Accounting for uncertainty in non-industrial private forest (NIPF) owner segmentation

Andrej Ficko & Andrej Boncina

University of Ljubljana, Biotechnical Faculty, Department of Forestry and Renewable Forest Resources, Vecna pot 83, 1000 Ljubljana, Slovenia

Main sources of uncertainty in NIPF owner segmentation:

In survey-based customer segmentation: two main sources:

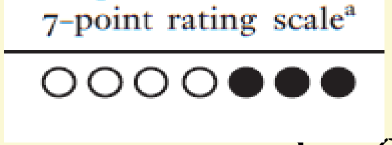
- 1) uncertainty about whether responses reflect the real opinion of a respondent or are biased (**respondent uncertainty**);
- 2) uncertainty of the researcher about the number of customer segments, their meaning, and customer membership (**analyst uncertainty**).

- **Analyst uncertainty** (e.g. Expectation Maximization (EM) clustering (Dempster et al., 1977):
  → Ficko & Boncina (2013). Forest Policy and Economics, 27, 34-43.
  → Book of abstracts (Umea conference).
- This presentation: focus on methods for accounting for **respondent uncertainty**.

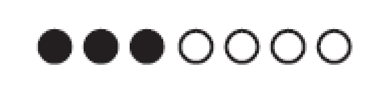
Responses may be biased!

Common response styles identified in social or marketing science literature:

1. the **acquiescence response style** (ARS) = the tendency to agree with the item irrespective of the content of that item;



2. the disacquiescence response style (DARS) = consistent disagreement with the items irrespective of their content;



3. and extreme responding (ERS) = a preference for extreme response categories.

How can respondent uncertainty in NIPF owner segmentation be accounted for?

1. **measured** directly with a follow-up rating question on certainty immediately after the valuation question (Shaikh et al., 200
   - e.g. On a scale of 1 to 10, how certa… answer to the previous [valuation] qu…

Several approaches!

Self-reported uncertainty!

2. **simulated** by skewing the distribution of the responses or by recoding the responses and… continuing with the procedures us… …istorted data

Conservative

Certainty???

How can respondent uncertainty in NIPF owner segmentation be accounted for?

3. **diagnosed** as latent response style behavior by means of several techniques (a review of VanVaerenberg & Thomas (20 developed mainly in behavioral, social, and research:

   – Structural equation modeling (SEM), e.g. Billiet & McClendon (2000).

