Ewa Zawojska, Mikołaj Czajkowski, Marek Giergiczny

University of Warsaw, Department of Economics
ezawojska@wne.uw.edu.pl

DOES THE NUMBER OF ALTERNATIVES MATTER FOR STATED PREFERENCES?

Ewa Zawojska, Mikołaj Czajkowski, Marek Giergiczny

University of Warsaw, Department of Economics
ezawojska@wne.uw.edu.pl

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 <u>state</u> 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
- A flexible method – enables valuation of goods in hypothetical situations.
- Commonly used to elicit public's preferences, especially towards non-market goods (as environmental goods)
- Important for effective allocation and management of resources

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}):

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

monetary attribute

non-monetary attributes

error term (deviations from the mean parameters estimates)

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

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.

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)

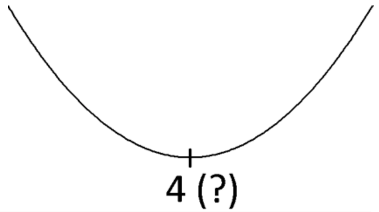
- Increased choice complexity may prompt respondents to avoid making choices at all.
- More alternatives increase the chances to find a satisfactory option, which makes the choice easier.

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.

Our research question

Does the number of alternatives matter for stated preferences?

With respect to two aspects:

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

Valuation of better tap water quality

PREVIOUS STUDIES

- **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?
- **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)

Our study design

- Discrete Choice Experiment
- Mail survey among residents of Milanówek (a city in the agglomeration of Warsaw, Poland)
- A hypothetical scenario: improvement of tap water quality in Milanówek

	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:
  - 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-squared 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 (GMXL) 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 the 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:

$$\delta_n \sim \mathrm{LN}\left(\underbrace{1+\phi' z_n}_{\text{how random/deterministic the respondents appear on average}}, \underbrace{\tau+\eta' z_n}_{\text{how differentiated each group of respondents is – do the respondents have similar scale parameters}}\right)$$

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.

Positive φ (higher scale) means less uncertainty in three-alternative respondents' choices on average.

Positive η (higher scale heterogeneity) means that three-alternative respondents are more diversified in terms of how predictable their choices are.

GMXL in WTP-space with scale covariates and parameters specific for the number of alternatives

	Means		Standard Deviations	
	Coef.	St. Err.	Coef.	St. Err.
Status quo × 2 alt	5.08***	0.66	5.61***	0.48
Status quo × 3 alt	6.42***	0.57	11.66***	0.72
Iron (-50%) × 2 alt	4.36***	0.65	2.36***	0.33
Iron (-50%) × 3 alt	4.03***	0.30	0.17	0.42
Iron (-75%) × 2 alt	4.31***	0.35	0.32	0.90
Iron (-75%) × 3 alt	5.26***	0.30	0.09	0.35
Iron (-95%) × 2 alt	4.79***	0.51	1.27***	0.46
Iron (-95%) × 3 alt	4.39***	0.29	2.33***	0.25
Chlorine (-80%) × 2 alt	2.79***	0.32	1.85***	0.33
Chlorine (-80%) × 3 alt	2.56***	0.22	3.69***	0.19
Hardness (-33%) × 2 alt	5.13***	0.56	0.47	0.83
Hardness (-33%) × 3 alt	4.35***	0.32	1.61***	0.35
Hardness (-50%) × 2 alt	5.44***	0.48	2.41***	0.43
Hardness (-50%) × 3 alt	6.24***	0.28	2.98***	0.30

GMXL parameters		
	Coef.	St. Err.
Scale variance	1.61***	0.18
Covariate of scale φ		
3 alt	1.09***	0.05
Covariate of scale variance η		
3 alt	1.45***	0.02

Model characteristics	
LL	-2997.63
Pseudo R^2	0.41
AIC/n	0.81
n	7497
k	33

GMXL in WTP-space with scale covariates and parameters specific for the number of alternatives

	Means		**Standard Deviations**	
	Coef.	St. Err.	Coef.	St. Err.
Status quo × 2 alt	5.08***	0.66	5.61***	0.48
Status quo × 3 alt	6.42***	0.57	11.66***	0.72
Iron (-50%) × 2 alt	4.36***	0.65	2.36***	0.33
Iron (-50%) × 3 alt	4.03***	0.30	0.17	0.42
Iron (-75%) × 2 alt	4.31***	0.35	0.32	0.90
Iron (-75%) × 3 alt	5.26***	0.30	0.09	0.35
Iron (-95%) × 2 alt	4.79***	0.51	1.27***	0.46
Iron (-95%) × 3 alt	4.39***	0.29	2.33***	0.25
Chlorine (-80%) × 2 alt	2.79***	0.32	1.85***	0.33
Chlorine (-80%) × 3 alt	2.56***	0.22	3.69***	0.19
Hardness (-33%) × 2 alt	5.13***	0.56	0.47	0.83
Hardness (-33%) × 3 alt	4.35***	0.32	1.61***	0.35
Hardness (-50%) × 2 alt	5.44***	0.48	2.41***	0.43
Hardness (-50%) × 3 alt	6.24***	0.28	2.98***	0.30

