VALUING TAP WATER QUALITY IMPROVEMENTS USING STATED PREFERENCE METHODS: DOES THE NUMBER OF DISCRETE CHOICE ALTERNATIVES MATTER?

Ewa Zawojska, Mikołaj Czajkowski, and Marek Giergiczny

University of Warsaw, Faculty of Economic Sciences

ewa.zawojska@uw.edu.pl

Stated preference methods in nonmarket valuation

- **Overall objective:**
  - Identify people's preferences, particularly for nonmarket goods, like clean air and nature conservation
  - Estimate economic values, for example, to assess the benefits of policy implementation
- **Applications:**
  - Input for benefit-cost analyses
  - Effective allocation and management of resources based on public preferences
  - Used in various settings, ranging from environmental to public health policy
- **Data:**
  - **Data collection**: Primarily utilize surveys to gather data
  - **Subject to skepticism**: Ongoing debate about the accuracy of survey responses in reflecting actual preferences under various conditions
- **Preference elicitation formats:**
  - Different formats are observed to generate diverse value estimates
  - Formats determine survey complexity, respondents' engagement, perceived incentives, etc.

Discrete choice experiments (DCEs)

- Most common among stated preference approaches
- Present a sequence of choice tasks to respondents
- Each task with a few choice options:

Policy option	No policy implemented (status quo, SQ)

→ Preferred for mitigating strategic responding and complexity

Policy option A	Policy option B	…	No policy implemented (status quo, SQ)

→ Preferred for efficiency reasons and improved preference matching

- How many choice options to include per task?

Numerous studies have empirically explored how the number of choice options affects decisions

From Weng et al. (2021, *Ecological Economics*, 182, 106904)

Differences across number-of-alternative studies.

	Number of non-SQ alternatives	Public or private good	Estimation method	Effects of non-SQ alternatives
DeShazo and Fermo (2002)	1 to 6 and 5 to 8	Public	Heteroskedastic multinomial logit model	Error variance first decreases, then increases, as alternatives increase
Arentze et al. (2003)	1, 2	Private	Multinomial logit model	No change on error variance, measured values of weights or goodness of fit
Hensher (2004)	1,2,3	Private	Uncorrelated mixed logit model	Significant difference on WTP
Caussade et al. (2005)	2,3,4	Private	Heteroskedastic logit model	Significant difference on error variances
Hensher (2006)	1,2,3	Private	Multinomial logit model	Significant difference on mean WTP, but not on variance of WTP
Boyle and Özdemir (2009)	1, 2	Public	Conditional logit model	Significant difference in preference parameters, but not scale
Rolfe and Bennett (2009)	1, 2	Public	Multinomial logit model	More significant coefficients and better model fit with SQ + 2
Rose et al. (2009)	2,3,4	Private	Mixed multinomial logit model	Significant impacts on WTP
Meyerhoff et al. (2015)	2,3,4	Public	Heteroskedastic logit model	Significant effect on scale, increases and then decreases
Collins and Vossler (2009)	1, 2	Private	Mixed logit model in WTP space	Significant differences on WTP and SQ bias
Volinskiy et al. (2009)	1, 2	Private	Hierarchical regression model	Subjects more likely to choose SQ and some attribute coefficients increase in SQ + 2
Zhang and Adamowicz (2011)	1st analysis 1, 2	Public	Conditional logit, random parameters logit, and mixed logit in WTP space models	Cannot pool SQ + 1 and SQ + 2 data, and more likely to choose SQ in SQ + 1 treatment
	2nd analysis 1, 2	Public	Conditional logit and random parameters logit models	Can pool SQ + 1 and SQ + 2 data with full interaction of context variables, mixed results for interaction with ASC, cannot pool with interaction with scale parameter – context variables address matching and complexity
Oehlmann et al. (2017)	2,3,4	Public	Mixed logit model with error components	Significant effect on choice of SQ alternative as number of alternatives increases and significant differences in WTP

From Weng et al. (2021, *Ecological Economics*, 182, 106904)

Differences across number-of-alternative studies.