Case study for a demonstration of 2 approaches

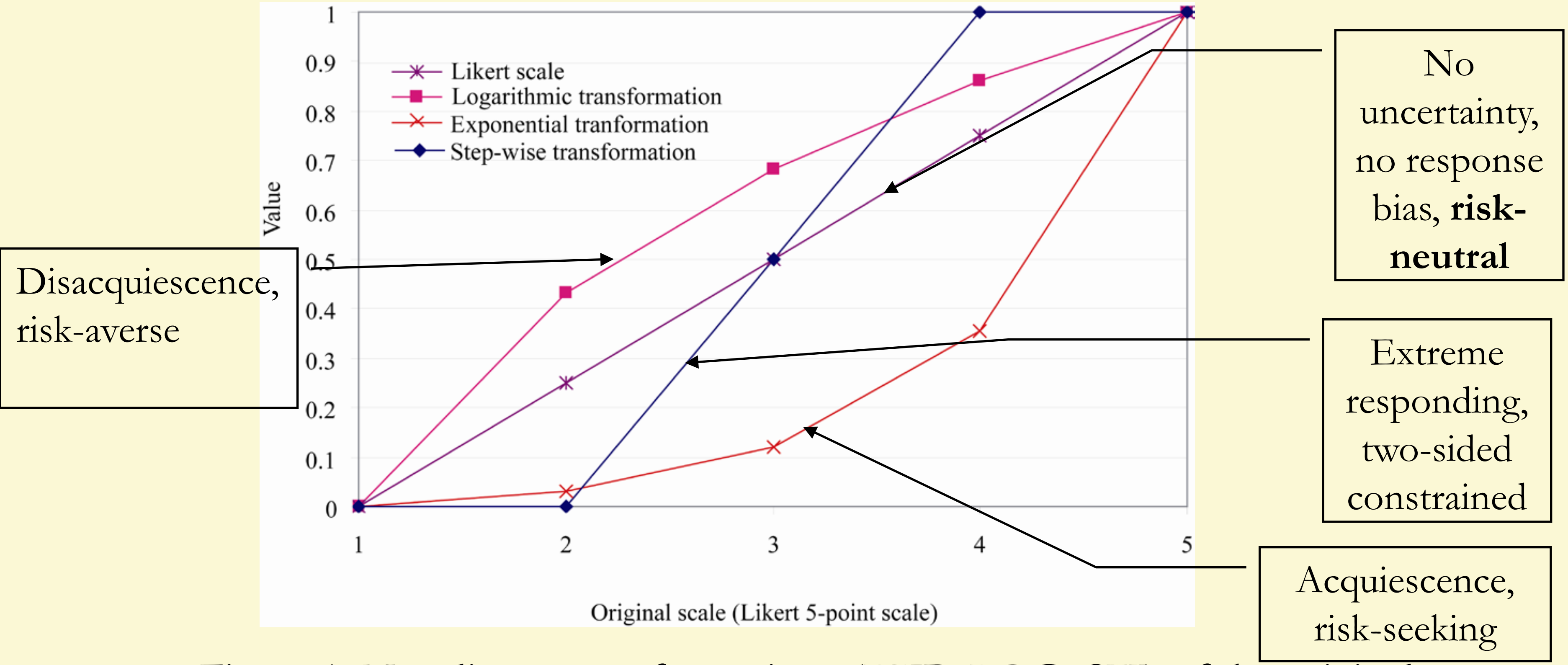
- N = 364 face-to-face interviews
- N = 19 items questioning about the relevance of certain types of information (Likert scale)

Manifest variables (v_i):	
v_1	Costs of forest operations
v_2	Profitability of forest management
v_3	Possibilities for hiring wood harvesting companies and the cost
v_4	Possibilities for mechanized harvesting
v_5	Bucking techniques
v_6	Wood prices and wood markets
v_7	Possible cut for each individual parcel
v_8	Silvicultural measures
v_9	Forest protection and bark beetle prevention
v_{10}	Current market price of forest land
v_{11}	Property boundaries
v_{12}	Locations of all parcels
v_{13}	Possibilities and costs of forest road building
v_{14}	Rights and duties of forest possession
v_{15}	Public rights on owner's holding
v_{16}	Game species and population densities
v_{17}	Management restrictions due to nature protection
v_{18}	Allowable cut
v_{19}	Contact with a person in charge of cutting approval

- Clustering of owners into decision-making types

1. Simulating the respondent uncertainty

- by skewing the distribution of the responses or by recoding the responses



- Figure 1: Non-linear transformations (EXP, LOG, SW) of the original response values

K-means clustering of owners: membership similarities

A) k-means clustering							Pair counting measure of similarity	Information theoretic measure of similarity
		LIN						
	Cl.	1	2	3	4	5		
		17,3	17,0	4,1	29,1	32,4	Cramer's V	Asymetric NMI
LOG	1	12,4	0,5	0	0	8,2	0.605*	0.376*
	2	3	3,8	1,1	5,5	1,9		
	3	0	0	1,9	0	0,3		
	4	1,9	1,4	1,1	23,6	22		
	5	0	11,3	0	0	0		
SW	1	7,7	0,8	0	0	7,4	0.452*	0,269*
	2	3,8	4,9	1,4	6,6	4,9		
	3	3,6	0,8	1,4	2,2	0,8		
	4	1,6	0,3	1,1	19,2	19,2		
	5	0,5	10,2	0,3	1,1	0		
EXP	1	9,6	1,6	2,2	4,4	4,7	0.405*	0.219*
	2	1,1	10,7	0	3,8	2,5		
	3	1,6	3,3	1,9	1,4	8,8		
	4	0	1,4	0	10,2	2,5		
	5	4,9	0	0	9,3	14		
RD	1	0,8	12,6	0	0,8	0	0.524*	0.330*
	2	10,4	0,3	0	0,8	8,8		
	3	1,6	0,8	2,2	0	3		
	4	3	2,5	1,4	8,5	2,5		
	5	1,4	0,8	0,5	19	18,1		

* **denotes approx. significance at p = 0.001**

Table 1. Similarities in the classification of NIPF owners (N=364) when respondent uncertainty is ignored or risk-neutral behavior is assumed(LIN) and the classification under 4 different assumptions of response styles.

