Ewa Zawojska, Mikołaj Czajkowski, Marek Giergiczny

DOES THE NUMBER OF ALTERNATIVES MATTER FOR STATED PREFERENCES?

Ewa Zawojska, Mikołaj Czajkowski, Marek Giergiczny

DOES THE NUMBER OF ALTERNATIVES MATTER FOR STATED PREFERENCES?

The research

- concerns the stated preference methodology;
- addresses the problem whether the number of alternatives provided in a single question impinges on the respondent's behaviour.

Stated preference method

- In specially designed surveys respondents state what they would do.
- Respondents are asked to choose their most preferred alternative from the provided set. Alternatives represent various policy scenarios which differ in the policy characteristics (attributes) including different costs (monetary attribute) related to the policy implementation.
- Contingent valuation
- This survey-based method is commonly used to elicit public's preferences.
- Knowledge of public's preferences allows for effective allocation and management of goods.
- A flexible method – enables valuation of goods in hypothetical situations.

An essential question:

Do people answer truthfully in stated preference surveys?

Theoretically suggested

Conditions for incentive compatibility

(Carson and Groves, 2007)

Incentive compatibility = Truthful preference revelation is the respondent's optimal strategy.

1. Respondents understand and answer the question being asked.
2. The payment mechanism is coercive (that is, imposes payment on all agents).
3. The survey is seen as a take-it-or-leave-it offer.
   (That is, already made choices do not influence any other offers that may be given.)
4. Respondents view the survey as consequential, which means:
   - their responses are seen as influencing agency's actions,
   - they care about the finally introduced solution.
5. The survey has the format of a single two-alternative question with one option being status quo "no change" (as suggested in the Gibbard-Satterthwaite theorem).

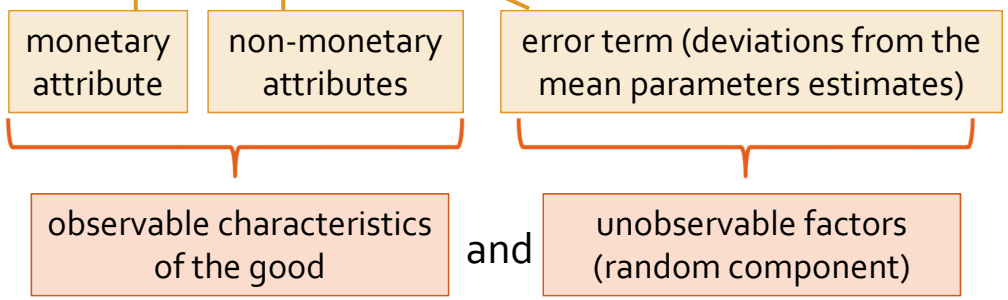
From the empirical perspective

Random Utility Model (McFadden, 1974)

FOUNDATION OF PREFRENCE MODELLING BASED ON DISCRETE CHOICE DATA

- Utility of consumer *n* from choosing alternative *j* in choice task *t* (U_{njt}):

$$U_{njt} = \alpha c_{njt} + bX_{njt} + e_{njt}$$



- A consumer derives utility from:

Empirical evidence on the role of the number of alternatives

Against the use of multiple alternatives

Xu et al. (2013)	Lab	In three-alternative tasks respondents choose their second most preferred option (private good).
Hensher (2004)	CAPI	The more complex the design, the higher stated values of travel time savings.
Hensher (2006)	CAPI	The more alternatives, the higher stated values of travel time savings (when not controlled for other design dimensions).
Rose et al. (2009)	CAPI	As the number of alternatives rises, Australian and Taiwanese respondents increasingly overstate their travel time savings, while Chilean understate.

- Lack of incentive compatibility – rationally no sense in voting for the most preferred alternative if it has no chance to win the voting.
- Increased choice complexity may prompt respondents to avoid making choices at all.

