A component-based Multidimensional Path Modelling R-package: THEME



X. BryI3M, Univ. Montpellier IIT. VerronITG - SEITA, Centre de recherche

1. Conceptual model of a situation = Thematic Model



1. Conceptual model of a situation = Thematic Model

Equation 1

Indicators describing new policy in city 1 year after election Economic,Social & Cultural indicators of city prior to local election

1. Conceptual model of a situation = Thematic Model



1. Conceptual model of a situation = Thematic Model



1. Conceptual model of a situation = Thematic Model



Numeric variables (so far)











2. The Path-Modelling problem





No bivariate correlation r(f,g)

g

g

 $r^2(g,f)=0$

f

h

g

1. Goodness of Fit of the Component Model

Pb: Every useful component partly depends on *all* others connected to it, directly or not...



No bivariate correlation r(f,g) Important partial effect of f on g, conditional on h



1. Goodness of Fit of the Component Model

Pb: Every useful component partly depends on *all* others connected to it, directly or not...



Proper (partial) effects cannot be correctly captured through global bivariate indicators.

 \Rightarrow THEME uses a Goodness-of-Fit criterion ψ capturing *multivariate* component-relationships



2. Components must capture interpretable variable structures

To be interpretable, components must be *structurally strong*, i.e. close to *observed variables bundles*



2. Components must capture interpretable variable structures

To be interpretable, components must be *structurally strong*, i.e. close to *observed variables bundles*



2. Components must capture interpretable variable structures

To be interpretable, components must be *structurally strong*, i.e. close to *observed variables bundles*



THEME uses an indicator of structural strength, $\phi \simeq$ closeness to bundles.

2. Components must capture interpretable variable structures

To be interpretable, components must be *structurally strong*, i.e. close to *observed variables bundles*

But What exactly is a bundle?

The question arises of the "locality" of the bundles of directions to focus on. We introduce a parameter l into ϕ , to tune the locality considered. Example:



2. Components must capture interpretable variable structures

To be interpretable, components must be *structurally strong*, i.e. close to *observed variables bundles*

But What exactly is a bundle?

The question arises of the "locality" of the bundles of directions to focus on. We introduce a parameter l into ϕ , to tune the locality considered. Example:



... one bundle? (*l* <<)

2. Components must capture interpretable variable structures

To be interpretable, components must be *structurally strong*, i.e. close to *observed variables bundles*

But What exactly is a bundle?

The question arises of the "locality" of the bundles of directions to focus on. We introduce a parameter l into ϕ , to tune the locality considered. Example:





2. Components must capture interpretable variable structures

To be interpretable, components must be *structurally strong*, i.e. close to *observed variables bundles*

But What exactly is a bundle?

The question arises of the "locality" of the bundles of directions to focus on. We introduce a parameter l into ϕ , to tune the locality considered. Example:



... four bundles? $(l \uparrow \uparrow)$

2. Components must capture interpretable variable structures

To be interpretable, components must be *structurally strong*, i.e. close to *observed variables bundles*

But What exactly is a bundle?

The question arises of the "locality" of the bundles of directions to focus on. We introduce a parameter l into ϕ , to tune the locality considered. Example:



... eight bundles, each one being a single direction? $(l \rightarrow \infty)$

2. Components must capture interpretable variable structures

To be interpretable, components must be *structurally strong*, i.e. close to *observed variables bundles*

But What exactly is a bundle?

The question arises of the "locality" of the bundles of directions to focus on. We introduce a parameter l into ϕ , to tune the locality considered. Example:



This ultimately depends on the data β Best *l* to be found through cross-validation.

... eight bundles, each one being a single direction? $(l \rightarrow \infty)$

3. Combining goodness of fit ψ and structural strength ϕ

The criterion to be maximised by a component *f*, given ALL others:



+1% on ϕ is compensated by -*s*% on ψ

Relative variations compensate at optimum

4. Algorithm \rightarrow component hierarchy

• The local-nesting (LocNes) principle:

In X_r , given all components in other groups:

 f_r^1 is the best component with respect to the criterion;



4. Algorithm \rightarrow component hierarchy

• The local-nesting (LocNes) principle:

In X_r , given all components in other groups:

- f_r^1 is the best component with respect to the criterion;
- f_r^2 is its best orthogonal complement



4. Algorithm \rightarrow component hierarchy

• The local-nesting (LocNes) principle:

In X_r , given all components in other groups:

- f_r^1 is the best component with respect to the criterion;
- f_r^2 is its best orthogonal complement
 - ... etc.