- Similarity between the original response-based clustering and the biased response-based clustering is low!
- If strong bias in the responses truly existed, taking the responses as unbiased would only reduce 21.9% to 37.6% of the uncertainty about the true clusters in the case of risk-seekers and risk avoiders, respectively.

K-means clustering of owners: PCs similarities

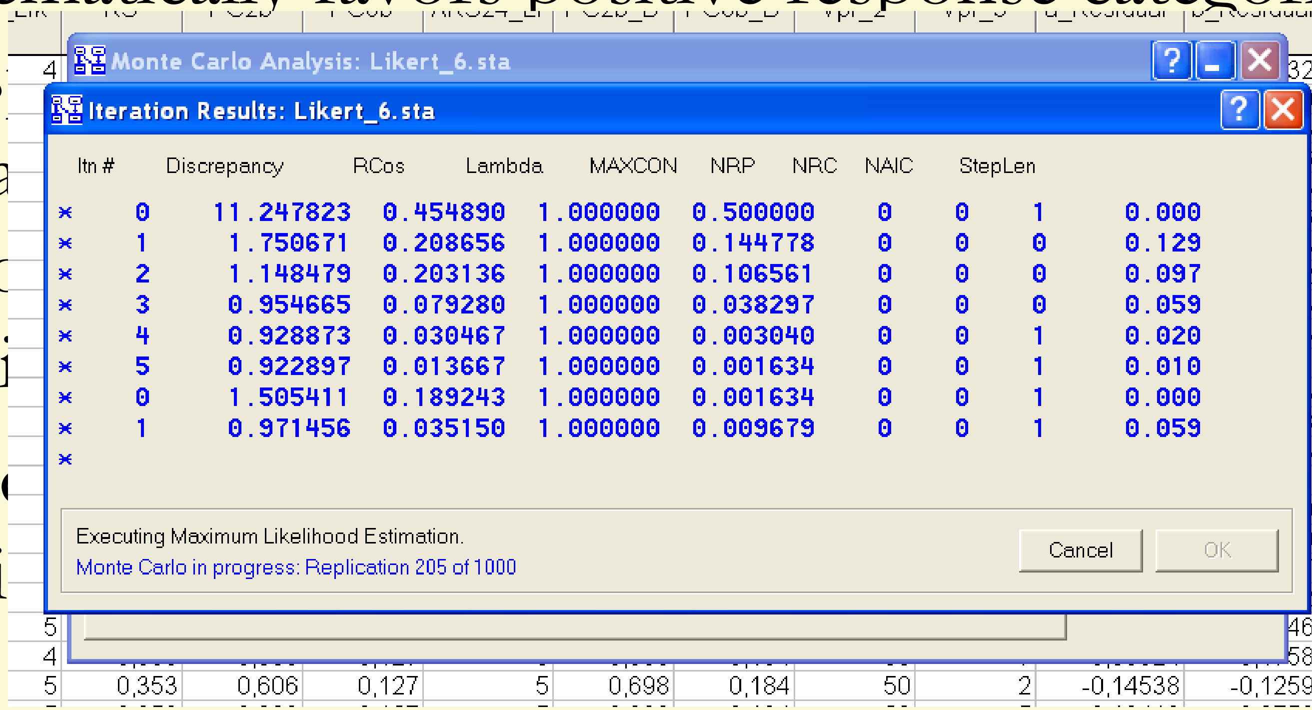
A) k-means clustering							
		Number of significantly different pairs of PCs at $p<0.05$					Percentage of significantly different PCs, pairs of all pairs
		LIN					
	Cl.	1	2	3	4	5	
LOG	1	n.s.	N/A	-	-	n.s.	3/84 = 3.6%
	2	n.s.	n.s.	1	n.s.	n.s.	
	3	-	N/A	n.s.	N/A	N/A	
	4	1	1	n.s.	n.s.	n.s.	
	5	-	n.s.	-	-	-	
SW	1	n.s.	n.s.	-	-	1	6/120 = 5.0%
	2	n.s.	n.s.	n.s.	n.s.	n.s.	
	3	n.s.	n.s.	n.s.	n.s.	n.s.	
	4	1	N/A	1	1	2	
	5	n.s.	n.s.	N/A	n.s.	-	
EXP	1	4	4	5	2	5	77/120 = 64.2%
	2	3	2	-	5	3	
	3	4	4	4	3	5	
	4	-	3	-	4	6	
	5	4	-	-	2	5	
RD	1	1	2	-	n.s.	-	46/120 = 38.3%
	2	5	N/A	N/A	2	5	
	3	1	1	n.s.	-	6	
	4	1	4	1	3	3	
	5	1	2	n.s.	4	3	