	Number of non-SQ alternatives	Public or private good	Estimation method	Effects of non-SQ alternatives
DeShazo and Fermo (2002)	1 to 6 and 5 to 8	Public	Hete skedastic multinomial logit model	Error variance rst decreases, then increases, as alternatives increase
Arentze et al. (2003)	1, 2			easured values of weights or goodness of fit
Hensher (2004)	1,2,3			Significant difference on WTP
Caussade et al. (2005)	2,3,4			Significant difference on error variances
Hensher (2006)	1,2,3			Significant difference on mean WTP, but not on variance of WTP
Boyle and Özdemir (2009)	1, 2			Significant difference in preference parameters, but not scale
Rolfe and Bennett (2009)	1, 2			More significant coefficients and better model fit with SQ + 2
Rose et al. (2009)	2,3,4			Significant impacts on WTP
Meyerhoff et al. (2015)	2,3,4			Significant effect on scale, increases and then decreases
Collins and Vossler (2009)	1, 2	Private	Mixed logit model in WTP space	Significant differences on WTP and SQ bias
Volinskiy et al. (2009)	1, 2	Private	Hierarchical regression model	Subjects more likely to choose SQ and some attribute coefficients increase in SQ + 2
Zhang and Adamowicz (2011)	1st analysis 1, 2	Public	Conditional logit, random parameters	Cannot pool SQ + 1 and SQ + 2 data, and more likely to choose SQ in SQ +
	2nd analysis 1, 2			ontext variables, nteraction with scale parameter – context variables address matching and complexity
Oehlmann et al. (2017)	2,3,4	Public	Mixed logit model with error components	Significant effect on choice of SQ alternative as number of alternatives increases and significant differences in WTP

- A variety of models are employed, yet there is a noticeable lack of comprehensive evaluations of the results' robustness across different model specifications
- Many studies utilize multinomial logit models, including their heteroskedastic variants

- Mixed results

- Can variations in model specifications account for the mixed results?
- How robust are the findings across different model specifications?

What we do

- Comparative analysis of DCE choices across two preference elicitation formats
- **Formats compared**:
  - One policy option and the status quo (1OPT+SQ)
  - Two policy options and the status quo (2OPT+SQ)
- **Objective**:
  - Investigate variations in DCE choices between the two formats using 22 distinct model specifications

Model specifications

- All based on the random utility framework (McFadden, 1974)
- Utility derived by consumer *n* choosing option *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

error term

monetary parameter

preference parameters

marginal willingness to pay parameters

scale coefficient – introduces heterogeneity into the variance of the error term

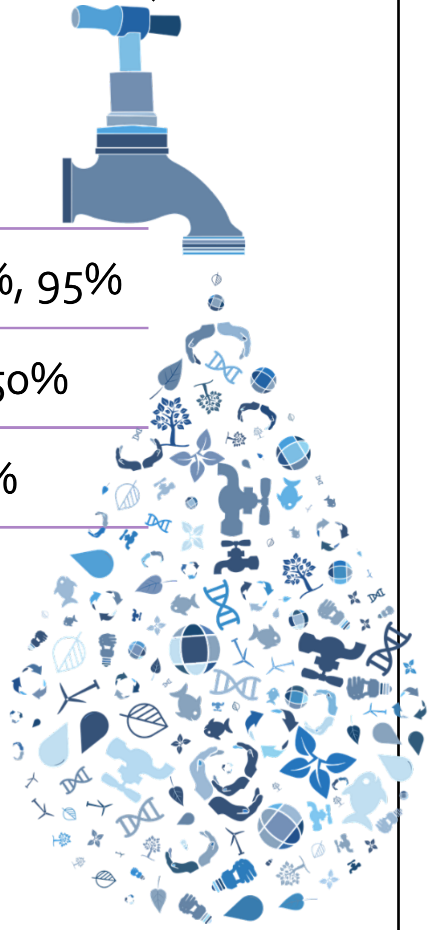
- The model specifications vary along the following dimensions:
  - multinomial logit model (**MNL**), mixed logit model with uncorrelated attributes (**MXL_un**), or mixed logit model with correlated attributes (**MXL**)
  - normal or log-normal distributions of the random parameters (with the monetary parameter always defined as log-normal and the status quo constant defined as normal)
  - preference space or willingness-to-pay (**WTP**) space
  - the DCE format (i.e., a dummy for 2OPT+SQ) explaining differences in scale (variance of the error term; **Xs**), means of the parameters (**Xm**), or both (**Xs+Xm**)

Data

- 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:
  - 1OPT+SQ treatment – 340 respondents
  - 2OPT+SQ treatment – 353 respondents
- 12 choice tasks per respondent