In favour of the use of multiple alternatives

Carson et al. (2011)	Lab	No significant differences in answers to two- and three-alternative tasks. Subjects rarely vote strategically.
Collins and Vossler (2009)	Lab	More deviations from the optimal choice in two-alternative tasks than in three-alternative tasks.
Arentze et al. (2003)	Field	No significant difference in the variance of the error term across two- and three-alternative tasks.
Ready et al. (1995)	Field	Better match of stated and true preferences when multiple alternatives used.
Rolfe, Bennett (2009)	Field	More robust models can be estimated on data from three-alternative tasks compared to two-alternative tasks. In two-alternative tasks a higher rate of "not sure" responses.

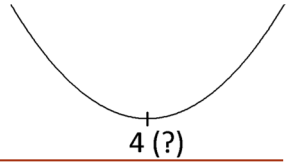
- Efficiency gains (more data in a cheaper way)
- More alternatives increase the chances to find a satisfactory option, which makes the choice easier.

Potential non-linearity of the impact of the number of alternatives

Evidence on the optimal number of alternatives

On the theoretical basis

Kuksov and Villas-Boas (2010)	• Given too many alternatives, a consumer has to engage in many searches to find a satisfactory fit, which may be too costly and make the consumer defer taking a choice. • Given too few alternatives, a consumer may not search, fearing that an acceptable choice is unlikely, and does not make a choice at all.



On the empirical basis

Caussade et al. (2005)	No systematic effect of the number of alternatives on the willingness to pay estimates. With respect to the variance of the error term in the utility function, a U-shaped relationship emerges – choices in four-alternative tasks possess the lowest variance in comparison to three- and five-alternative tasks.
DeShazo and Fermo (2002)	The variance of the error term in the utility function follows a U-shaped pattern – up to a threshold number of alternatives, the variance decreases and later it increases.
Meyerhoff et al. (2014)	Across sets with three, four and five alternatives, the lowest variance of the error term in the utility function is obtained for a four-alternative choice task.

RESEARCH QUESTION

Does the number of alternatives matter for stated preferences?

With respect to two aspects:

- Willingness to pay (WTP) estimates
- The variance of the error term in the utility function

Hypothesis 1.

WTP estimates derived from consumers' stated preferences differ significantly depending on the number of alternatives per choice task.

Hypothesis 2.

Variance of the error term in the consumer's utility function is higher for two-alternative choice tasks than for three-alternative choice tasks.

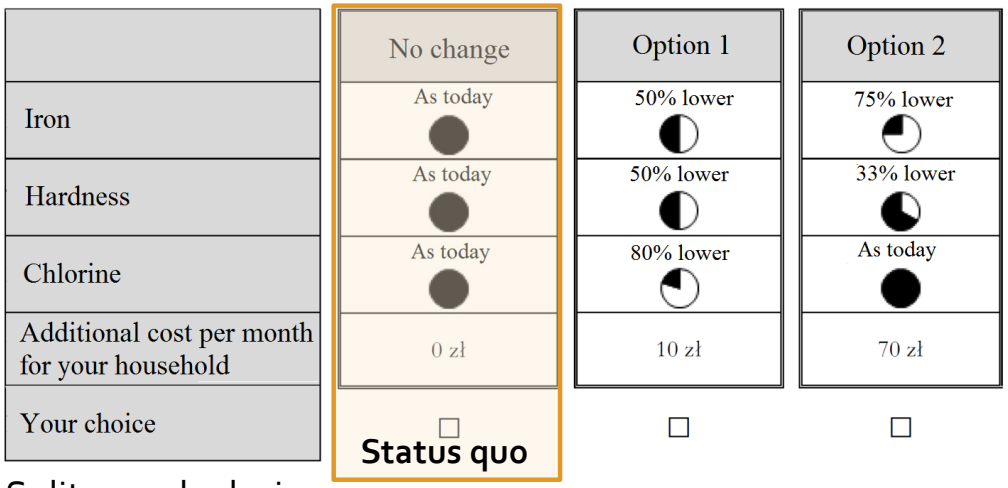
Valuation of better tap water quality

PREVIOUS STUDIES

- The averting behaviour method
  - values derived from averting (or defensive) actions taken by consumers (e.g. purchasing bottled water) to avoid negative consequences of bad tap water quality
  - typically to assess health risk reduction (Abdalla, 1990; Dupont and Jahan, 2012; Um, Kwak and Kim, 2002)
  - but are the consumers' actions interpreted as defensive indeed defensive?
- The contingent valuation method
  - more flexible – valuation of hypothetical scenarios
  - valuation of health risk reduction (Adamowicz, Dupont, Krupnick and Zhang, 2011; Cho, Easter, McCann and Homans, 2005)
  - valuation of improvements of physical tap water characteristics: chlorine odour, chlorine taste, water turbidity, calcium carbonate stains, water colour (Day et al., 2012; Scarpa, Thiene and Hensher, 2012)

