Multilevel Component Analysis applied to the measurement of a complex product experience

Boucon, C.A., Petit-Jublot, C.E.F., Groeneschild C., Dijksterhuis, G.B.

Unilever

Outline

- ✓ Background
- ✓ Introduction to Simultaneous Component Analysis (SCA) and Multi Level Component Analysis (MLCA)
- ✓ Technical details
- ✓ Analysis and results
- ✓ Conclusions & outlook

Measuring product experience of food

✓ Food is a complex, multisensory experience

Total Product Experience
Multisensory

✓ Questionnaire to capture multisensory experience in consumers

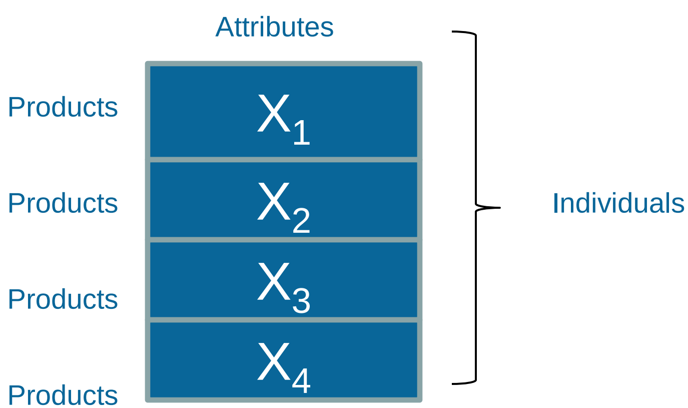
- Based on literature review: Berlyne's work on exploratory behaviour and aesthetics, choice/preference theory by Dember and Earl
- Covering different aspects of product experience: manipulation, preparation, consumption
- Evaluative variables related to complexity, aesthetics, usage, novelty
- 33 items, line scale, left and right anchored (example below)

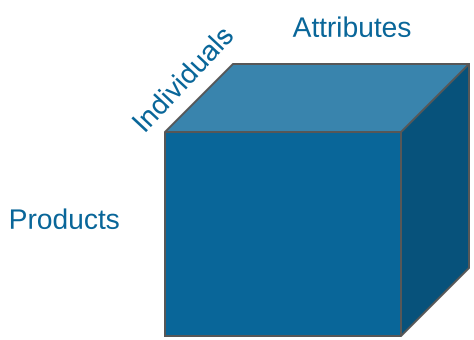
Difficult to open |————————————| *Easy to open*

Old |————————————| *New*

Study design

- ✓ Evaluation of product experience in milk tea
  - 220 subjects
  - Central Location Test (China)
- ✓ Stimuli
  - Four products following a 2x2 (Packaging, Flavour) factorial design
  - Warming up with dummy
  - Preparation included in evaluation
- ✓ Randomised per subject to correct for order and carry over
- ✓ Results in multilevel (multiway) multivariate data





Analysis of multi-level data in sensory science

✓ Common approach to analyse multivariate data from multiple subjects (panel) in sensory

- Average score per product (or LS estimates after ANOVA)
- PCA (biplot visualisation)

✓ Multilevel, multivariate data with consumers

- Not very common
- Averaging does not make sense as consumers are not trained and may vary widely in their perception or interpretation of the attributes
- Advanced alternatives (e.g. MFA, GPA, STATIS) focus on finding a consensus in terms of products

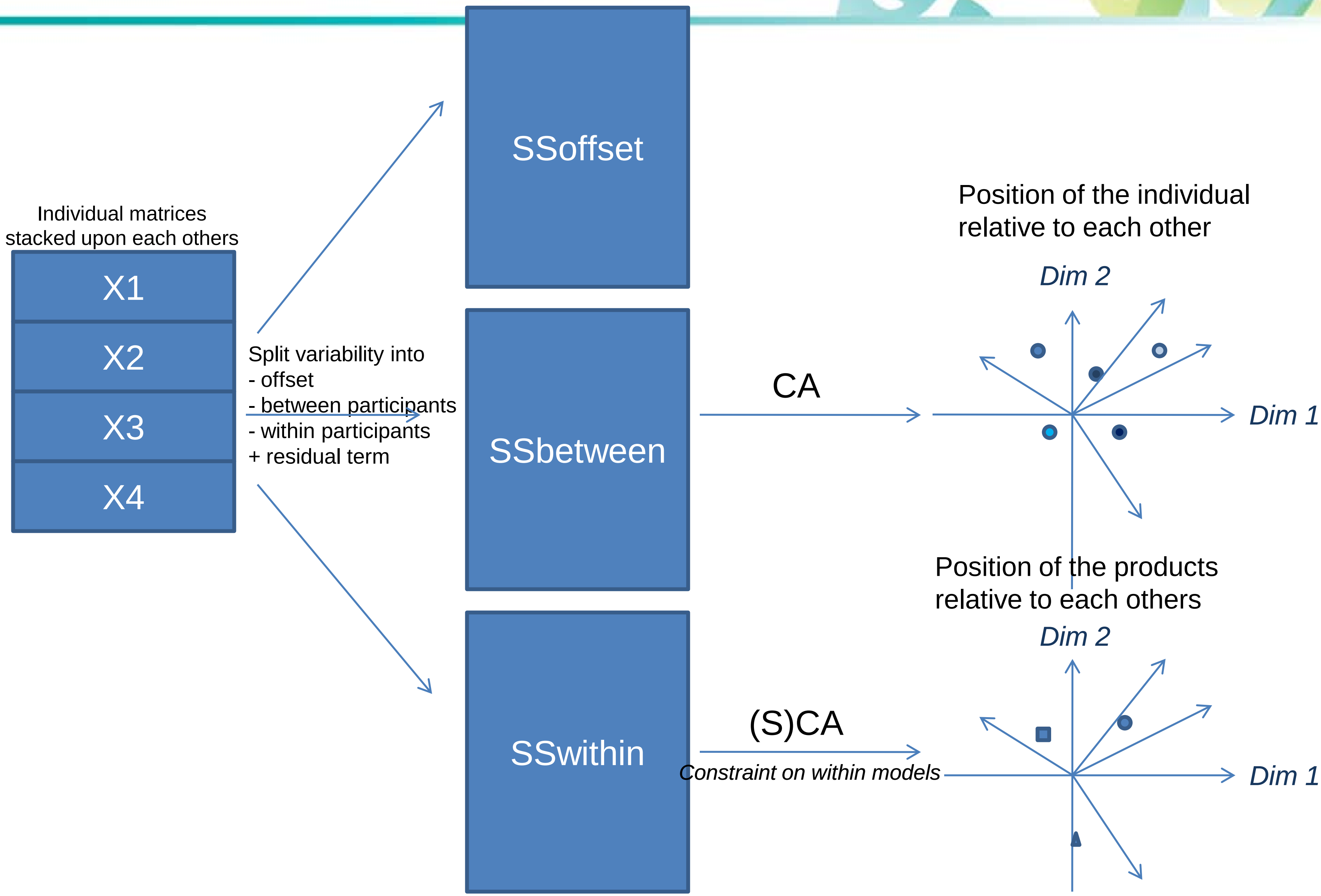
⇨ Method that estimate a common component model but would allow to reflect the individual differences and take into account the hierarchical nature of the data

