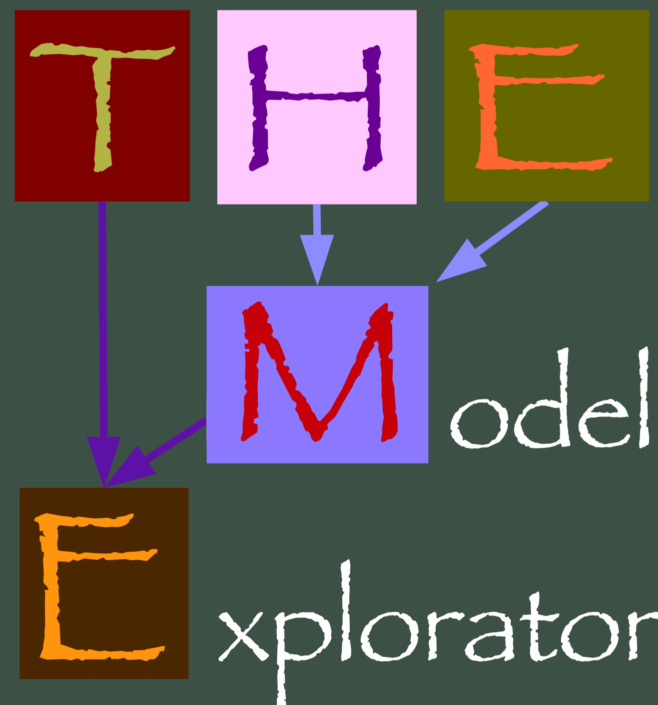
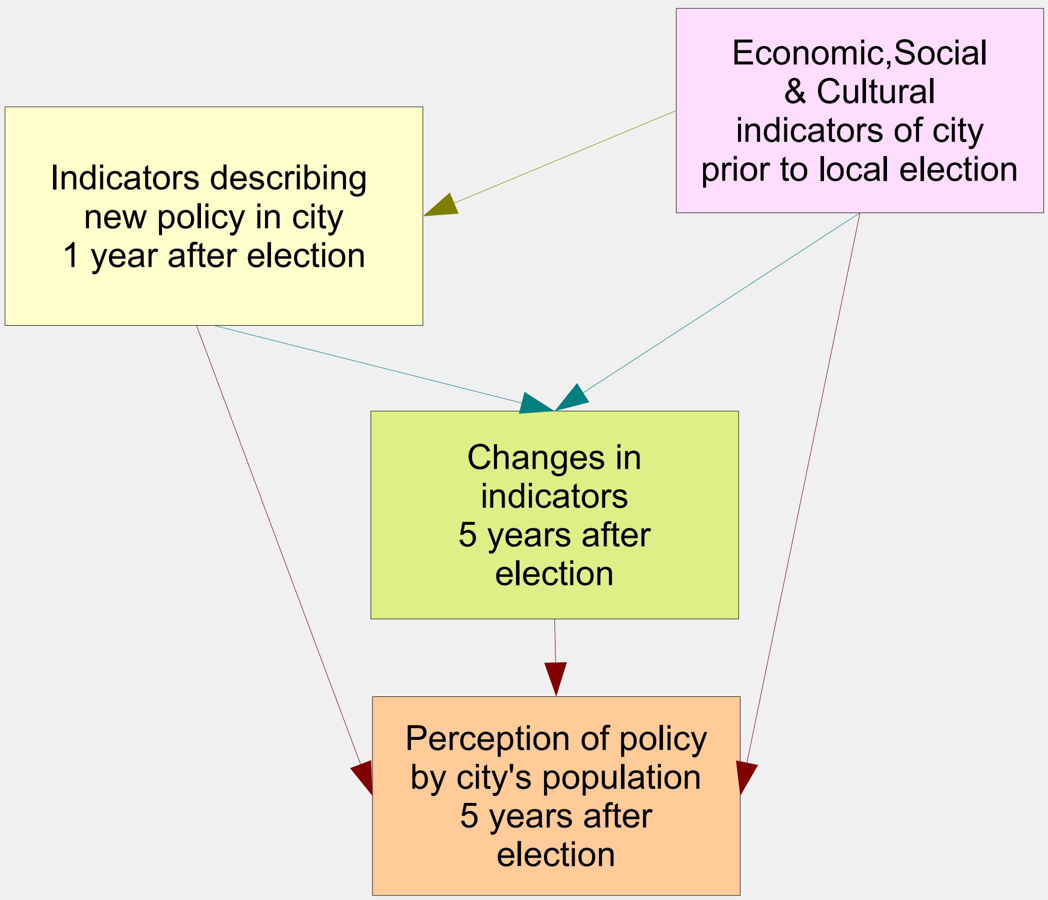
A component-based
Multidimensional Path Modelling R-package: THEME



X. Bry *I3M, Univ. Montpellier II*
T. Verron *ITG - SEITA, Centre de recherche*

Data and Problem:

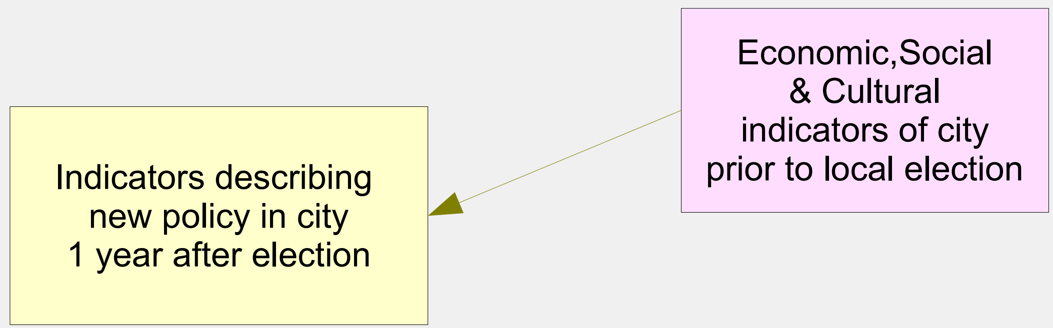
1. Conceptual model of a situation = Thematic Model



Data and Problem:

1. Conceptual model of a situation = Thematic Model

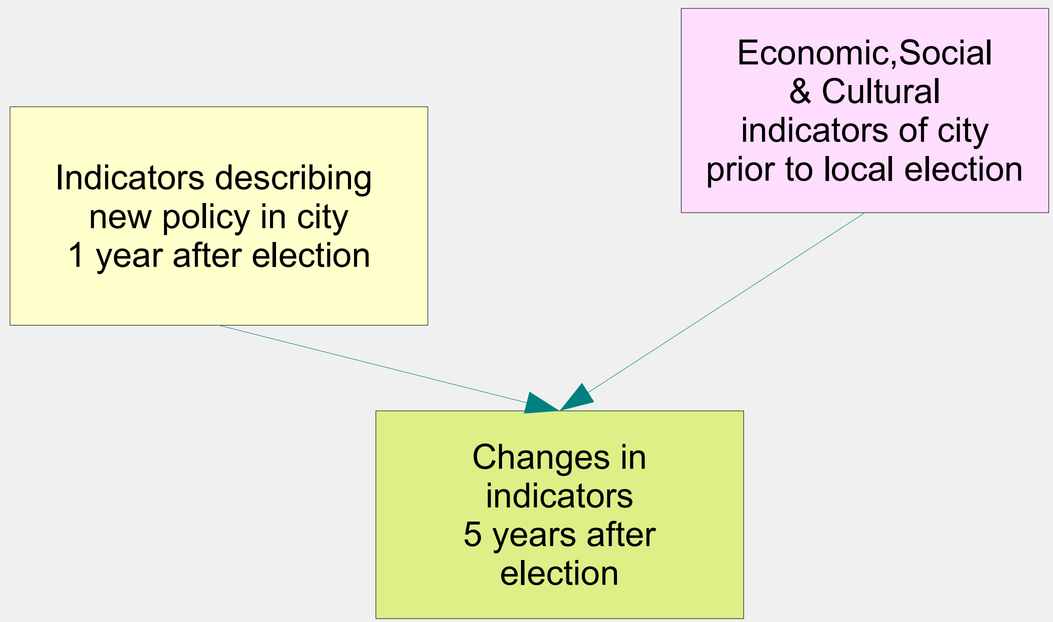
Equation 1



Data and Problem:

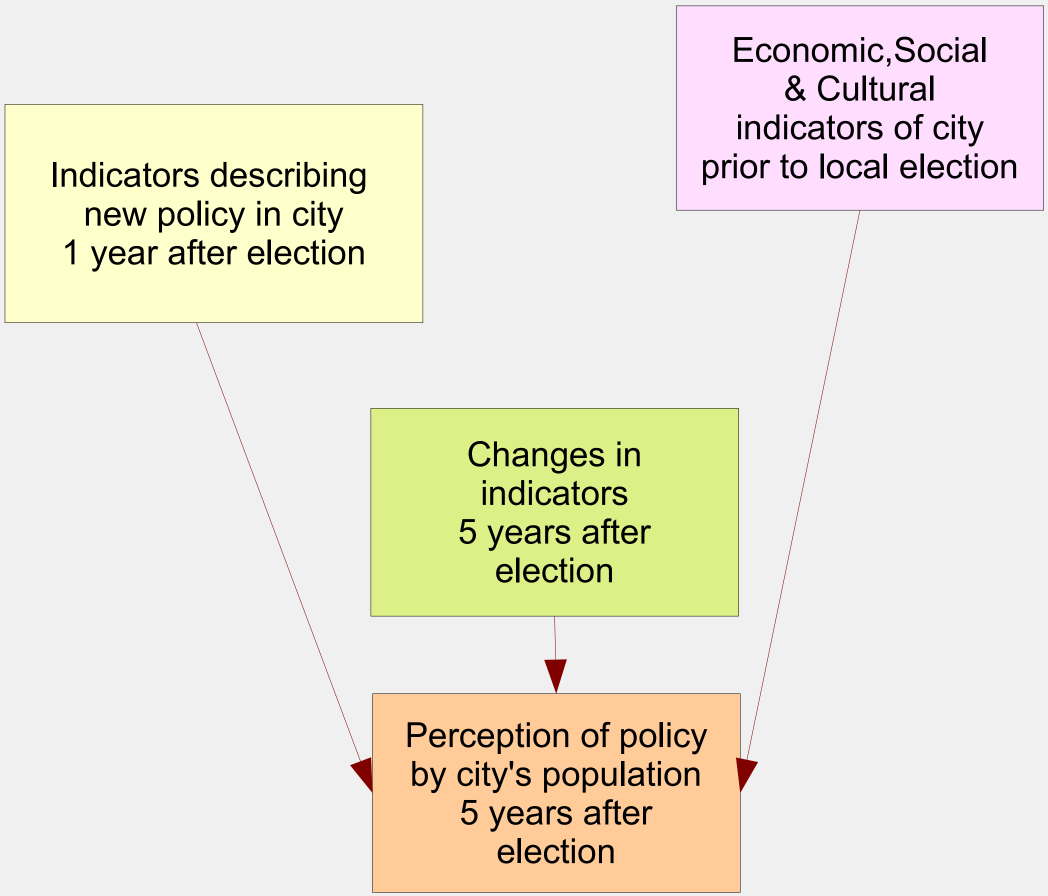
1. Conceptual model of a situation = Thematic Model

Equation 2



Data and Problem:

1. Conceptual model of a situation = Thematic Model

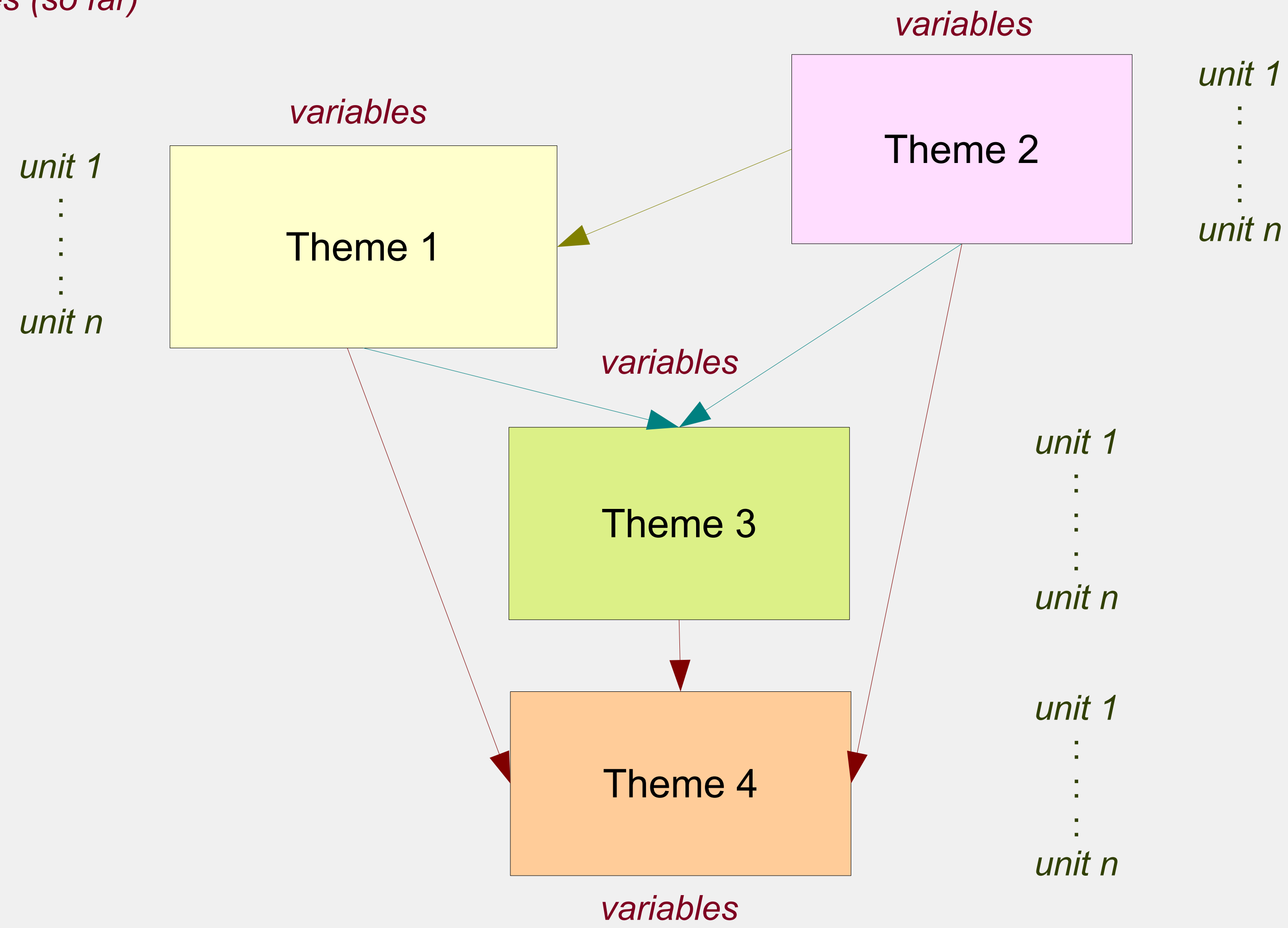


Equation 3

Data and Problem:

1. Conceptual model of a situation = Thematic Model

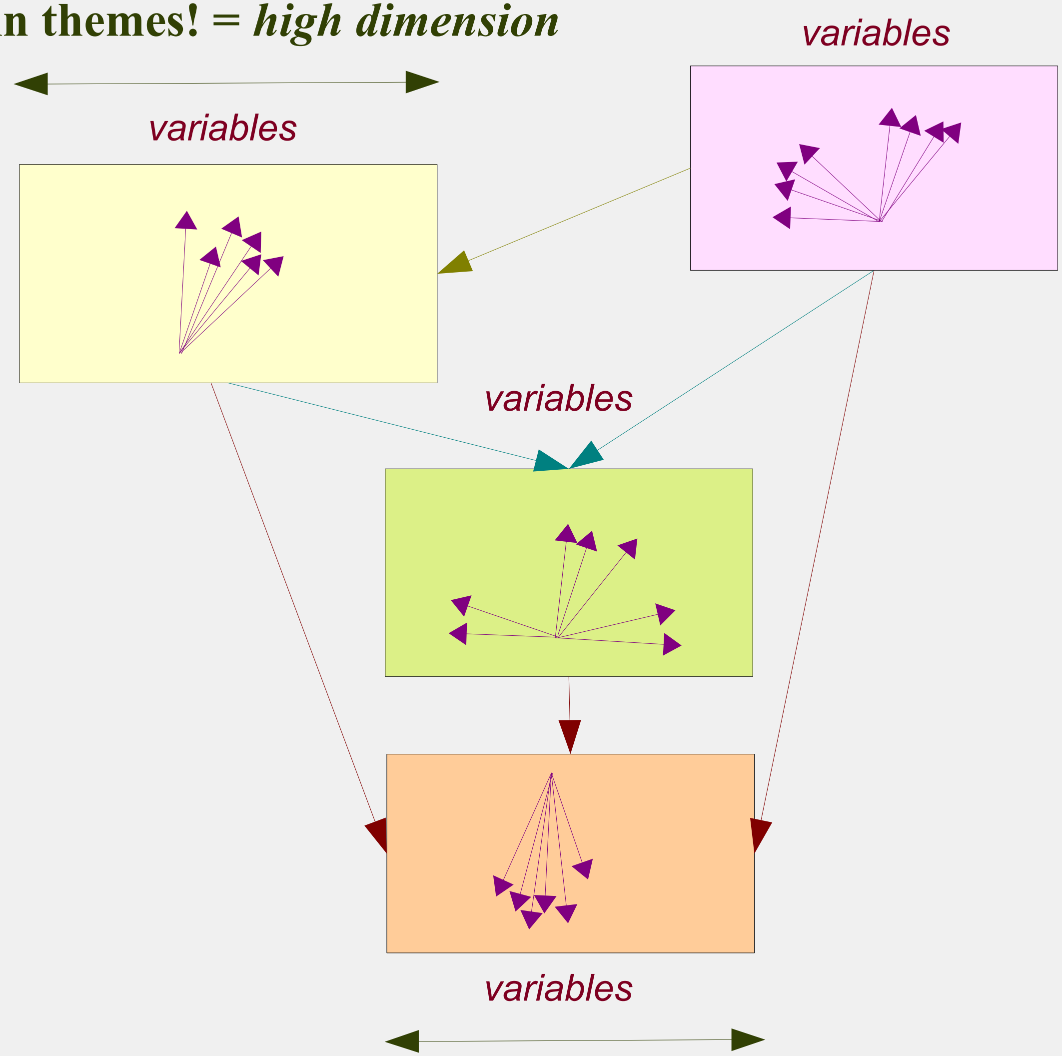
Numeric variables (so far)



Data and Problem:

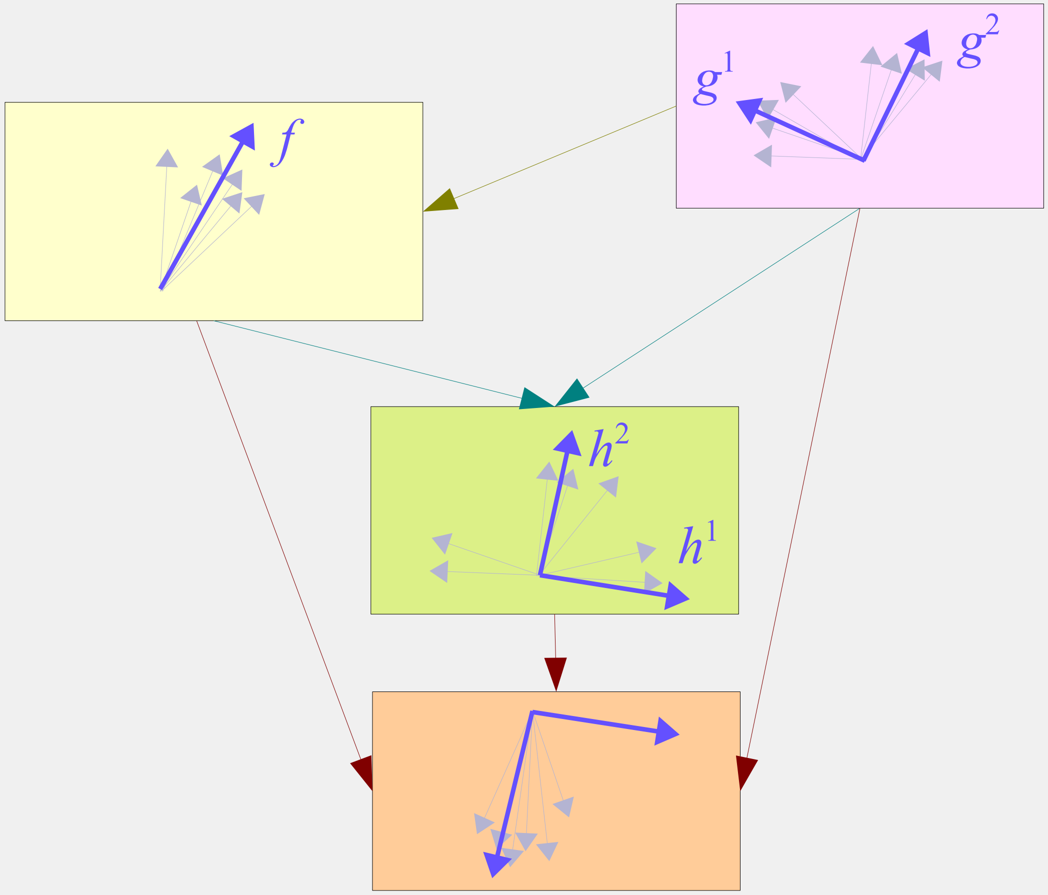
2. The Path-Modelling problem

- **Too many variables in themes! =** ***high dimension***



Data and Problem:

2. The Path-Modelling problem

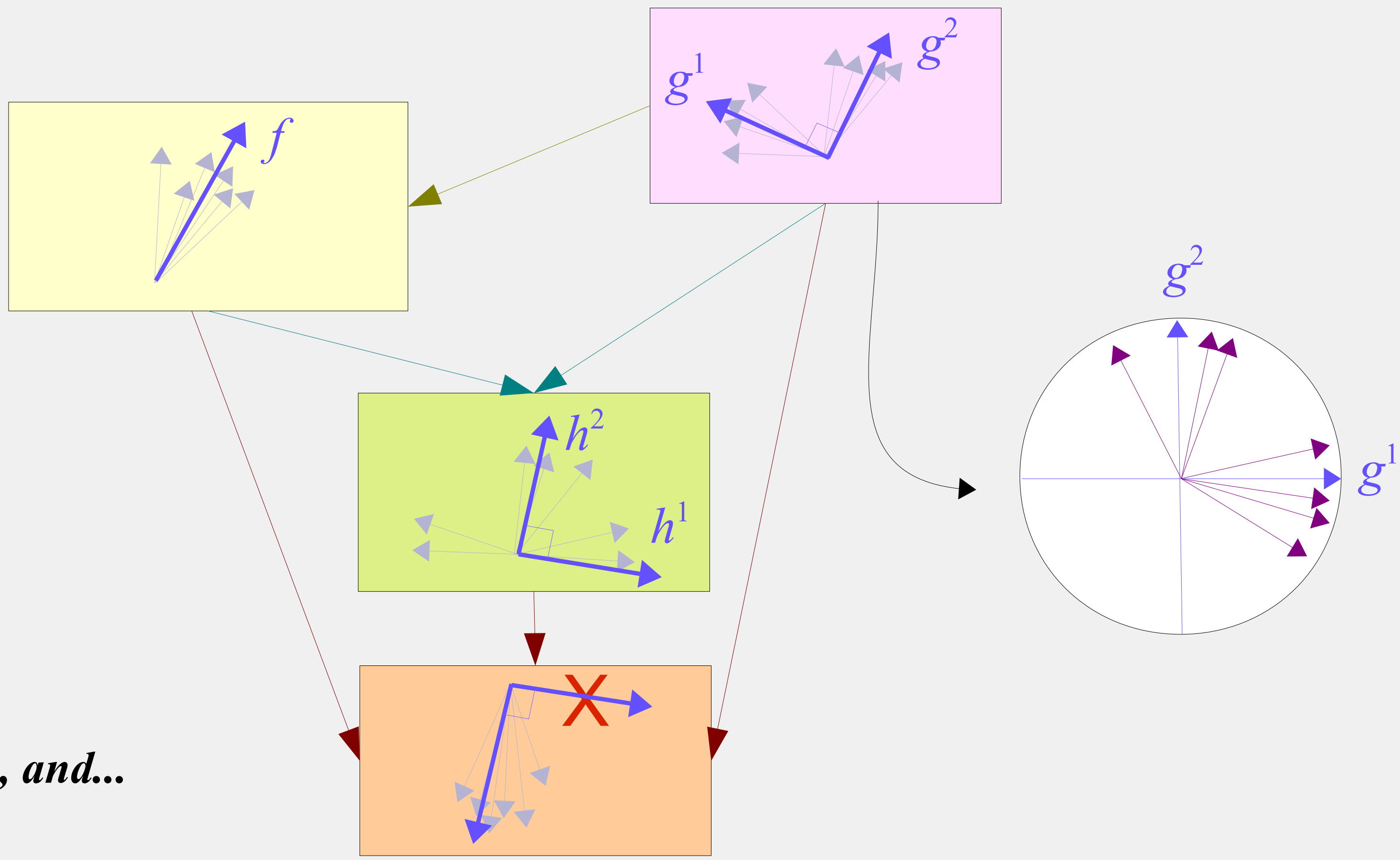


⇒ ***Reduce dimension in each theme ...***

through a few *Thematic Components*

Data and Problem:

2. The Path-Modelling problem



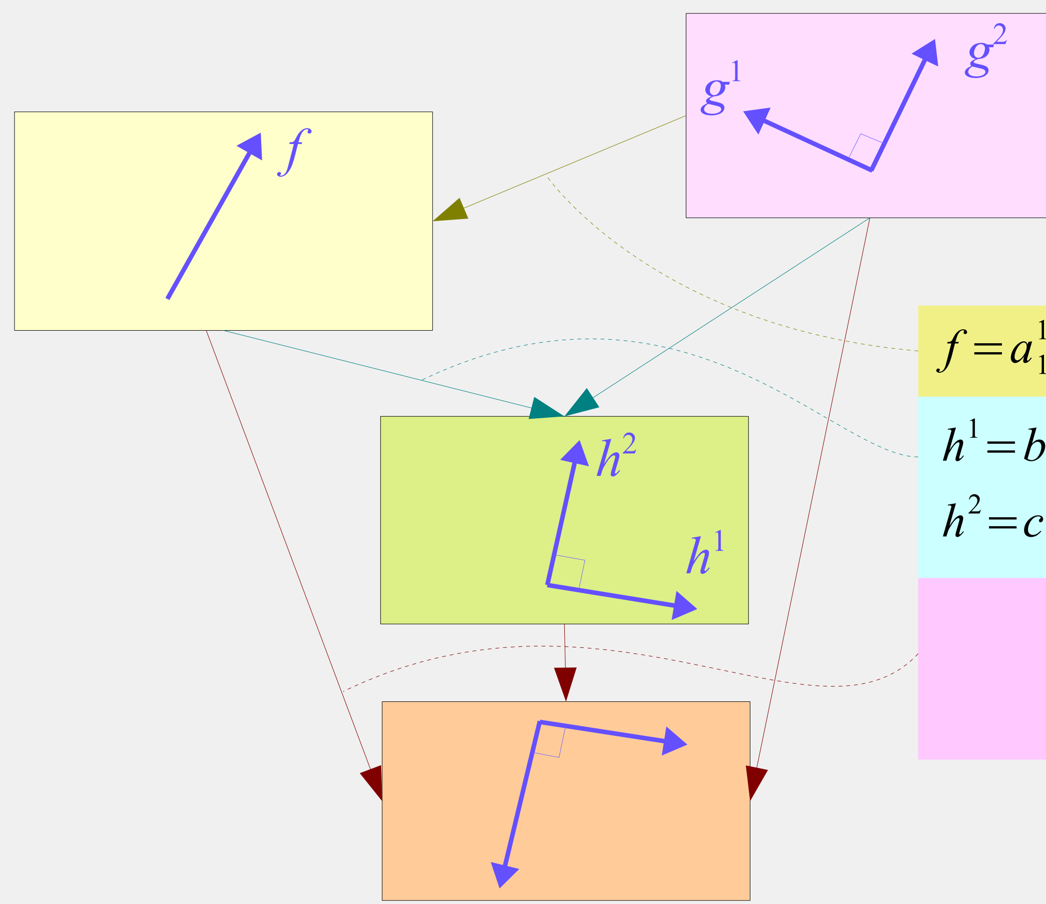
⇒ Reduce dimension in each theme …

through a few
Thematic Components

… non-redundant, strong, and…

Data and Problem:

2. The Path-Modelling problem



$$f = a_1^1 g^1 + a_2^1 g^2 + \varepsilon_f$$

$$h^1 = b_0 f + b_1 g^1 + b_2 g^2 + \varepsilon_{h1}$$

$$h^2 = c_0 f + c_1 g^1 + c_2 g^2 + \varepsilon_{h2}$$

etc.