4. Algorithm \rightarrow component hierarchy

• The local-nesting (LocNes) principle:

In X_r , given all components in other groups:

- f_r^1 is the best component with respect to the criterion;
- f_r^2 is its best orthogonal complement

... etc.

And the algorithm loops over groups X_r until convergence.



Х

5. Backward component selection

• How to select the suitable number of components in each group?

Local nesting makes *backward selection* natural and easy: Start with too many components per theme, so as to capture real partial effects.

so as to capture real
$$X_s$$
 $f_s^1 f_s^2 \dots f_s^{Ks}$
 X_r $f_r^1 f_r^2 \dots f_r^{Kr}$
 X_u $f_u^1 f_u^2 \dots f_u^{Ku}$
 X_t $f_t^1 f_t^2 \dots f_t^{Kt}$

5. Backward component selection



5. Backward component selection



5. Backward component selection

• How to select the suitable number of components in each group?

Local nesting makes *backward selection* natural and easy: X_{s} f_{s}^{1} f_{s}^{2} \dots f_{s}^{Ks} $X_r \qquad f_r^1 \quad f_r^2 \quad \dots \quad X^r$ The last component in each theme best complements *all* other components \rightarrow • measure the gain it brings; $X_{u} \mid f_{u}^{1} \quad f_{u}^{2} \quad \dots \quad f_{u}^{Ku}$ • compare this gain to that of all last components; $X_p \quad f_p^1 \quad f_p^2 \quad \dots \quad f_p^{Kp}$ • eliminate the component bringing the smallest gain; $X_t \qquad f_t^1 \quad f_t^2 \quad \dots \quad f_t^{Kt}$ • re-estimate the model, etc.

5. Backward component selection



5. Backward component selection



5. Backward component selection



5. Backward component selection


5. Backward component selection



5. Backward component selection



5. Backward component selection



5. Backward component selection



5. Backward component selection



5. Backward component selection



5. Backward component selection



5. Backward component selection



5. Backward component selection



5. Backward component selection



5. Backward component selection



5. Backward component selection



5. Backward component selection



5. Backward component selection



5. Backward component selection



5. Backward component selection



5. Backward component selection

• How to select the suitable number of components in each group?

Local nesting makes *backward selection* natural and easy:



5. Backward component selection

• How to select the suitable number of components in each group?

Local nesting makes *backward selection* natural and easy:



5. Backward component selection

• How to select the suitable number of components in each group?

Local nesting makes *backward selection* natural and easy:



5. Backward component selection



1. The main window

	THEME (Version 08-02-2014)
	Run
	Data / Design Selection / Validation Advanced Options
	Data
Data Input	Calibration set
	Design
	Number of equations 1 👻
	Number of groups 3 -
Output 5	Save
	Save in C:/Resultats

2. From raw data to Thematic Model

• *Data file* = ASCII-file with tab separator: data_VDKM0_6groupes.txt

SAMPLE_NAME	code	Tchem_1	Tchem_2	Tchem_3	Tchem_4	Tchem_5	Tchem_6
cig1	6	1.54	0.67	1.85	42.38	3.56	89
cig2	2	0.4	0.92	1.95	42.18	2.31	66
cig3	1	0.56	0.75	1.8	43.23	2.74	123
cig4	5	0.97	0.96	1.83	41.27	2.79	96
cig5	5	0.66	0.85	1.47	41.37	2.29	119
cig6	1	0.89	0.77	2.03	42.2	2.88	142
cig29	4	0.87	0.77	1.89	40.72	2.75	177

Variables

Obs.

2. From raw data to Thematic Model

• *Data file* = ASCII-file with tab separator: data_VDKM0_6groupes.txt

SAMPLE_NAME	code	Tchem_1	Tchem_2	Tchem_3	Tchem_4	Tchem_5	Tchem_6
cig1	6	1.54	0.67	1.85	42.38	3.56	89
cig2	2	0.4	0.92	1.95	42.18	2.31	66
cig3	1	0.56	0.75	1.8	43.23	2.74	123
cig4	5	0.97	0.96	1.83	41.27	2.79	96
cig5	5	0.66	0.85	1.47	41.37	2.29	119
cig6	1	0.89	0.77	2.03	42.2	2.88	142
•••							
cig29	4	0.87	0.77	1.89	40.72	2.75	177

Variables

• Design of the thematic model:

6 themes 2 equations

Obs.