Introduction to MCA

- ✓ Simultaneous Components Analysis (SCA)
  - Generalization of PCA developed (ten Berge, Kiers, van der Stel,1992) for situations where same variables are measured in two or more populations
  - Applied e.g. in social sciences (same questionnaire applied to different populations)
  - Common loadings maximizing explained variance in each groups
- ✓ Extension to model multivariate time series (Timmerman & Kiers, 2003)
  - Shows evolution of latent structure in time
  - Common loadings
  - Different degree of constraints imposed on scores matrices
- ✓ Generalisation of SCA to multi-level data (Timmerman, 2006)
  - Decomposition of data into within and between part
  - Separate (S)CA to model between and within part

Principle

Application to our product experience data



Principle

✓ Split the different sources of variability (ANOVA) for each variable j

$$SS_{total,\,j} = SS_{offset,\,j} + SS_{between\ participant,\,j} + SS_{within\ participant,\,j} + SS_{error,\,j}$$

✓ Component model for each of the part

$$Y_i = 1_{K_i} m' + 1_{K_i} f'_{ib} B'_b + F_{iw} B'_{iw} + E_i$$

offset per variable

error term

between subject CA

within subject CA

- in model 0 (MLCA): F_{iw} and B_{iw} differ for each individual subject
- in model 1 to 4 (MLSCA): $B_{iw} = B_{w}$, with different constraints on the variance-covariance structure of F_{iw}

MLSCA-P: $\frac{1}{K_i} F_{iw}' F_{iw} = \Phi_i$

MLSCA-PF2: $\frac{1}{K_i} F_{iw}' F_{iw} = D_i \Phi D_i$

MLSCA-IND: $\frac{1}{K_i} F_{iw}' F_{iw} = D_i^2$

MLSCA-ECP: $\frac{1}{K_i} F_{iw}' F_{iw} = \Phi$

Within subject model

✓ Five alternative with increasing degree of constraint, based on the alternatives proposed in SCA

Model	Name	Loadings	Covariances	Variances
0	MLCA	free	free	free
4	MLSCA-P	equal	free	free
3	MLSCA-PF2	equal	equal across subjects	free
2	MLSCA-IND	equal	equal to 0	free
1	MLSCA-ECP	equal	equal across subjects	equal across subjects

⇨ Compare how the different models fit the data and how they can be interpreted in terms of consumers' perception

Selecting & comparing models

- ✓ Selecting the right model
  - Fit: Variance accounted for within part & between part of the data
  - Stability: assessed by means of a split-half procedure
  - Degree of complexity and interpretability
- ✓ Split-half procedure
  - Random split between participants
  - Comparison between models (loadings) for both halves
  - Repeat n=100 times
  - Average over n repetitions
- ✓ Interpretation
  - Compare loadings matrices
  - Visualisation (biplots)
  - Assess agreement between subject by comparing loadings and/or scores

Comparing models

✓ Two indices quantifying similarities between matrices

- Tucker congruence coefficient (φ)

$$\varphi = \frac{tr(XY')}{\sqrt{tr[(XX')]tr[(YY')]}}$$

  – introduced to measure similarity of two factorial configurations
  – apply to matrices (e.g. loadings or scores matrices) of same dimensions
  – takes values between -1 and 1 (φ=0 no correlation, |φ|=1 perfect correlation)
  – applied after (procrustes) rotation and scaling of the factor solution: φ_{rot}

- RV-coefficient (Robert & Escoufier, 1976)

$$RV = \frac{tr(\widetilde{XX}'\widetilde{YY}')}{\sqrt{tr[(\widetilde{XX}')^2]tr[(\widetilde{YY}')^2]}} \quad \text{where } \widetilde{XX}' = XX' \quad \text{(original)}$$
$$\widetilde{XX}' = [XX' - diag\,(XX')] \quad \text{(modified)}$$

  – orientation independent
  – allows for different number of variables
  – usually used to compare sample configurations (scores)
  – modified version independent of sample size (Smilde, 2009) and takes values between -1 and 1 (φ=0 no correlation, |φ|=1 perfect correlation)

Results
Fit and model selection

✓ Fit and stability of the model

- Between part

Number of components	1	2	3	4	5
VAF (%)	21	27	32	34	35
Mean congruency coefficient	0.98	0.95	0.96	0.85	0.88

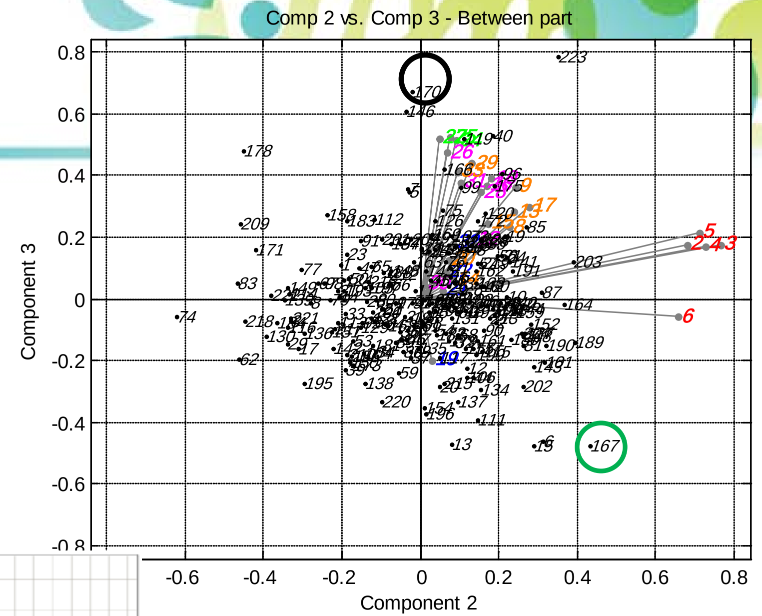
- Within part

Number of components		1	2	3
VAF (%)	0 (Unconstrained)	40	50	54
	4 (Loadings)	25	32	35
	3 (Loadings, cov)	25	31	33
	2 (Loadings, cov=0)	25	30	32
	1 (Loadings & var-cov)	19	21	22