Study design

- Discrete Choice Experiment
- Mail survey among residents of Milanówek conducted in 2013
- A hypothetical scenario: improvement of tap water quality in Milanówek



Attribute levels

Reduction by 50%, 75%, 95%
Reduction by 33%, 50%
Reduction by 80%

- Split sample design:
  - 403 respondents in a two-alternative treatment
  - 401 respondents in a three-alternative treatment
- 12 choice tasks per respondent

Two- and three-alternative samples – do they differ?

- Wilcoxon-Mann-Whitney test of equality of distributions

	Sample means		
	2 alt	3 alt	p-value
Years lived in Milanówek	32.69	32.68	0.73
Age	51.59	51.36	0.93
Household size	2.841	2.816	0.90
Immature household members	0.4543	0.4898	0.93
Litres of used bottled water per month	22.15	20.84	0.26

The null hypothesis of equality of distributions cannot be rejected.

Samples do not differ with respect to these characteristics.

- Chi-square test of equality of proportions

	p-value
Share of males	0.14
Education	0.16
Income	0.12

The null hypothesis of equality of proportions cannot be rejected.

Samples do not differ with respect to these characteristics.

ECONOMETRIC APPROACH

Generalised Mixed Logit in WTP-space

- Based on the Random Utility Model (McFadden, 1974)
- Discrete Choice Model in WTP-space with random parameters and scale heterogeneity
- Utility derived by consumer *n* choosing alternative *j* in choice task *t* (U_{njt}):

$$U_{njt} = \delta_n \left(\alpha_n c_{njt} + b_n X_{njt} \right) + \varepsilon_{njt} = \delta_n \alpha_n \left(c_{njt} + \beta_n X_{njt} \right) + \varepsilon_{njt}$$

monetary attribute

non-monetary attributes

Gumbel distributed error term with variance normalised to $\pi^2 / 6$

consumer-specific, log-normally distributed (random) parameter

consumer-specific, normally distributed (random) parameters

money-metric marginal utilities of attributes (willingness to pay)

consumer-specific, log-normally distributed scale coefficient – introduces heterogeneity into the variance of the error term

ECONOMETRIC APPROACH

Generalised Mixed Logit with scale covariates

- In the basic model (MNL), the error term is assumed to be independent, identically distributed (the same variance for all observations). The scale heterogeneity model allows the variance of the error term to vary across respondents.
- Scale
  - The inverse of the variance of the error term in the utility function
  - Introduces perceived randomness into consumers' choices
  - The higher the scale, the less random consumers' choices (more predictable from the modeller's perspective)
- The generalised model assumes the individual scale to be a random variable.
- Possible systematic differences in the mean scale and in its variance:

$$\delta_n \sim \mathrm{LN}\left(1+\phi' z_n, \tau+\eta' z_n\right)$$

z_n – a treatment-related covariate (a dummy for the three-alternative treatment)

how random/deterministic the respondents appear on average

how differentiated each group of respondents is – do the respondents have similar scale parameters

- It allows for greater flexibility in accounting for scale differences between groups of respondents.

$$\delta_n = \exp(\bar{\delta} + \theta' z_n + \exp(\lambda' z_n)\tau\upsilon_n)$$

Positive ϑ (higher scale) means less uncertainty in respondents' choices on average.

Positive λ (higher scale heterogeneity) represents respondents who are more diversified in terms of how predictable their choices are.

VERIFICATION OF HYPOTHESIS 1

Do marginal WTP differ depending on the number of alternatives?

MODEL 1.

