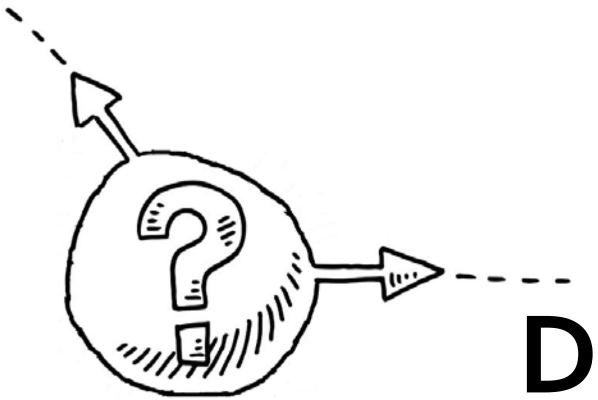


DOES THE NUMBER OF DISCRETE CHOICE ALTERNATIVES MATTER FOR STATED PREFERENCES?

Mikołaj Czajkowski, Marek Giergiczny, Ewa Zawojska

University of Warsaw, Department of Economics

ezawojsk@utk.edu

Stated preference methods

- Used to determine public's preferences, especially towards non-market goods
- Survey-based
- Provide estimates of benefits for cost-benefit analysis
- Help in effective allocation and management of resources
- BUT much skepticism whether survey responses reflect actual preferences

Stated preference methods

- Used to determine public's preferences, especially towards non-market goods
- Survey-based
- Provide estimates of benefits for cost-benefit analysis
- Help in effective allocation and management of resources
- BUT much skepticism whether survey responses reflect actual preferences

When do people answer truthfully in stated preference surveys?

Conditions for incentive compatibility

(Carson and Groves 2007, Vossler et al. 2012, Carson et al. 2014)

Incentive compatibility = Revealing true preferences is the respondent's optimal strategy.

1. The survey is seen as a take-it-or-leave-it offer.
2. The survey involves a yes-no answer on a single project.
   (the Gibbard-Satterthwaite theorem)
3. The authority can enforce payment (coercive payment).
4. The survey is perceived as consequential: Respondents believe that
   - their responses will influence decisions related to the outcome in question,
   - they will be required to pay for that outcome if it is implemented.

Conditions for incentive compatibility

(Carson and Groves 2007, Vossler et al. 2012, Carson et al. 2014)

Incentive compatibility = Revealing true preferences is the respondent's optimal strategy.

1. The survey is seen as a take-it-or-leave-it offer.
2. The survey involves a yes-no answer on a single project.
   (the Gibbard-Satterthwaite theorem)
3. The authority can enforce payment (coercive payment).
4. The survey is perceived as consequential: Respondents believe that
   - their responses will influence decisions related to the outcome in question,
   - they will be required to pay for that outcome if it is implemented.

Later advancements:
- A sequence of questions (Vossler et al. 2012)
- Open-ended format (Holladay and Vossler 2016)

Should we care about the conditions for incentive compatibility?

- Are they important in practice?
- The vast majority of field stated preference surveys do not satisfy the conditions.
- The conditions place important limitations on the survey design.
- Trade-off between incentive compatibility and statistical efficiency.
- BUT our literature review of validity tests of the stated preference methods (Zawojska and Czajkowski, 2015) suggests that:
  - when the conditions are fulfilled, no divergence between stated preferences and true preferences is observed;
  - when they are not fulfilled, many studies report divergence.

Conditions for incentive compatibility

(Carson and Groves 2007, Vossler et al. 2012, Carson et al. 2014)

Incentive compatibility = Revealing true preferences is the respondent's optimal strategy.

1. The survey is seen as a take-it-or-leave-it offer.
2. The survey involves a yes-no answer on a single project. (the Gibbard-Satterthwaite theorem)
3. The authority can enforce payment (coercive payment).
4. The survey is perceived as consequential: Respondents believe that
   - their responses will influence decisions related to the outcome in question,
   - they will be required to pay for that outcome if it is implemented.

Does the number of choice alternatives matter?

