent n chooses the alternative i if he/she considers the subset $\mathbf{C}_m$
$P_{\mathbf{C}_m n}$ = The probability that the respondent n considers the subset $\mathbf{C}_m$
$\mathbf{C}_n$ = The set of alternatives available for the respondent n
Γ_n = The whole set of possible subset of $\mathbf{C}_n$
$\mathbf{C}_m$ = A particular subset of Γ_n

SWOT of the Stated Preference Methods

Strengths

- The context is under control
- The complexity is optimally designed for your model

Weaknesses

- The design is a prior which can be wrong !
- The models' hypotheses

Opportunities

- *The world as it could be*
- The users' imagination

Threats

- *All other things being equal*
- Declarative $\neq$ Observed

Multinomial Response

	Car option	**Public Transport Option**	**Bicycle Option**
Time and its variability :	Between 25 and 45 minutes according to the traffic flow	Between 25 and 30 minutes according to the waiting time	Around 45 minutes (speed : 15 km/h)
Connections :		No connection needed	
Price :	Fuel prices 50% higher Diesel : 1.659 € / l Euro 95 : 2.114 € / l	2.4 € / trip with a monthly pass	Cyclist indemnity 1 € / trip
Amenities on your way :	Supermarket with bakery	(none)	Supermarket with bakery
Your preference :	☐ Car Option	☐ Public Transport Option	☐ Bicycle Option

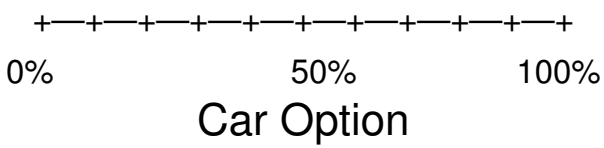
Multinomial Response and None Option

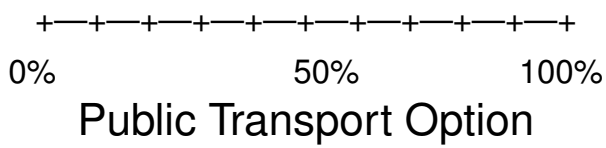
	Car option	**Public Transport Option**	**Bicycle Option**	
Time and its variability :	Between 25 and 45 minutes according to the traffic flow	Between 25 and 30 minutes according to the waiting time	Around 45 minutes (speed : 15 km/h)	
Connections :		No connection needed		
Price :	Fuel prices 50% higher Diesel : 1.659 € / l Euro 95 : 2.114 € / l	2.4 € / trip with a monthly pass	Cyclist indemnity 1 € / trip	
Amenities on your way :	Supermarket with bakery	(none)	Supermarket with bakery	
Your preference :	☐ Car Option	☐ Public Transport Option	☐ Bicycle Option	☐ No Choice

Binary Response with Likert Scales

	Car option	**Public Transport Option**
Time and its variability :	Between 25 and 45 minutes according to the traffic flow	Between 25 and 30 minutes according to the waiting time
Connections :		No connection needed
Price :	Fuel prices 50% higher Diesel : 1.659 € / l Euro 95 : 2.114 € / l	2.4 € / trip with a monthly pass
Amenities on your way :	Supermarket with bakery	(none)

Your evaluation :





Binary Response with Cross Likert Scale

	Car option	**Public Transport Option**
Time and its variability :	Between 25 and 45 minutes according to the traffic flow	Between 25 and 30 minutes according to the waiting time
Connections :		No connection needed
Price :	Fuel prices 50% higher Diesel : 1.659 €/ l Euro 95 : 2.114 €/ l	2.4 €/ trip with a monthly pass
Amenities on your way :	Supermarket with bakery	(none)

In favor of the Car Option : 100% 50% 0%

+—+—+—+—+—+—+—+—+—+—+—+

In favor of the PT Option : 0% 50% 100%

Unlabeled Alternatives

	Drug 1	**Drug 2**	**Drug 3**	
Medicine :	Original on prescription	Generic on prescription	Original on prescription	
Price per month :	55 € per m.	35 € per m.	75 € per m.	
Regular side effects :	Headaches	Diarrhea	None	
Services :	None	A recipe book	1 hour with a personal coach for free	
Your preference :	☐ Drug 1	☐ Drug 2	☐ Drug 3	☐ No Choice

Unlabeled Alternatives with Likert Scales

	Drug 1	**Drug 2**	**Drug 3**
Medicine :	Original on prescription	Generic on prescription	Original on prescription
Price per month :	55 € per m.	35 € per m.	75 € per m.
Regular side effects :	Headaches	Diarrhea	None
Services :	None	A recipe book	1 hour with a personal coach for free
Your evaluation :	+−+−+−+−+ 0% 100% Drug 1	+−+−+−+−+ 0% 100% Drug 2	+−+−+−+−+ 0% 100% Drug 3