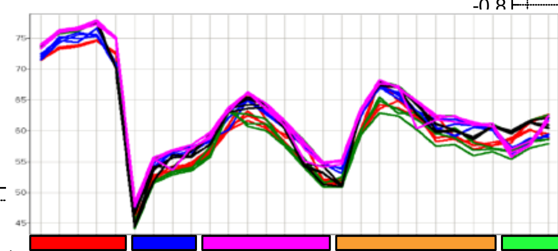
Number of components		1	2	3
Mean congruency coefficient	0 (Unconstrained)	-	-	-
	4 (Loadings)	0.9920	0.9830	0.9766
	3 (Loadings, cov)	0.9919	0.5935	0.7576
	2 (Loadings, cov=0)	0.9922	0.9719	0.9707
	1 (Loadings & var-cov)	0.9935	0.9699	0.9791

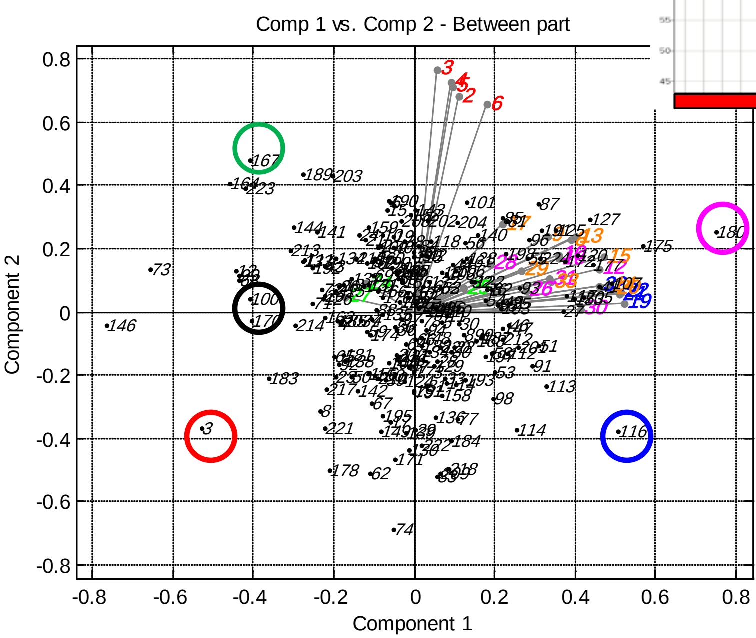
⇨ Number of dimensions: between part: Qb=3, within part: Qw=2

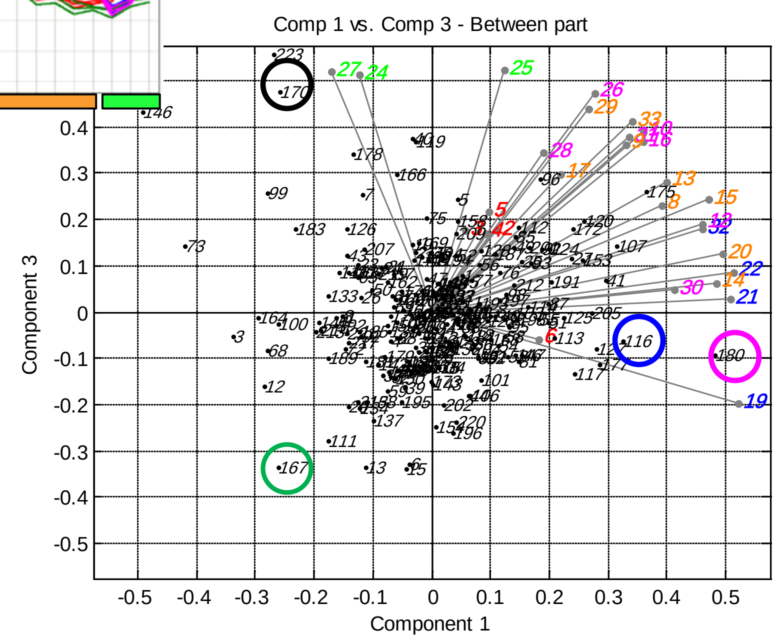
Between part (rotated)



- ✓ Interpretation of the questionnaire scale usage, overall perception of milk tea
  - Dim 1: Novelty, Aesthetics & Complexity
  - Dim 2: Usage
  - Dim 3: Familiarity







Within part

Model 0: unconstrained

✓ Individual PCA (rotated)

- Large variability between individual
  - Loadings: $\varphi_{rot} = 0.34$ (median)
  - Scores: $\varphi_{rot} = 0.76$, RV= 0.64, RVM = 0.39 (median)