Random Utility Model (McFadden, 1974)

FOUNDATION OF PREFERENCE MODELLING BASED ON DISCRETE CHOICE DATA

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

$$U_{njt} = \underbrace{\underbrace{\alpha c_{njt}}_{\text{monetary attribute}} + \underbrace{bX_{njt}}_{\text{non-monetary attributes}}}_{\text{observable characteristics of the good}} + \underbrace{e_{njt}}_{\substack{\text{error term (deviations from the}\\ \text{mean parameters' estimates)}}}$$

- A consumer derives utility from: observable characteristics of the good and unobservable factors (random component)

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 chances to win.
- Increased choice complexity may prompt respondents to avoid making choices at all.

In favor of the use of multiple alternatives

Carson et al. (2011)	Lab	No significant differences in answers to two- and three-alternative tasks.
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 and Bennett (2009)	Field	More robust models on three-alternative data than on two-alternative. A higher rate of "not sure" responses in two-alternative tasks.

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

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 chances to win.

In favor of the use of multiple alternatives

Carson et al. (2011)	Lab	No significant differences in answers to two- and three-alternative tasks.
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 and Bennett (2009)	Field	More robust models on three-alternative data than on two-alternative. A higher rate of "not sure" responses in two-alternative tasks.

- Efficiency gains (more data in a cheaper way).

- Increased choice complexity may prompt respondents to avoid making choices at all.
- More alternatives increase the chances to find a satisfactory option, which makes the choice easier.

Possibly a non-linear impact of the number of alternatives

Evidence on the optimal number of alternatives

On the theoretical basis

Kuksov and Villas-Boas (2010)

- Many alternatives – a consumer has to engage in many searches to find a satisfactory fit; it may be too costly and make the consumer defer taking a choice.
- Few alternatives – a consumer may not search, fearing that an acceptable choice is unlikely, and does not make a choice at all.

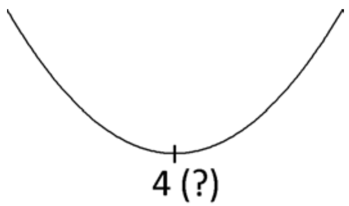
On the empirical basis

Caussade et al. (2005)

DeShazo and Fermo (2002)

Meyerhoff et al. (2014)

A <u>U-shaped pattern</u> of the variance of the error term – up to a threshold number of alternatives (usually 4), the variance decreases and later increases.



OUR RESEARCH QUESTION

Does the number of alternatives matter for stated preferences?

With respect to the two aspects:

1. Do **willingness to pay** (WTP) estimates derived from two- and three-alternative responses differ?
2. Does **the variance of the error term** in the utility function differ for the estimates based on two- and three-alternative data?

Our discrete choice experiment

- A mail survey among residents of Milanowek (a city in the agglomeration of Warsaw, Poland)
- A hypothetical scenario: improvement of tap water quality in Milanowek

	No change	Option 1	Option 2
Iron	As today	50% lower	75% lower
Hardness	As today	50% lower	33% lower
Chlorine	As today	80% lower	As today
Additional cost per month for your household	0 zł	10 zł	70 zł
Your choice	☐	☐	☐

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

- Split sample design:
  - Two-alternative treatment – 403 respondents
  - Three-alternative treatment – 401 respondents
- 12 choice tasks per respondent

Our discrete choice experiment

- A mail survey among residents of Milanowek (a city in the agglomeration of Warsaw, Poland)
- A hypothetical scenario: improvement of tap water quality in Milanowek

	No change	Option 1	Option 2
Iron	As today	50% lower	75% lower
Hardness	As today	50% lower	33% lower
Chlorine	As today	80% lower	As today
Additional cost per month for your household	0 zł	10 zł	70 zł
Your choice	☐	☐	☐

Status quo

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

- Split sample design:
  - Two-alternative treatment – 403 respondents
  - Three-alternative treatment – 401 respondents
- 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 Milanowek	32.69	32.68	0.73
Age	51.59	51.36	0.93
Household size	2.841	2.816	0.90
Household members below 18 years old	0.4543	0.4898	0.93
Litres of bottled water consumed per month	22.15	20.84	0.26

- Chi-squared test of equality of proportions

	p-value
Gender	0.14
Education	0.16
Income	0.12

The null hypothesis of equality cannot be rejected.