Table 1. Similarities in the classification of NIPF owners (N=364) when respondent uncertainty is ignored or risk-neutral behavior is assumed (LIN) and the classification under 4 different assumptions of response styles.

Pairwise comparison of means of clustering variables confirmed the extent of the influence of risk-seeking behavior to cluster assignment; more than 60% of all pairs of clustering variables significantly differed when EXP was compared to LIN !

2. Diagnosing the respondent uncertainty

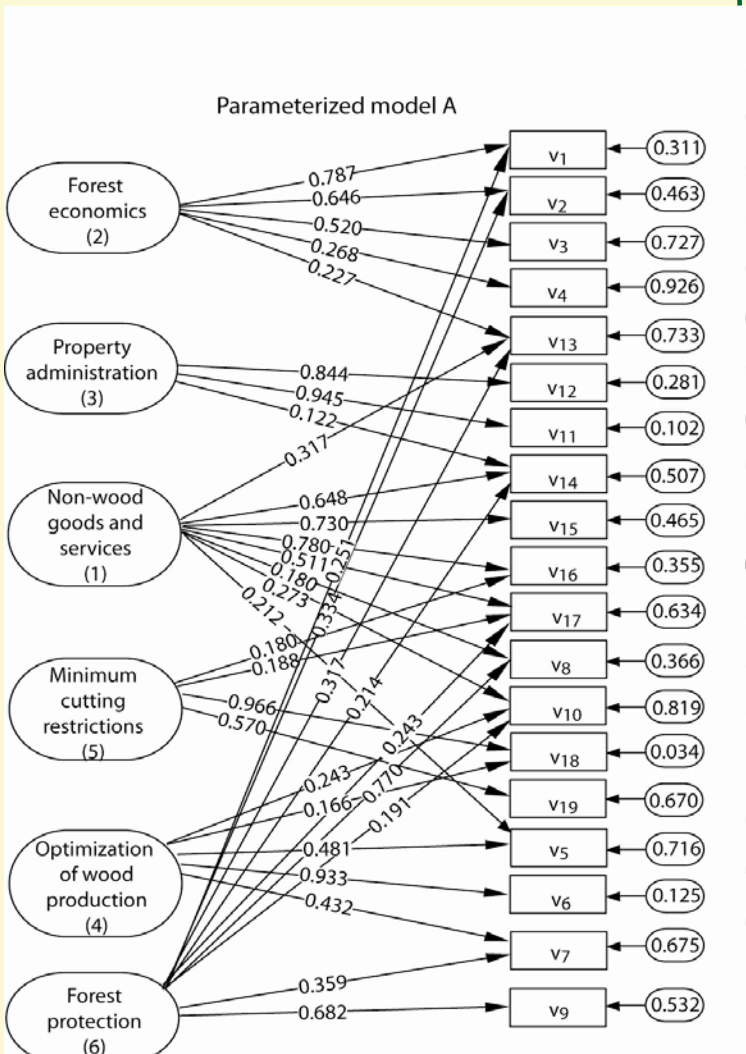
- If a substantial number of respondents systematically favors positive response categories irres… beha… facto… actor (Billi…
- Stru… e in Stati…