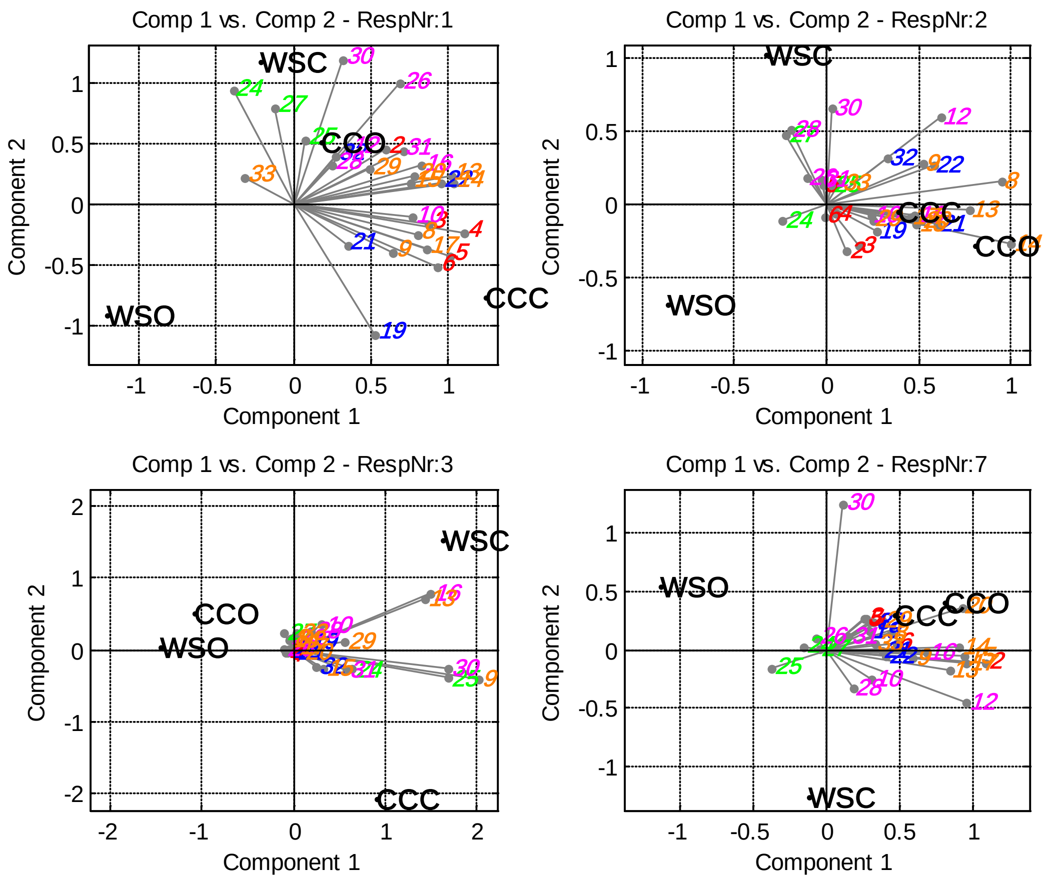
Loadings

φ_{rot}	1	2	3	7
1	-	0.64	0.21	0.58
2	-	-	0.32	0.64
3	-	-	-	0.18
7	-	-	-	-

Scores

RV	1	2	3	7
1	-	0.65	0.73	0.71
2	-	-	0.54	0.99
3	-	-	-	0.55
7	-	-	-	-

RVM	1	2	3	7
1	-	0.42	0.46	0.52
2	-	-	0.20	0.98
3	-	-	-	0.19
7	-	-	-	-

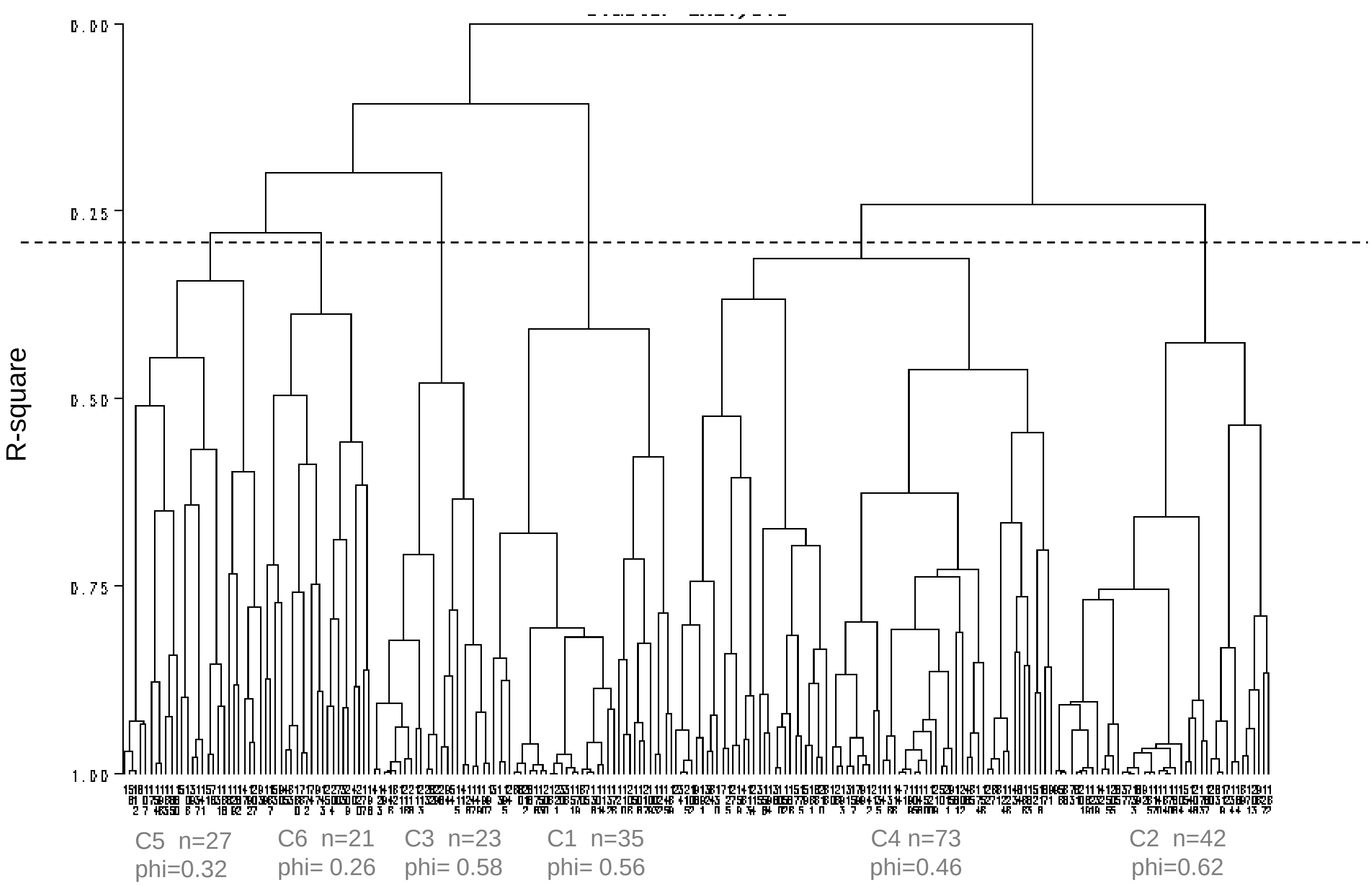


Improving interpretation

Unconstrained model

✓ Segmentation

- Cluster analysis (Ward's method) based on similarity of individual loading matrices (as measured by φ_{rot})