The samples do not differ with respect to these characteristics.

ECONOMETRIC APPROACH

Generalized 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, normally distributed scale coefficient – introduces heterogeneity into the variance of the error term

How do we test the role of the number of alternatives?

Impact on the variance of the error term

Impact on the willingness-to-pay estimates

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

scale coefficient

preference parameters (willingness to pay) – coefficients on the dummies for each improvement (e.g., reduction of iron by 50%)

- Scale – the inverse of the variance of the error term
- Shows how random choices of the respondents are
- The higher the scale, the less random the consumers' choices (more predictable)
- We test if the scale depends on a treatment dummy

Three model specifications

- Model 1 with preference parameters equal for both treatments
- Model 2 with the means of preference parameters interacted with a treatment dummy
- Model 3 with treatment-specific preference parameters

The impact of the number of alternatives

- Model 1 with preference parameters equal for both treatments
- Model 2 with the means of preference parameters interacted with a treatment dummy
- Model 3 with treatment-specific preference parameters

The treatment dummy explaining scale – not significant, no significant differences in scale

	Likelihood ratio test statistics	Degrees of freedom	P-value
Model 1 vs. Model 2	2.9017	7	0.8939
Model 1 vs. Model 3	195.9970	107	0.0000
Model 2 vs. Model 3	193.0953	100	0.0000

The impact of the number of alternatives

- Model 1 with preference parameters equal for both treatments
- Model 2 with the means of preference parameters interacted with a treatment dummy
- Model 3 with treatment-specific preference parameters

The treatment dummy explaining scale – not significant, no significant differences in scale

	Likelihood ratio test statistics	Degrees of freedom	P-value
Model 1 vs. Model 2	2.9017	7	0.8939
Model 1 vs. Model 3	195.9970	107	0.0000
Model 2 vs. Model 3	193.0953	100	0.0000

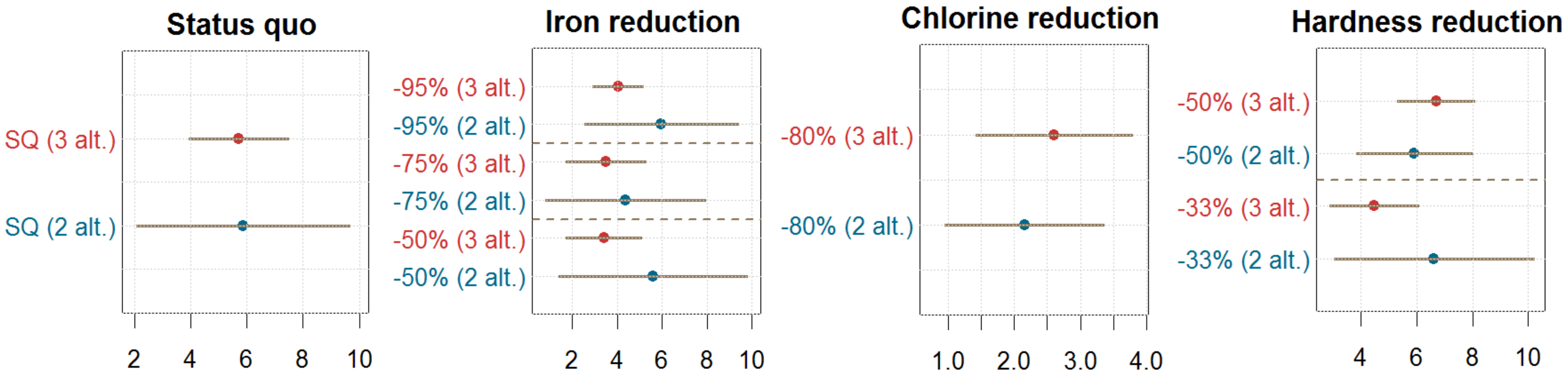
The impact of the number of alternatives