2. From raw data to Thematic Model

• *Data file* = ASCII-file with tab separator: data_VDKM0_6groupes.txt

	SAMPLE_NAME	code	Tchem_1	Tchem_2	Tchem_3	Tchem_4	Tchem_5	Tchem_6
	cig1	6	1.54	0.67	1.85	42.38	3.56	89
	cig2	2	0.4	0.92	1.95	42.18	2.31	66
	cig3	1	0.56	0.75	1.8	43.23	2.74	123
bs.	cig4	5	0.97	0.96	1.83	41.27	2.79	96
	cig5	5	0.66	0.85	1.47	41.37	2.29	119
	cig6	1	0.89	0.77	2.03	42.2	2.88	142
	cig29	4	0.87	0.77	1.89	40.72	2.75	177
	TGC	cci	1	3 🔍	1 %,	<u>,</u> 3	1	1

Variables

Group Coding (0 = variable not used)

Thematic

• Design of the thematic model:

6 themes 2 equations



2. From raw data to Thematic Model

	Kun
	Data / Design Selection / Validation Advanced Options
	Data
Data Innut	Calibration set C:/Users/FRAMTEV/Desktop/data THEME/VDKAM0_THEMEextSEER.txt
	Validation set
L	
	Design
	Number of equations 2 =
	Number of groups 6
	Save
	Save in C:/Resultats

2. From raw data to Thematic Model

	THEME (Version 08-02-2014)
	Run
	Data / Design Selection / Validation Advanced Options
	Data
	Calibration set C:/Users/FRAMTEV/Desktop/data THEME/VDKAM0_THEMEextSEER.txt Validation set
Model design	Number of equations 2 Design Number of groups 6
	Save
	Save in C:/Resultats

2. From raw data to Thematic Model

Number of components in groups

data & model design									X
in									
#comp. Eq.1 Eq.2	NA	G-1	G-2 1 •	G-3 1 •	G-4 1 •	G-5 1 •	G-6 1 •		*
	NA	G-1	G-2	G-3	G-4	G-5	G-6	4	*
SAMPLE_NAME	æ	С	С	С	С	С	0		
Tchem_1	0	œ	С	С	С	С	0		I.
Tchem_2	0	С	С	•	0	0	С		=
Tchem_3	C	œ	С	С	С	С	С		
Tchem_4	0	0	С		0	0	С		
Tchem_5	0	œ	С	С	С	0	0		
Tchem_6	0	œ	С	С	0	0	0		
Tchem_7	0	æ	С	С	С	0	0		
Tchem_8	C	œ	С	С	С	С	0		
Tchem_9	С	œ	С	С	С	C	С		
Tchem_10	С	œ	С	С	С	С	0		
Tchem_11	0	æ	С	С	0	0	С		
Tchem_12	0	æ	С	С	С	0	С		
Tchem_13	0	æ	С	С	0	0	0		
Tchem_14	C	œ	С	С	С	0	С		
Tchem_15	0	œ	С	С	С	С	0		
Tchem_16	0	С	С		0	0	0		
Tchem_17	0	æ	С	С	С	С	0		
Tchem_18	С	æ	С	С	C	С	C		

2. From raw data to Thematic Model

Number of components in groups Role of groups in equation (explanatory=X, dependent =Y)

😰 data & model design								
Run								
#comp. Eq.1 Eq.2	NA	G-1 2 X •	G-2 2 X •	G-3 2 X •	G-4 2 Y • X •	G-5 2 X •	G-6 2 ▼	
•	NA	G-1	G-2	G-3	G-4	G-5	G-6	
SAMPLE_NAME	œ	С	C	С	С	С	С	
Tchem_1	С	(•	0	С	С	С	С	
Tchem_2	С	С	С	(•	0	С	0	
Tchem_3	С	œ	С	0	0	С	0	
Tchem_4	С	С	С	œ	С	С	С	
Tchem_5	С	œ	C	0	0	С	0	
Tchem_6	С	œ	0	0	0	С	0	
Tchem_7	С	œ	С	С	С	С	0	
Tchem_8	С	œ	С	С	С	С	0	
Tchem_9	С	œ	С	С	С	С	0	
Tchem_10	С	œ	С	С	0	С	С	
Tchem_11	С	æ	С	С	С	С	С	
Tchem_12	С	æ	0	С	С	0	С	
Tchem_13	C	œ	0	0	0	0	0	
Tchem_14	С	œ	0	0	0	0	0	
Tchem_15	С	œ	0	С	С	0	С	
Tchem_16	С	С	0	œ	С	С	С	
Tchem_17	С	œ	0	0	0	0	С	
Tchem_18	0		С	0	0	C	С	