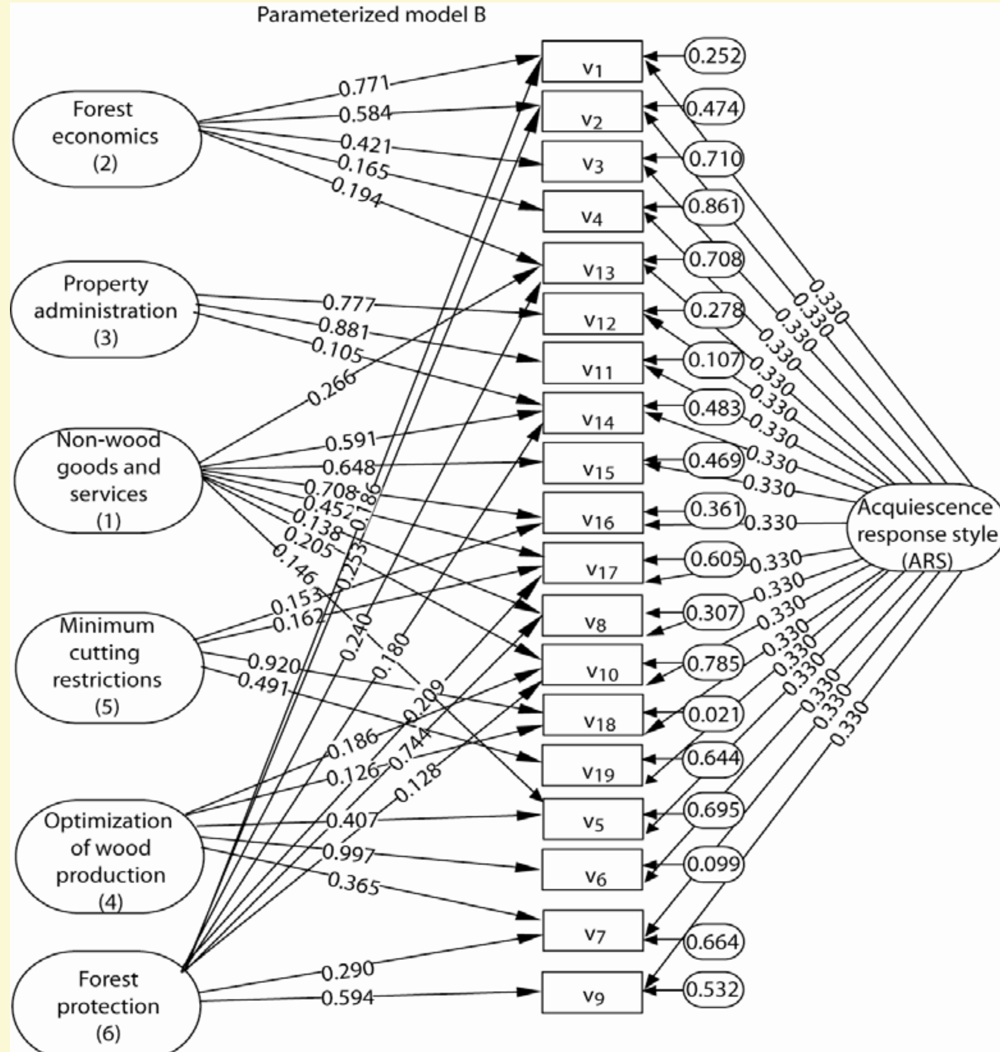


	Itn #	Discrepancy	RCos	Lambda	MAXCON	NRP	NRC	NAIC	StepLen
*	0	11.247823	0.454890	1.000000	0.500000	0	0	1	0.000
*	1	1.750671	0.208656	1.000000	0.144778	0	0	0	0.129
*	2	1.148479	0.203136	1.000000	0.106561	0	0	0	0.097
*	3	0.954665	0.079280	1.000000	0.038297	0	0	0	0.059
*	4	0.928873	0.030467	1.000000	0.003040	0	0	1	0.020
*	5	0.922897	0.013667	1.000000	0.001634	0	0	1	0.010
*	0	1.505411	0.189243	1.000000	0.001634	0	0	1	0.000
*	1	0.971456	0.035150	1.000000	0.009679	0	0	1	0.059
*									

Structural equation modeling (SEM)