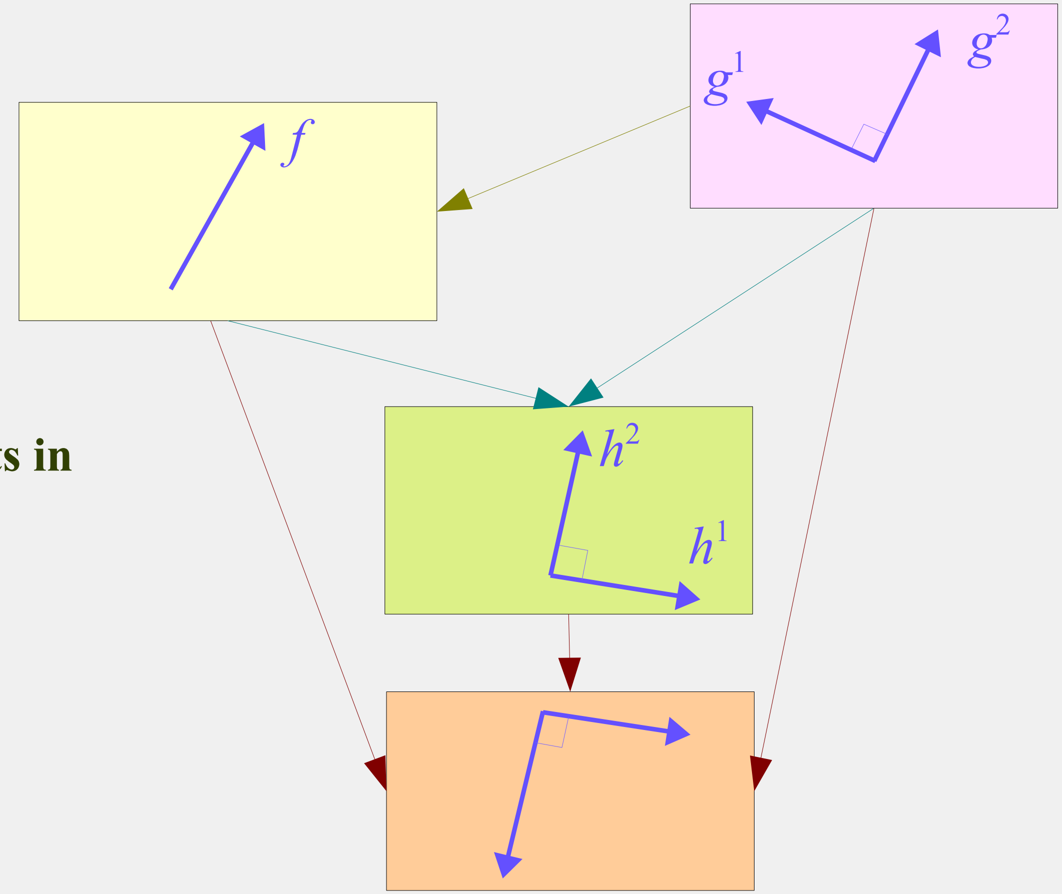
⇒ Reduce dimension in each theme …

through a few ***Thematic Components***

… ***satisfying the model*****.**

Data and Problem:

2. The Path-Modelling problem

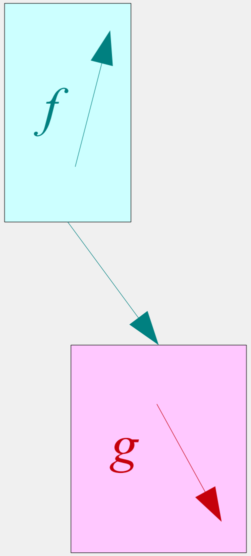


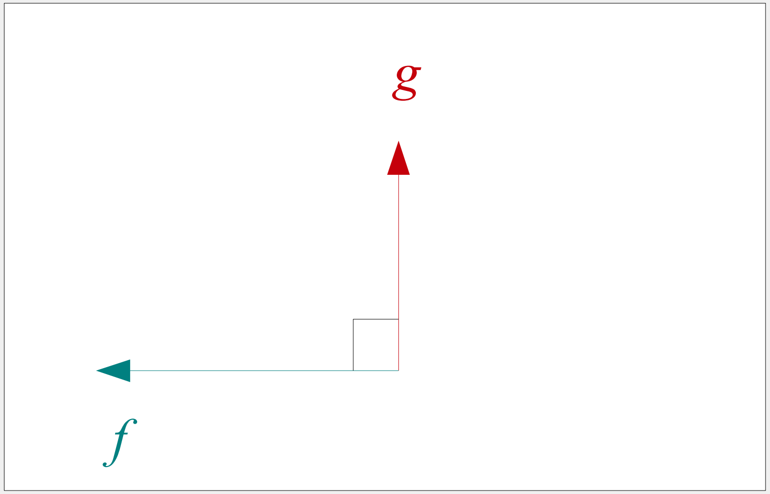
⇒ **How many components in each theme?**

... and which?

How THEME works

2. The Path-Modelling problem





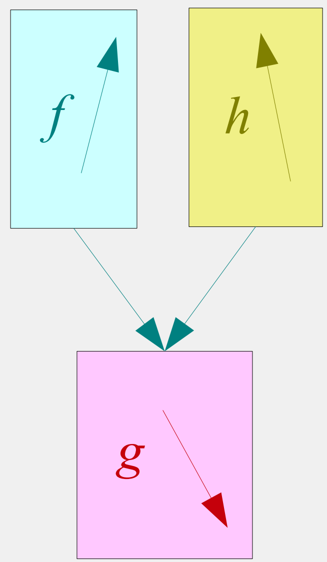
$r^2(g, f) = 0$

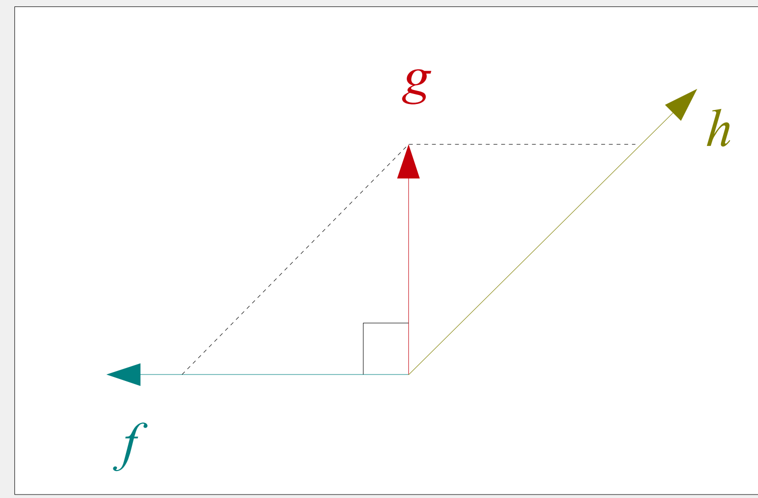
Pb: Every useful component partly depends on *all* others connected to it, directly or not...

No bivariate correlation r(f,g)

How THEME works

1. Goodness of Fit of the Component Model





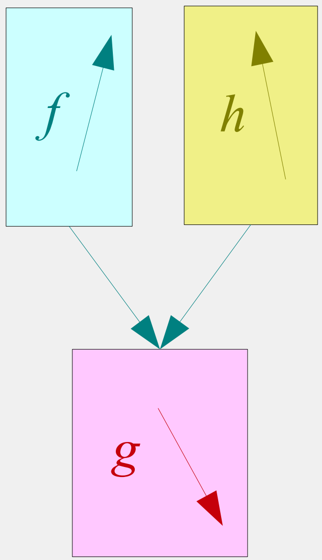
$R^2(g \mid f,h) = 1$

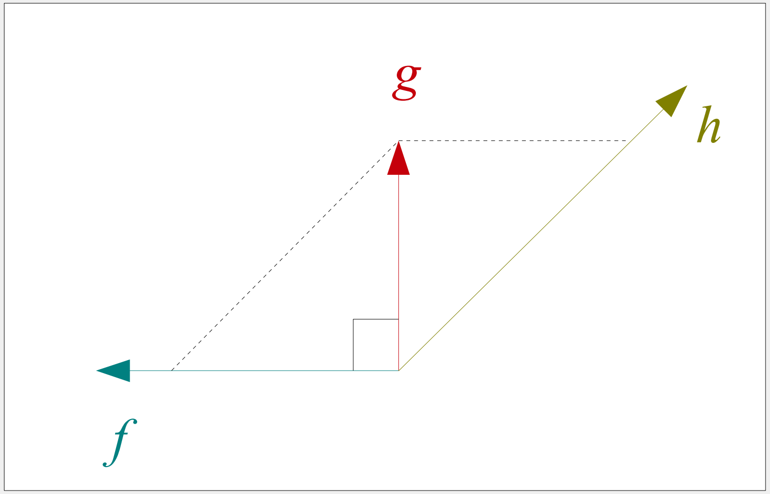
Pb: Every useful component partly depends on *all* others connected to it, directly or not...

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

How THEME works

1. Goodness of Fit of the Component Model





$R^2(g \mid f,h) = 1$

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

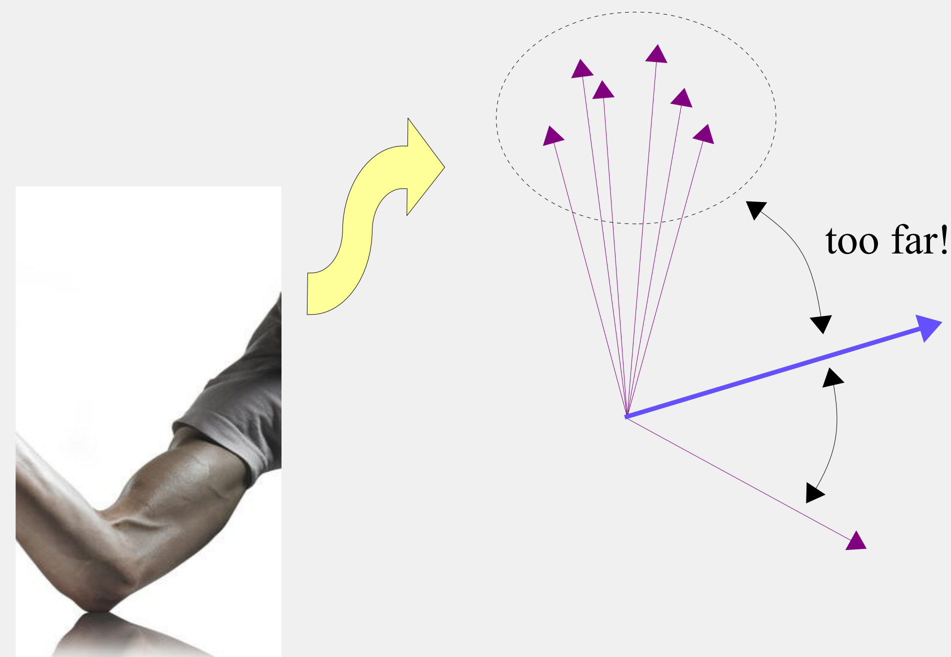
Proper (partial) effects cannot be correctly captured through global bivariate indicators.

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

How THEME works

2. Components must capture interpretable variable structures

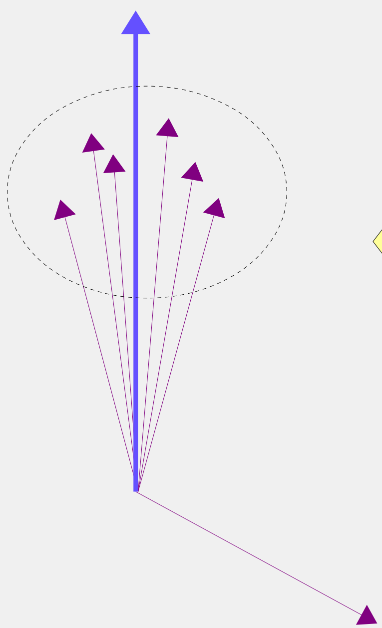
To be interpretable, components must be *structurally strong*,
i.e. close to *observed variables bundles*

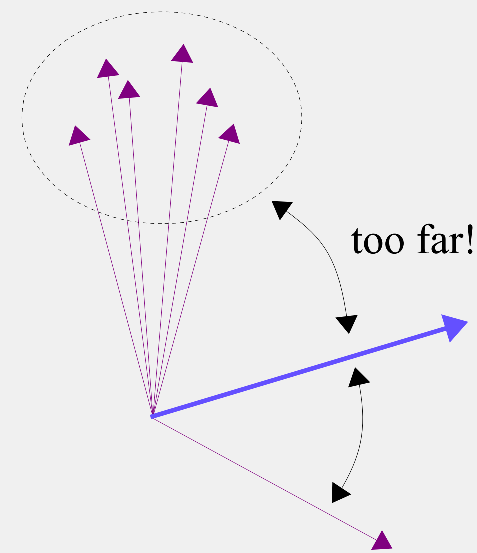


How THEME works

2. Components must capture interpretable variable structures

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

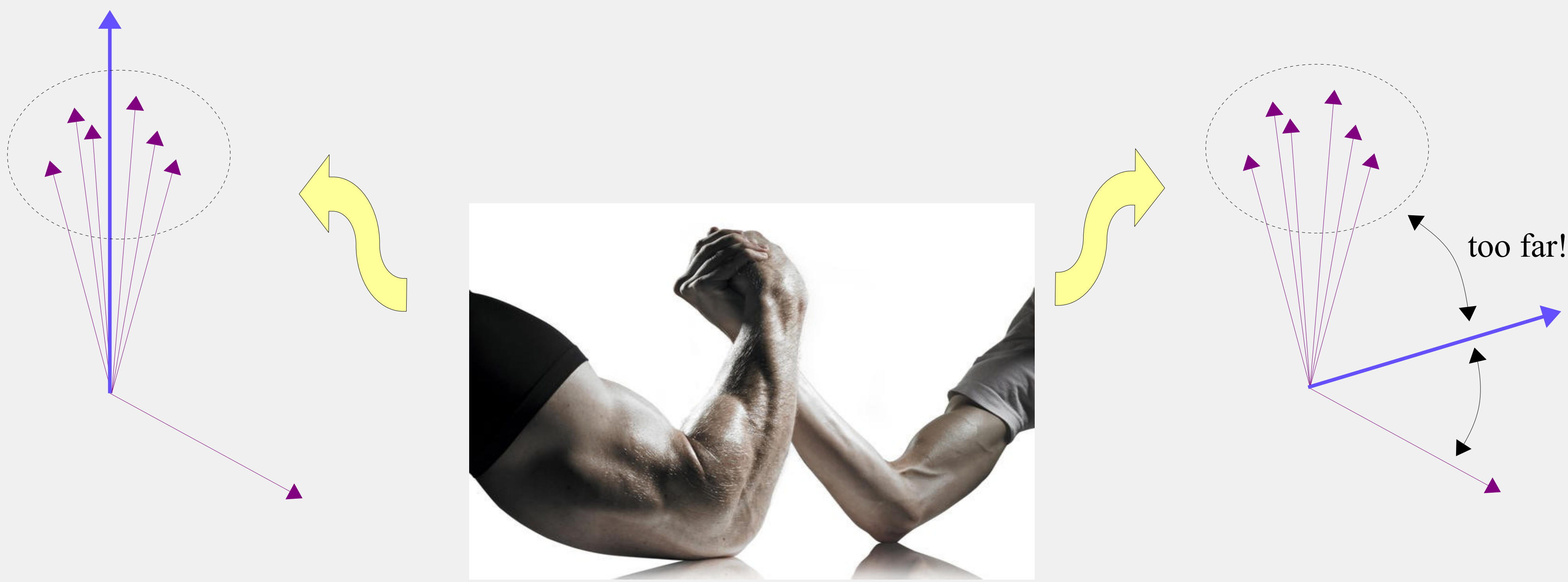




How THEME works

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.

How THEME works

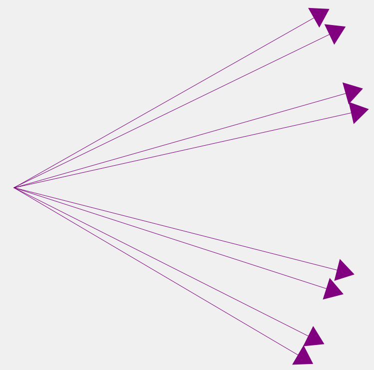
2. Components must capture interpretable variable structures

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:



How THEME works

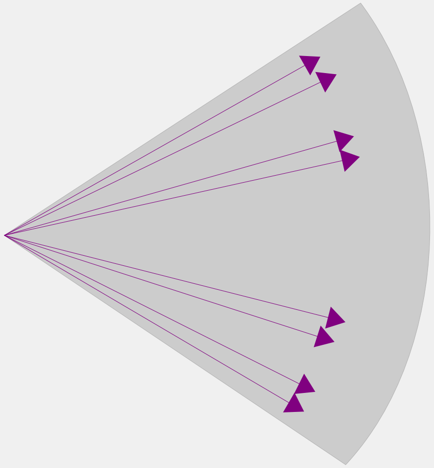
2. Components must capture interpretable variable structures

**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:



How THEME works

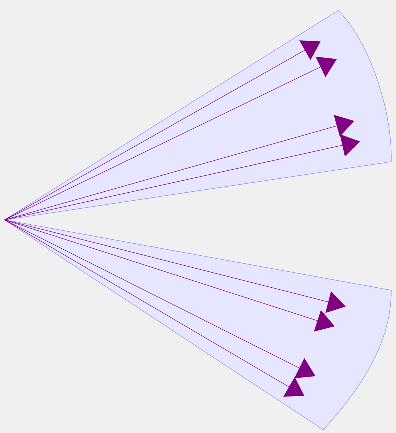
2. Components must capture interpretable variable structures

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:



... two bundles? (l ↑)

How THEME works

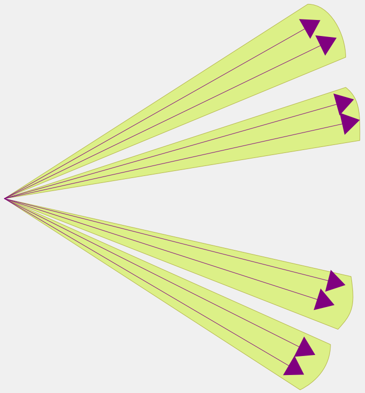
2. Components must capture interpretable variable structures

**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 ↑↑)

How THEME works

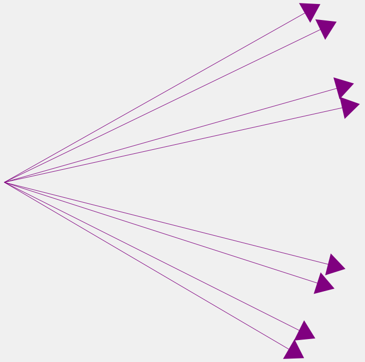
2. Components must capture interpretable variable structures

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? ($l \to \infty$)

How THEME works

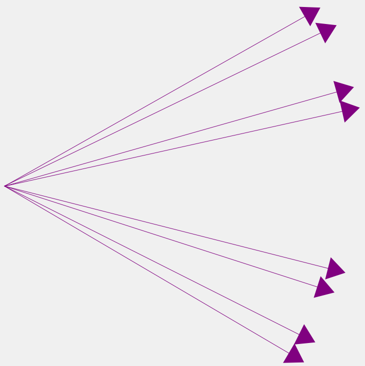
2. Components must capture interpretable variable structures

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? ($l \to \infty$)

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

How THEME works

3. Combining goodness of fit ψ and structural strength φ

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

$$\psi(f)\,\phi_l(f)^s$$

GoF Str. importance given to the SR relative to the GoF .

+1% on ϕ is compensated by -s% on ψ

Relative variations compensate at optimum

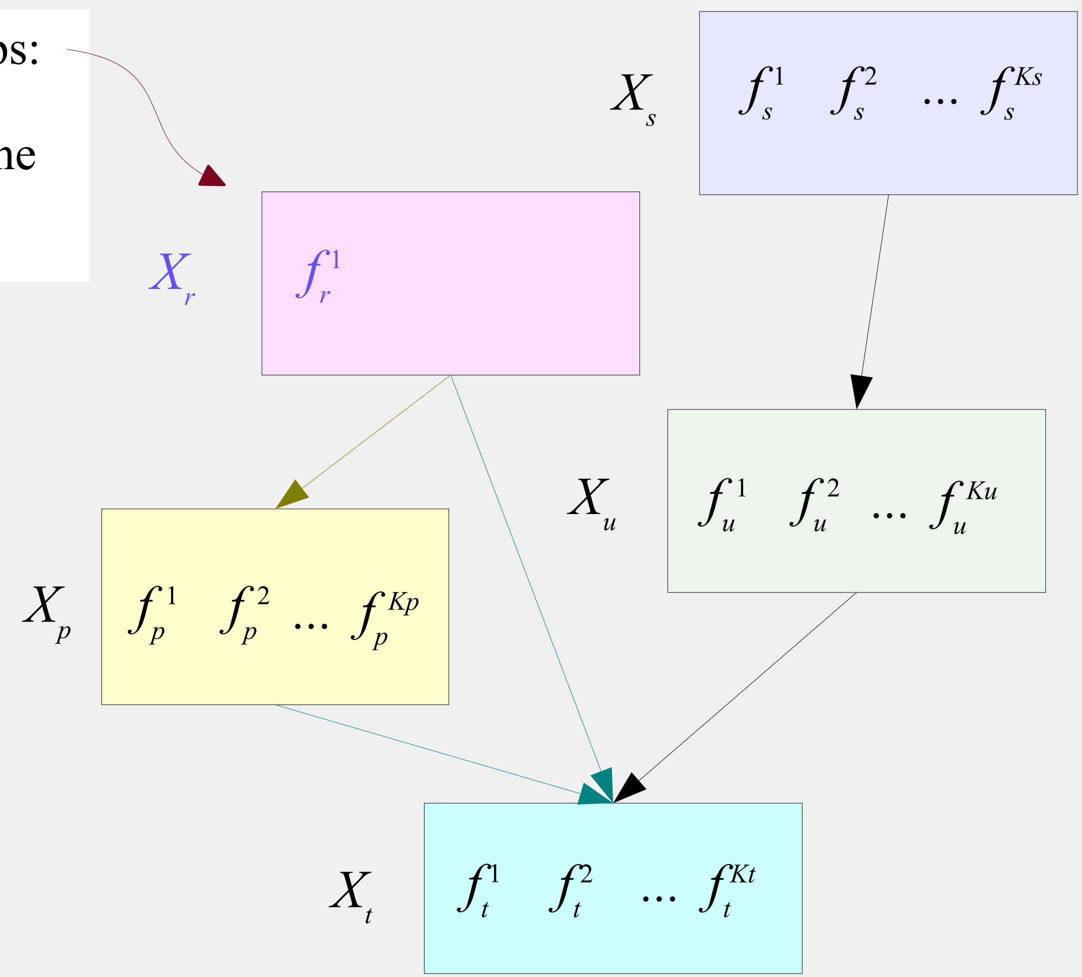
How THEME works

4. Algorithm → 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;



How THEME works

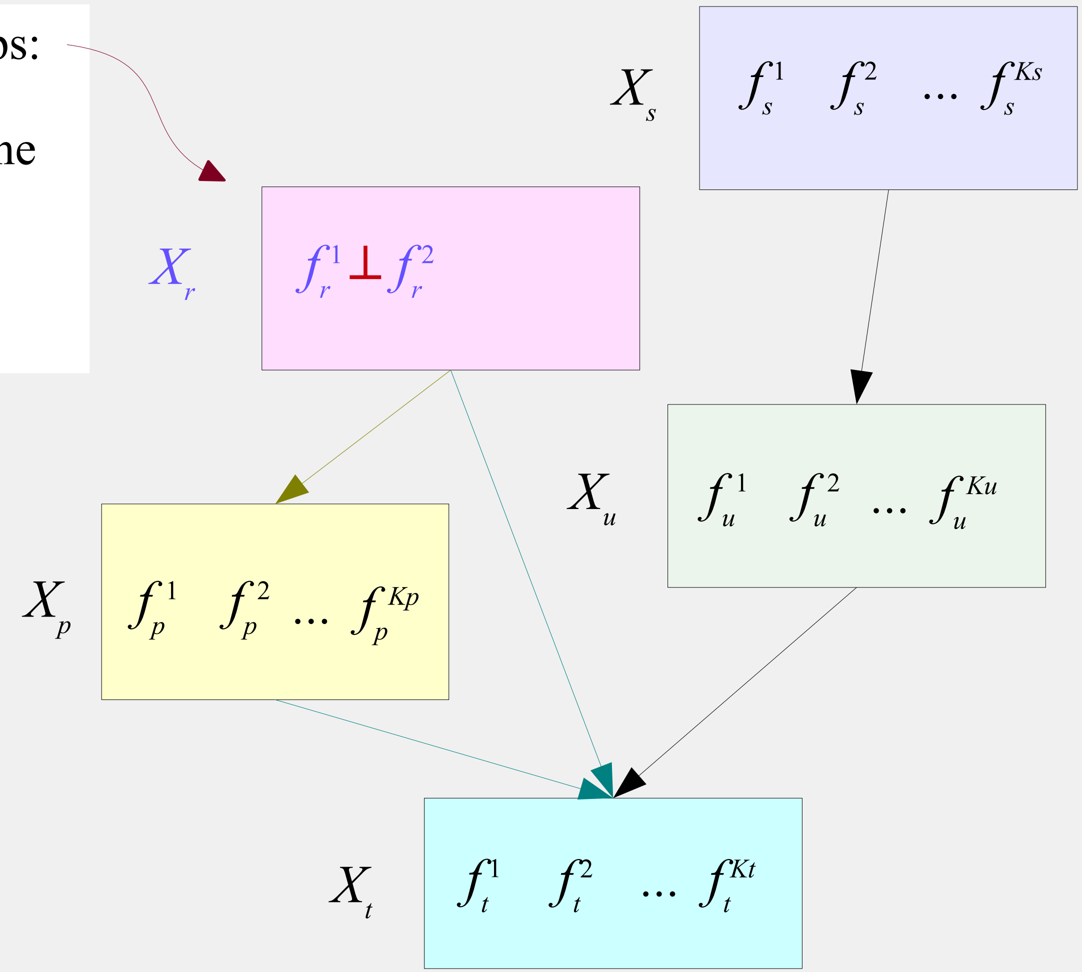
4. Algorithm → component hierarchy

- **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



How THEME works

4. Algorithm → component hierarchy

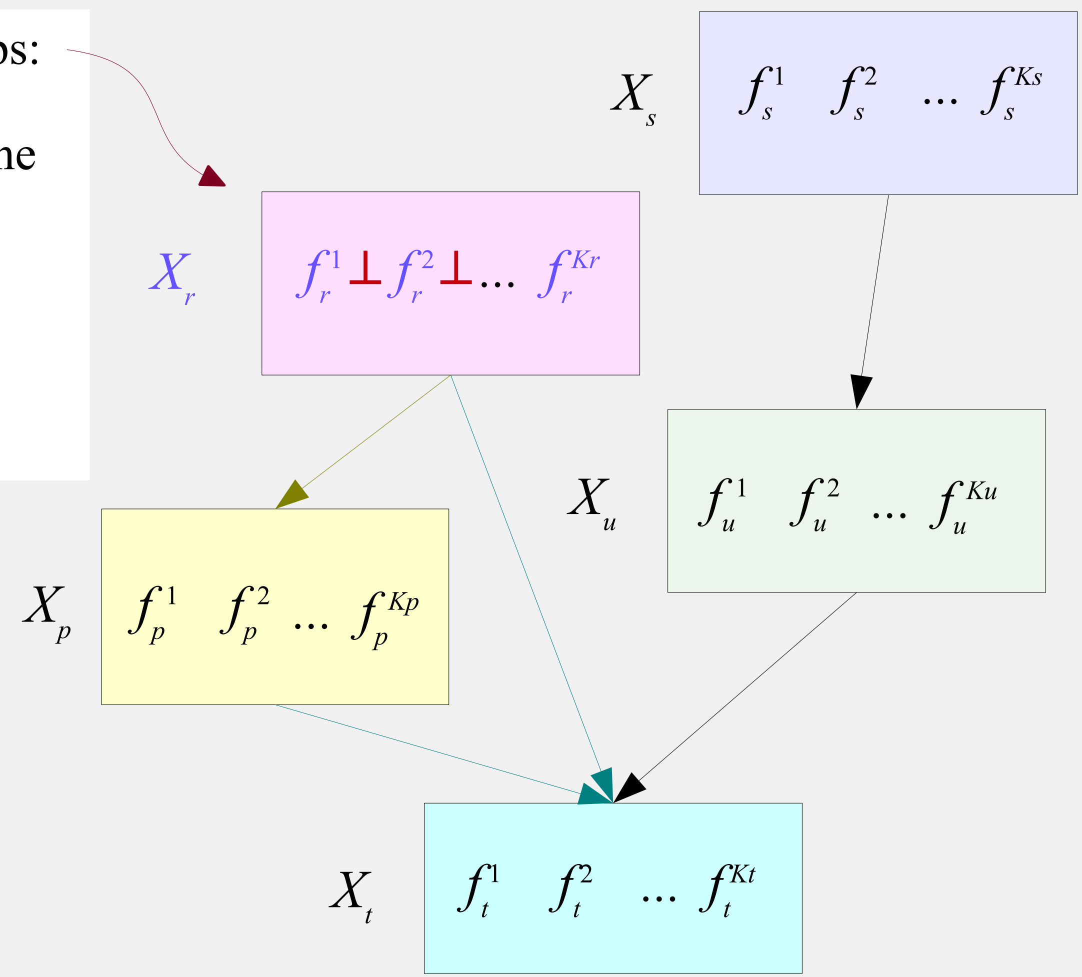
- **The local-nesting (LocNes) principle:**

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.