2. From raw data to Thematic Model

Number of components in groups Role of groups in equation (explanatory=X, dependent =Y)

If TGC line in datafile, pre-filled. Else, interactive design:

#comp.	NA	G-1 2	G-2 2	G-3 2	G-4 2	G-5 2	G-6 2	
Eq.1		X -	X -	X -	γ –	•	-	
Eq.2		•	•	•	X 💌	X -	▼	
								Þ
1978 1010 1020 2000	NA	G-1	G-2	G-3	G-4	G-5	G-6	
SAMPLE_NAME	ſ	0	С	С	0	С	C	
Tchem_1	С	œ	0	С	С	С	0	
Tchem_2	С	С	0	•	С	0	С	
Tchem_3	С	œ	0	С	С	С	0	
Tchem_4	С	0	С	ſ	С	С	С	
Tchem_5	C	æ	0	0	0	0	0	
Tchem_6	С	œ	С	С	С	С	0	
Tchem_7	С	æ	С	С	С	С	0	
Tchem_8	С	œ	0	С	С	C	C	
Tchem_9	С	œ	0	C	0	C	С	
Tchem_10	С	œ	0	C	С	С	С	
Tchem_11	С	æ	С	С	С	С	С	
Tchem_12	С	æ	С	С	С	С	С	
Tchem_13	0	æ	0	C	0	0	0	
Tchem_14	С	æ	0	С	С	0	С	
Tchem_15	С	œ	С	С	С	С	С	
Tchem_16	С	С	С	•	С	С	С	
Tchem 17	С	æ	C	С	С	С	С	
Tchem 18	C	œ	C	C	C	C	C	

3. Setting the selection & validation parameters

Component selection	/ Design Selection / Validation Advanced Options Component selection kward selection no ▼ ance 1 ▼ Model validation
Component selection	/ Design Selection / Validation Advanced Options Component selection kward selection no ▼ ance 1 ▼ Model validation
Component Selection	Component selection
selection Bac Bala	ikward selection no - ance 1 - Model validation
Cro	Model validation
Cro	very validation (Validation) 0 -

3. Setting the selection & validation parameters



4. Setting the structural strength and goodness of fit parameters

(THEME (Version 08-02-2014)	
	Run	
	Data / Design Selection / Validation Advanced Options	
Structural strength parameters GoF {	Balance Mode Structural strength: THEME- for THEME-COST: = 2 Adjustement quality Multiplicative S= 1 Multiplicative	
)

5. Launching estimation

Kun		
Data / Design Selection / Va	alidation Advanced Options	
Balance Mode	Mode A 🔹	
Structural strength: THEME-	COST 🔻	
for THEME-COST: I=	2 ▼ s= 1 ▼	
Adjustement quality	Multiplicative -	

6. Waiting for results





6. Waiting for results



Model (2_2_2_1_2)	 Model (2_2_2_2_2)
Execution of THEME-SEER	Coefficient estimations

6. Waiting for results





222212)		Coefficie	ent estimations
Model (2_2	22222)		
6. Waiting for results



6. Waiting for results















7. Reaping results



Axis 1

7. Reaping results

Getting ALL the results as an object





Help-files yet to be written...

Soon available on the CRAN

THEME - Bry, Verron ; Rencontres R 2014

THE END

Thank you, all

Bry X., Verron T., Redont P. (2010) : *Multidimensional Exploratory Analysis of a Structural Model using a class of generalized covariance criteria*, COMPSTAT 2010, Proceedings, Springer.

Bry X., Redont P,. Verron T., Cazes P. (2012) : *THEME-SEER: a multidimensional exploratory technique to analyze a structural model using an extended covariance criterion*, Journal of chemometrics, 26, pp 158-169.