	Two-alternative treatment		Three-alternative treatment	
	Mean (SE)	SD (SE)	Mean (SE)	SD (SE)
Status quo	5.8834*** (1.9195)	7.2904*** (2.3909)	5.7004*** (0.8861)	11.0032*** (1.4410)
Iron -50%	5.6059*** (2.1168)	5.4310*** (1.8271)	3.3985*** (0.8299)	4.5739*** (0.8180)
Iron -75%	4.3652** (1.7940)	5.4945*** (1.5515)	3.4969*** (0.8853)	6.6086*** (0.8738)
Iron -95%	5.9614*** (1.7312)	5.9965*** (1.5079)	4.0400*** (0.5561)	4.6180*** (0.5138)
Chlorine -80%	2.1510*** (0.6100)	5.4932*** (1.1694)	2.5991*** (0.5973)	4.3528*** (0.4201)
Hardness -33%	6.6156*** (1.8176)	7.5041*** (1.9096)	4.4679*** (0.7944)	4.9875*** (0.6936)
Hardness -50%	5.9210*** (1.0470)	10.1080*** (2.1199)	6.6968*** (0.6900)	5.8320*** (0.5426)

Model characteristics	
Log likelihood	-2878.37
McFadden pseudo R^2	0.43
AIC/n	0.81
No. of observations (n)	7497
No. of parameters	152

Do the WTP estimates differ significantly?

Mean WTP estimates with 95% confidence intervals [EUR]



- The intervals for each attribute overlap.
- Narrower intervals for the three-alternative-based estimates.

Do the standard errors differ in the number of alternatives?

- Coefficient of variation of an estimate (VC) = $\frac{\text{Standard error of the estimate}}{\text{Value of the estimate}}$

	VC for the mean		**VC for the SD**	
	Two-alternative	Three-alternative	Two-alternative	Three-alternative
Status quo	0.33	0.16	0.33	0.13
Iron -50%	0.38	0.24	0.34	0.18
Iron -75%	0.41	0.25	0.28	0.13
Iron -95%	0.29	0.14	0.25	0.11
Chlorine -80%	0.28	0.23	0.21	0.10
Hardness -33%	0.27	0.18	0.25	0.14
Hardness -50%	0.18	0.10	0.21	0.09
Cost	1.37	0.44	0.24	0.16
Average	**0.44** >	**0.22**	**0.26** >	**0.13**

- Smaller standard errors of the three-alternative-based estimates.
- Responses to three-alternative choice tasks gives more precise estimates.

Conclusions

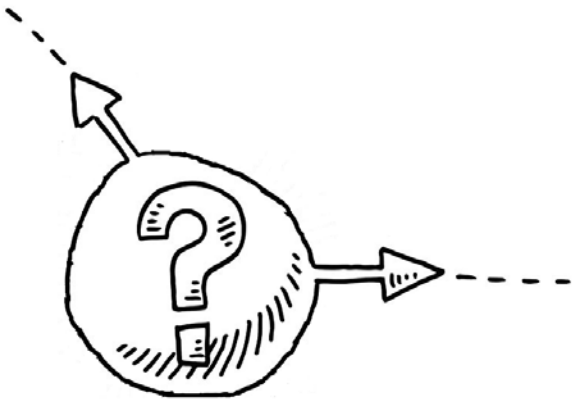
- Marginal WTP do not differ significantly across two- and three-alternative choice tasks.
- No significant differences in scale (the variance of the error term in the utility function).
- Three-alternative-based parameter have smaller standard errors. –> More precise WTP estimates.

Although the use of two-alternatives questions is theoretically suggested, in a field study we find that **three-alternative choice tasks might provide efficiency gains** in preference modelling, while not biasing the results.

Why?

Strategic manipulation in preference disclosure might appear difficult

- under task complexity,
- under uncertainty about preferences of others,
- under uncertainty about the voting rule.



Ewa Zawojska

University of Warsaw, Department of Economics
ezawojsk@utk.edu