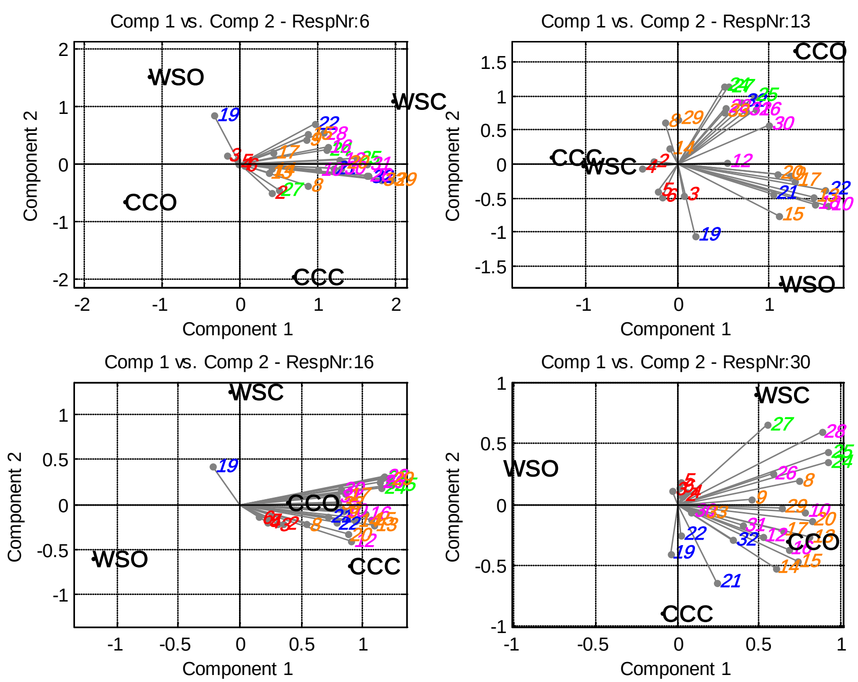


Segmentation
Unconstrained model

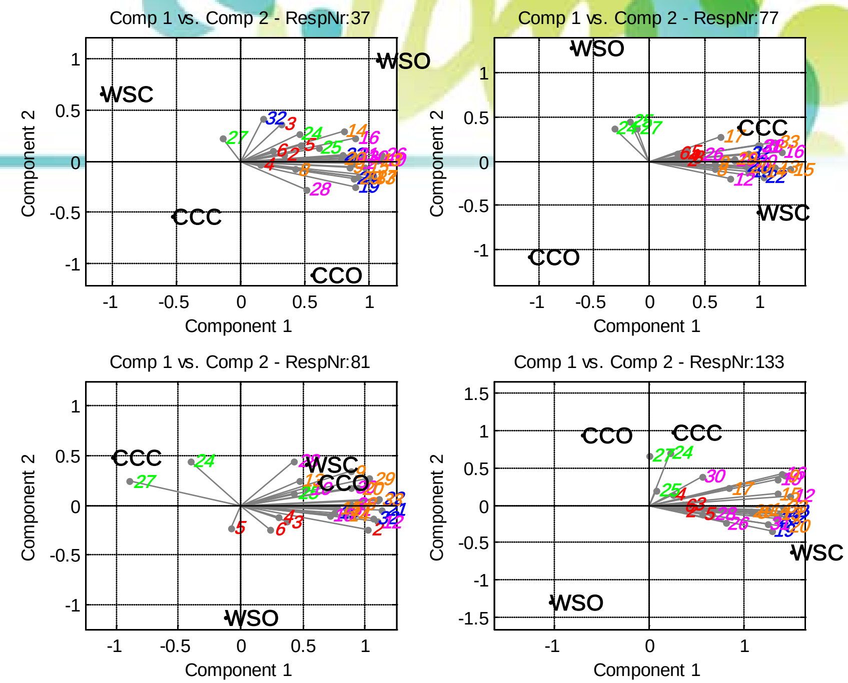
✓ Example of cluster membership

– Visualisation

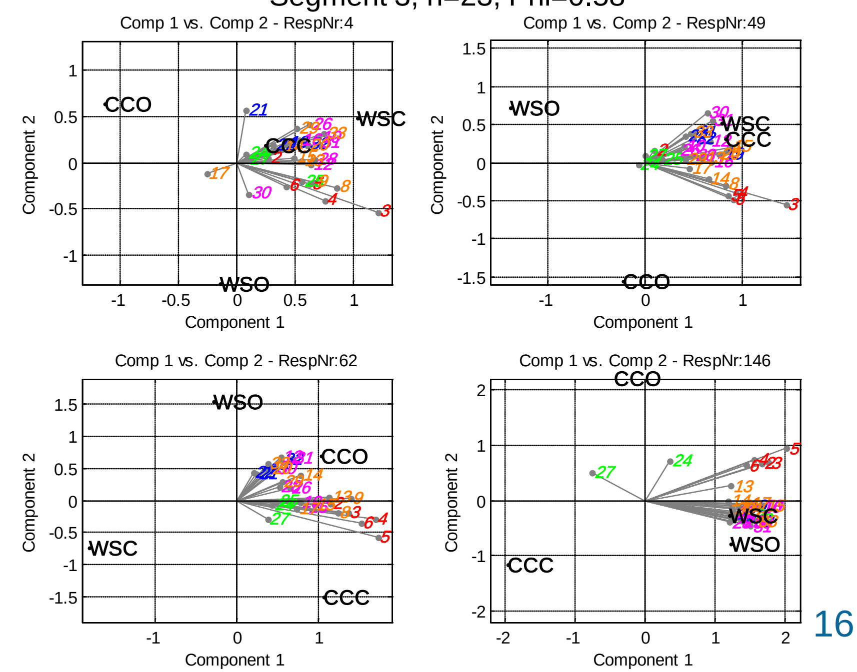
Segment 1, n=35, phi = 0.54



Segment 2, n=42, Phi=0.62



Segment 3, n=23, Phi=0.58



Within part

Model 2: constrained loadings and covariance = 0

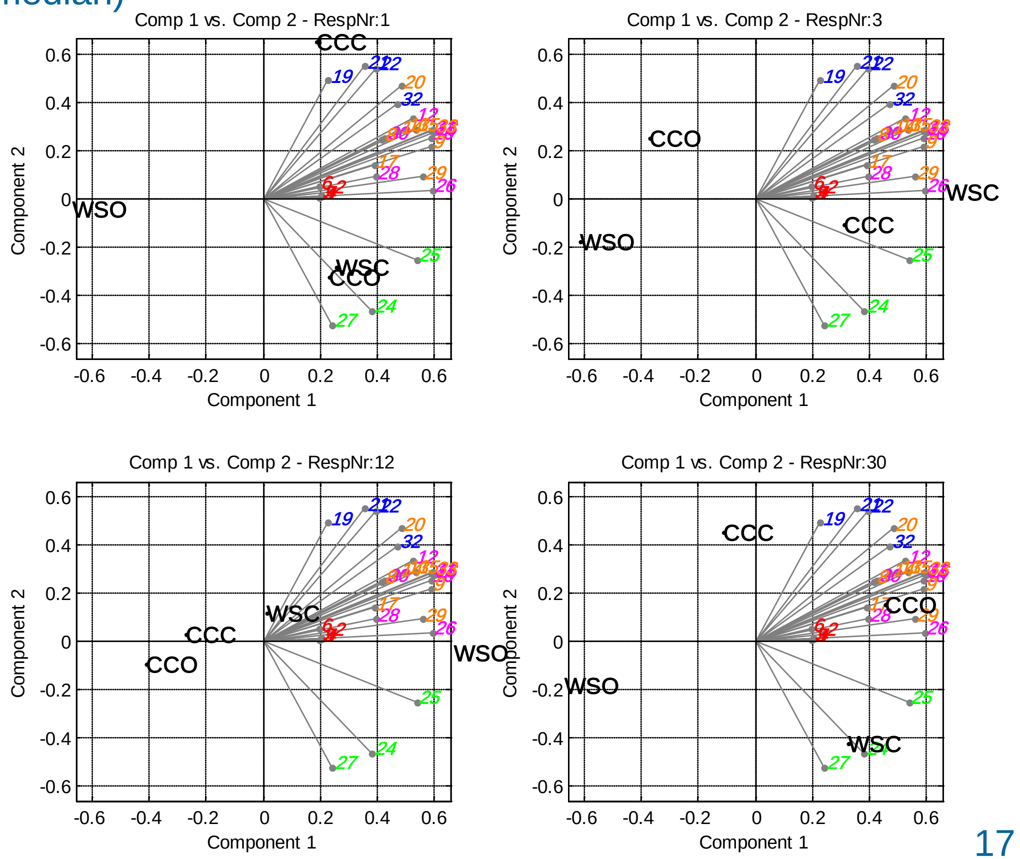
- ✓ Same loadings for all subjects
  - Dim1: Aesthetics & Complexity   Dim 2: Novelty/Familiarity
- ✓ Agreement between subjects:
  - – φ_{rot} = 0.69, RV=0.50, RVM= 0.28 (median)