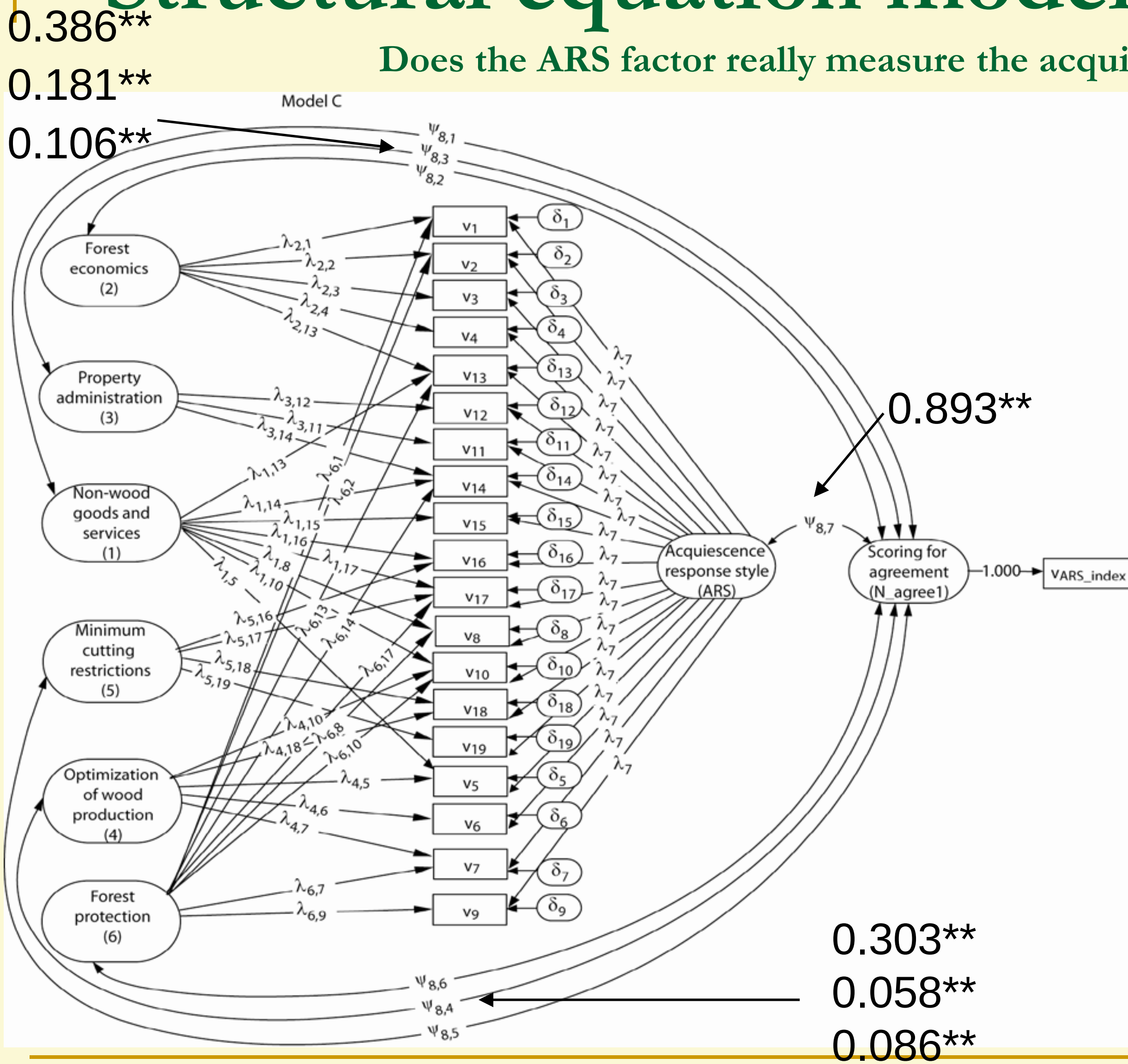
Which model replicates the correlation matrix better?





Structural equation modeling (SEM)

Does the ARS factor really measure the acquiescence?



- Correlation between the ARS factor and the variable measuring the frequency of the very important and rather important response category selection **must be high**.