How THEME works

4. Algorithm → component hierarchy

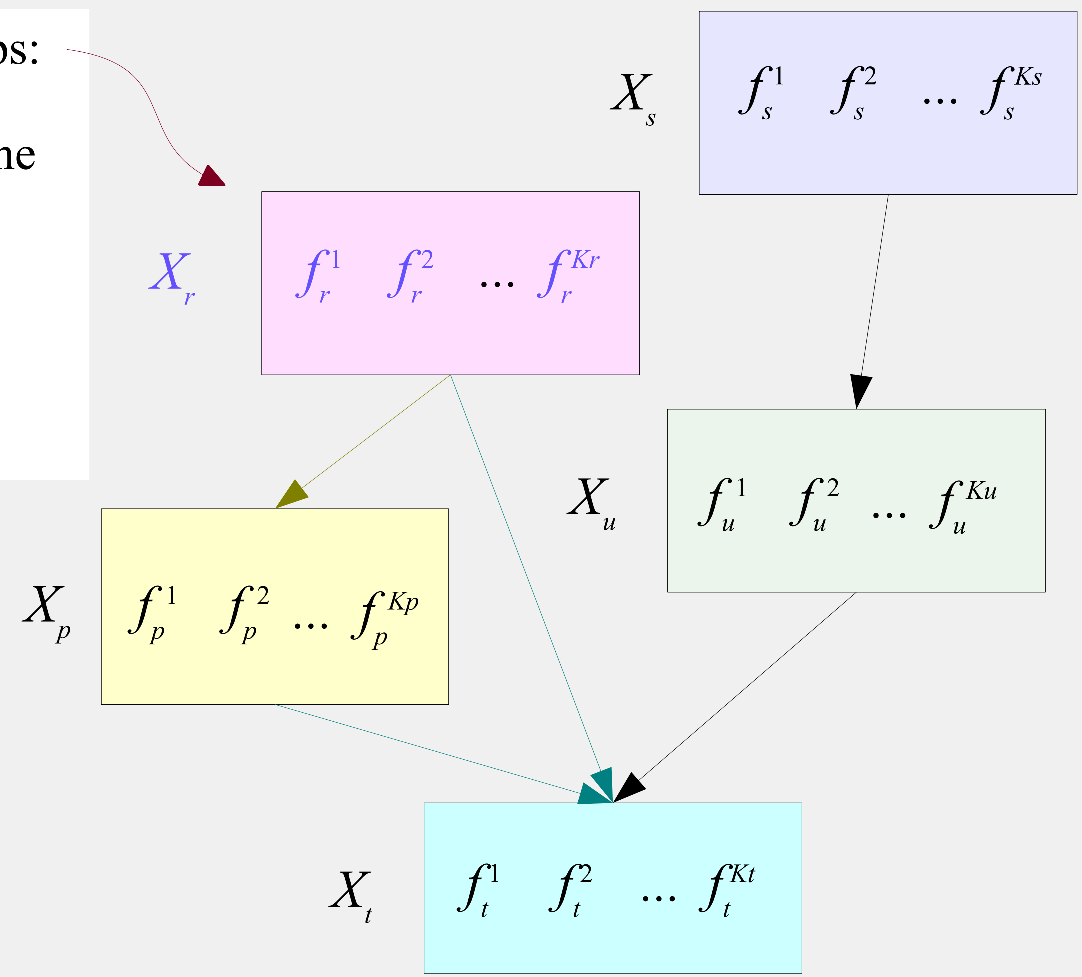
- **The local-nesting (LocNes) principle:**

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.



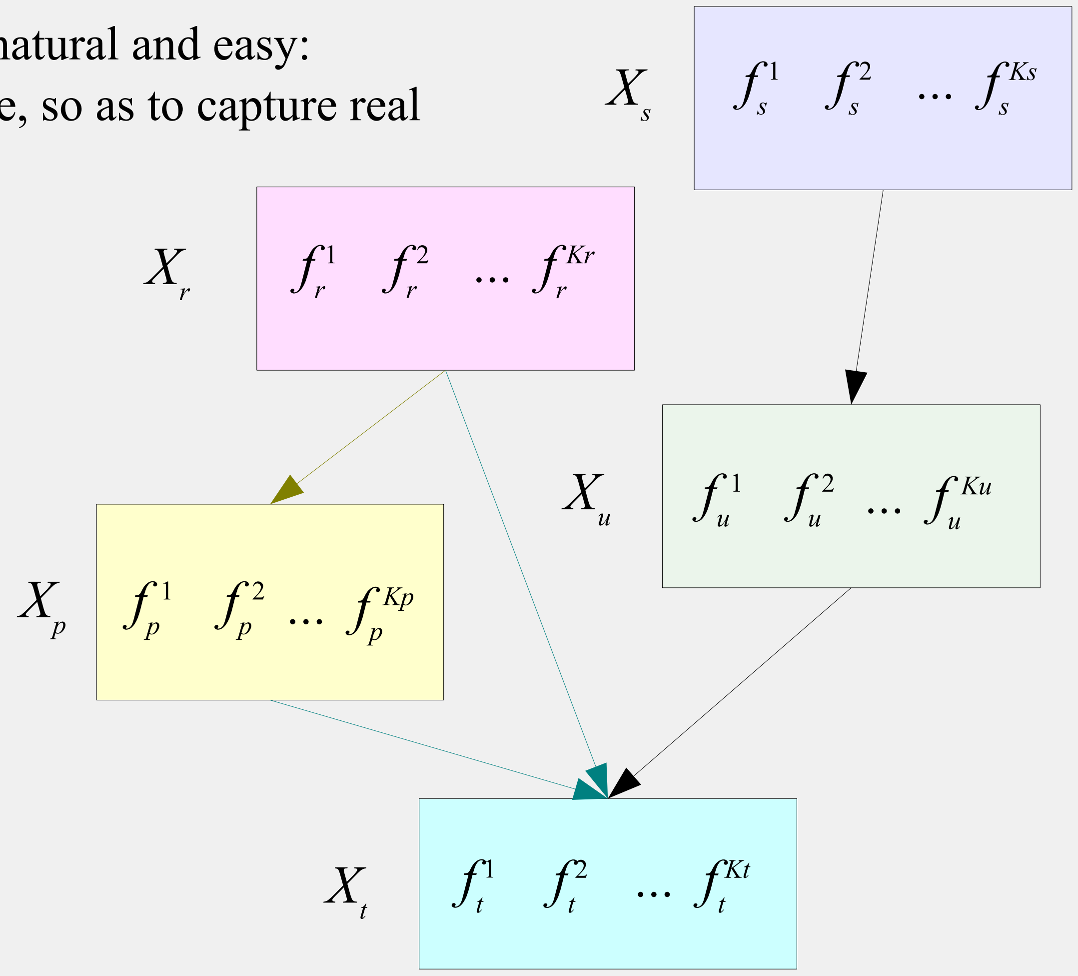
And the algorithm loops over groups X_r until convergence.

How THEME works

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.



How THEME works

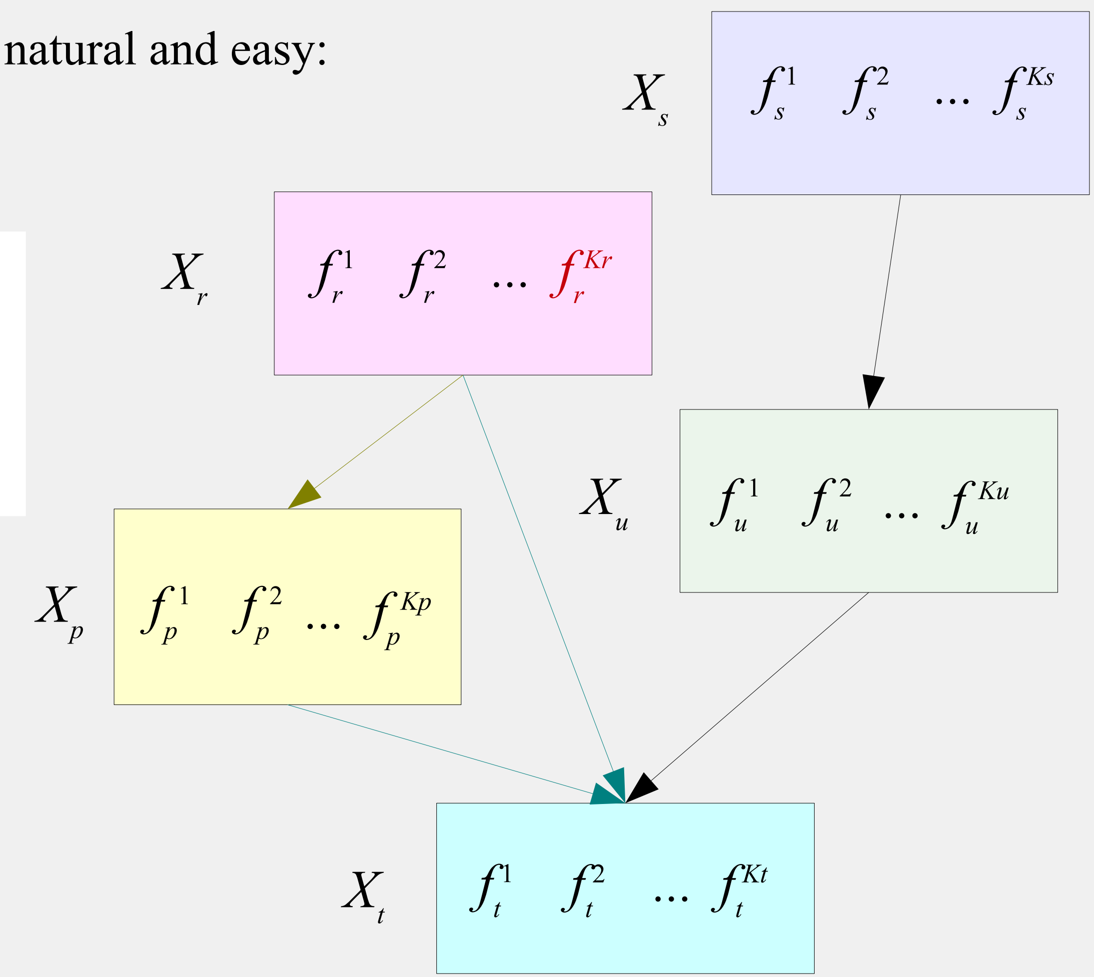
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:

The last component in each theme best complements *all* other components →

- measure the gain it brings;

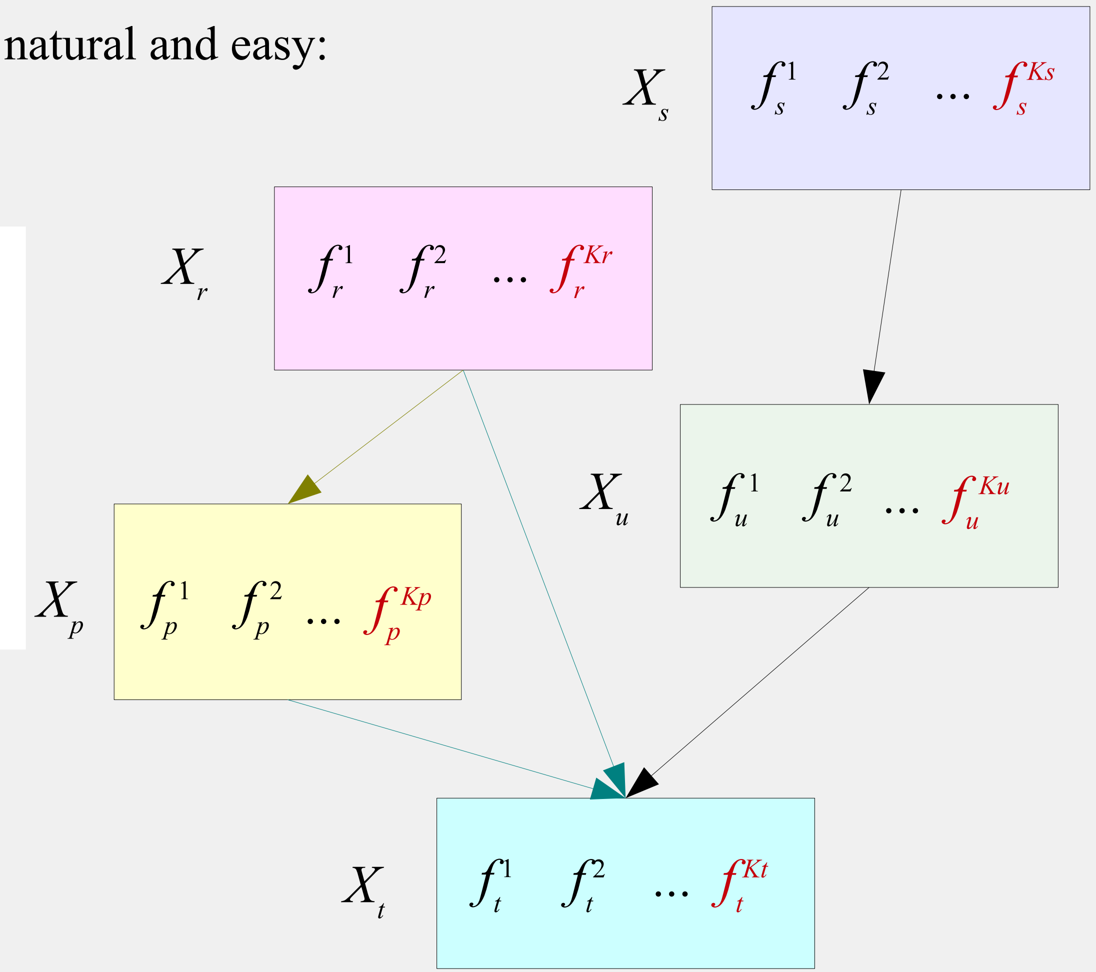


How THEME works

5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:



The last component in each theme best complements *all* other components →

- measure the gain it brings;

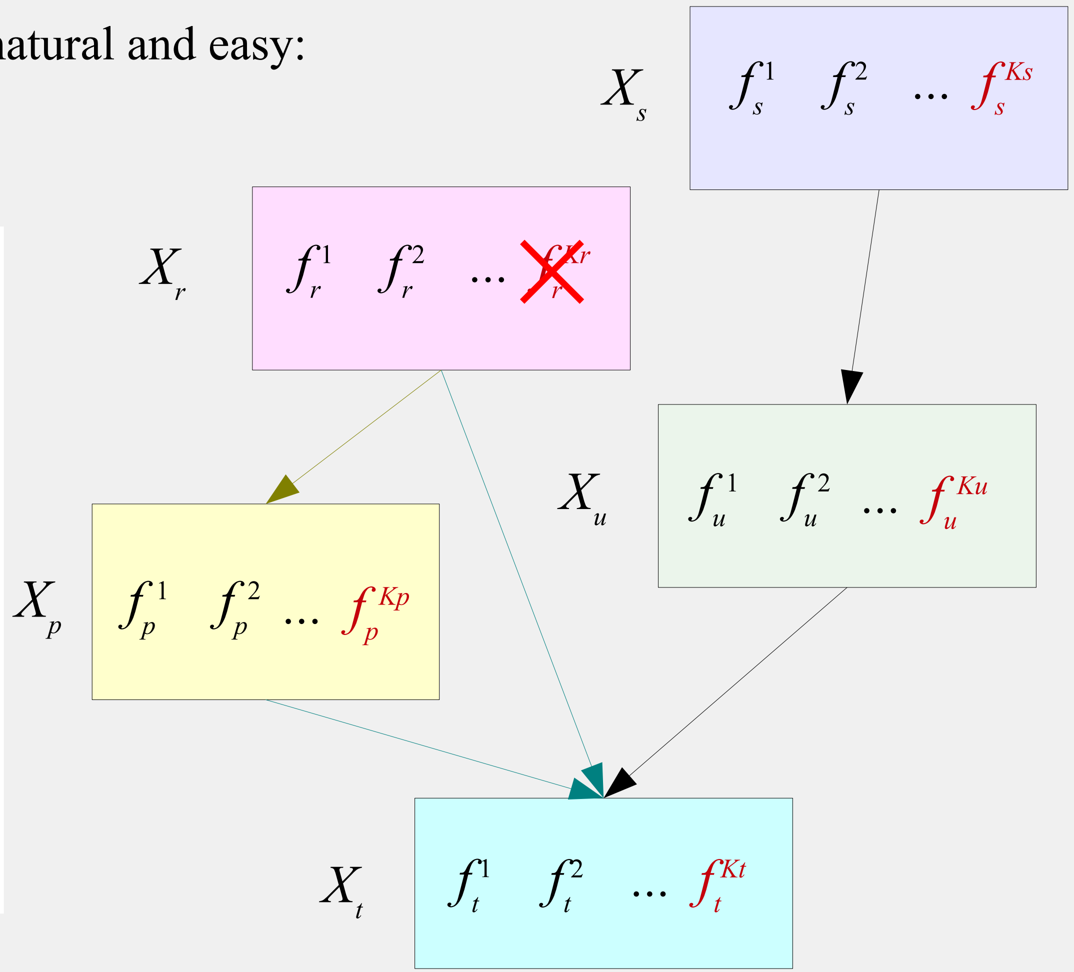
- compare this gain to that of all last components;

How THEME works

5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:



The last component in each theme best complements *all* other components →

- measure the gain it brings;
- compare this gain to that of all last components;
- eliminate the component bringing the smallest gain;
- re-estimate the model, etc.

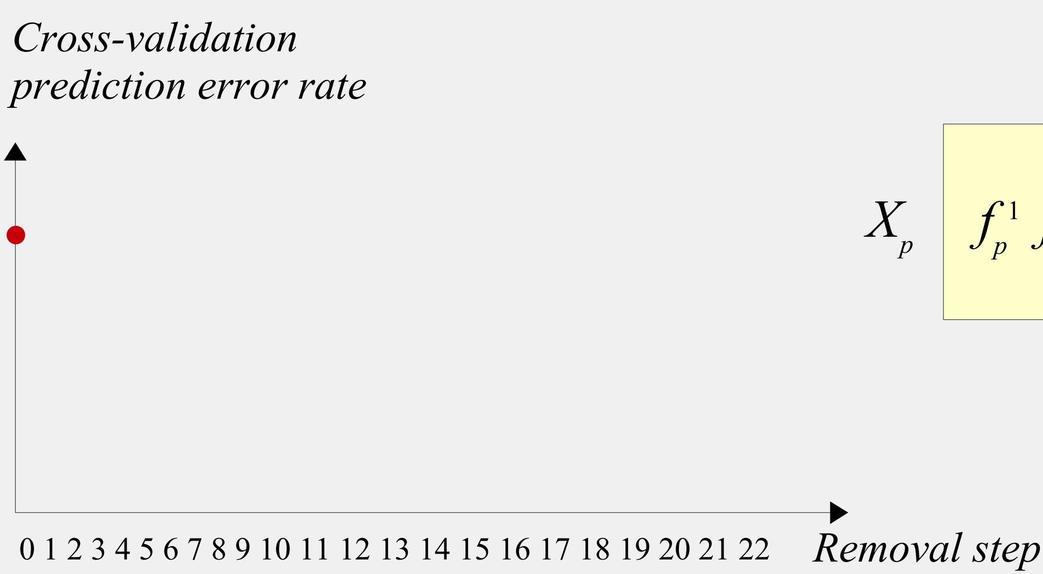
How THEME works

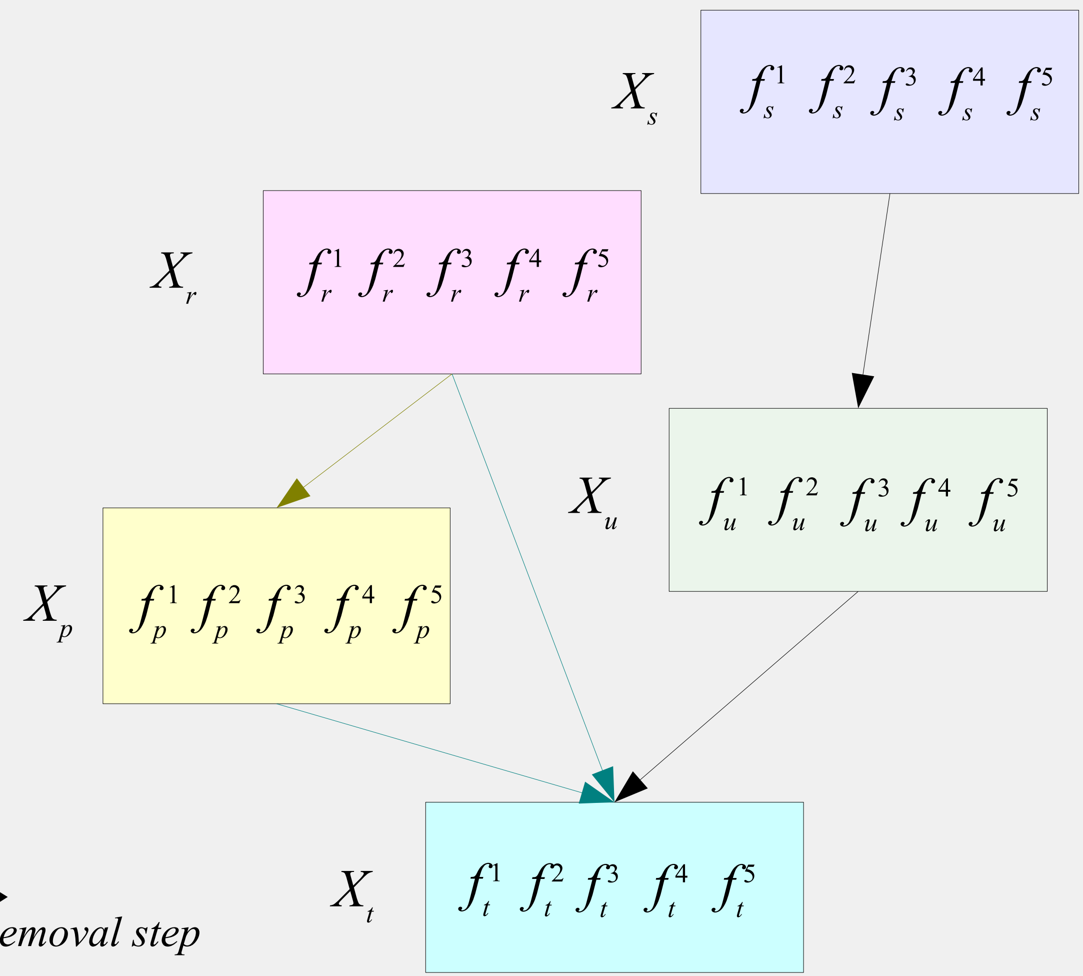
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:

Example:



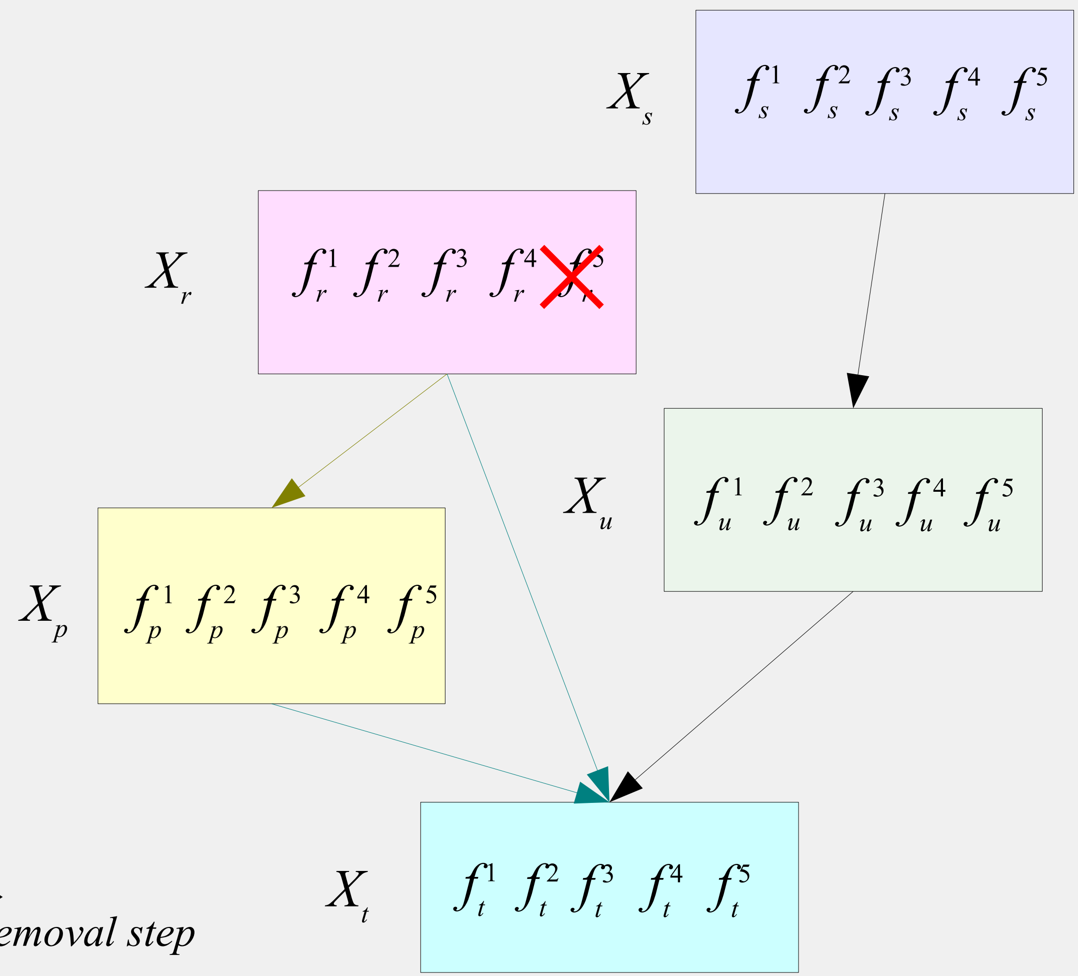


How THEME works

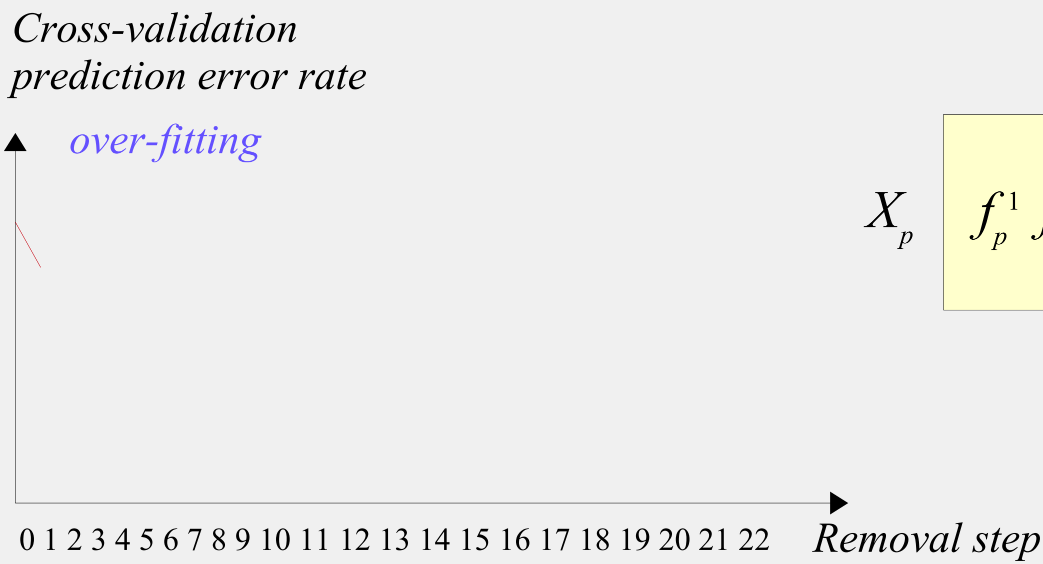
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:



Example:

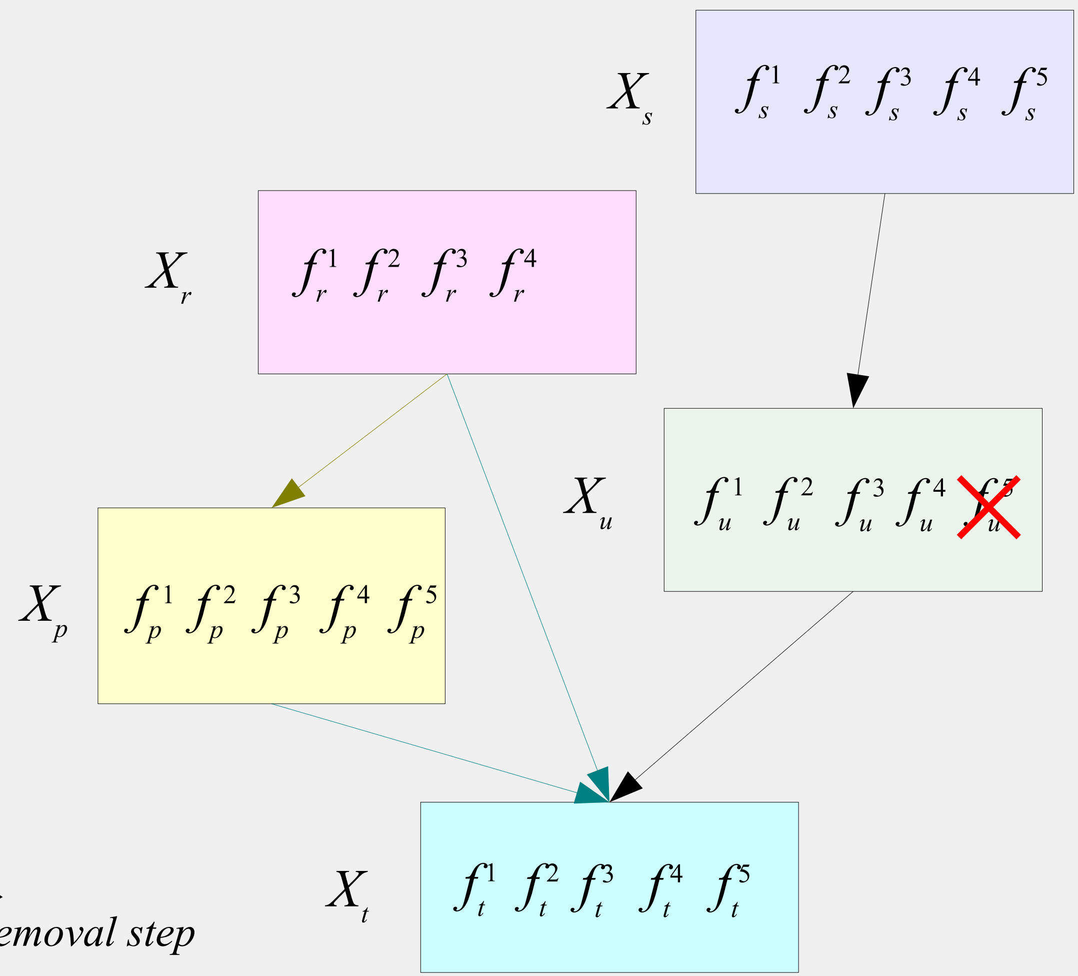


How THEME works

5. Backward component selection

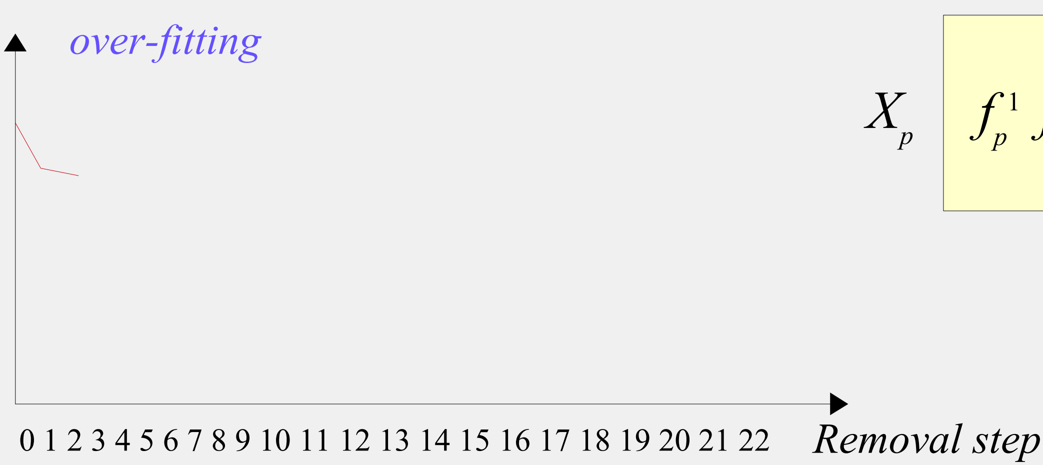
- **How to select the suitable number of components in each group?**

Local nesting makes ***backward selection*** natural and easy:



Example:

Cross-validation prediction error rate

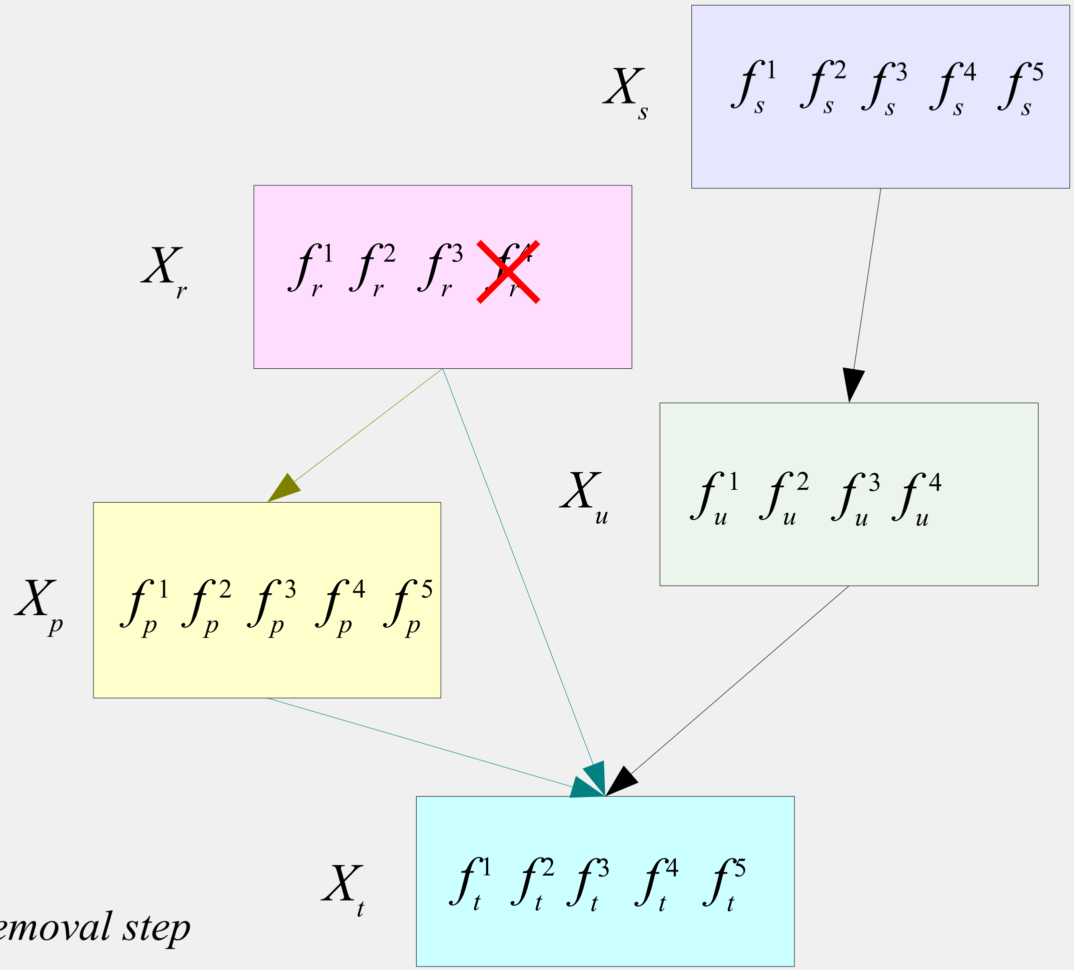


How THEME works

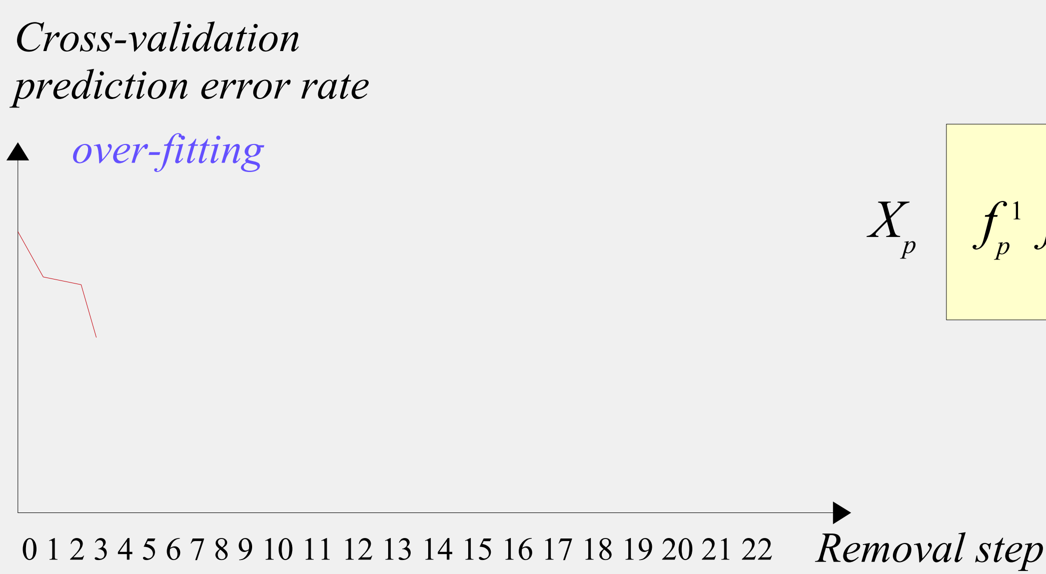
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:



Example:



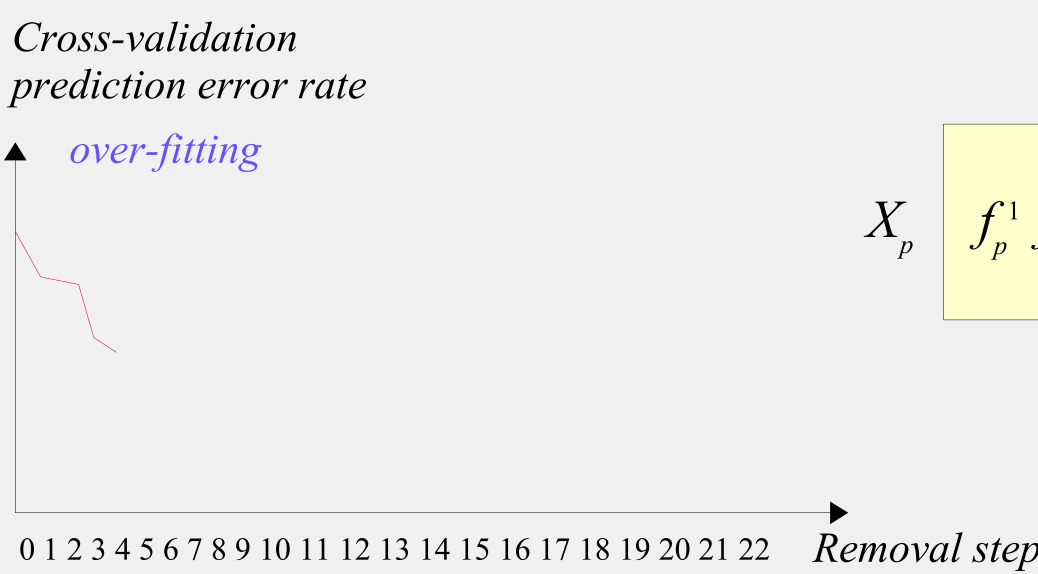
How THEME works

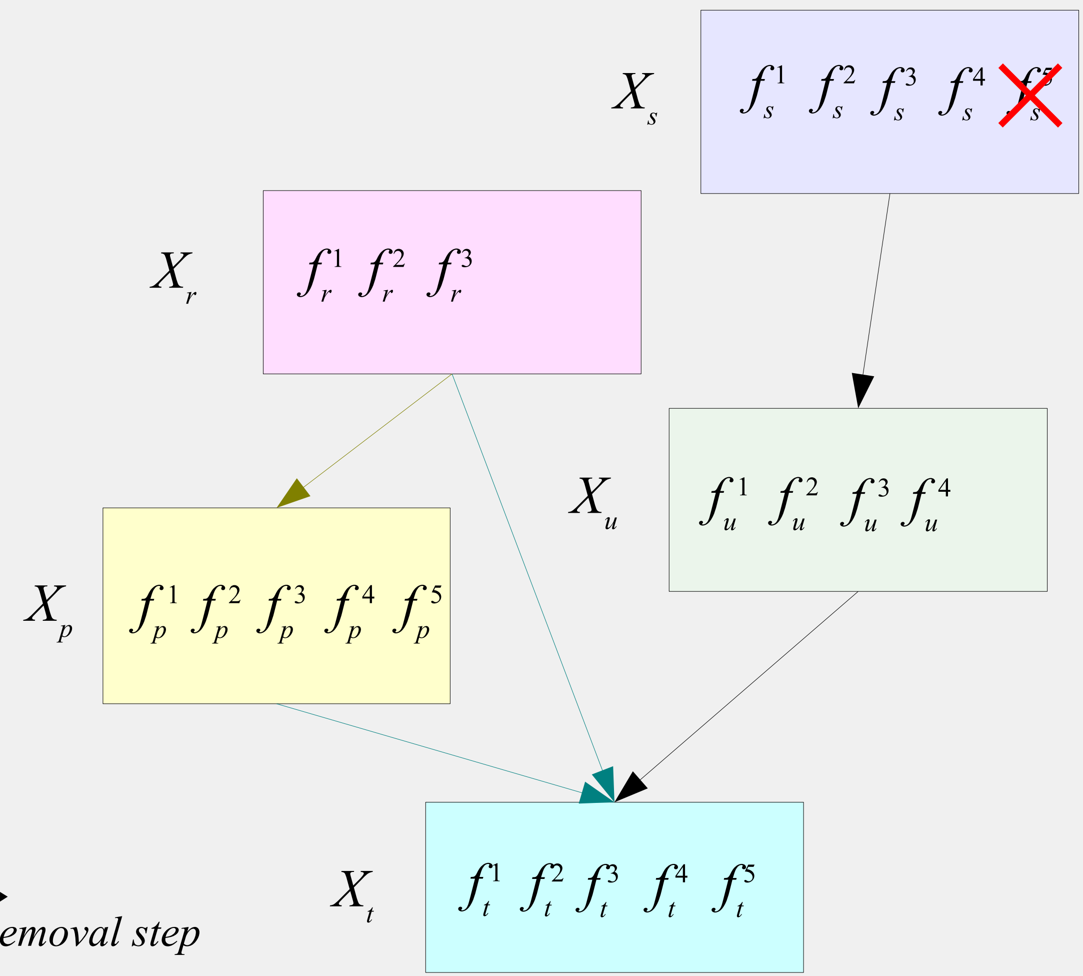
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:

Example:



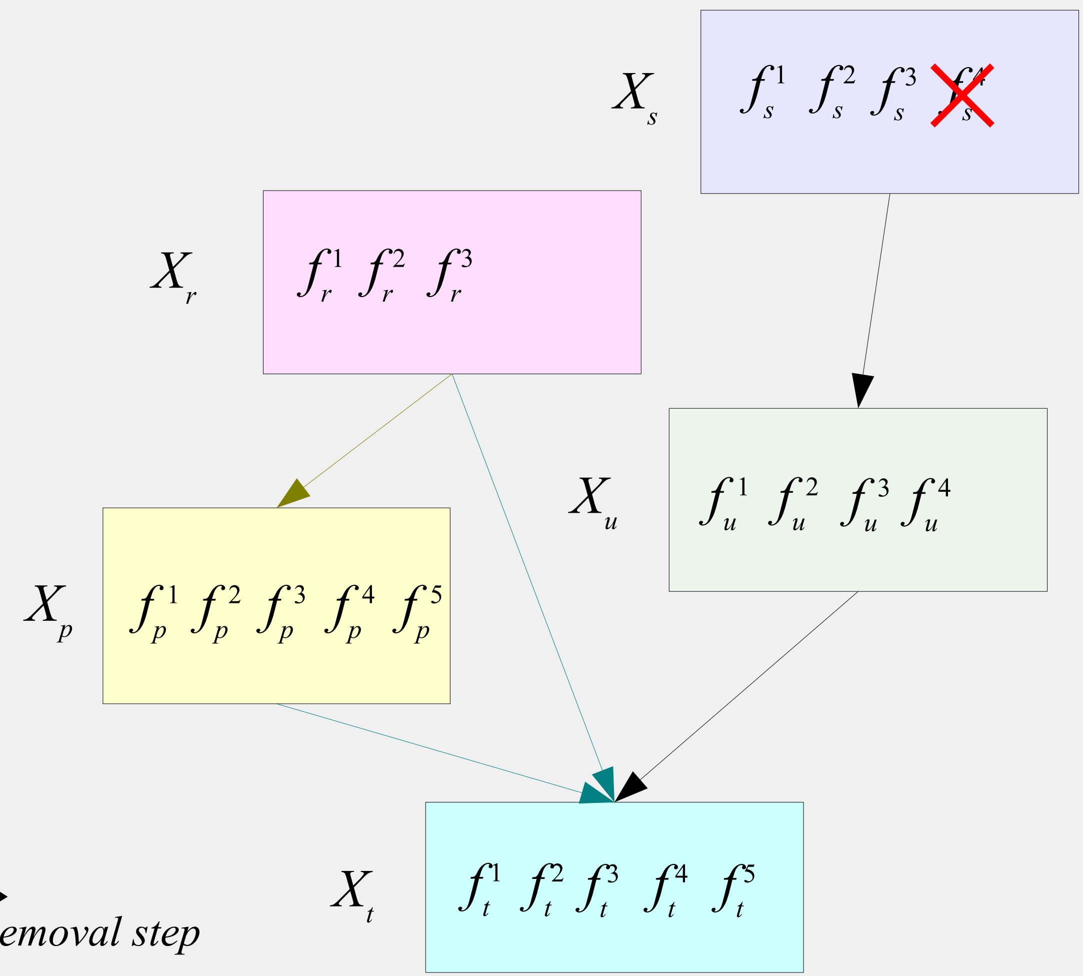


How THEME works

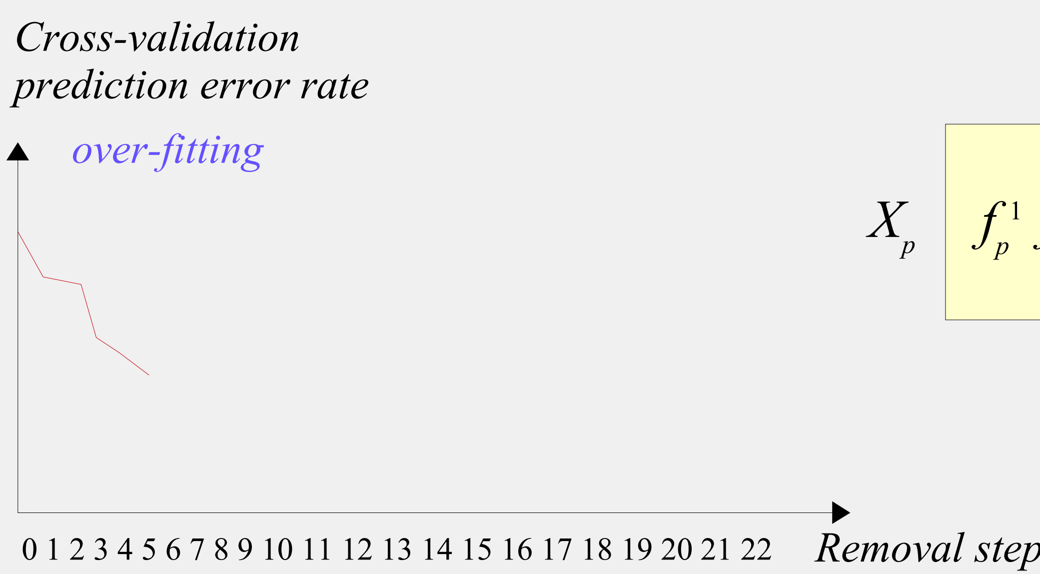
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:



Example:

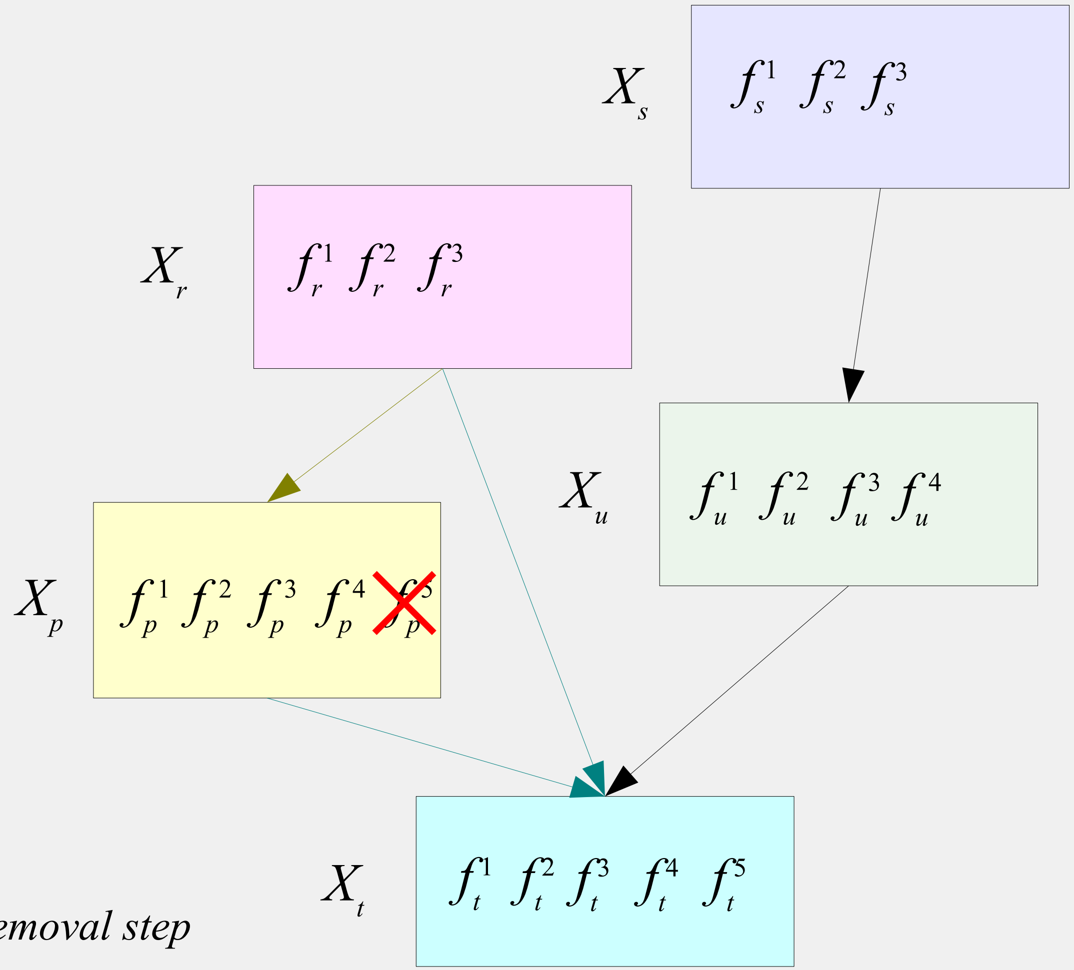


How THEME works

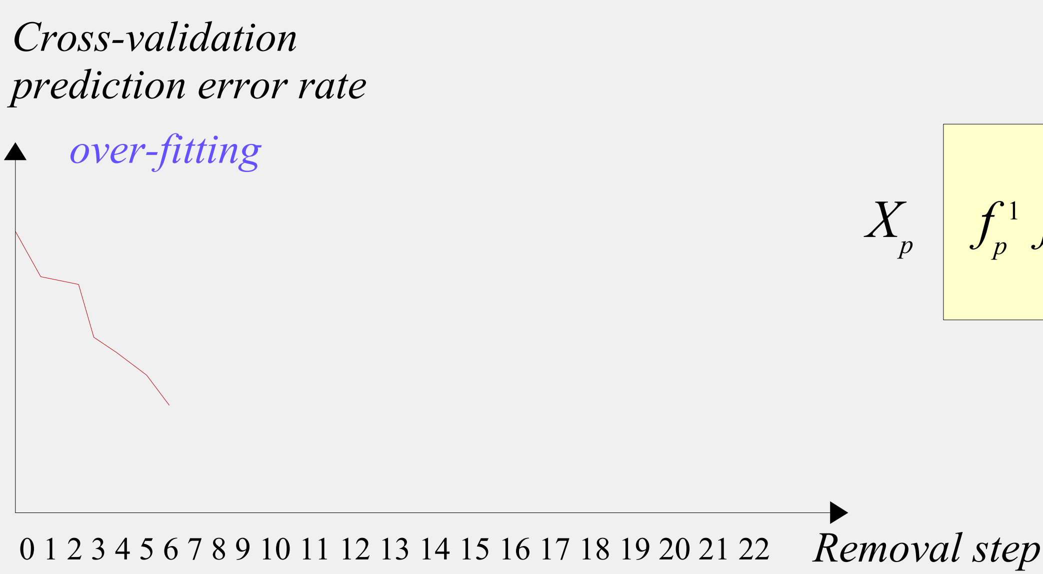
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:



Example:

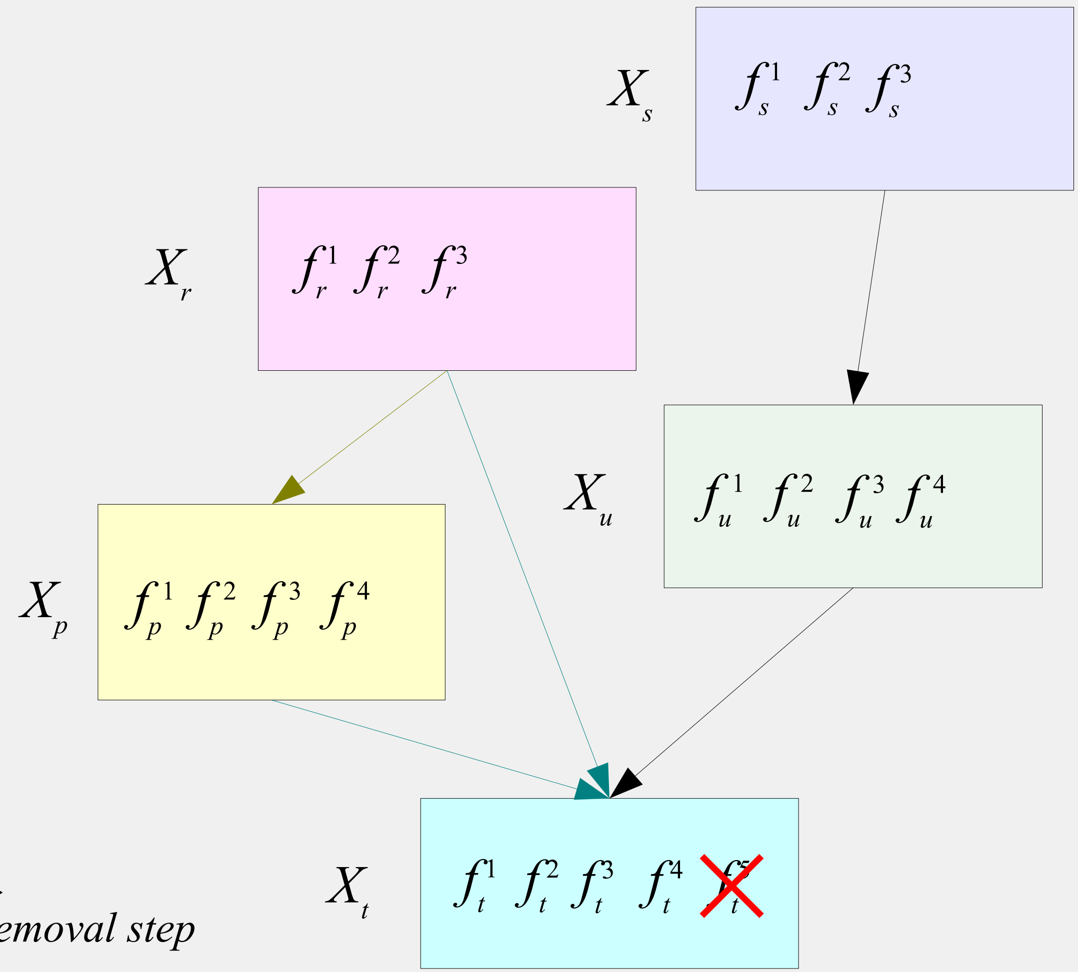


How THEME works

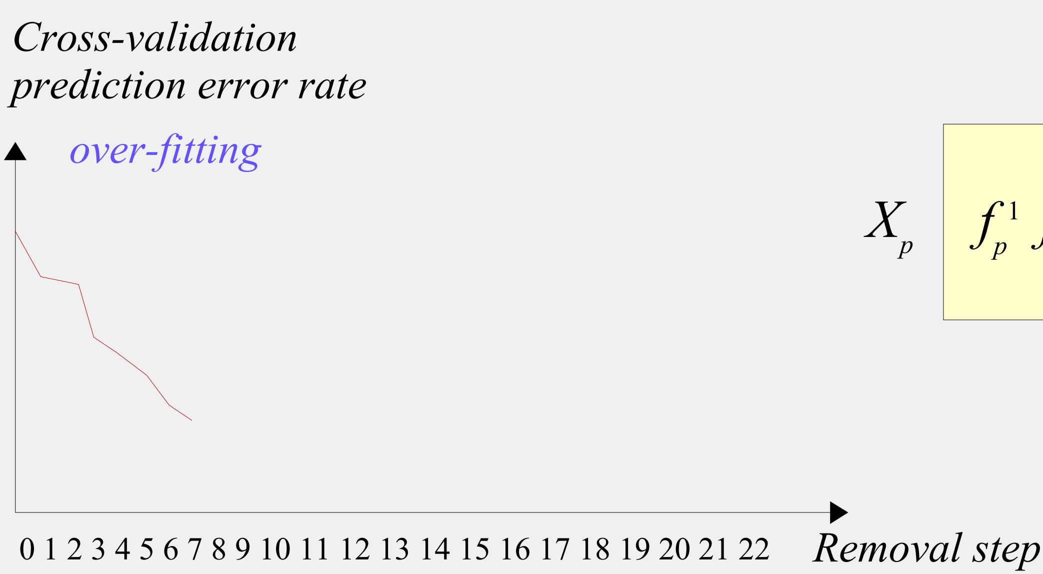
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:



Example:



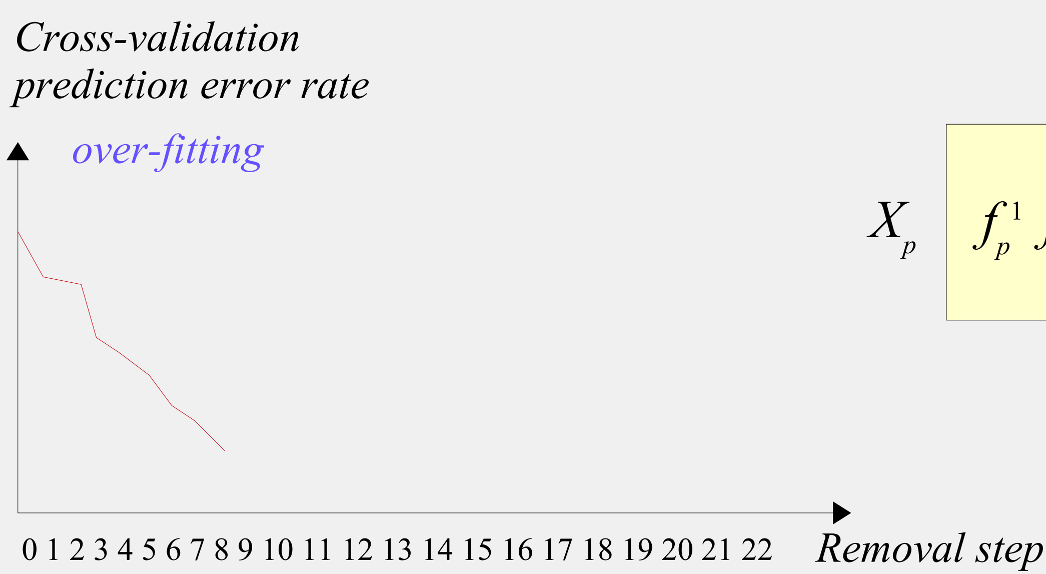
How THEME works

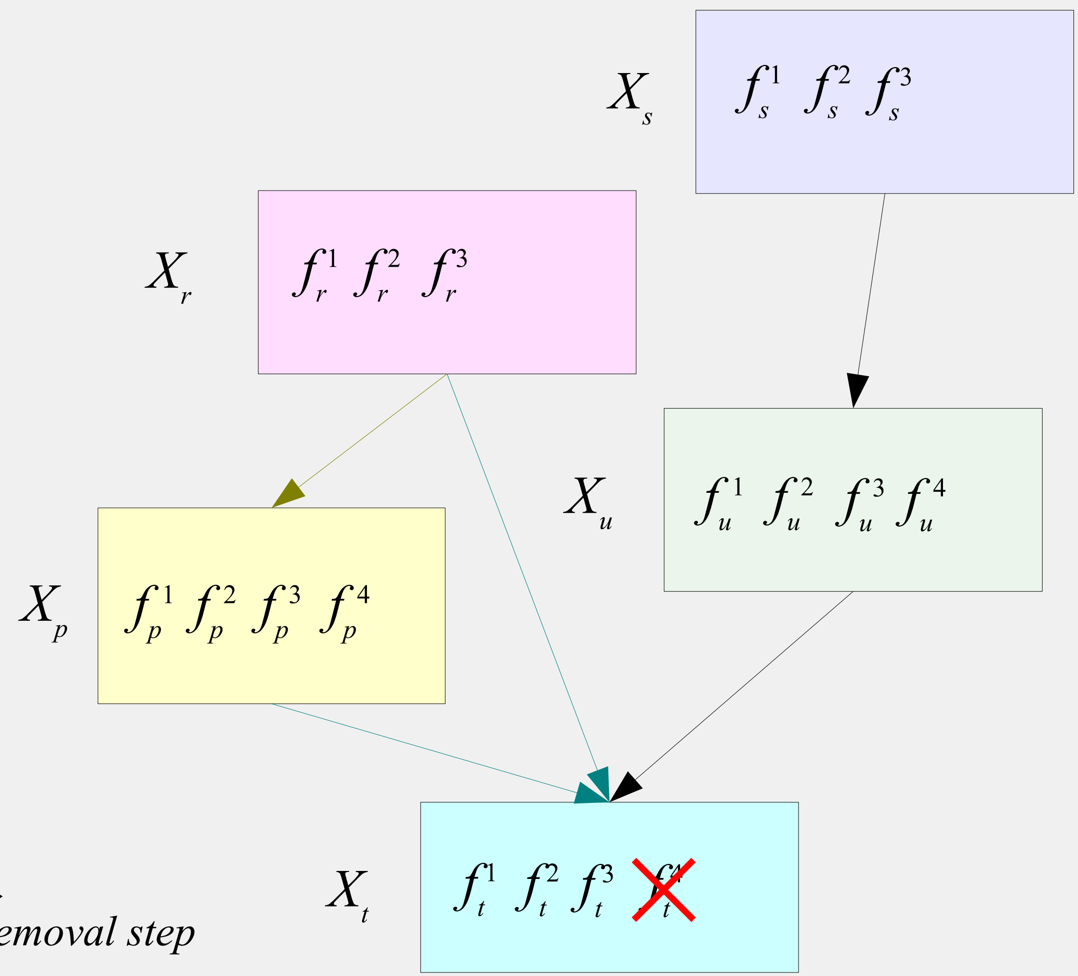
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:

Example:





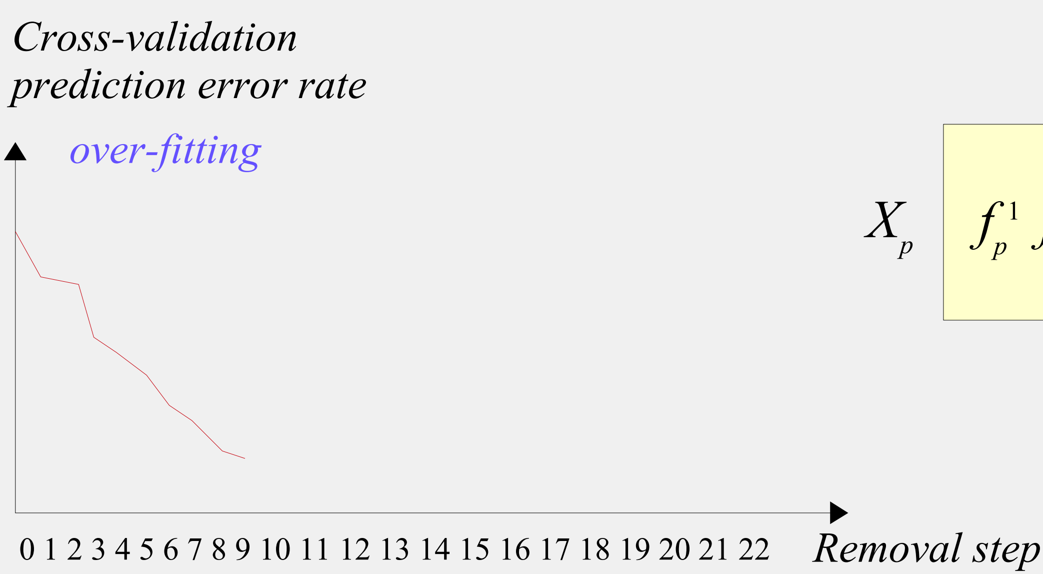
How THEME works

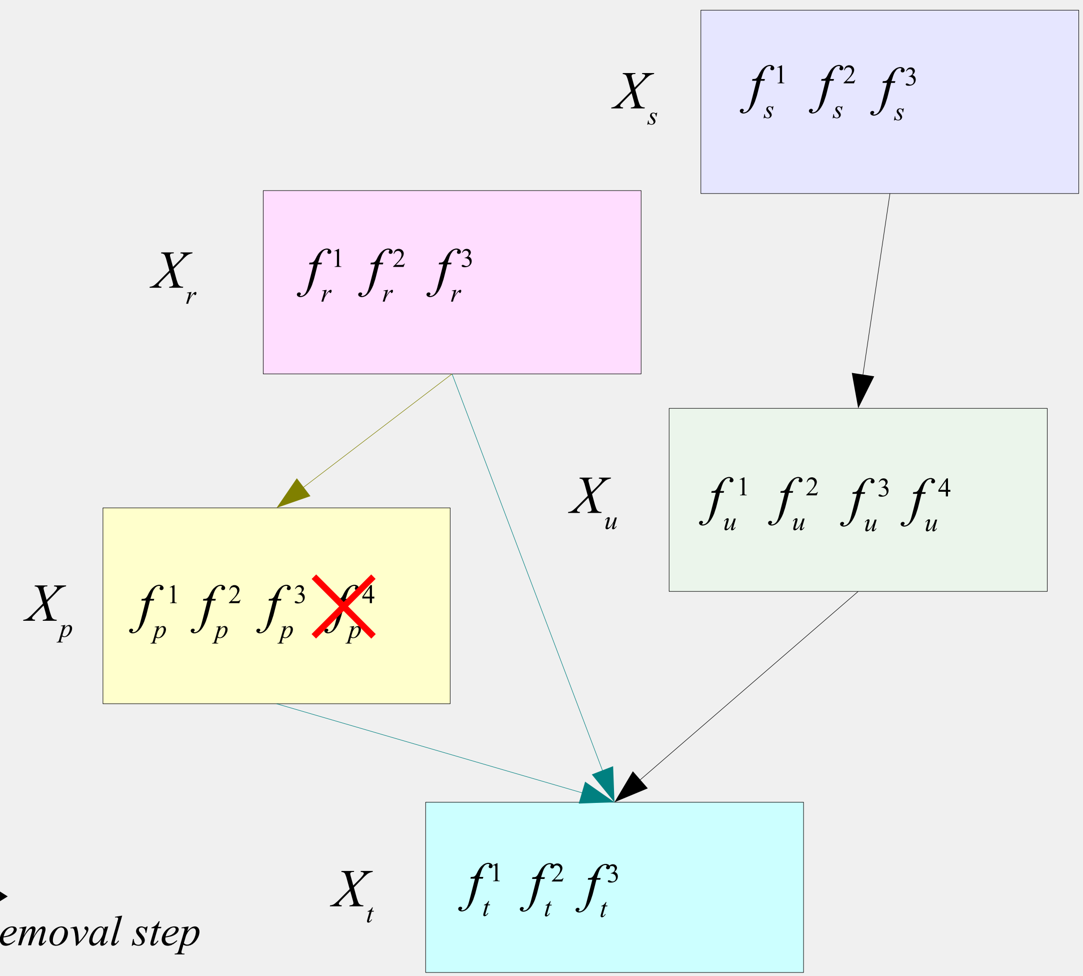
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:

Example:





How THEME works

5. Backward component selection

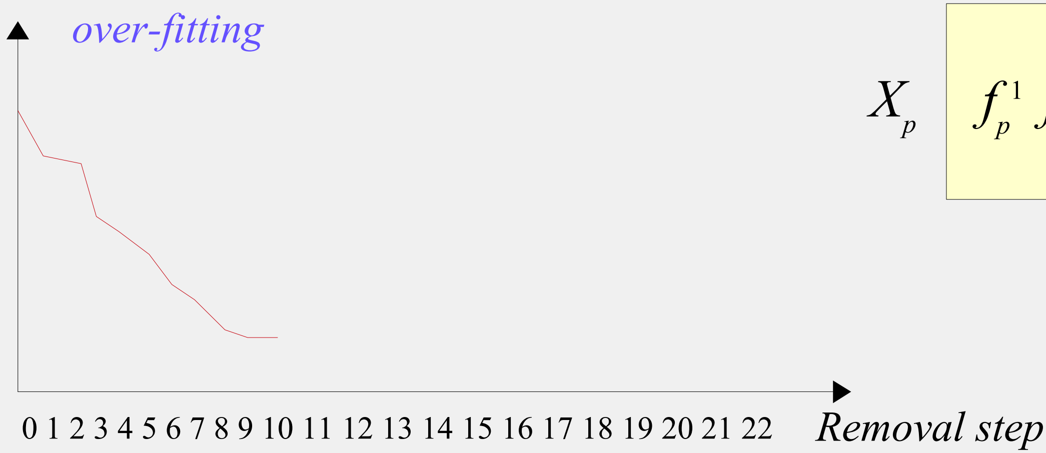
- **How to select the suitable number of components in each group?**

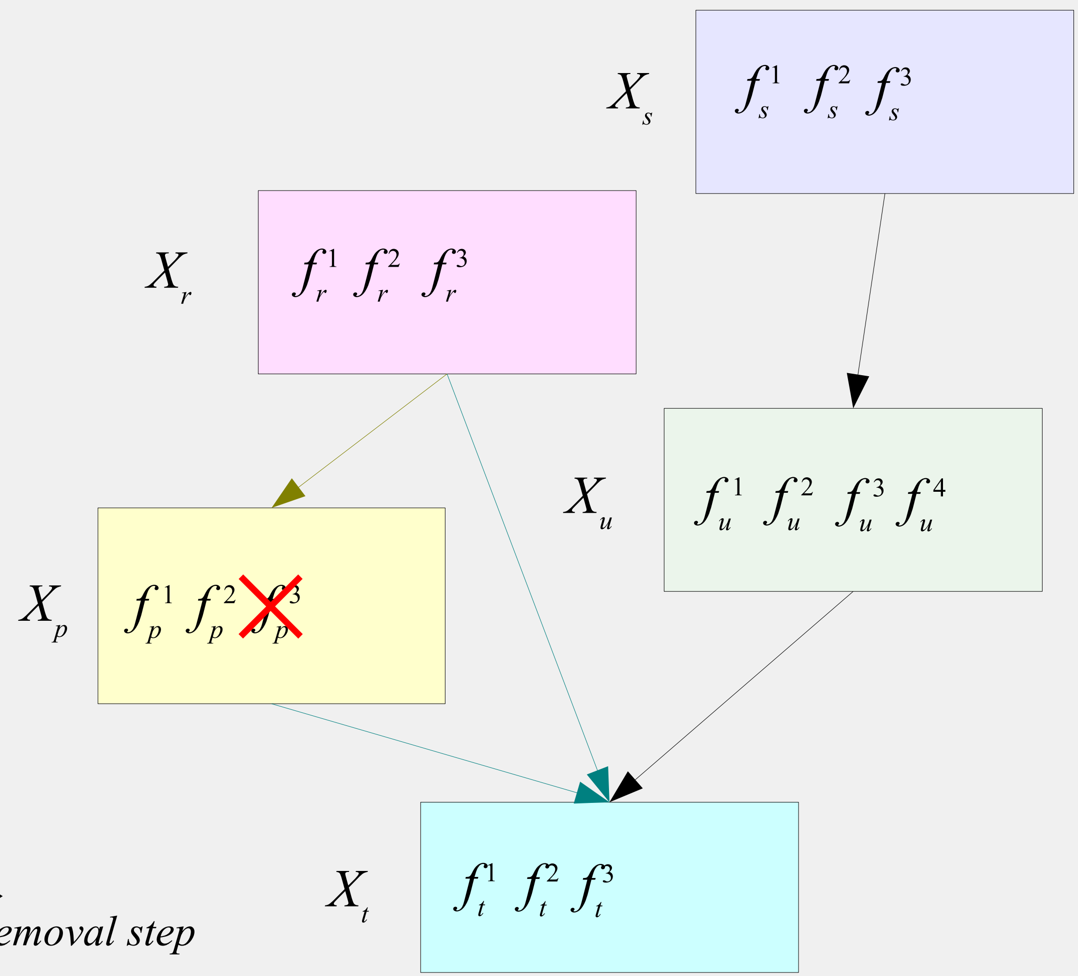
Local nesting makes ***backward selection*** natural and easy:

Example:

Cross-validation prediction error rate

over-fitting





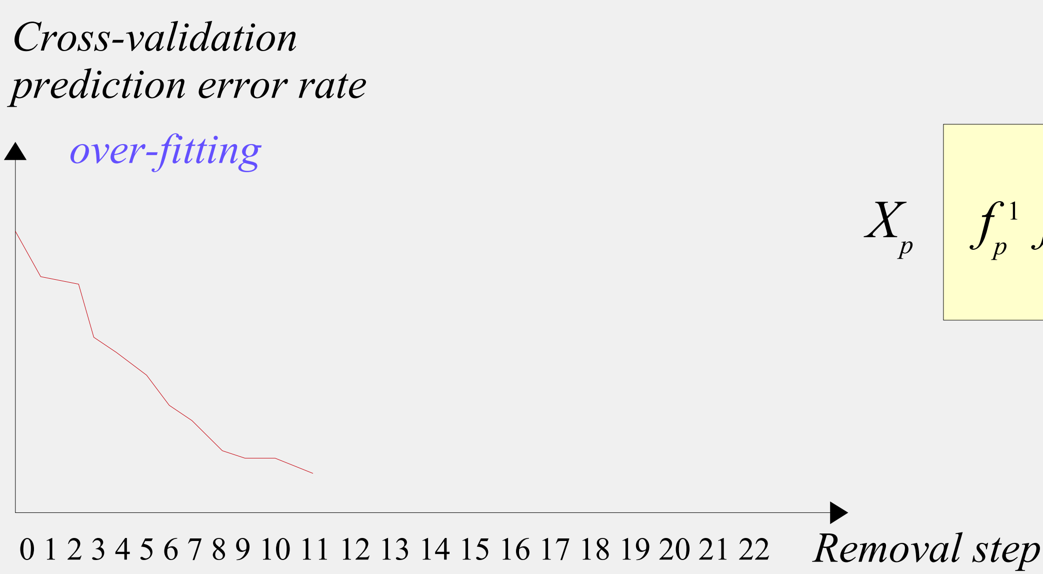
How THEME works

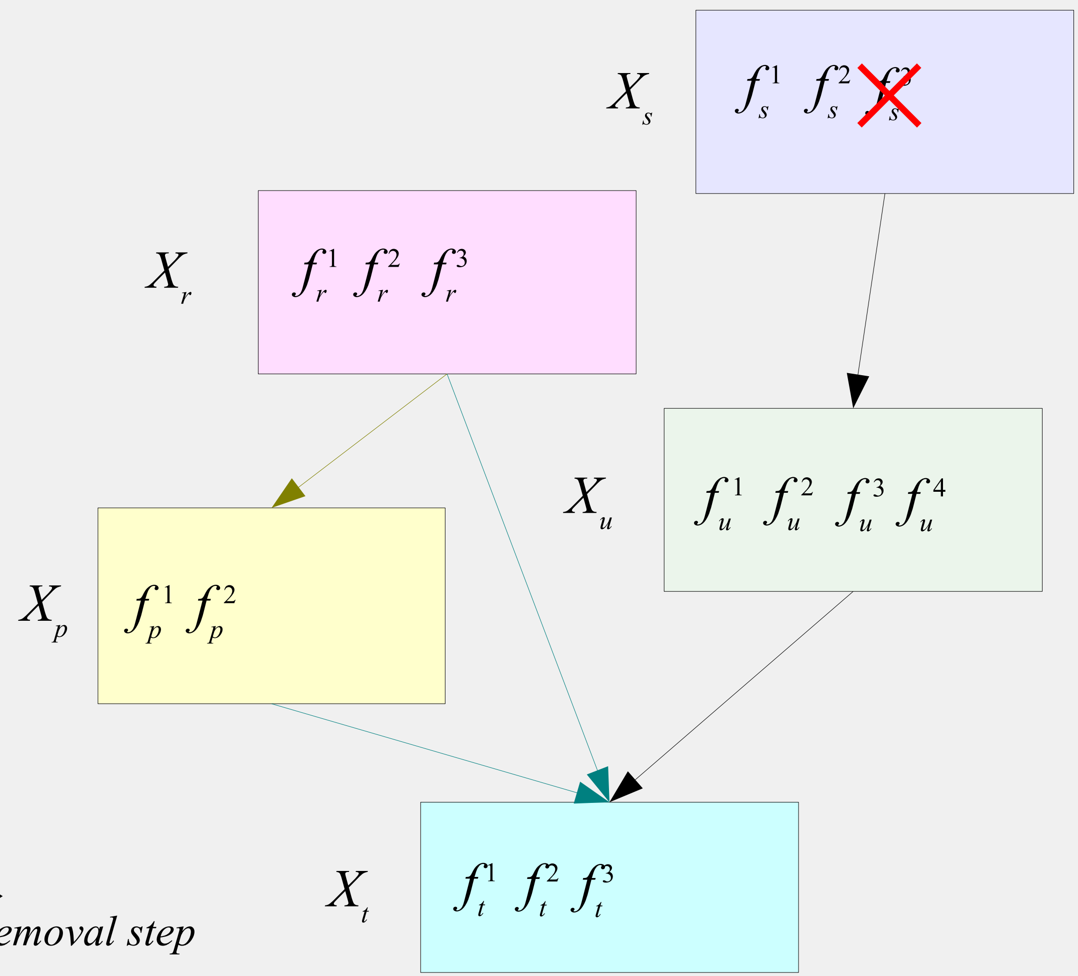
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:

Example:



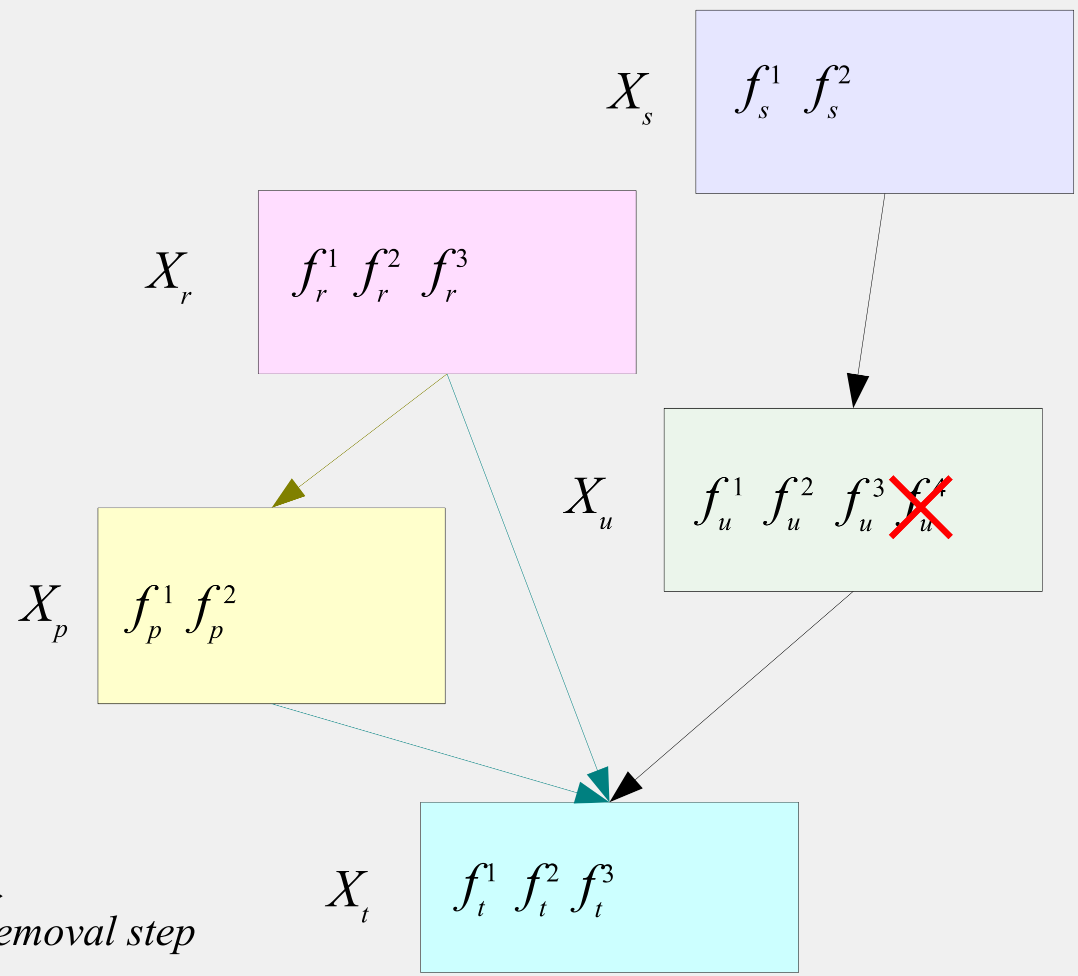


How THEME works

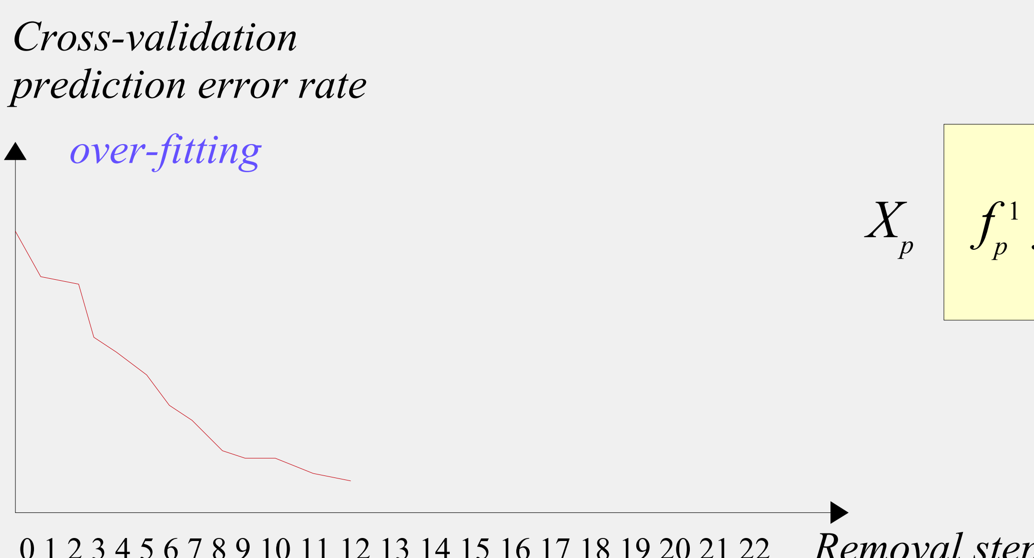
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:



Example:

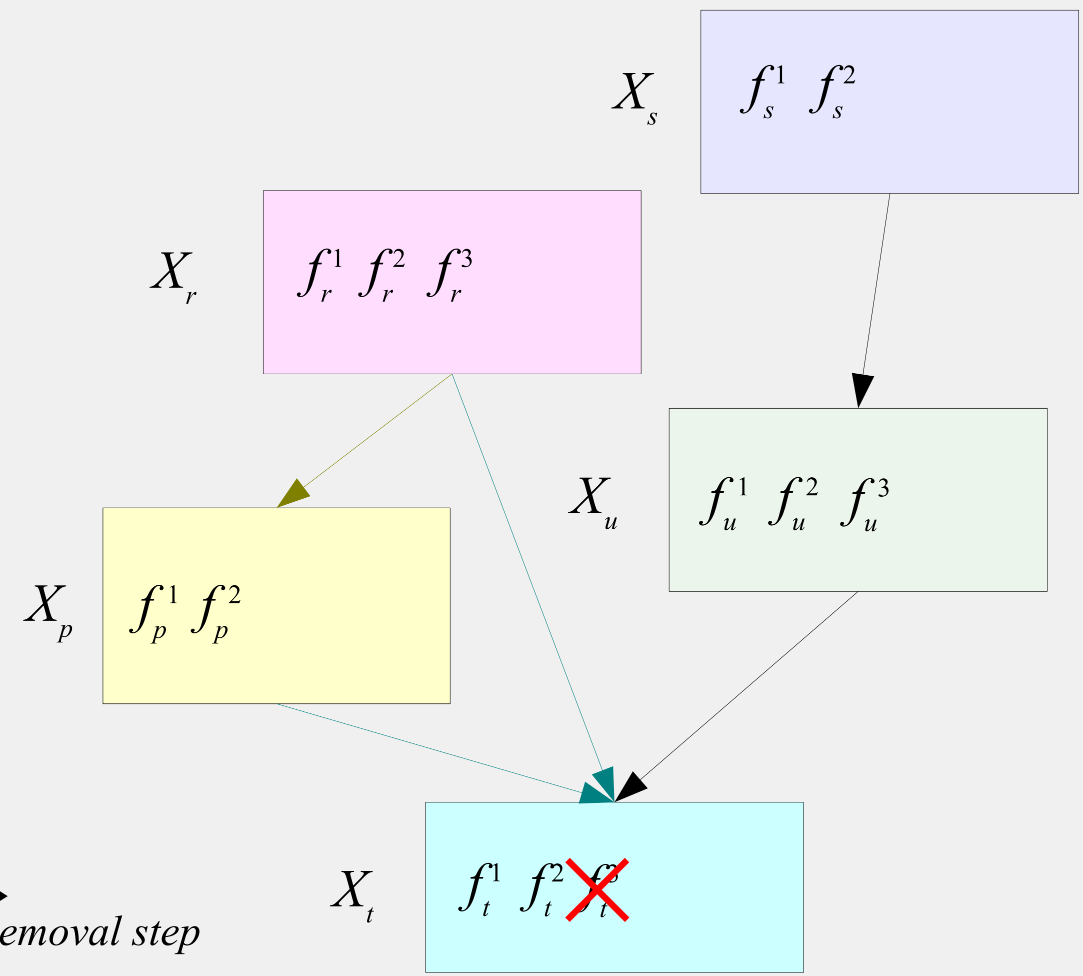


How THEME works

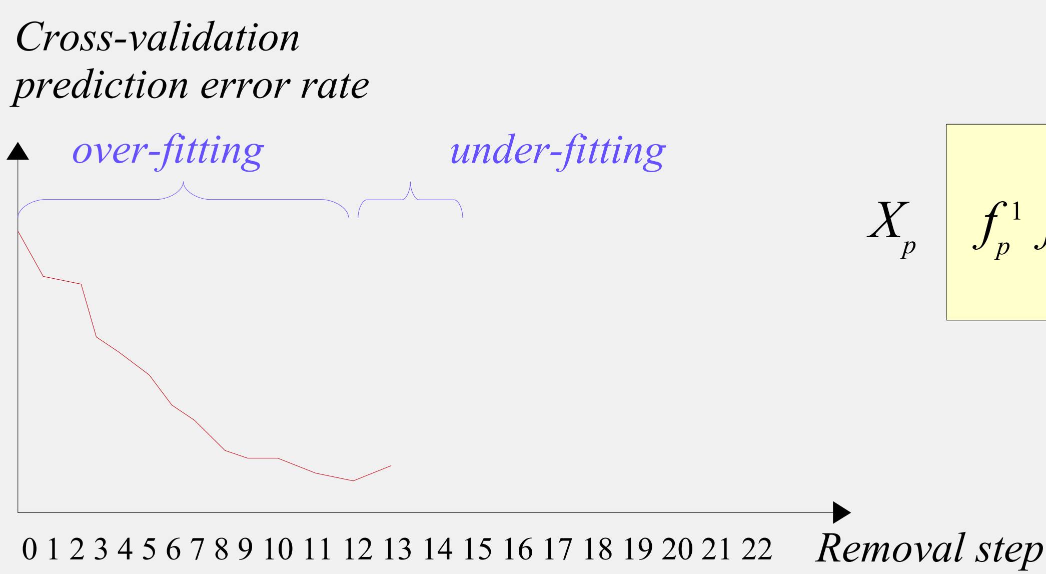
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:



Example:

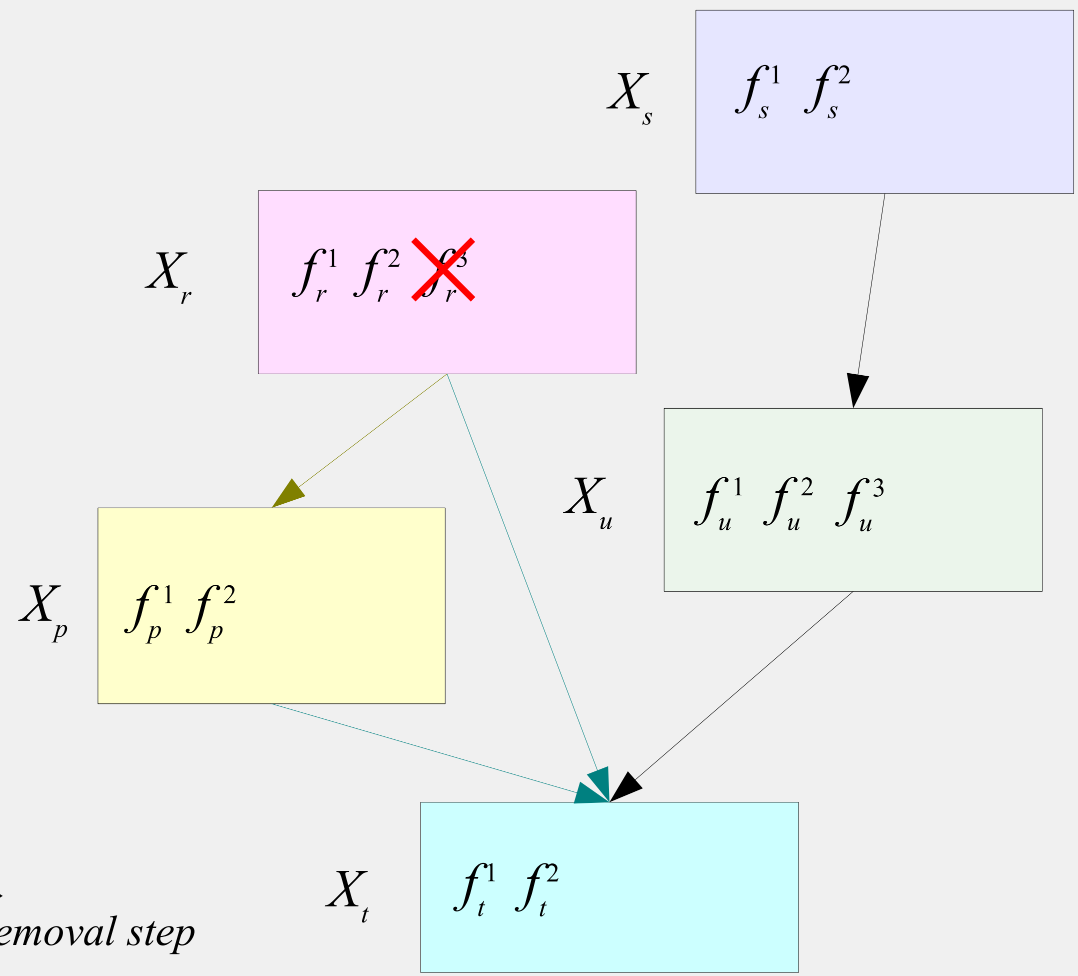


How THEME works

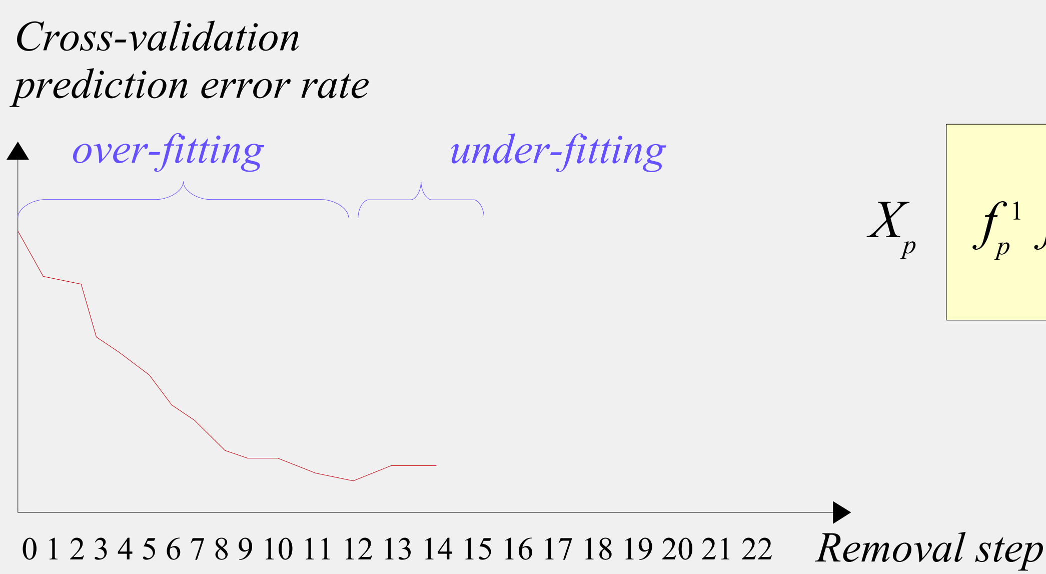
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:



Example:



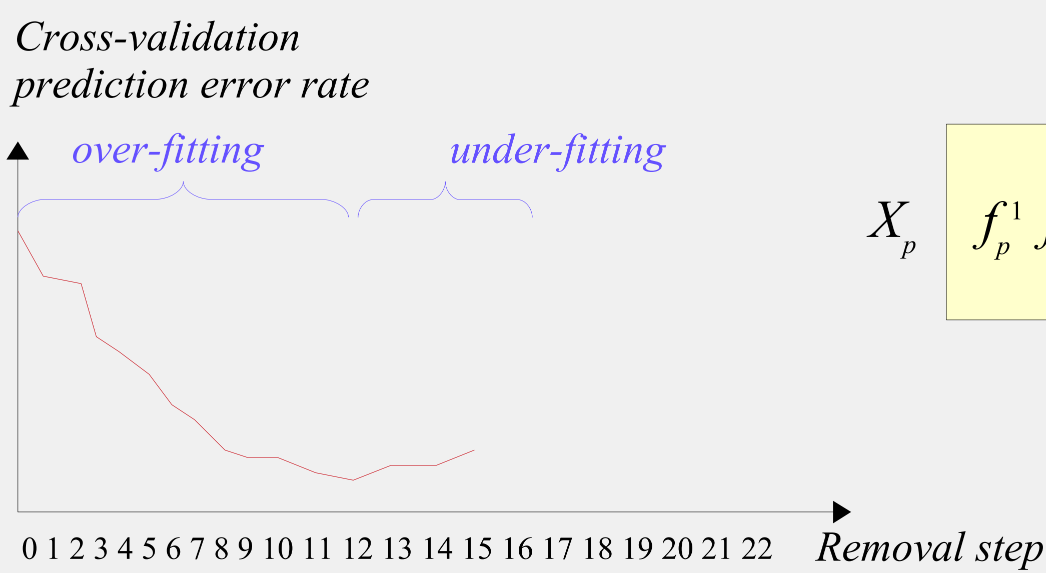
How THEME works

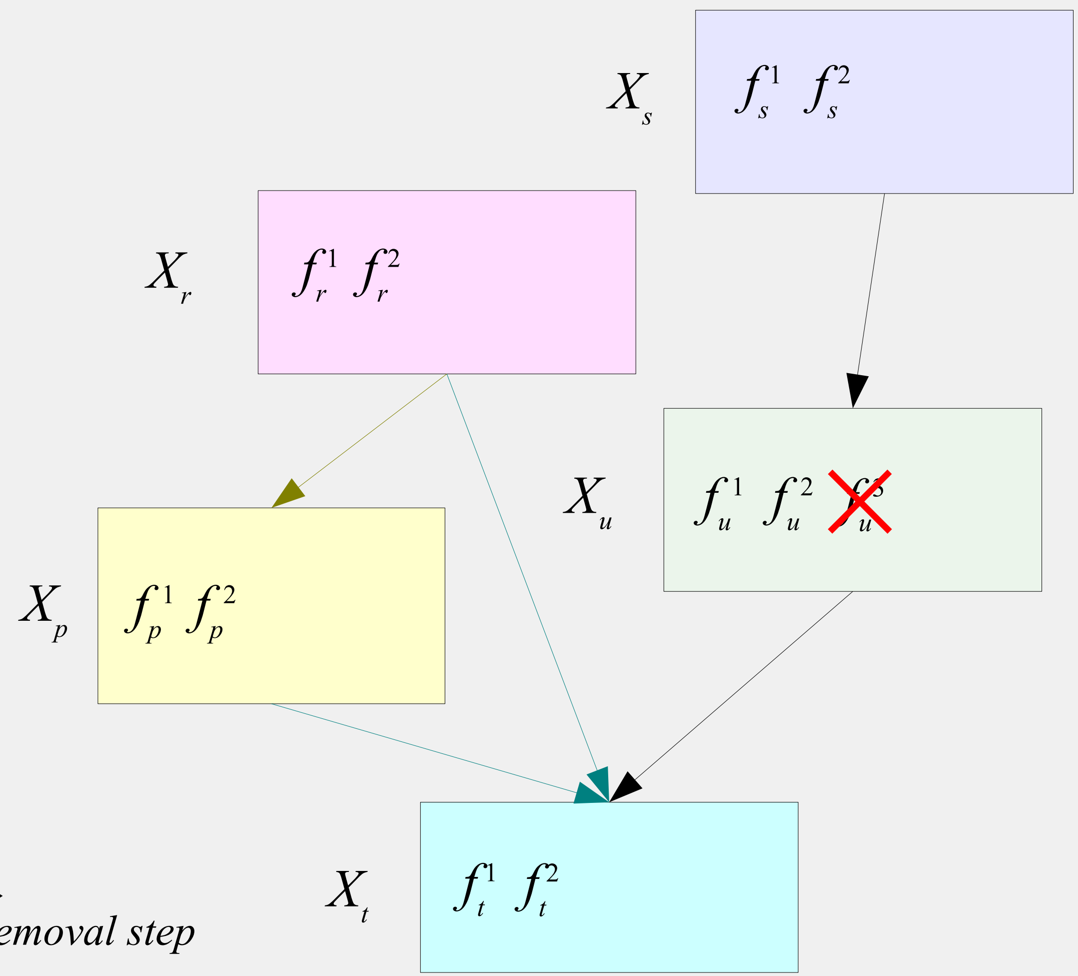
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:

Example:





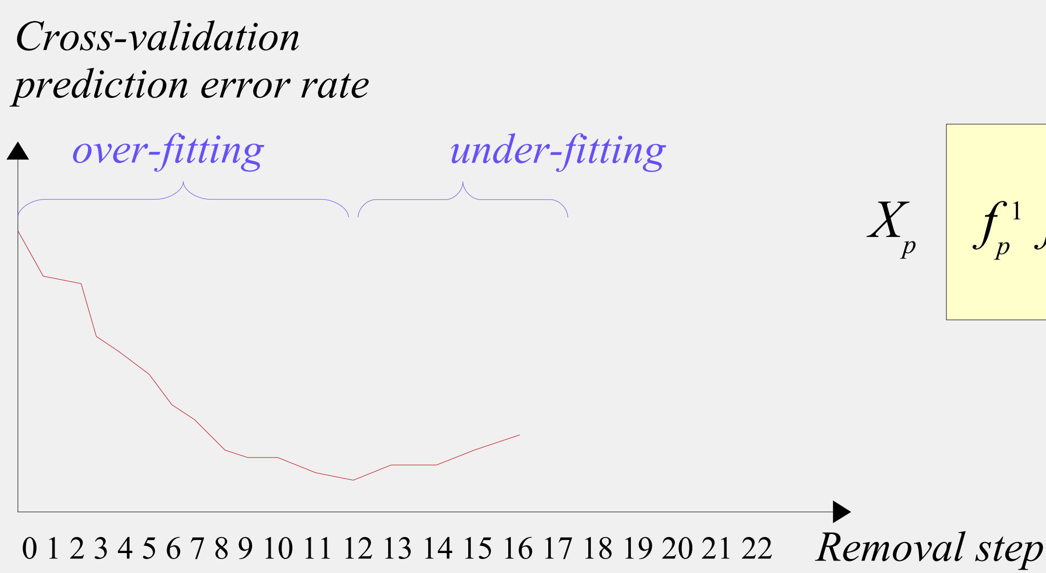
How THEME works

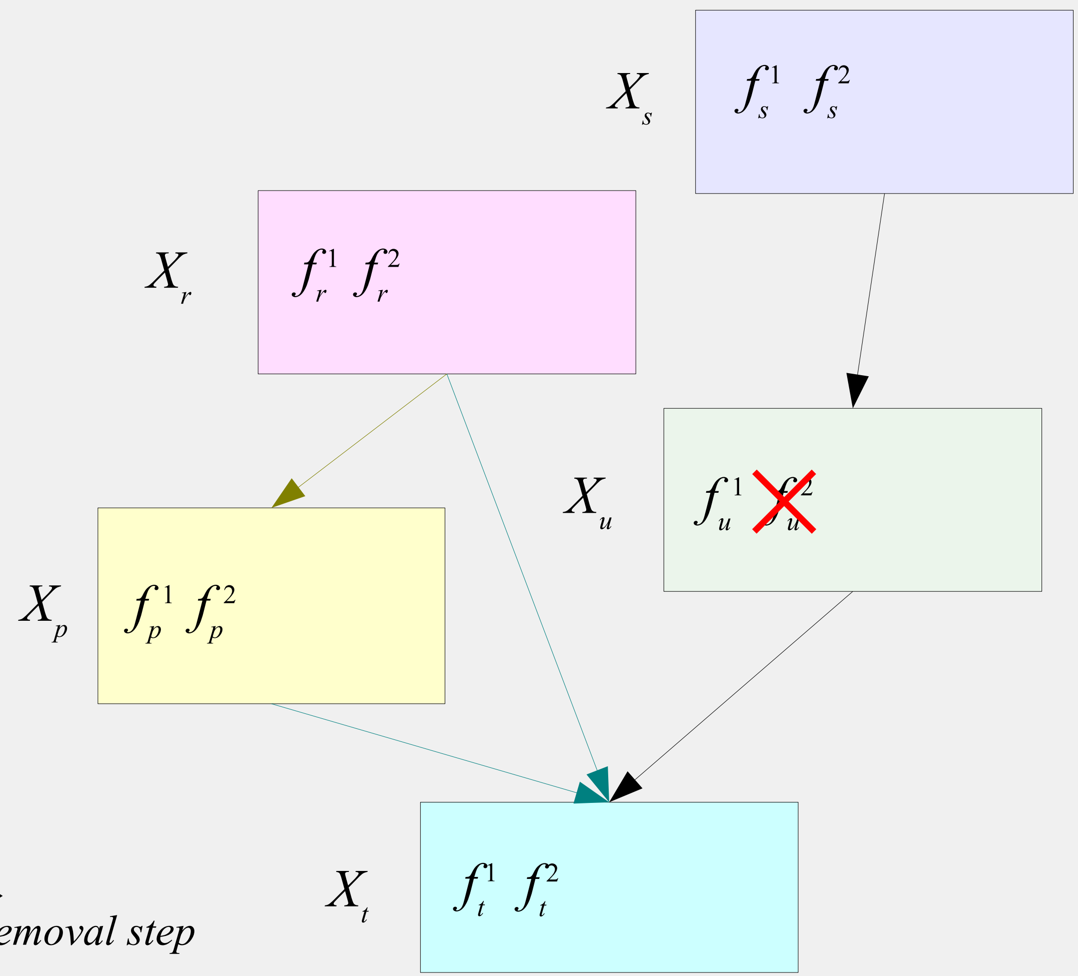
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:

Example:



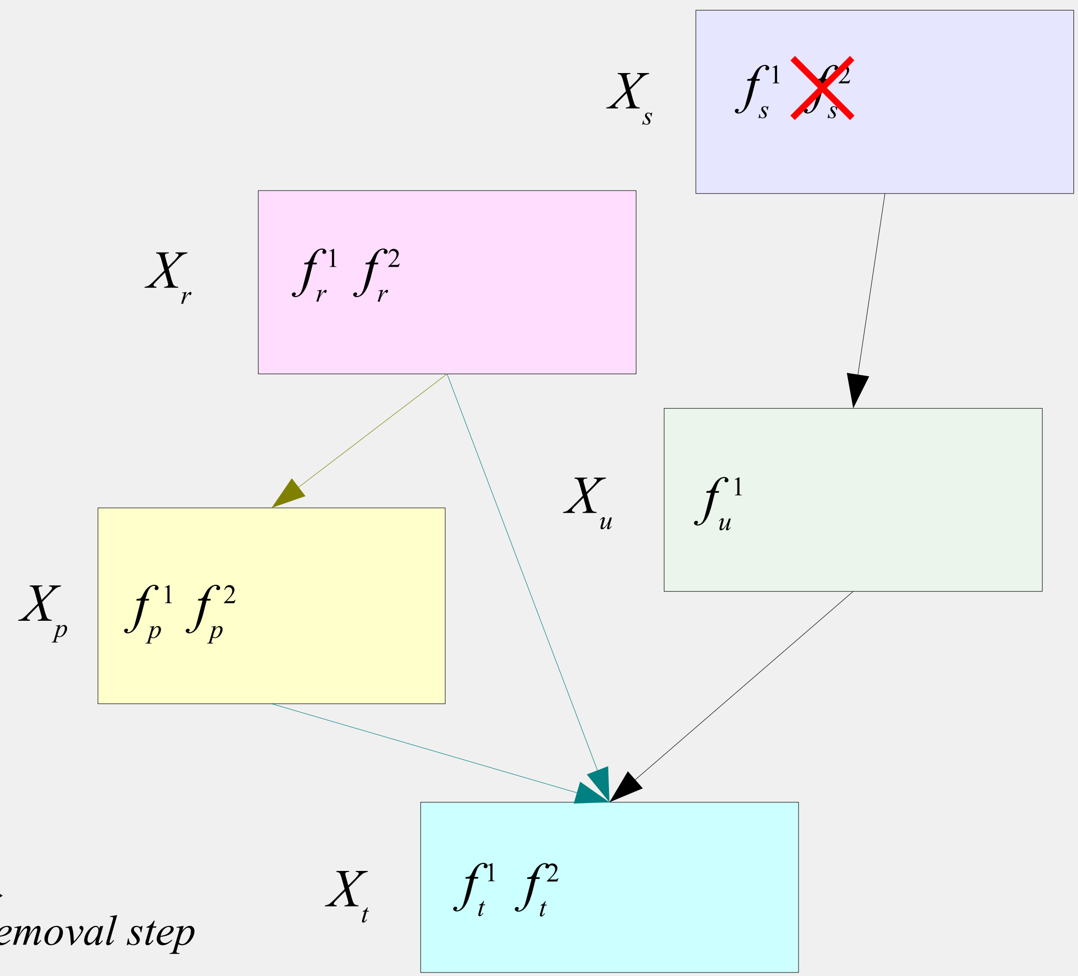


How THEME works

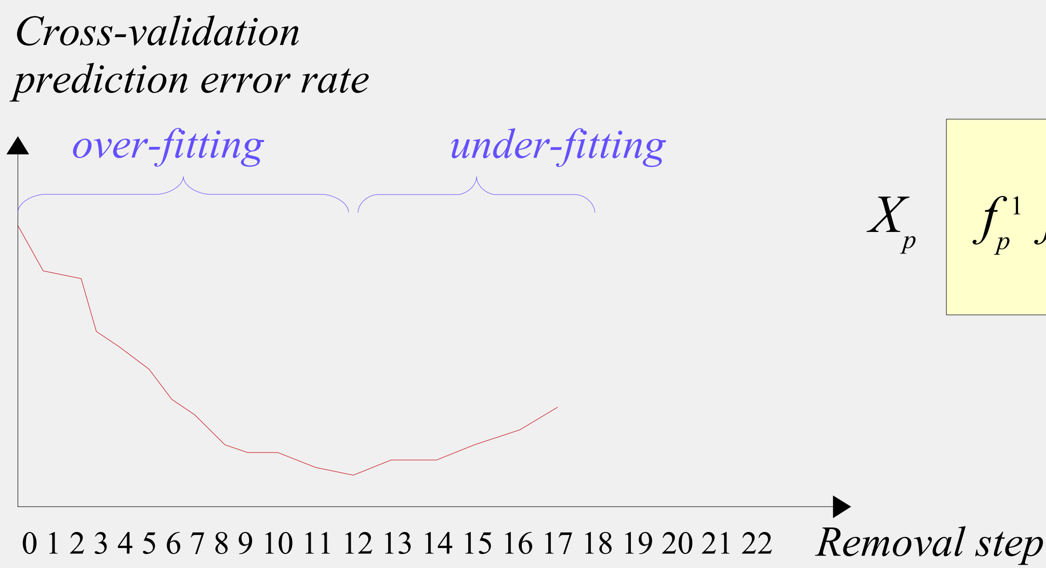
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:



Example:



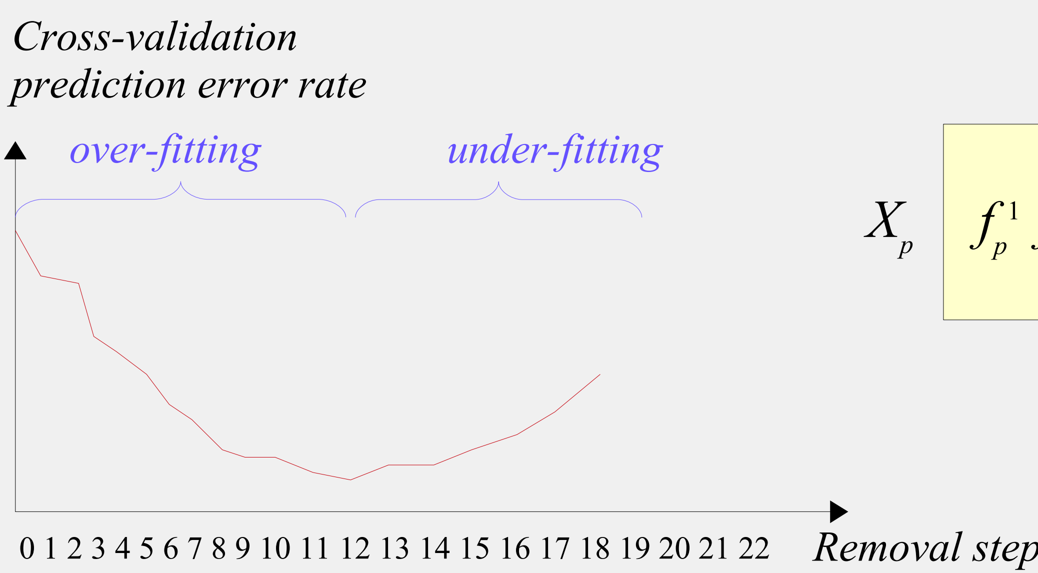
How THEME works

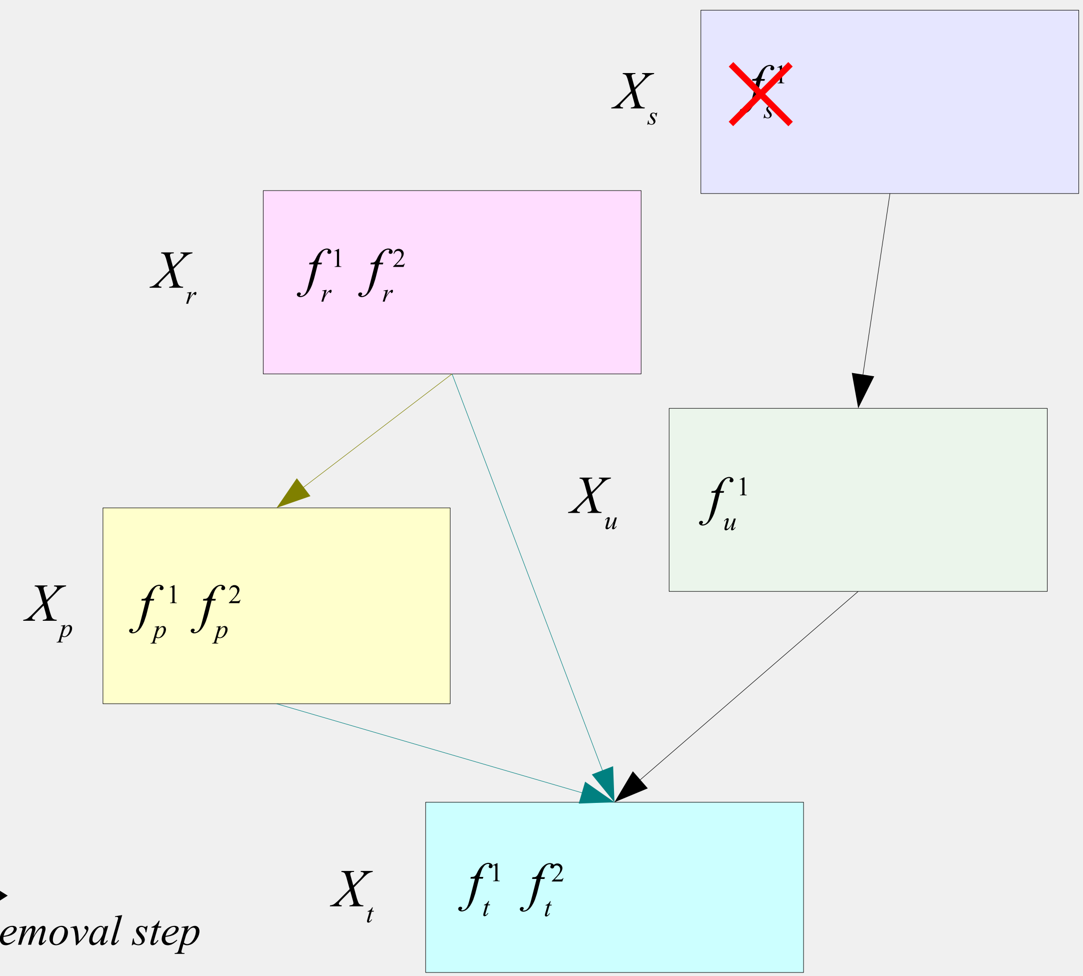
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:

Example:





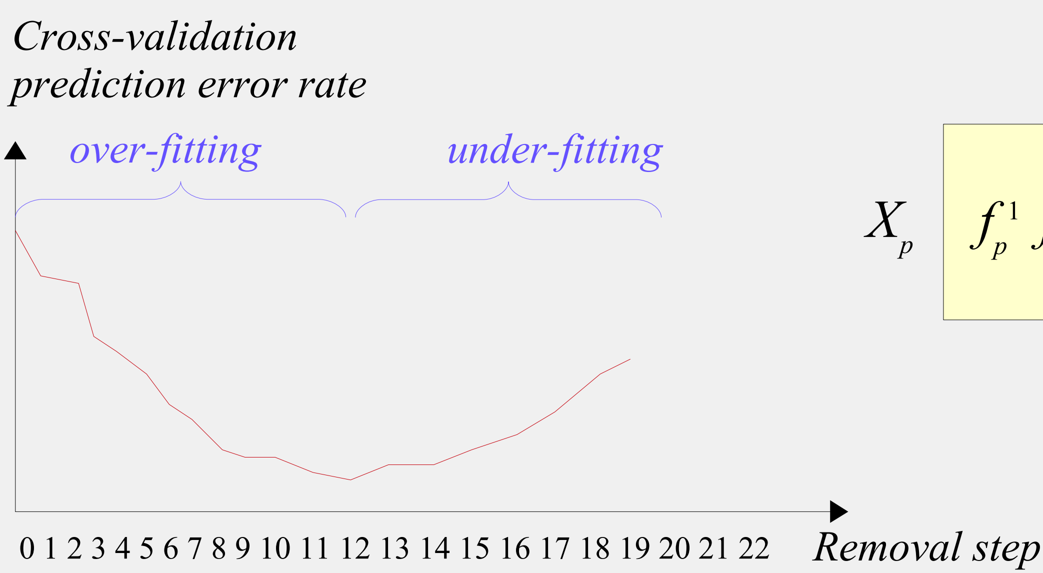
How THEME works

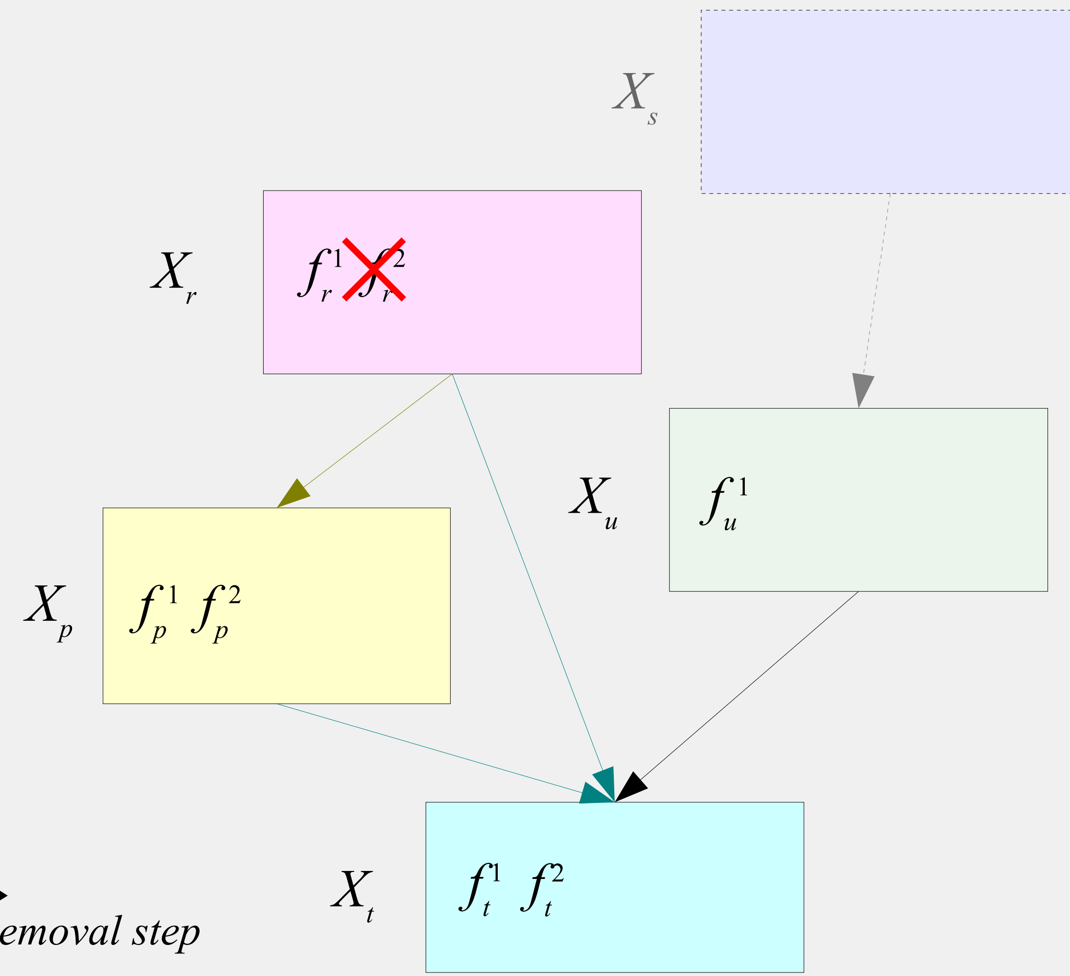
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:

Example:





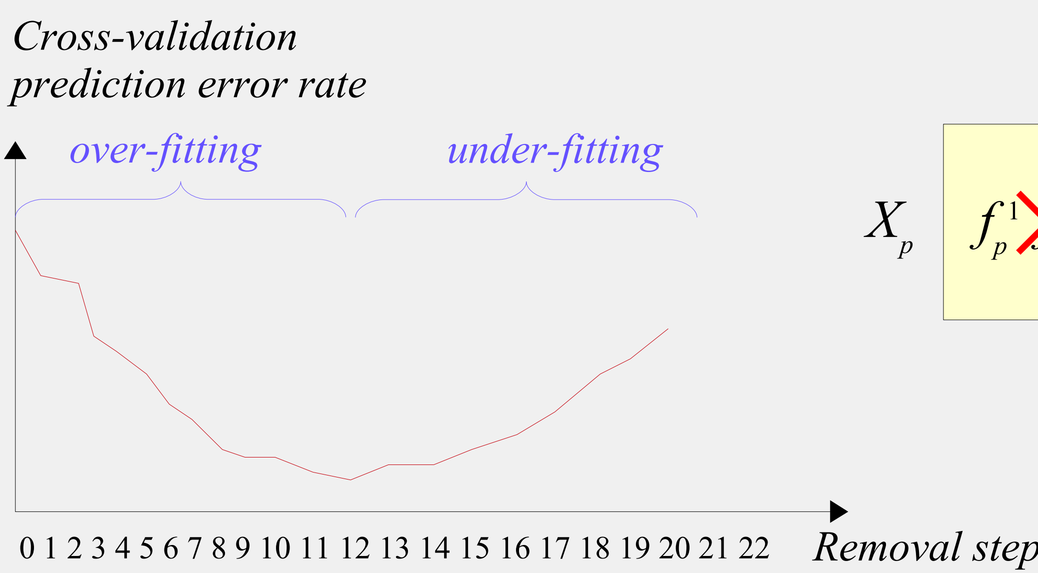
How THEME works

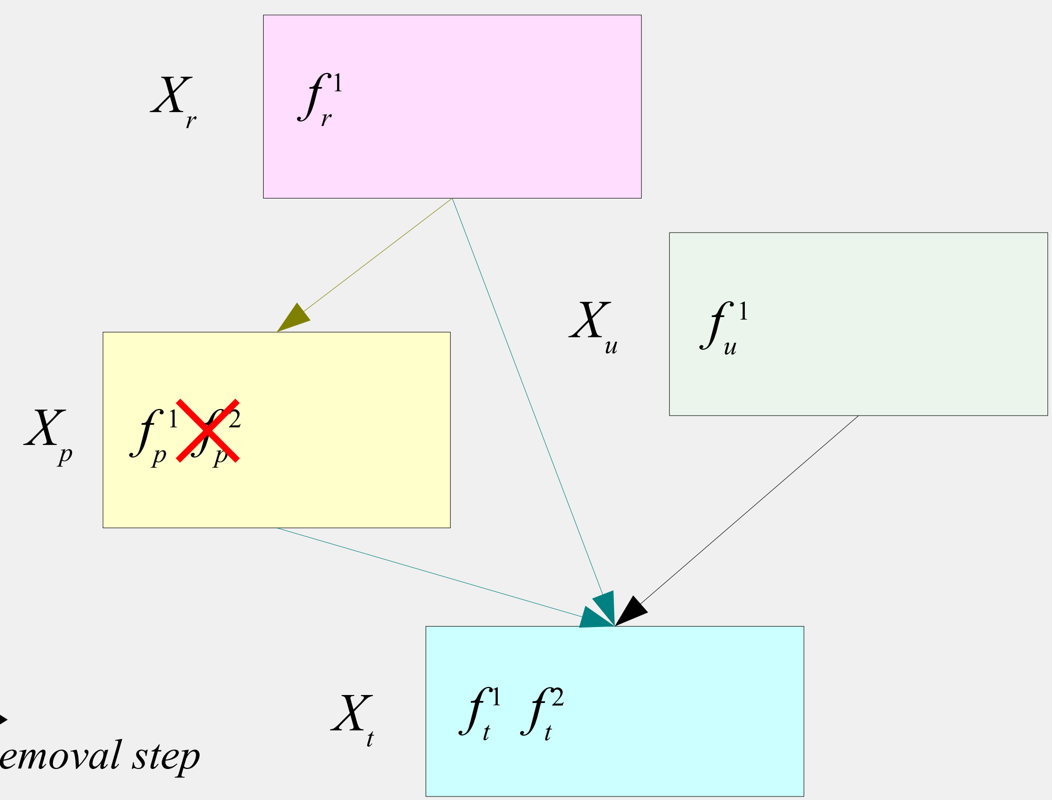
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:

Example:





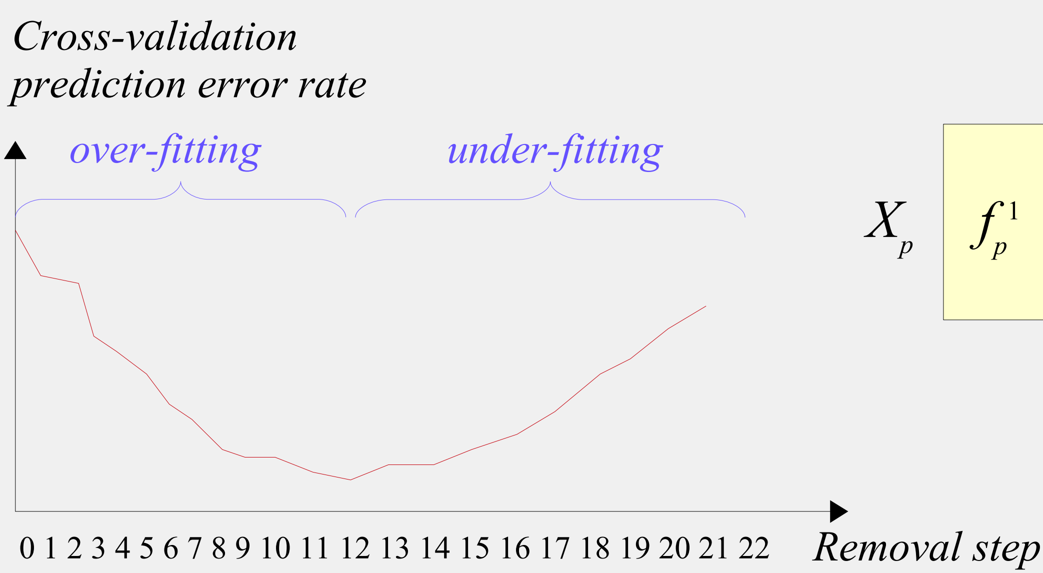
How THEME works

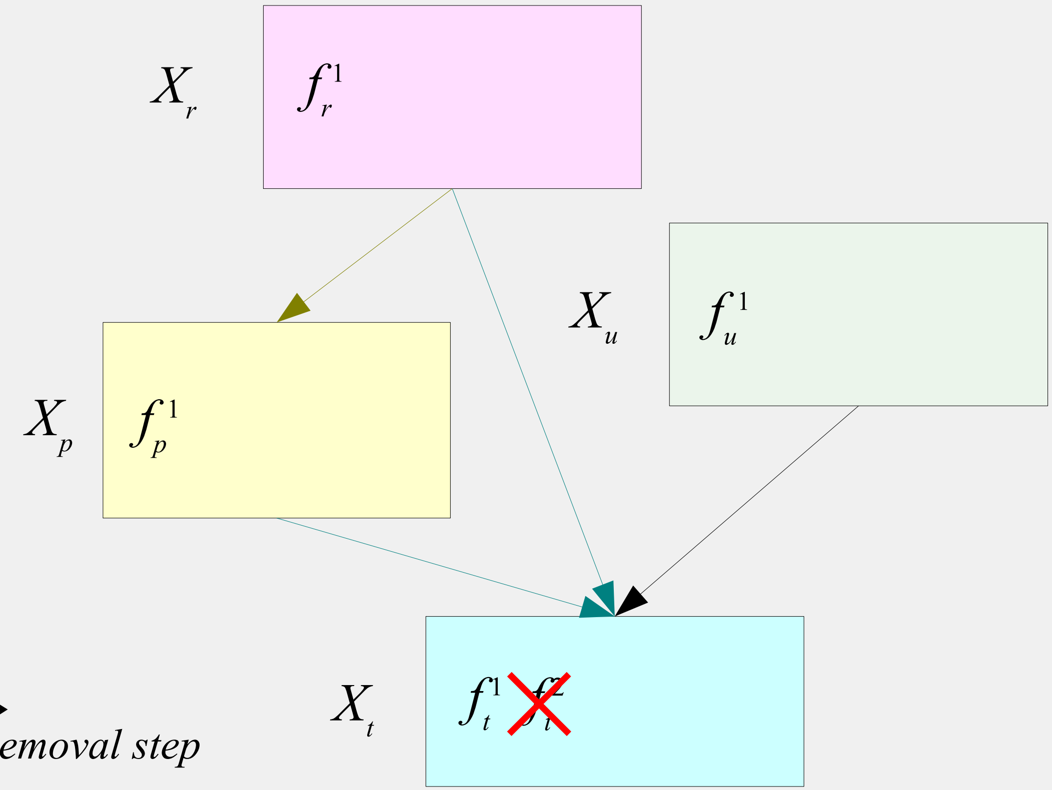
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:

Example:





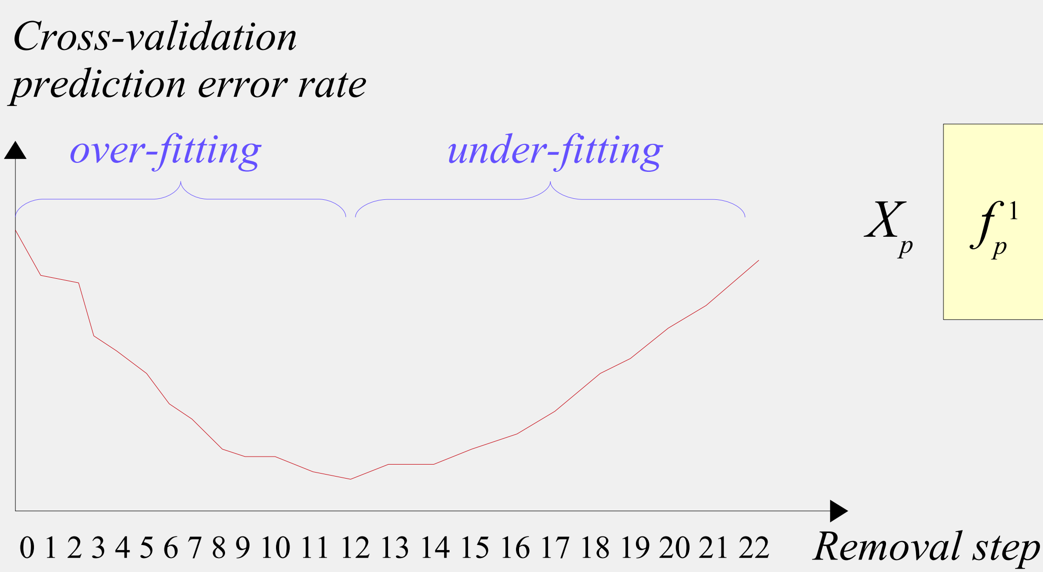
How THEME works

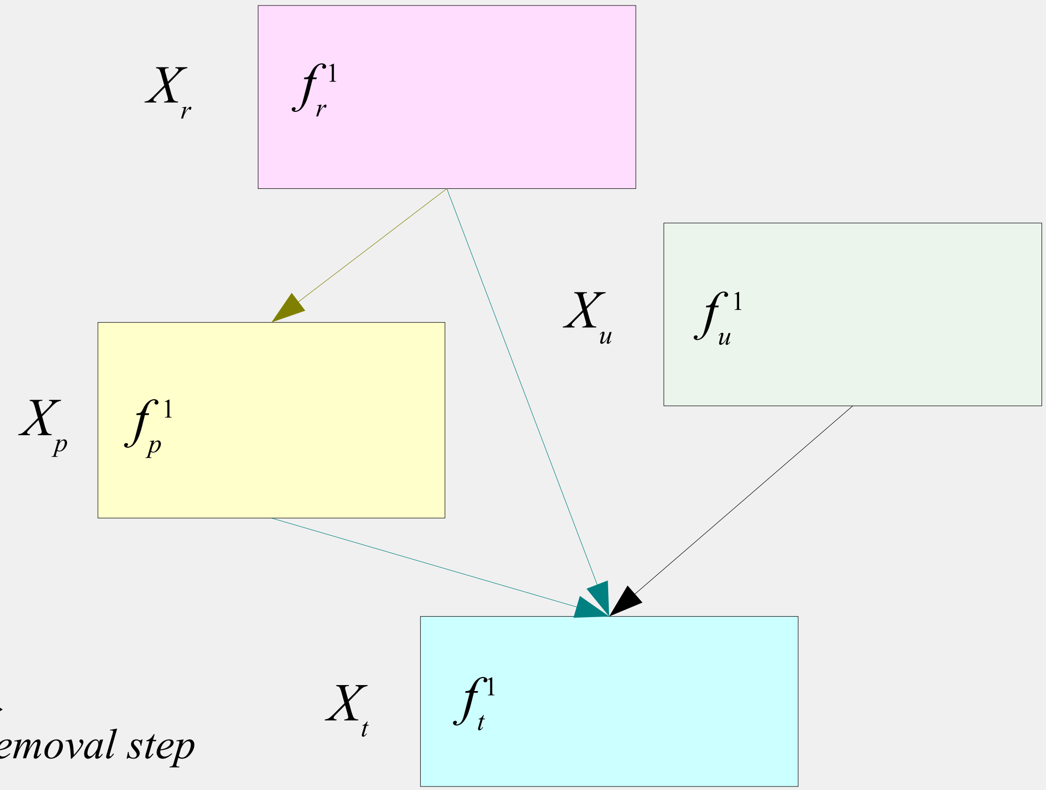
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:

Example:





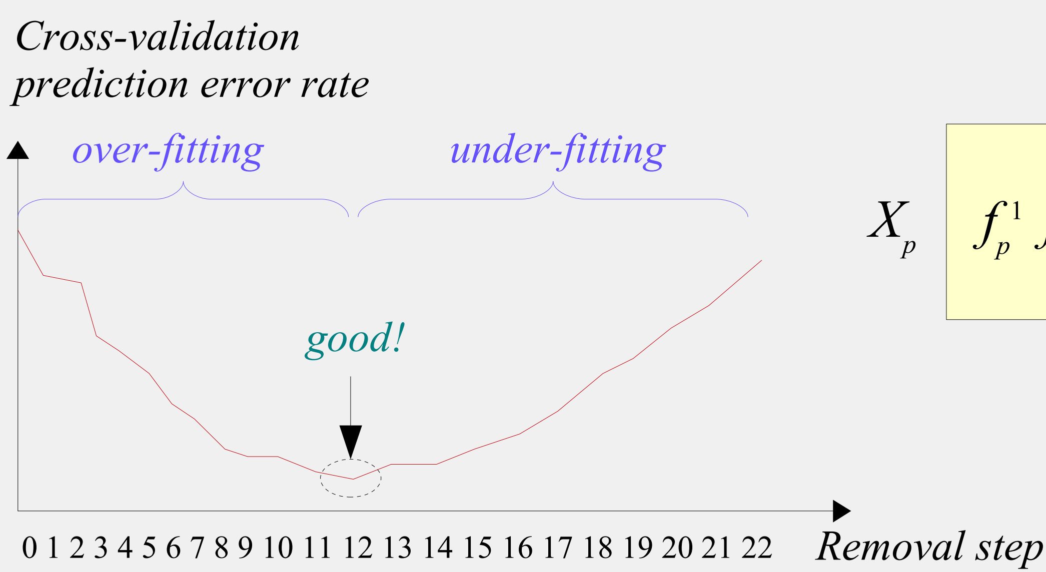
How THEME works

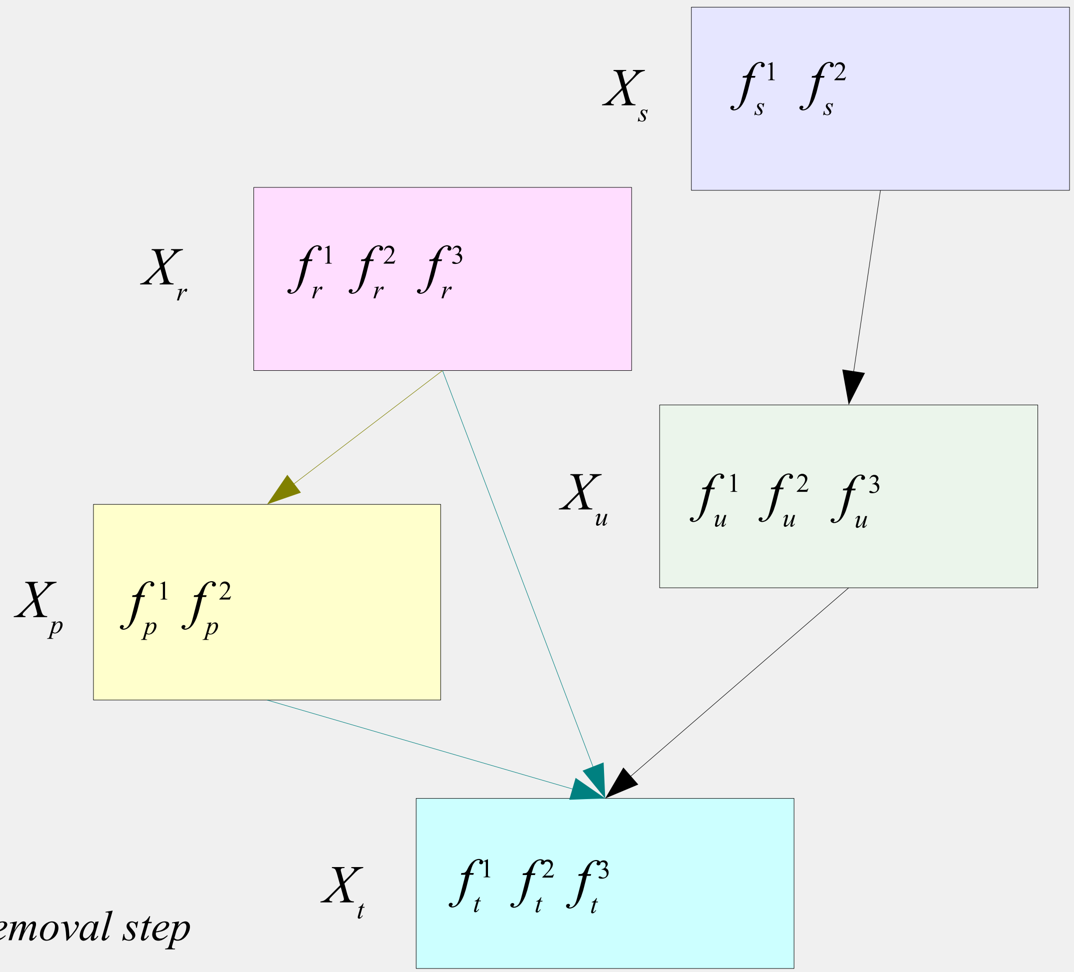
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:

Example:



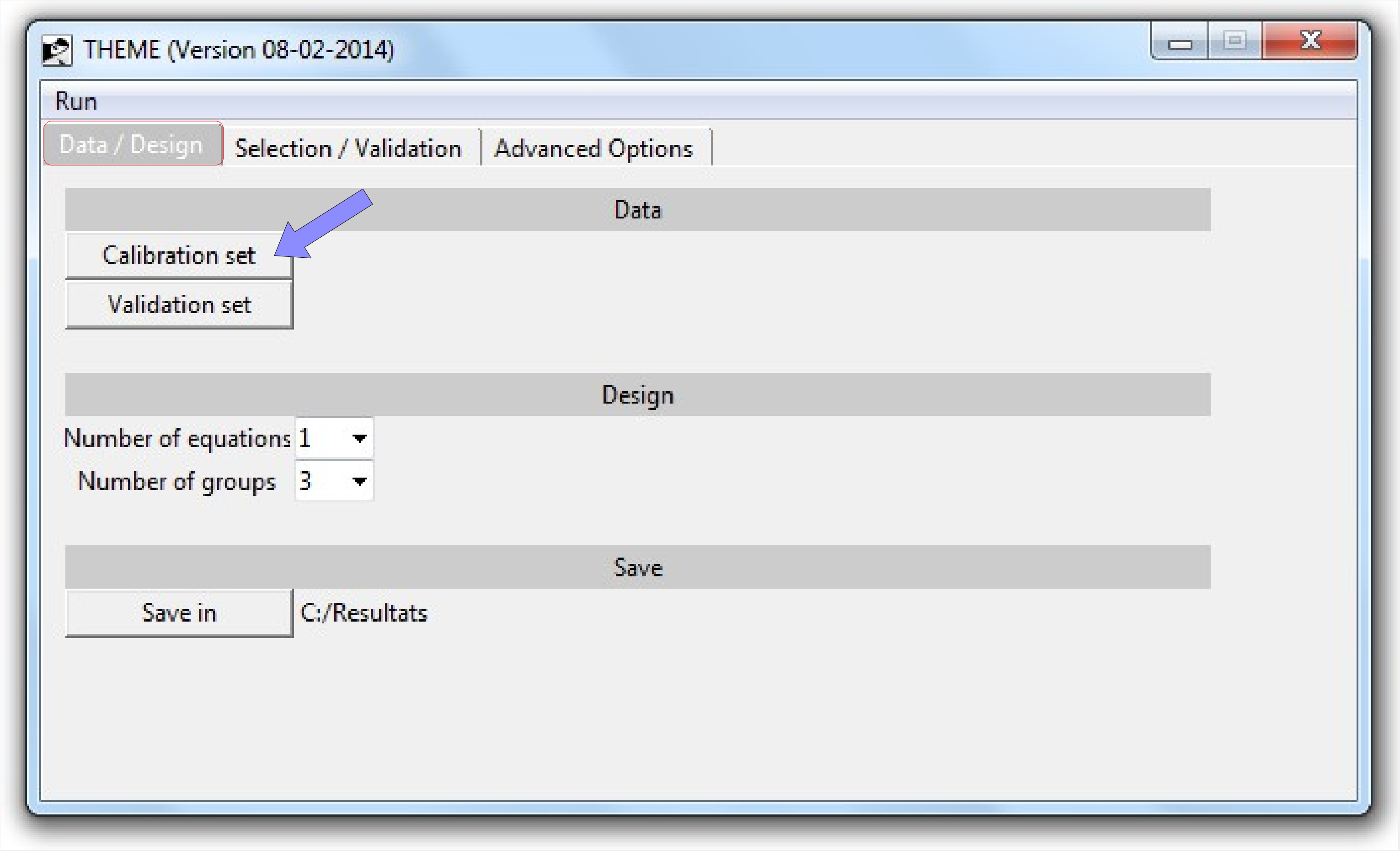


How to operate the THEME R-software?

1. The main window

Data Input

Output



How to operate the THEME R-software?

2. From raw data to Thematic Model

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

Variables

Obs.

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

How to operate the THEME R-software?