Scores

φ	1	3	12	30
1	-	0.34	-0.61	0.81
3	-	-	-0.38	0.36
12	-	-	-	-0.69
30	-	-	-	-

RV	1	3	12	30
1	-	0.42	0.61	0.81
3	-	-	0.23	0.27
12	-	-	-	0.75
30	-	-	-	-

RVM	1	3	12	30
1	-	0.12	0.35	0.69
3	-	-	-0.15	-0.27
12	-	-	-	0.61
30	-	-	-	-

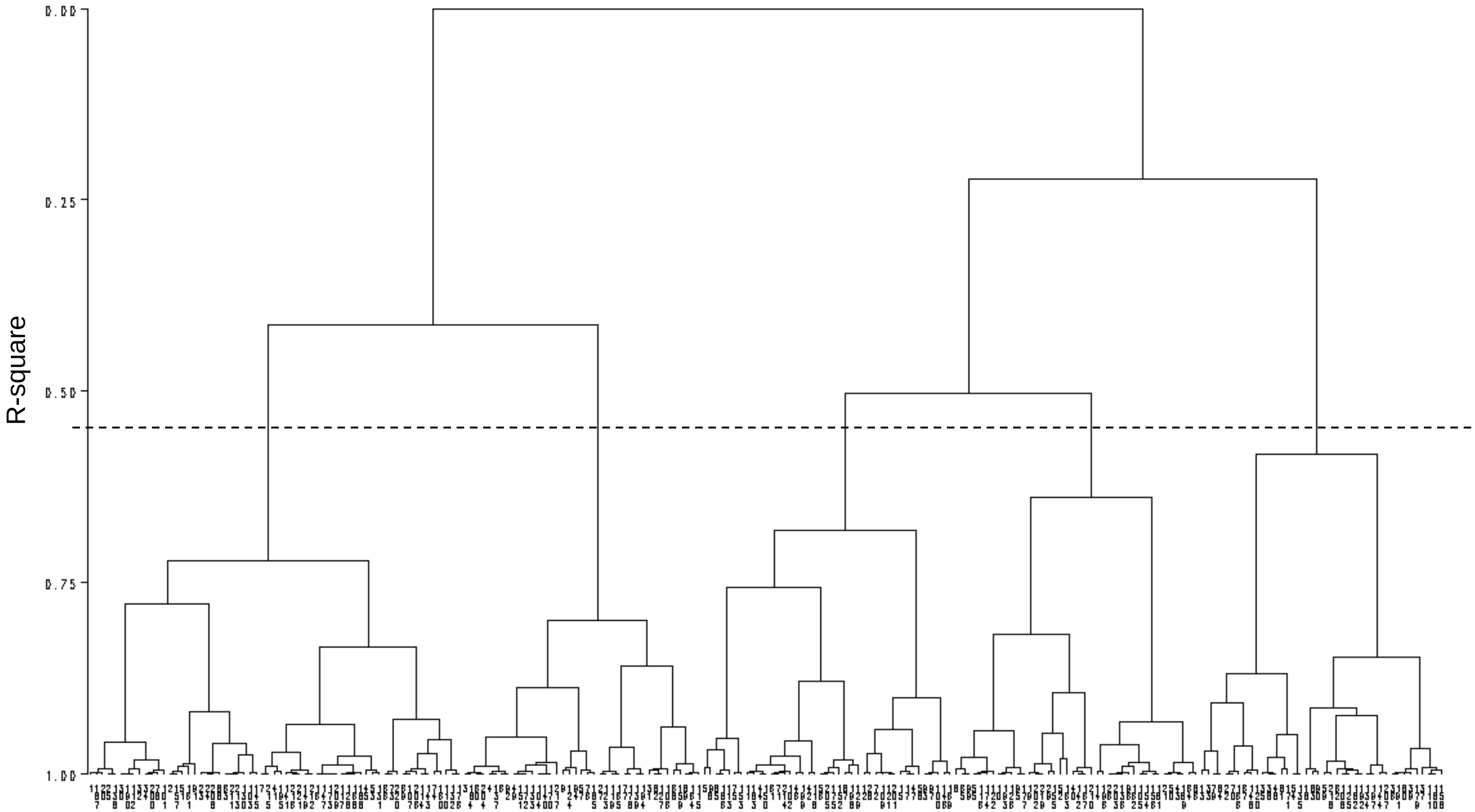


Improving interpretation

Model 2: constrained loadings and covariance = 0

✓ Segmentation

- Cluster analysis (Ward’s method) based on similarity of product configuration in the common space (as measured by unrotated φ)