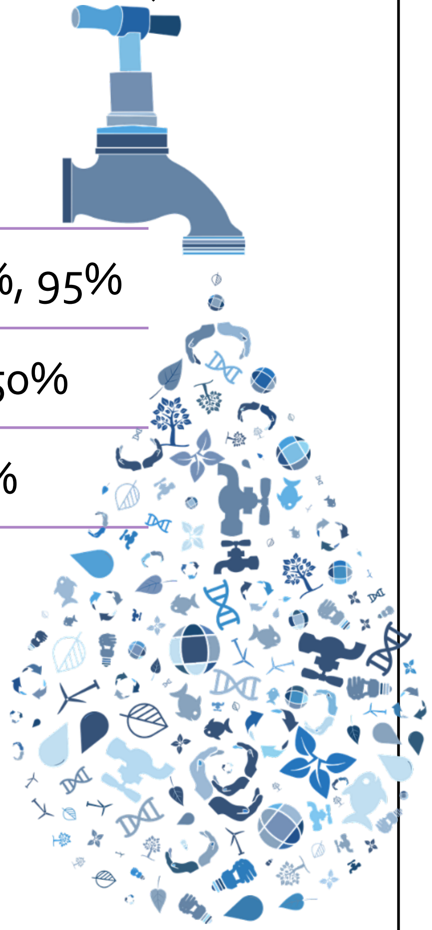
Data

- 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 (SQ)

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



- Split-sample design:
  - 1OPT+SQ treatment – 340 respondents
  - 2OPT+SQ treatment – 353 respondents
- 12 choice tasks per respondent

Do the split samples differ in observed characteristics?

- Wilcoxon rank-sum test of equality of distributions

	Sample means		
	1OPT+SQ	2OPT+SQ	p-value
Years lived in Milanowek	31.7	32.0	0.90
Age	50.4	50.0	0.78
Household size	2.9	2.9	0.81
Household members below 18 years old	0.46	0.50	0.99
Litres of bottled water consumed per month	24.9	23.1	0.96

- Chi-squared test of equality of proportions

	p-value
Male	0.30
Education (4 categories)	0.23
Income (7 categories)	0.15

The null hypothesis of equality cannot be rejected.

The samples do not differ with respect to these characteristics.

Fit of the model specifications to the data – Log-likelihood values

DCE format explaining:		Preference space		WTP space	
		Normal	Log-normal	Normal	Log-normal
scale	MNL Xs	-4444.48		-4444.48	
	MXL_un Xs	-2952.69	-2960.58	-3080.85	-3073.12
	MXL Xs	-2849.77	-2826.02	-2941.18	-2848.78
means	MNL Xm	-4436.57		-4436.57	
	MXL_un Xm	-2948.44	-2955.33	-3075.30	-3071.46
	MXL Xm	-2843.73	-2823.85	-2927.65	-2839.95
both	MXL_un Xs+Xm	-2939.71	-2950.58		
	MXL Xs+Xm	-2839.64	-2816.09		

- In WTP space, the log-normal distribution fits better
- In preference space, the normal distribution fits better for MXL_un and the log-normal distribution fits better for MXL

Let us see an example...

Fit of the model specifications to the data – Log-likelihood values

DCE format explaining:		Preference space		WTP space	
		Normal	Log-normal	Normal	Log-normal
scale	MNL Xs	-4444.48		-4444.48	
	MXL_un Xs	-2952.69	-2960.58	-3080.85	-3073.12
	MXL Xs	-2849.77	-2826.02	-2941.18	-2848.78
means	MNL Xm	-4436.57		-4436.57	
	MXL_un Xm	-2948.44	-2955.33	-3075.30	-3071.46
	MXL Xm	-2843.73	-2823.85	-2927.65	-2839.95
both	MXL_un Xs+Xm	-2939.71	-2950.58		
	MXL Xs+Xm	-2839.64	-2816.09		

- In WTP space, the log-normal distribution fits better
- In preference space, the normal distribution fits better for MXL_un and the log-normal distribution fits better for MXL

Let us see an example…

One specification for illustration: MXL Xs+Xm in preference space with log-normal distributions