Unlabeled Alternatives with Likert Scales and choice

	Drug 1	**Drug 2**	**Drug 3**	
Medicine :	Original on prescription	Generic on prescription	Original on prescription	
Price per month :	55 € per m.	35 € per m.	75 € per m.	
Regular side effects :	Headaches	Diarrhea	None	
Services :	None	A recipe book	1 hour with a personal coach for free	
Your preference :	☐ Drug 1	☐ Drug 2	☐ Drug 3	☐ No Choice
Your evaluation :	+−+−+−+−+ 0% 100%	+−+−+−+−+ 0% 100%	+−+−+−+−+ 0% 100%	

Dropping or Adding Alternatives and Attribute

The drop : Using prior information about the respondent to drop an alternative or an attribute from the scenarios of his/her questionnaire

	Car as driver	Car passenger	Bus	Bicycle
Factor A :	A1	A1	A4	A2
Factor B :	B2	B1		B1
Factor C :	C3	C1	C1	C2
Factor D :	D1		D1	
	☐ Car as driver	☐ Car passenger	☐ Bus	☐ Bicycle

Dropping or Adding Alternatives and Attribute

The add : Using prior information about the respondent to add an alternative or an attribute to the scenarios of his/her questionnaire

	Car as driver	Car passenger	Bus	Bicycle
Factor A :	A1	A1	A4	A2
Factor B :	B2	B1		B1
Factor C :	C3	C1	C1	C2
Factor D :	D1		D1	
	☐ Car as driver	☐ Car passenger	☐ Bus	☐ Bicycle

Sequential Design

Step 0 : Common optimal design for $Y = \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2$

Step 1 : After n_1 respondents $\rightarrow$ first estimation of our model

Ex : Estimated variance of β_1 $>>$ Estimated variance of β_2

Step 2 : Bayesian sequential design to focus on β_1 and β_{12}

Step 3 : After n_2 respondent...

Thank you for your attention !

- Bateman I. *et al.* (2002) *Economic Valuation with Stated Preference Techniques : A Manual*, Cheltenham : Edward Elgar Publishing
- Bhat C. (1995) *A Heteroscedastic Extreme Value Model of Intercity Travel Mode Choice*, Transportation Research Part B, Vol. 29, Issue 6, pp.471-83
- Ben-Akiva M. (1973) *Structure of Passenger Travel Demand Models*, M.I.T. Ph.D. Dissertation ; *cited in* Verboven F. (1996) *The Nested Logit Model and Representative Consumer Theory*, Economic Letters, Vol. 50, pp.57-63
- Ben-Akiva M., Bolduc D., Bradley M. (1993) *Estimation of Travel Choice Models with Randomly Distributed Values of Time*, Transportation Research Record, 1413, pp.8897
- Beuthe M., Bouffioux C., Krier C., Mouchart M. (2008) *A Comparison of Conjoint, Multi-Criteria, Conditional Logit and Neural Network Anal- yses for Rank-Ordered Preference Data* ; *in* Ben-Akiva, M. et al. (2008) *Recent Developments in Transport Modeling : Lessons for the Freight Sector*, Amsterdam: Elsevier p.157-178
- Erdem T., Swait, J. (1998) *Brand equity as a signaling phenomenon*, Journal of Consumer Psychology, Vol. 7, Issue 2, p.131-157
- Louviere J., Hensher D., Swait J. (2000) *Stated Choice Methods : Analysis and Applications*, Cambridge : Cambridge University Press
- McFadden D. (1978) *Model the choice of residential location* ; *in* Karlvist A., Ludvist L., Snickars F., Weibull J. eds. (1978) *Spatial interaction theory and planning models (studies In Regional Science And Urban Economics)*, Amsterdam : Elsevier Science, pp.75-96
- McFadden D. (1979) *Quantitative Methods for Analyzing Travel Behaviour of Individuals: Some Recent Developments* ; *in* Hensher D., Stopher P. (1979) *Behavioural travel modelling*, London : Croom Helm, pp.279-318.
- McFadden D., Train K. (1996) *Mixed MNL models for discrete response*, Journal of Applied Econometrics, Vol. 15, Issue 5, pp.447-470
- Street D., Burgess L. (2007) *The Construction of Optimal Stated Choice Experiments : Theory and Methods*, Wiley Series in Probability and Statistics, Hoboken : John Wiley and Sons
- Swait J. (1984) *Probabilistic choice set formation in transportation demand models*, unpublished PhD thesis, Department of Civil Engineering, MIT ; *cited in* Louviere, J. *et al.*, 2000.
- Swait J., Adamowicz W. (1996) *The Effect of Choice Environment and Task Demands on Consumer Behavior : Discriminating Between Contribution and Confusion*, Alberta : University of Alberta, Department of Rural Economy, Staff Paper 96-09
- Swait J., Stacey E.C. (1996) *Consumer brand assessment and assessment confidence in models of longitudinal choice behavior* ; *presented at* The 1996 INFORMS Marketing Science Conference, 710 March, Gainesville, Florida.