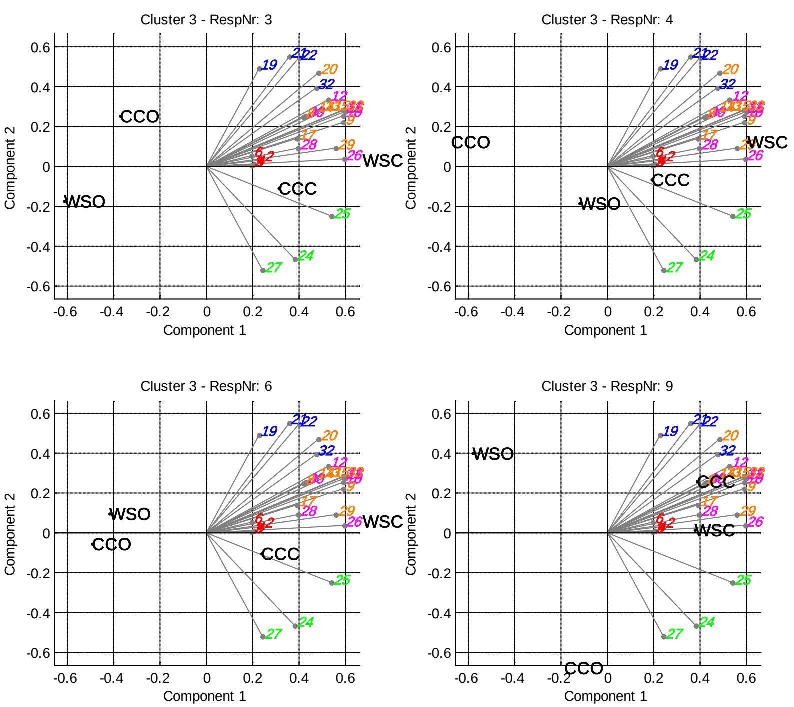


C2 n=61 phi=0.48 | C3 n=39 phi=0.56 | C5 n= 41 phi=0.36 | C1= n=40 phi=0.48 | C4 n=40 phi=0.29

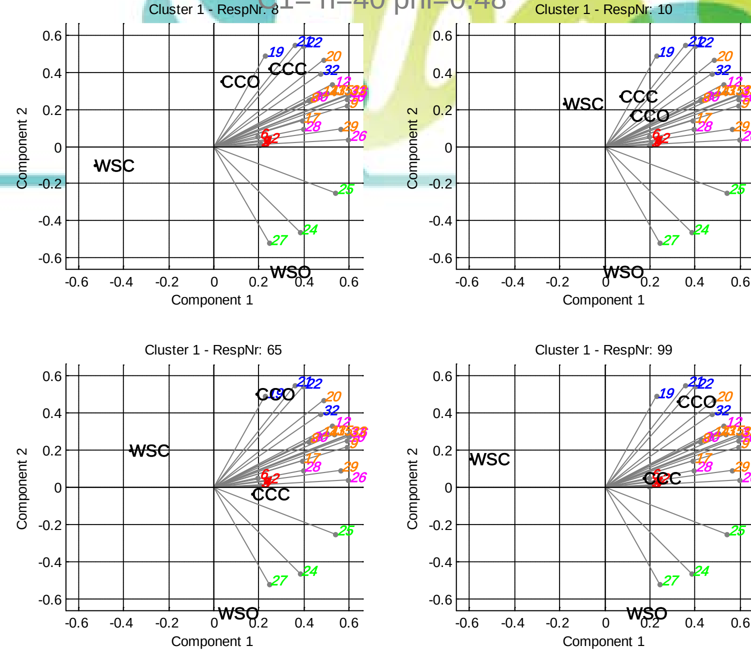
Segmentation (Model 2)

- ✓ Example of cluster membership
  - – visualisation

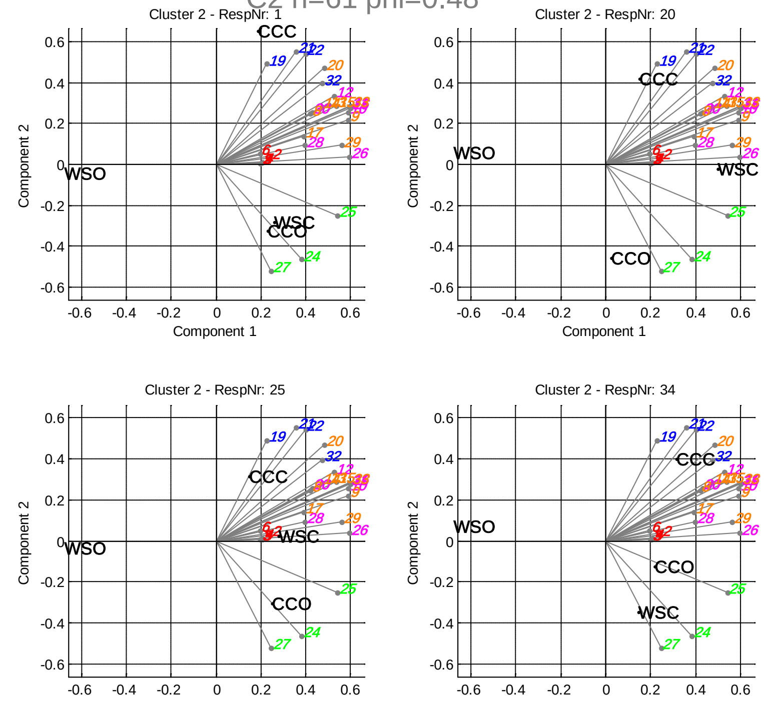
C3 n=39 phi=0.56



C1= n=40 phi=0.48



C2 n=61 phi=0.48



Summary

- ✓ ML(S)CA proved useful approach to model our data
  - Takes account of the hierarchical structure of the data
  - Offers the possibility to impose a common factor structure across subjects
  - Allows to compare different levels of constraints for the individual models
- ✓ Congruence & RV-coefficients are useful
  - in selecting and comparing models
  - in interpreting the solution (measure of similarity between individual configuration & input for segmentation)
  - the best index depends on the purpose of the comparison/segmentation
- ✓ Unconstrained model provides the best fit but the separate interpretation of the individual within loadings matrices can be very inefficient and difficult to reveal intra-individual similarities
- ✓ Imposing SCA constraints on the within part of the data
  - Models 4 & 2 perform best: lead to comparable VAF, stability and interpretation; model 2 best VAF/complexity ratio
  - Model 3 with 2 components is unstable compared to the rest but model with only one component has reduced fit and interpretability
  - Model 1: drop in fit indicates that same variance not suitable for our data

Relation to other methods

- ✓ Framework for comparison in van Deun et al (2009)
  - MFA: common object model i.e. look for a common configuration of the product , pre-processing (scaling by individual), weight individual matrices by amount of redundant information (1st eigenvalue)
  - STATIS: common object mode, no specific pre-processing, larger weight on matrices with cross-products(RV) most similar to others (compromise)
  - GPA: common object mode, pre-processing taken into account by translation or scaling transformation, all individual matrices equally weighted in consensus
  - ML(S)CA: common variable mode i.e. seeks for a common set of underlying components, pre-processing: normalising per respondent taken care by offset & between part of model, all individual matrices equally weighted in solution
- ✓ Timmermans (2006) also makes the parallel with multiway methods and multi-level SEM
  - Tucker-1 model equivalent to SCA-P model
  - Tucker-2, -3 & PARAFAC more than one mode is reduced into a component matrix; possible alternatives for within part of the model
  - Existing multilevel SEM constrain within covariance matrices to be equal for all participants