GMXL IN WTP-SPACE WITH PARAMETERS SPECIFIC FOR THE NUMBER OF ALTERNATIVES PER CHOICE TASK

	Means		Standard Deviations	
	Coef.	St. Err.	Coef.	St. Err.
Status quo × 2 alt	5.72***	0.65	5.58***	0.28
Status quo × 3 alt	6.20***	0.63	12.49***	0.70
Iron (-50%) × 2 alt	4.41***	0.60	2.19***	0.32
Iron (-50%) × 3 alt	4.05***	0.31	0.00	0.00
Iron (-75%) × 2 alt	4.27***	0.37	0.40	0.43
Iron (-75%) × 3 alt	4.62***	0.36	0.00	0.00
Iron (-95%) × 2 alt	5.01***	0.46	0.87***	0.24
Iron (-95%) × 3 alt	4.66***	0.34	1.97***	0.23
Chlorine (-80%) × 2 alt	2.71***	0.38	2.37***	0.20
Chlorine (-80%) × 3 alt	2.26***	0.30	4.18***	0.19
Hardness (-33%) × 2 alt	5.52***	0.56	0.09	0.17
Hardness (-33%) × 3 alt	4.49***	0.37	2.31***	0.31
Hardness (-50%) × 2 alt	5.52***	0.39	2.26***	0.33
Hardness (-50%) × 3 alt	6.33***	0.32	2.76***	0.29

GMXL parameters	Coef.	St. Err.
Scale variance	1.17***	0.05
Covariate of scale		
3 alt	-0.21	0.14
Covariate of scale variance		
3 alt	0.06***	0.02

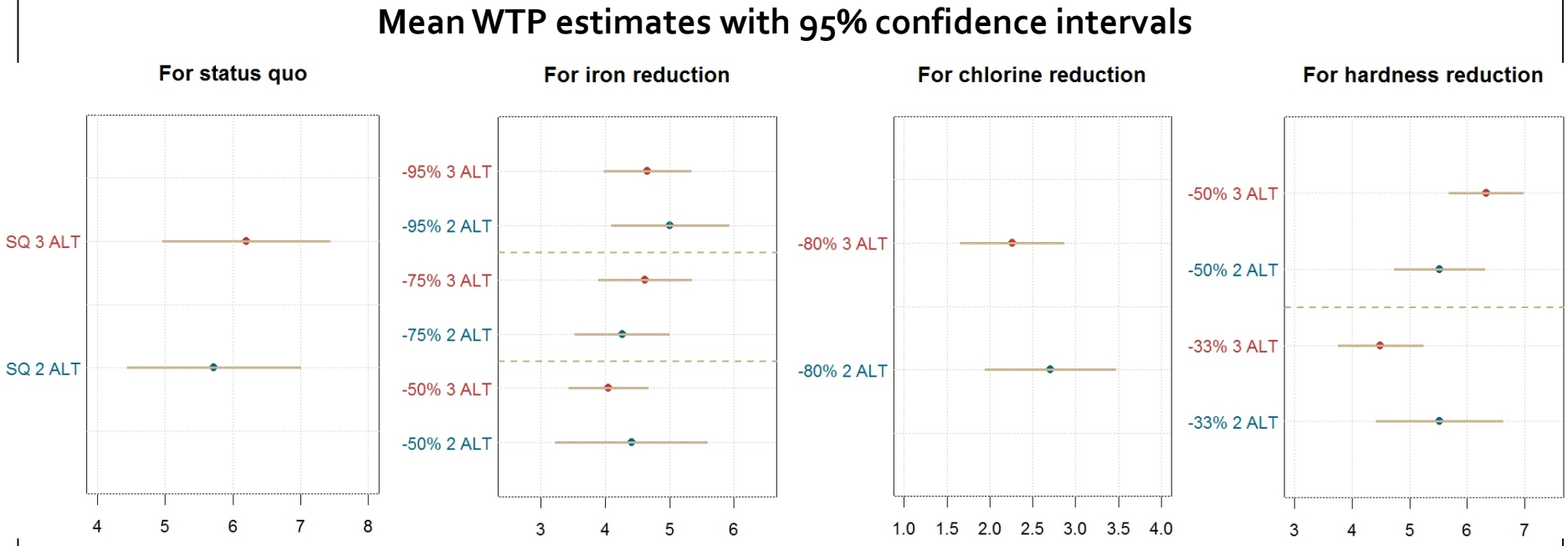
Model characteristics	
LL_0	-5058.61
LL	-3037.31
Pseudo R^2	0.40
AIC/n	0.82
n	7497
k	33

VERIFICATION OF HYPOTHESIS 1

Do marginal WTP differ depending on the number of alternatives?