Xm

	Dist.	Means		Standard deviations		Means interacted with 2OPT+SQ	
Status quo	n	0.21***	(0.05)	0.57***	(0.06)	0.11**	(0.05)
Iron 80 µg/l (-50%)	l	0.37***	(0.06)	0.9***	(0.29)	-0.02***	(0)
Iron 40 µg/l (-75%)	l	0.36***	(0.06)	0.98***	(0.36)	0.13***	(0.02)
Iron 8 µg/l (-95%)	l	0.45***	(0.07)	1.09***	(0.32)	-0.08***	(0.01)
Chlorine 0.8 µg/l (-80%)	l	0.28***	(0.04)	0.51***	(0.12)	-0.05***	(0.01)
Hardness 20°f (-33%)	l	0.43***	(0.07)	0.76***	(0.18)	-0.03***	(0)
Hardness 15°f (-50%)	l	0.58***	(0.08)	0.91***	(0.19)	-0.06***	(0.01)
Cost (EUR)	l	1.65***	(0.27)	4.65***	(1.45)	-0.56***	(0.09)
Covariates of scale							
2OPT+SQ		0.41***	(0.13)				

Xs

Results from preference-space models

Effects of the 2OPT+SQ on scale and marginal utilities (reference: 1OPT+SQ)

	Scale		**Cost sensitivity**		**Status quo**	
	Normal	Log-normal	Normal	Log-normal	Normal	Log-normal
MNL Xs	insign.		n/a	n/a	n/a	n/a
MXL_un Xs	+	insign.	n/a	n/a	n/a	n/a
MXL Xs	+	+	n/a	n/a	n/a	n/a
MNL Xm	n/a		-		insign.	
MXL_un Xm	n/a	n/a	insign.	-	-	-
MXL Xm	n/a	n/a	insign.	-	insign.	insign.
MXL_un Xs+Xm	+	+	-	-	-	-
MXL Xs+Xm	+	+	-	-	insign.	+

- Effects on scale:
  - In most cases, higher scale (lower variance of the error term) in 2OPT+SQ
  - But we do not find the result in the (commonly used) MNL
- Effects on cost sensitivity: In most cases, lower cost sensitivity in 2OPT+SQ
- Effects on the status quo parameter: Inconsistent

Results from preference-space models

Effects of the 2OPT+SQ on marginal utilities (reference: 1OPT+SQ)

	MNL Xm	MXL_un Xm		MXL Xm		MXL_un Xs+Xm		MXL Xs+Xm	
		Normal	Log-norm.	Normal	Log-norm.	Normal	Log-norm.	Normal	Log-norm.
Status quo	insign.	-0.1209	-0.1065	insign.	insign.	-0.1044	-0.0878	insign.	0.1052
Iron -50%	insign.	-0.1050	-0.1173	insign.	0.0449	-0.1348	-0.1426	-0.0920	-0.0159
Iron -75%	insign.	insign.	0.0060	insign.	0.1729	-0.0572	-0.0519	insign.	0.1259
Iron -95%	-0.0366	-0.9347	-0.1017	insign.	-0.0297	-0.1362	-0.1414	-0.0964	-0.0758
Chlorine -80%	-0.0275	insign.	-0.0205	insign.	-0.0086	-0.0734	-0.0567	-0.0623	-0.0491
Hardness -33%	-0.0576	-0.8479	-0.0767	insign.	0.0414	-0.1350	-0.1240	insign.	-0.0334
Hardness -50%	insign.	insign.	-0.0570	insign.	0.0534	-0.1341	-0.1279	-0.0757	-0.0590

- Effects on the marginal utilities:
  - In most cases, there is a statistically significant divergence
  - Especially for the preferred specifications
  - However, the directions of the effects vary

→ Lack of convergent validity of the marginal utility parameters

Comparison of WTP estimates across the formats

Effects of the 2OPT+SQ on marginal WTP (reference: 1OPT+SQ)

	MNL Xm	MXL_un Xm		MXL Xm	
		Normal	Log-norm.	Normal	Log-norm.
Status quo	insign.	-0.2144	insign.	insign.	0.0952
Iron -50%	insign.	insign.	insign.	insign.	insign.
Iron -75%	insign.	insign.	insign.	insign.	0.3277
Iron -95%	insign.	-0.1356	insign.	insign.	insign.
Chlorine -80%	insign.	insign.	insign.	insign.	insign.
Hardness -33%	insign.	insign.	insign.	insign.	0.1819
Hardness -50%	insign.	insign.	insign.	0.0941	insign.