GMXL parameters		
	Coef.	St. Err.
Scale variance	1.61***	0.18
Covariate of scale φ		
3 alt	1.09***	0.05
Covariate of scale variance η		
3 alt	1.45***	0.02

Model characteristics	
LL	-2997.63
Pseudo R^2	0.41
AIC/n	0.81
n	7497
k	33

Less uncertainty in three-alternative choices on average

Three-alternative respondents are more diversified in terms of how predictable their choices are.

GMXL in WTP-space with scale covariates and parameters specific for the number of alternatives

	Means		Standard Deviations	
	Coef.	St. Err.	Coef.	St. Err.
Status quo × 2 alt	5.08***	0.66	5.61***	0.48
Status quo × 3 alt	6.42***	0.57	11.66***	0.72
Iron (-50%) × 2 alt	4.36***	0.65	2.36***	0.33
Iron (-50%) × 3 alt	4.03***	0.30	0.17	0.42
Iron (-75%) × 2 alt	4.31***	0.35	0.32	0.90
Iron (-75%) × 3 alt	5.26***	0.30	0.09	0.35
Iron (-95%) × 2 alt	4.79***	0.51	1.27***	0.46
Iron (-95%) × 3 alt	4.39***	0.29	2.33***	0.25
Chlorine (-80%) × 2 alt	2.79***	0.32	1.85***	0.33
Chlorine (-80%) × 3 alt	2.56***	0.22	3.69***	0.19
Hardness (-33%) × 2 alt	5.13***	0.56	0.47	0.83
Hardness (-33%) × 3 alt	4.35***	0.32	1.61***	0.35
Hardness (-50%) × 2 alt	5.44***	0.48	2.41***	0.43
Hardness (-50%) × 3 alt	6.24***	0.28	2.98***	0.30

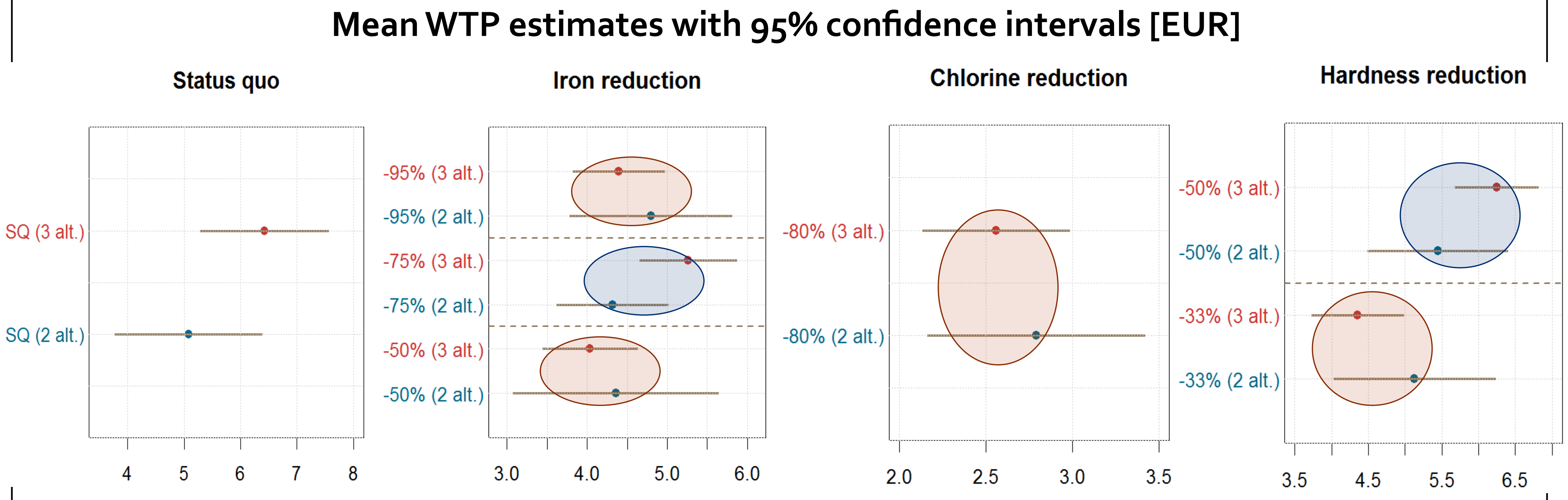
GMXL parameters

	Coef.	St. Err.
Scale variance	1.61***	0.18
Covariate of scale φ		
3 alt	1.09***	0.05
Covariate of scale variance η		
3 alt	1.45***	0.02

Model characteristics

LL	-2997.63
Pseudo R^2	0.41
AIC/n	0.81
n	7497
k	33

Do the WTP estimates differ significantly?