MODEL 1.

GMXL IN WTP-SPACE WITH PARAMETERS SPECIFIC FOR THE NUMBER OF ALTERNATIVES PER CHOICE TASK



VERIFICATION OF HYPOTHESIS 1

Do marginal WTP differ depending on the number of alternatives?

MODEL 2.

GMXL IN WTP-SPACE WITH EQUAL PARAMETERS FOR DIFFERENT NUMBERS OF ALTERNATIVES PER CHOICE TASK

	Means		Standard Deviations	
	Coef.	St. Err.	Coef.	St. Err.
Status quo	4.81***	0.38	9.72***	0.71
Iron (-50%)	4.21***	0.22	0.00	0.00
Iron (-75%)	5.10***	0.23	0.90***	0.20
Iron (-95%)	4.13***	0.16	2.04***	0.19
Chlorine (-80%)	2.25***	0.15	3.42***	0.12
Hardness (-33%)	4.21***	0.20	0.00	0.00
Hardness (-50%)	5.68***	0.20	3.42***	0.18

Likelihood-ratio test
test statistics: 6.12
p-value: 0.96

Model 2 is chosen.

The restrictions (equal WTP derived from two- and three-alternative responses) do not lower significantly the goodness of model fit.

Marginal WTP do not differ significantly for choice tasks with different numbers of alternatives.

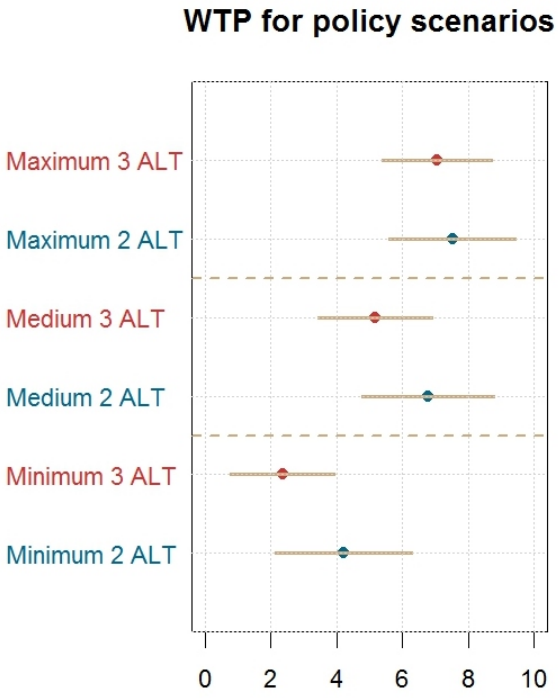
GMXL parameters		
	Coef.	St. Err.
Scale variance	1.22***	0.13
Covariate of scale		
3 alt	0.02	0.14
Covariate of scale variance		
3 alt	0.37***	0.02

Model characteristics	
LL_0	-5058.61
LL	-3040.37
Pseudo R^2	0.40
AIC/n	0.82
n	7497
k	19

VERIFICATION OF HYPOTHESIS 1

Do welfare measures for policy scenarios differ depending on the number of alternatives?

	2 alt		3 alt	
	WTP	St. Err.	WTP	St. Err.
Program maximum ○ -95% Iron ○ -80% Chlorine ○ -50% Hardness	7.52	0.97	7.05	0.84
Program medium ○ -75% Iron ○ -80% Chlorine ○ -33% Hardness	6.77	1.01	5.18	0.87
Program minimum ○ -50% Iron ○ -33% Hardness	4.21	1.04	2.35	0.79



VERIFICATION OF HYPOTHESIS 2

How does the error variance (scale) in the utility function differ depending on the number of alternatives?