- Results of WTP-space models (the format interactions with the means):
  - Inconsistent effects on WTP for the status quo
  - Inconsistent effects on marginal WTP for the attributes – While lack of differences dominates, there are three significant differences in the most preferred specification (as based on the log-likelihood value)
- Simulated WTP:
  - Preference-space models: Even fewer differences in simulated marginal WTP
  - Simulated WTP for two programs (small and large): Hardly any differences; The only difference across the DCE formats for the small program in MNL

Summary of our main findings

1. Higher scale (lower variance of the error term) in 2OPT+SQ
2. Lower cost sensitivity in 2OPT+SQ
3. Many statistically significant (and often inconsistent across model specifications) differences in marginal utility parameters across the 2OPT+SQ and 1OPT+SQ formats

Main takeaway:

There are few differences in WTP estimates across the examined DCE formats, considering the variability driven by model specifications

Few differences in WTP estimates: Does it not matter which format to use?

Other considerations for the DCE format choice:

Potentially in favor of 1OPT+SQ:

- incentive compatibility
- task complexity
- fatigue
- ...

Potentially in favor of 2OPT+SQ:

- precision of the estimates
- preference learning
- preference matching
- ...

Few differences in WTP estimates: Does it not matter which format to use?

Other considerations for the DCE format choice:

Potentially in favor of 1OPT+SQ:

- incentive compatibility
- task complexity
- fatigue
- ...

Potentially in favor of 2OPT+SQ:

- precision of the estimates
- preference learning
- preference matching
- ...

- Potentially captured through the standard errors of the WTP estimates
- In many cases, the standard errors are lower in 2OPT+SQ
- But not unambiguously, particularly for the specifications preferred based on the log-likelihood values
- Hence, unclear precision gains from 2OPT+SQ

Few differences in WTP estimates: Does it not matter which format to use?

Other considerations for the DCE format choice:

Potentially in favor of 1OPT+SQ:

- incentive compatibility
- task complexity
- fatigue
- ...

Potentially in favor of 2OPT+SQ:

- precision of the estimates
- preference learning
- preference matching
- ...

- Potentially captured through standard deviations (coefficients of variation) of the WTP estimates
- The assessment is based on the standard deviations from separate models for the two formats
- Most often the standard deviations are larger in 2OPT+SQ, which can signal higher complexity

Few differences in WTP estimates: Does it not matter which format to use?

Other considerations for the DCE format choice:

Potentially in favor of 1OPT+SQ:

- incentive compatibility
- task complexity
- fatigue

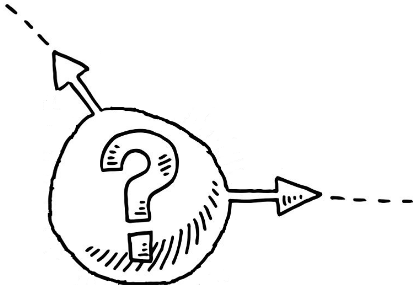
Potentially in favor of 2OPT+SQ:

- precision of the estimates
- preference learning
- preference matching

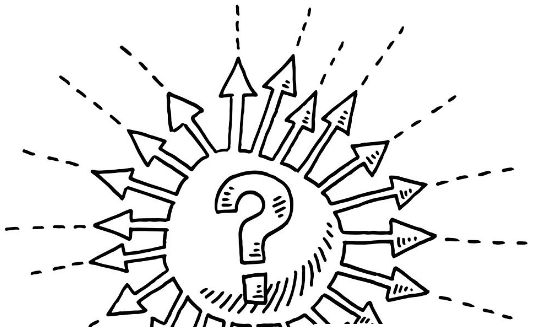
- Potentially captured through status quo choices
- If additional options help people identify a desirable one, likely fewer SQ choices. If additional options increase complexity, likely more SQ choices.
- In 9 (out of 12) choice tasks, more SQ choices in 1OPT+SQ. On average, 72% of SQ choices per task in 1OPT+SQ and 62% in 2OPT+SQ → Perhaps an indication of preference matching

Conclusions

- **Impact of model specifications**: Variations in model specifications can contribute to the mixed results observed in the existing studies
- **Convergent validity issues**: There is a noticeable lack of convergent validity in marginal utility parameters between the two DCE formats
- **WTP estimates**: While minimal differences are observed in WTP estimates between the 2OPT+SQ and 1OPT+SQ formats, designers of DCEs must consider other factors, such as the risk of strategic responses and the complexity of tasks



THANK YOU!



Ewa Zawojska

University of Warsaw, Faculty of Economic Sciences

ewa.zawojska@uw.edu.pl