- After adding the ARS factor to model A, the **model fit improved** (χ^2=404.45, df=136). The difference in the χ^2 statistics between model A and B is highly signficant, $p < 0.001$.
- **Fit indices** of B vs. A are higher: RMSEA=0.07, GFI=0.93, AGFI=0.90, CFI=0.94, TLI=0.93, χ^2/df=3.0).
- The model with the ARS factor (**Model B**) explains the data significantly **better** than the model with content factors only (Model A). ➔ **acquiescence response style diagnosed**.

Correcting the raw data for acquiescence (an improved approach)

- Acquiescen… the items
- We genera… …del A and 100…
- We selecte… …o the model … correlation…

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14					
V1	1	.55	.41	.20	.09	.12	.14	.29	.27	.15	.15	.13	.33	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
V2	.53	1	.33	.20	.07	.08	.16	.27	.24	.11	.15	.13	.28	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
V3	.38	.31	1	.20	.10	.11	.10	.11	.14	.11	.16	.14	.22	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
V4	.18	.17	.16	1	.10	.06	.10	.14	.10	.06	.07	.07	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
V5	-.02	-.04	.00	.00	1	.53	.31	.11	.10	.20	[illegible]	[illegible]	[illegible]	.20	.18	.19	.17	.23	.14
V6	.03	-.02	.01	-.04	.51	1	.51	.12	.08	[illegible]	[illegible]	.12	.11	.12	.09	.10	.13	.31	.12
V7	.07	.10	.00	-.01	.26	.49	1	.28	.22	.16	[illegible]	.18	.23	.17	.08	.14	.20	.24	.12
V8	.25	.23	-.01	.03	.01	.01	.22	1	.58	.21	.18	.17	.41	.32	.17	.22	.28	.11	.08
V9	.23	.21	.03	-.01	-.01	-.02	.18	[illegible]	[illegible]	[illegible]	.15	.15	.31	.24	.11	.19	.22	.09	.12
V10	.06	.04	.00	-.05	.16	.23	.13	.17	.19	1	.03	.03	.28	.23	.27	.32	.19	.19	.14
V11	.04	.05	.05	-.04	.01	.06	.07	.07	.03	-.08	1	.81	.21	.30	.11	.15	.13	.06	.03
V12	.02	.03	.03	-.04	.00	.01	.08	.06	.04	-.08	.81	1	.18	.28	.11	.15	.11	.05	.06
V13	.29	.25	.14	.08	.03	[illegible]	[illegible]	.38	.27	.23	.09	.06	1	.33	.23	.31	.24	.13	.10
V14	.05	-.01	-.06	[illegible]	[illegible]	[illegible]	.08	.28	.18	.20	.22	.20	.29	1	.45	.50	.37	.10	.07
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.13	-.01	-.04	.10	.00	.24	.00	.00	.17	.43	1	.56	.33	.11	.10
[illegible]	[illegible]	[illegible]	[illegible]	-.06	.14	-.01	.03	.15	.08	.29	.03	.03	.24	.48	.56	1	.39	.26	.14
[illegible]	[illegible]	[illegible]	.01	-.01	.10	.03	.12	.22	.15	.15	.01	.00	.18	.33	.29	.36	1	.32	.20
[illegible]	[illegible]	[illegible]	.02	-.02	.16	.26	.16	-.01	-.03	.11	-.07	-.08	.01	-.01	.01	.18	.25	1	.56
[illegible]	[illegible]	[illegible]	-.02	.03	.04	.02	.01	-.04	.01	.04	-.09	-.07	-.02	-.04	.01	.07	.13	.55	1

Correlation matrix of a dataset perfectly fitting to model with the ARS factor (B)

B - A

Correlation matrix of a dataset perfectly fitting to model with the content factors only (A)

Correcting the raw data for acquiescence (an improved approach)

- Subtracting the correlation matrix of model A from the correlation matrix of model B (B-A) = the effect of ARS on correlations (Net ARS matrix).
- The correlation matrix of raw data minus the Net ARS matrix = correlation matrix corrected for acquiescence.
- PCA analysis with the correlation matrix corrected for acquiescence.
- Monte Carlo generation of the 364 responses with the desired corrected correlations between the 19 items: not successful in 1000 attempts (the resulted correlations not accurate enough).