Model 2. GMXL in WTP-space with scale covariates

	Means		Standard Deviations	
	Coef.	St. Err.	Coef.	St. Err.
Status quo	4.81***	0.38	9.72***	0.71
Iron (-50%)	4.21***	0.22	0.00	0.00
Iron (-75%)	5.10***	0.23	0.90***	0.20
Iron (-95%)	4.13***	0.16	2.04***	0.19
Chlorine (-80%)	2.25***	0.15	3.42***	0.12
Hardness (-33%)	4.21***	0.20	0.00	0.00
Hardness (-50%)	5.68***	0.20	3.42***	0.18

Model characteristics	
LL_0	-5058.61
LL	-3040.37
Pseudo R^2	0.40
AIC/n	0.82
n	7497
k	19

GMXL parameters	Coef.	St. Err.
Scale variance	1.22***	0.13
Covariate of scale		
3 alt	0.02	0.14
Covariate of scale variance		
3 alt	0.37***	0.02

Model 3. GMXL in WTP-space without scale covariates

	Means		Standard Deviations	
	Coef.	St. Err.	Coef.	St. Err.
Status quo	1.26***	0.01	0.00	0.00
Iron (-50%)	3.28***	0.01	0.00	0.00
Iron (-75%)	3.77***	0.03	0.17***	0.03
Iron (-95%)	3.33***	0.02	0.47***	0.01
Chlorine (-80%)	1.62***	0.01	2.65***	0.02
Hardness (-33%)	3.38***	0.01	1.60***	0.01
Hardness (-50%)	4.22***	0.01	3.33***	0.00

GMXL parameters	Coef.	St. Err.
Scale variance	3.31***	0.02

Model characteristics	
LL_0	-5058.61
LL	-3283.78
Pseudo R^2	0.35
AIC/n	0.88
n	7497.00
k	17.00

Likelihood-ratio test
test statistics: 486.83
p-value: 0.00

Model 2 is better.

Do the standard errors for marginal WTP differ depending on the number of alternatives?

- Coefficient of variation of a parameter estimate = the standard error of the parameter estimate / the estimated parameter

Coefficients of variation		
	2 alt	3 alt
Status quo	0.11	0.10
Iron (-50%)	0.13	0.08
Iron (-75%)	0.09	0.08
Iron (-95%)	0.09	0.07
Chlorine (-80%)	0.14	0.13
Hardness (-33%)	0.10	0.08
Hardness (-50%)	0.07	0.05
Average	0.11	0.08

- WTP estimates for each attribute and the status quo option have smaller standard errors in the three-alternative treatment as compared to the two-alternative treatment.
- Data from three-alternative choice tasks gives more precise estimates.

Conclusions

- Marginal WTP values for each attribute and the status quo option are not significantly different across the two- and three-alternative choice tasks.
- For typical policy scenarios considered, consumers in three-alternative choice tasks state lower (however, not significantly lower) WTP than in two-alternative choice tasks.
- The mean scale (the mean variance of the error term) in the utility function is not different between two- and three-alternative choice tasks.
- The variance of scale is higher in three-alternative choice tasks – those respondents are more diversified in terms of how predictable (from the modeller's perspective) their choices are.
- Although the use of two-alternatives questions is theoretically suggested, in a field study we find
  1. no significant differences in WTP values elicited in two- and three-alternative choice tasks,
  2. no significant differences in the variance of consumers' choices across two- and three-alternative choice tasks,
  3. that estimates based on three-alternative data have smaller standard errors.

Three-alternative choice tasks might provide efficiency gains in preference modelling, while not biasing the results.

How does this study correspond to my PhD research?

My PhD research goal is to **verify empirically** (on the basis of field stated preference surveys) **the importance of** the theoretically suggested **incentive compatibility conditions**:

1. Respondents view the survey as consequential, which means:
   - their responses are seen as influencing agency's actions,
   - they care about the finally introduced solution.
2. The survey has the format of a single two-alternative question with one option being status quo.

In particular, the research questions are:

- Is incentive compatibility indeed necessary for respondents' truthful preference elicitation?
- Which of the conditions (consequentiality, number of choice situations, number of choice alternatives) plays a major role in ensuring truthful preference revelation?
- What is the extent of the bias resulting from violating these conditions in field stated preference research?