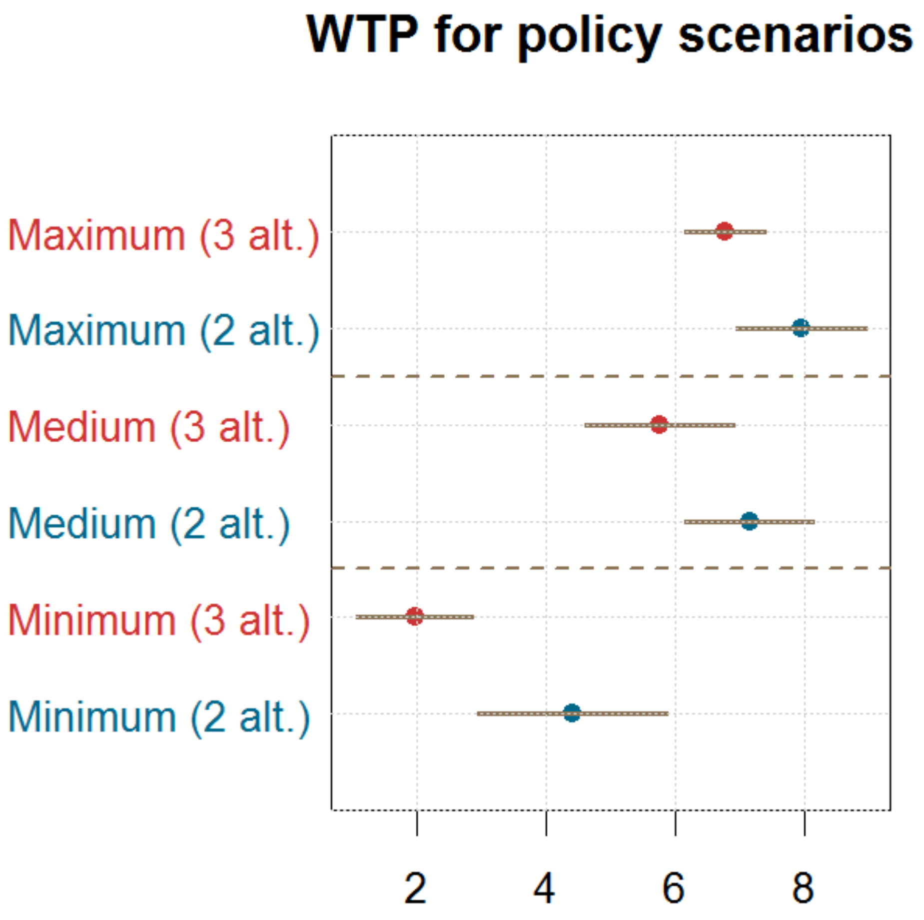


Lower mean WTP in the three-alternative treatment

Higher mean WTP in the three-alternative treatment

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

	2 alt.		3 alt.	
	WTP	St. Err.	WTP	St. Err.
Program maximum ○ -95% Iron ○ -80% Chlorine ○ -50% Hardness	7.95	0.50	6.77	0.30
Program medium ○ -75% Iron ○ -80% Chlorine ○ -33% Hardness	7.15	0.50	5.75	0.58
Program minimum ○ -50% Iron ○ -33% Hardness	4.41	0.74	1.96	0.49



Do the standard errors for marginal WTP differ in 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	0.13	0.09
Iron (-50%)	0.15	0.07
Iron (-75%)	0.08	0.06
Iron (-95%)	0.11	0.07
Chlorine (-80%)	0.11	0.08
Hardness (-33%)	0.11	0.07
Hardness (-50%)	0.09	0.05
Average	0.11	0.07

- WTP estimates for each attribute and the status quo option have smaller standard errors in the three-alternative data than in the two-alternative data.
- Responses to three-alternative choice tasks gives more precise estimates.

Conclusions

- With respect to the WTP values:
  - **Marginal WTP** estimates **do not differ significantly** across 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.
  - WTP estimates based on **three-alternative** data have **smaller standard errors**.
- With respect to scale (the variance of the error term in the utility function):
  - Higher mean scale in **three-alternative choice** tasks – **less uncertainty** in three-alternative respondents' choices on average
  - Higher variance of scale in three-alternative choice tasks – three-alternative 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 that three-alternative choice tasks might provide efficiency gains in preference modelling, while not biasing the results.
- Possibly, because respondents do not engage in analysing which alternatives are most likely adopted (no strategic voting), like they do in election voting when expectations are well-formed.

Thank you for attention

Ewa Zawojska

University of Warsaw, Department of Economics

ezawojska@wne.uw.edu.pl