Conclusions

	(a) Raw data[b]						(b) Data corrected for acquiescence					
	Content factors						Content factors					
Manifest variable	1	2	3	4	5	6	1	2	3	4	5	6
V_1	-0.03	**0.81**	0.11	0.01	0.00	0.21	-0.07	**0.82**	0.03	-0.01	-0.02	0.16
V_2	0.07	**0.72**	0.14	0.10	-0.03	0.29	0.03	**0.77**	0.07	0.08	-0.03	0.25
V_3	0.12	**0.72**	0.05	0.00	0.12	-0.24	0.02	**0.69**	-0.02	-0.08	0.03	-0.34
V_4	0.24	0.48	-0.23	0.15	0.08	-0.10	0.13	0.42	-0.26	0.00	-0.07	-0.21
V_5	0.21	-0.03	0.05	**0.71**	0.12	0.04	0.20	-0.09	-0.01	**0.69**	0.06	-0.04
V_6	0.00	0.07	0.11	**0.85**	0.10	0.01	-0.05	0.02	0.04	**0.86**	0.08	-0.06
V_7	0.06	0.03	0.19	**0.58**	-0.02	0.43	-0.01	0.04	0.11	**0.63**	-0.01	0.41
V_8	0.21	0.05	0.00	0.14	0.12	**0.78**	0.19	0.06	-0.04	0.08	0.04	**0.78**
V_9	0.05	0.13	-0.07	0.04	0.14	**0.77**	-0.02	0.10	-0.11	-0.04	0.03	**0.78**
V_{10}	0.31	0.18	0.14	0.41	-0.22	0.20	0.32	0.11	0.04	0.42	-0.21	0.19
V_{11}	0.14	0.06	**0.90**	0.10	0.12	0.03	0.08	0.02	**0.93**	0.05	0.06	-0.01
V_{12}	0.09	0.08	**0.90**	0.09	0.09	-0.03	0.03	0.02	**0.93**	0.02	0.02	-0.06
V_{13}	0.35	0.29	0.22	-0.07	-0.06	0.45	0.34	0.32	0.11	-0.08	-0.09	0.43
V_{14}	**0.72**	0.08	0.22	0.04	-0.11	0.28	**0.73**	0.06	0.19	0.04	-0.14	0.25
V_{15}	**0.80**	0.03	0.08	0.08	0.12	-0.01	**0.81**	-0.03	0.03	0.03	0.05	-0.05
V_{16}	**0.82**	0.05	0.08	0.06	0.17	0.03	**0.84**	-0.02	0.02	0.02	0.13	-0.01
V_{17}	**0.62**	0.03	0.08	0.10	0.18	0.26	**0.62**	0.01	0.02	0.06	0.20	0.24
V_{18}	0.15	0.00	0.13	0.15	**0.84**	0.01	0.10	-0.03	0.04	0.14	**0.87**	-0.06
V_{19}	0.06	0.06	0.14	-0.01	**0.81**	0.19	0.02	0.01	0.05	-0.07	**0.85**	0.11
Eigenvalue	4.5	1.8	1.6	1.6	1.4	1.3	3.3	2.2	1.8	1.8	1.5	1.5
Cumulative variance explained (%)	23.6	33.3	41.9	50.1	57.4	64.1	17.3	28.8	38.3	47.6	55.7	63.3

[a]Bolded loading indicates a value greater than 0.50.

[b]PCA when acquiescence is left in the responses (Ficko & Boncina, 2013a)

- Acquiescence had no effect on substantive construct in this case.
- The cumulative variance explained decreased from 64.1% to 63.3% when the responses were corrected for the ARS.

Lessons – learned & research recommendations

- Response style can threaten the validity of clustering results → unvalid typologies
- Researchers: Pay attention to stimuli of response style (e.g. survey design, looking for socially desirable behavior, personal characteristics)
- Use of advanced methods (e.g. probabilistic clustering, mixed-methods, other methods from social sciences) → more simple typologies, owners not forced into a priori groups
- Towards more harmonized approach to NIPF owner segmentation to compare the typologies also statistically.

Thank you all for the opportunity to work with the FORSYS community and to share your experiences, and for accepting this presentation.