References

Multilevel components analysis, M.E. Timmerman, *British Journal of Mathematical and Statistical Psychology* (2006), 59, 301-320

Four simultaneous component models for the analysis of multivariate time series from more than one subject to model intraindividual and interindividual differences, M.E. Timmerman, H.K. Kiers, *Psychometrika*, 2003, 68 (1), 105-121

Simultaneous Components Analysis, J.M.F. Ten Berge, H.A.L. Kiers and V van der Stel, *Statistica Applicata*, 1992, 4(4), 277-392

A unifying tool for linear multivariate statistical methods: the RV-Coefficient, P. Robert and Y. Escouffier, *Applied Statistics*, 1976, 25(3), 257-265

Matrix correlations for high-dimensional data: the modified RV-coefficient, A.K. Smilde et al., *Bioinformatics*, 2009, 25(3),401-405

A structured overview of simultaneous component based data integration, K. Van Deun et al., *BMC Bioinformatics*, 2009, 10:246

Acknowledgements

Henk Kiers and Marieke Timmermans
Ivana Stanimirova
My colleagues & co-authors

Backup slide: MLCA algorithm

- ✓ Minimizing the SS_{res} using a OLS approach

$$F(m, f_{ib}, B_b, F_{iw}, B_{iw}) = \sum_{i=1}^{I} \left\| Y_i - 1_{K_i} m' + 1_{K_i} f'_{ib} B'_b + F_{iw} B'_{iw} \right\|^2 \quad \text{where} \quad \begin{array}{l} \sum_{i=1}^{I} K_i f_{ib} = 0_{Q_b} \\ 1'_{K_i} F_{iw} = 0'_{Q_{iw}} \end{array}$$

- ✓ Offset, between and within part solved separately by minimizing

(1) $$f_1(m) = \sum_{i=1}^{I} \left\| \mathrm{Ysup} - 1_{K_i} \mathrm{m}' \right\|^2$$

where Ysup denotes a supermatrix with the Y_i stacked upon each others

⇨ Solved by taking m = vector containing the observed mean scores computed for all participants and products

(2) $$f_2(\mathrm{Fsup_b}, B_b) = \left\| \mathrm{Ysup} - 1_{K_i} f'_{ib} B'_b \right\|^2$$

where Fsup denotes a supermatrix with the $1_{Ki} f'_{ib}$ stacked upon each others

(3) $$f_3(F_{iw}, B_{iw}) = \sum_{i=1}^{I} \| Y_i - F_{iw} B'_{iw} \|^2$$

⇨ Both (2) and (3) solved based on singular value decomposition

Backup slide: MLSCA algorithm

✓ Minimizing the SS_{res} using an OLS approach

$$G(m, f_{ib}, B_b, F_{iw}, B_w) = \sum_{i=1}^{I} \left\| Y_i - 1_{K_i} m' + 1_{K_i} f'_{ib} B'_b + F_{iw} B'_w \right\|^2$$

✓ Offset, between and within part solved separately by minimizing

- Offset and between part, see previous slide
- Within part solved based on ALS algorithm described in Kiers, ten Berge & Bro (1999)

$$g_1(F_{iw}, B_w) = \sum_{i=1}^{I} \left(\frac{1}{n} Y'_i 1' Y_i \right) + \sum_{i=1}^{I} \| \mathrm{J} Y_i - F_{iw} B'_w \|^2$$

subject to constraint on covariance of matrices F_{iw} of the specific SCA model

MLSCA-P: $\frac{1}{K_i} F_{iw}{}' F_{iw} = \Phi_i$

MLSCA-PF2: $\frac{1}{K_i} F_{iw}{}' F_{iw} = \mathrm{D}_i \Phi \mathrm{D}_i$

MLSCA-IND: $\frac{1}{K_i} F_{iw}{}' F_{iw} = \mathrm{D}_i{}^2$

MLSCA-ECP: $\frac{1}{K_i} F_{iw}{}' F_{iw} = \Phi$

Within subject model

- ✓ Model 0 (MLCA): individual PCA for each subject; the space describing the products and their position therein is different for each subject; similar to like GPA (Paey’s approach) except that we are not trying to rotate the results to get a consensus (might be an idea)

MLSCA

- ✓ Model 4 (SCA-P): the space describing the position of the products is the same but the position of the product may vary per individual; this is similar to Tucker-1 and performing PCA on the stacked matrices (under certain conditions)
- ✓ Model 3 (SCA-PF2): the space describing the position of the products is the same and the relative position is constrained to be the same for each subject; the variance may differ per subject; related to PARAFAC
- ✓ Model 2 (SCA-IND): the components are constrained to be uncorrelated for each individual
- ✓ Model 1 (SCA-ECP): most constrained model where variance is constrained to be the same for all subjects; might be less relevant to our data

⇨ Interesting to compare how the different models fit the data and how they can be interpreted in terms of consumers’ perception

Additional issues

✓ Pre-processing

- Centring across or per subject not necessary since offset and between subject terms are modelled explicitly
- Normalisation per variable (over other modes) recommended
  - Eliminate artificial scale differences between variables
  - No further lost of source of variability, factor model preserved
  - Arguable in our situation: might choose not to standardised at all, since difference in variability between variable might reflect perceived differences

✓ Rotational freedom

- Between part: insensitive to orthogonal and oblique rotation
- Within part:
  - Model 0, 1 & 4: insensitive to orthogonal and oblique rotation
  - Model 2 & 3: unique solutions
- Normalisation of component scores to facilitate comparisons

Results
Agreement between subjects

✓ Overview (median)

	Model 0	Model 4	Model 3	Model 2	Model 1
Description	Unconstrained	Constrained loadings	Constrained loadings and cov	Constrained loadings and cov=0	Constrained loadings and var-cov matrices
Loadings	φ, φ_{rot} 0.14, 0.34	-	-	-	-
Scores	φ, φ_{rot}, RV, RVM 0.12,0.76,0.64,0.39	φ, φ_{rot}, RV, RVM 0.07,0.63,0.41,0.19	φ, φ_{rot}, RV, RVM 0.10, 0.58, 0.31, 0.10	φ, φ_{rot}, RV, RVM 0.05,0.69,0.50,0.28	φ, φ_{rot}, RV, RVM 0.03, 0.75, 0.63, 0.36

✓ Most suitable index depends on objective

- Model 0: compare rotated configuration since individual models unconstrained
  - moderate agreement on loadings
  - seemingly high agreement on scores but not higher than chance given the small number of samples
- Model 1 to 0: compare scores directly (unrotated) makes sense since loadings are constrained to be equal
  - very low agreement
  - higher level of constraint improves agreement on relative position of products
  - model 3 falls out of this trend