2. From raw data to Thematic Model

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

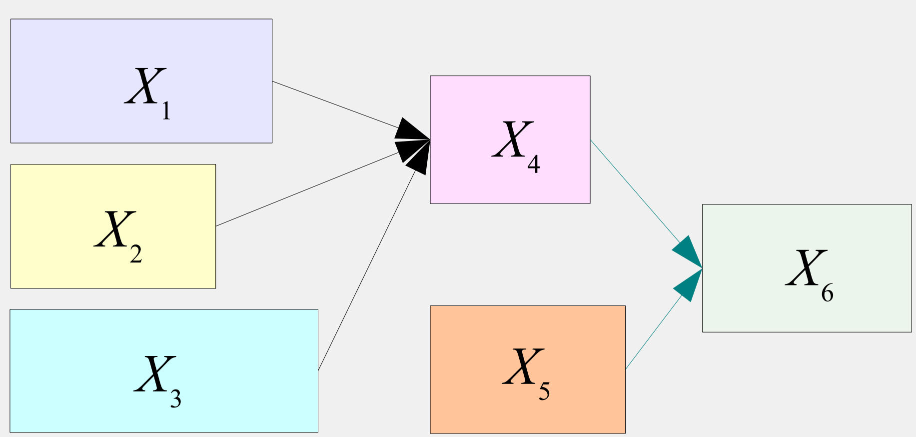
Variables

Obs.

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

- ***Design of the thematic model:***

6 themes
2 equations



How to operate the THEME R-software?

2. *From raw data to Thematic Model*

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

Variables

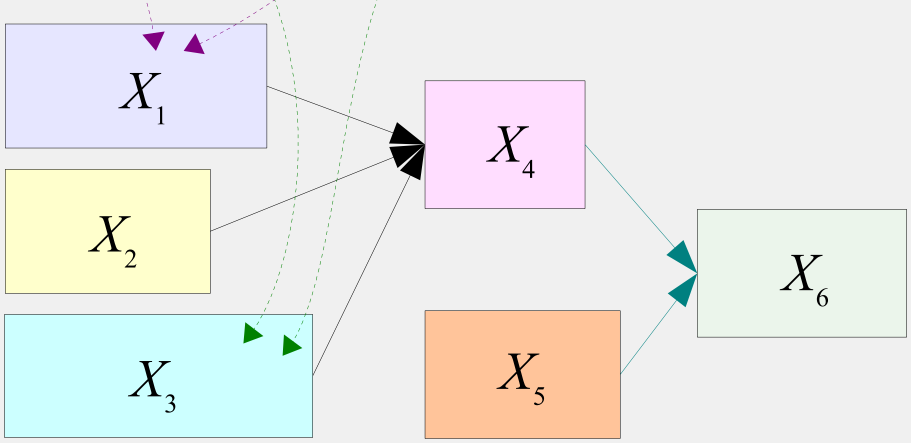
Obs.

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
TGC	cci	1	3	1	3	1	1

Thematic Group Coding →
(0 = *variable not used*)

- ***Design of the thematic model:***

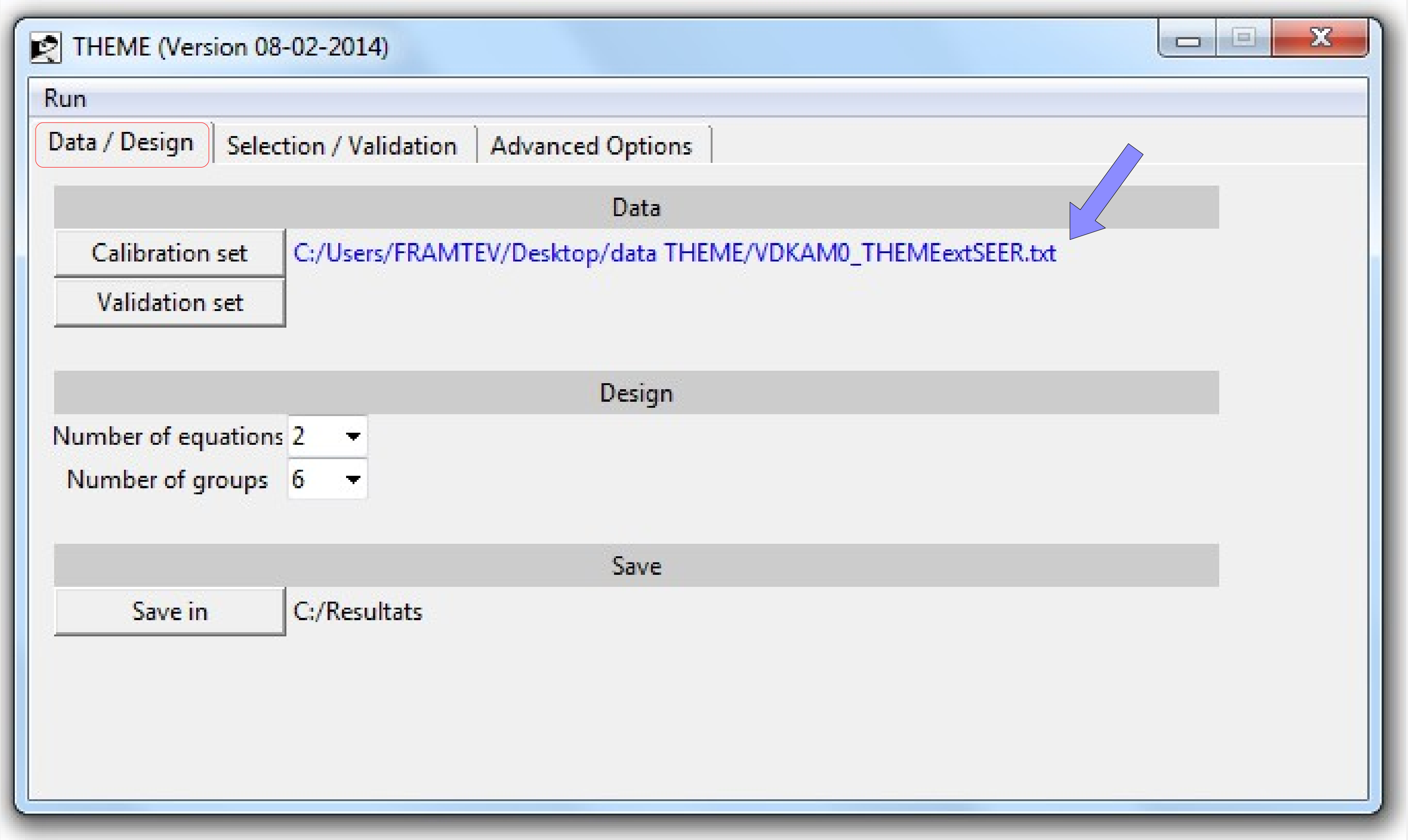
6 themes
2 equations



How to operate the THEME R-software?

2. From raw data to Thematic Model

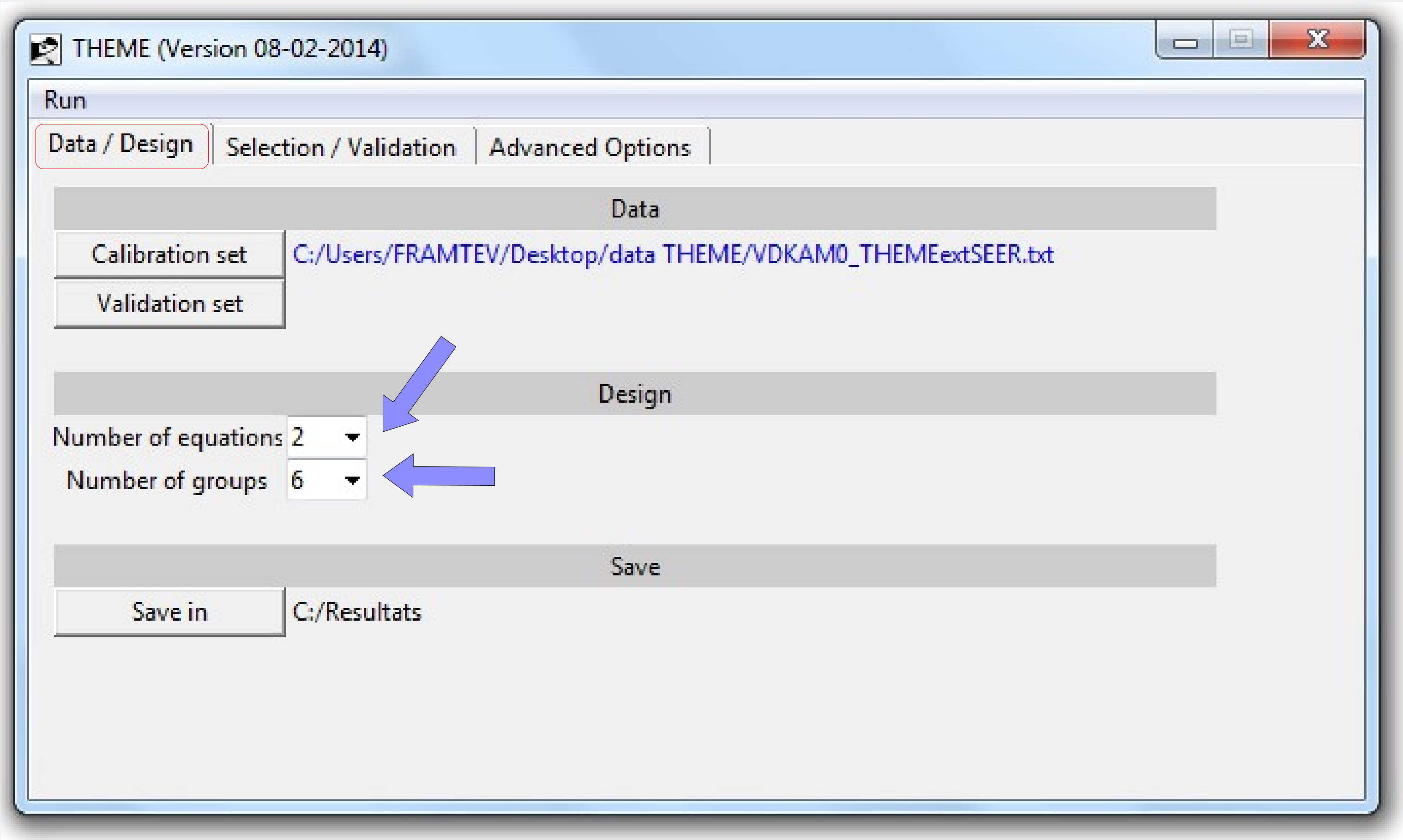
Data Input



How to operate the THEME R-software?

2. From raw data to Thematic Model

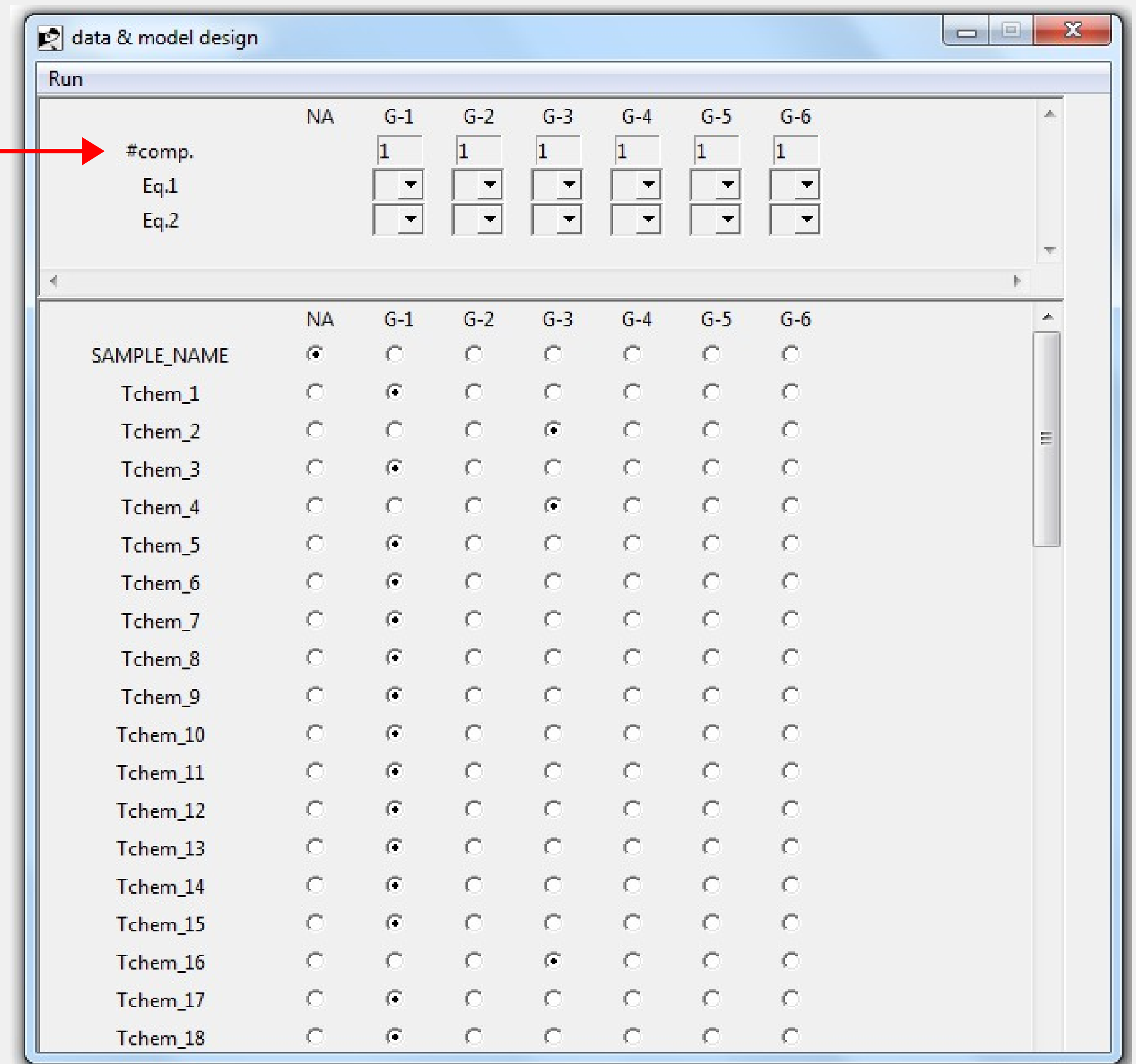
Model design



How to operate the THEME R-software?

2. From raw data to Thematic Model

Number of components in groups

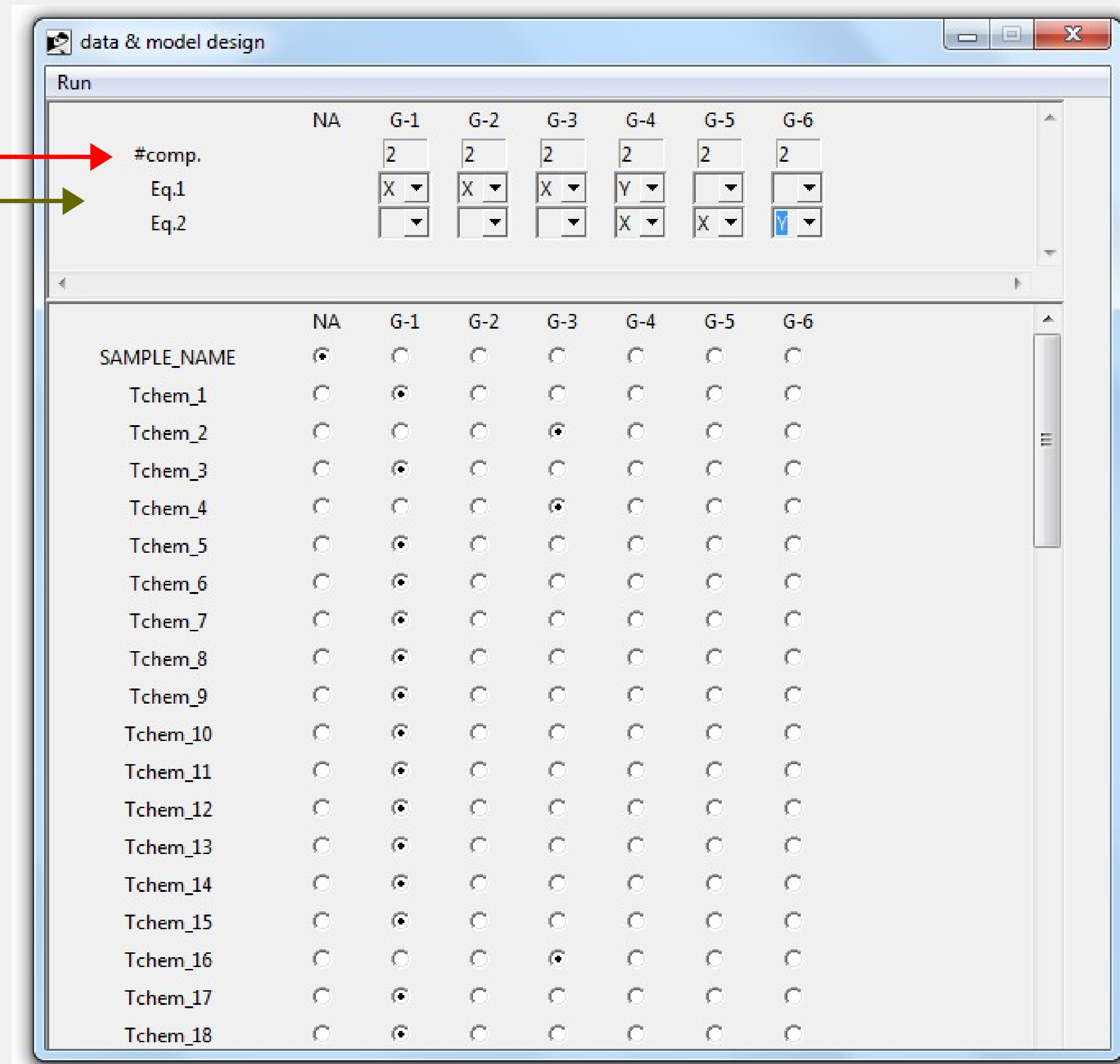


How to operate the THEME R-software?

2. From raw data to Thematic Model

Number of components in groups

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



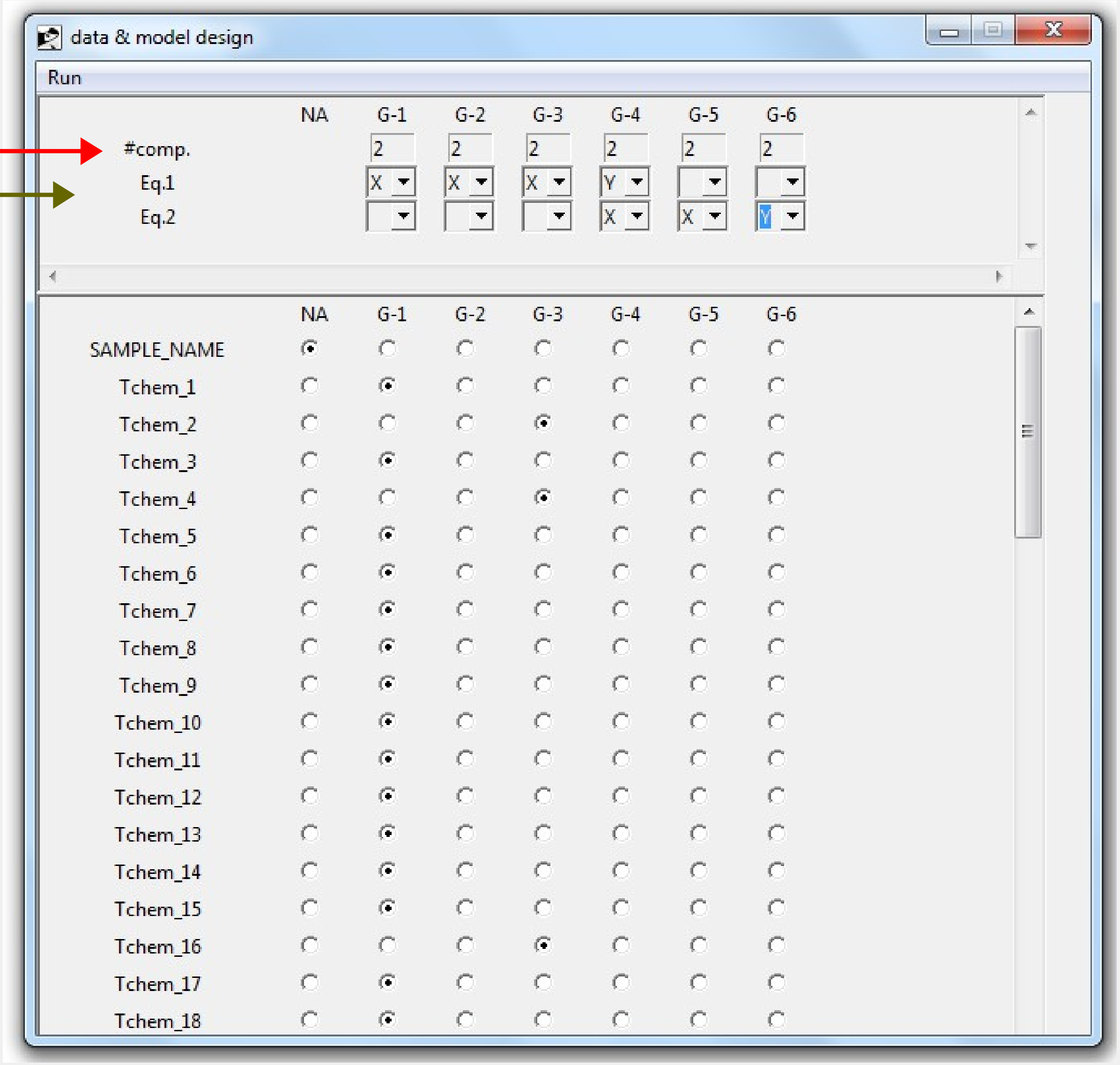
How to operate the THEME R-software?

2. From raw data to Thematic Model

Number of components in groups

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

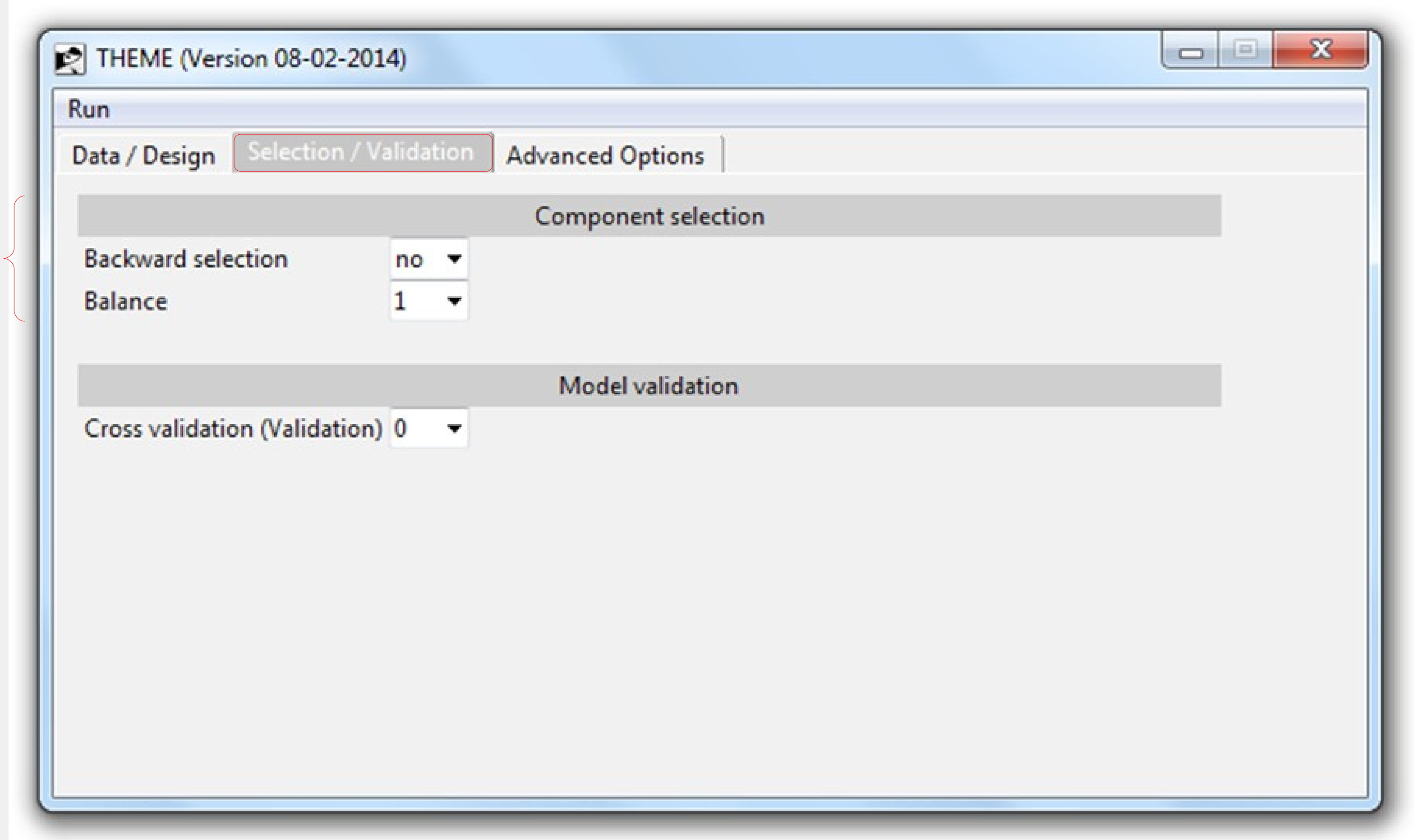
If TGC line in datafile, pre-filled. Else, interactive design:



How to operate the THEME R-software?

3. Setting the selection & validation parameters

Component selection

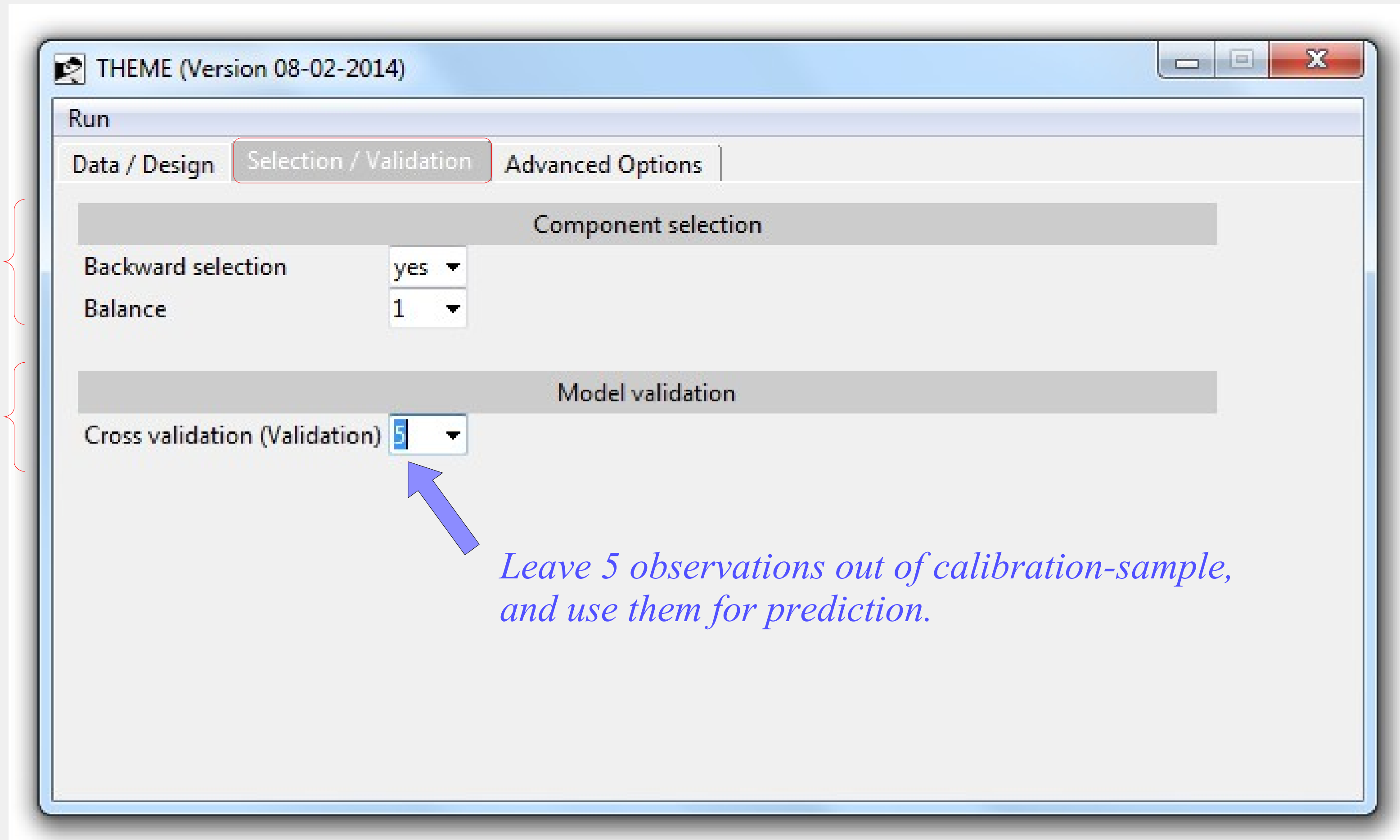


How to operate the THEME R-software?

3. Setting the selection & validation parameters

Component selection

Model cross-validation



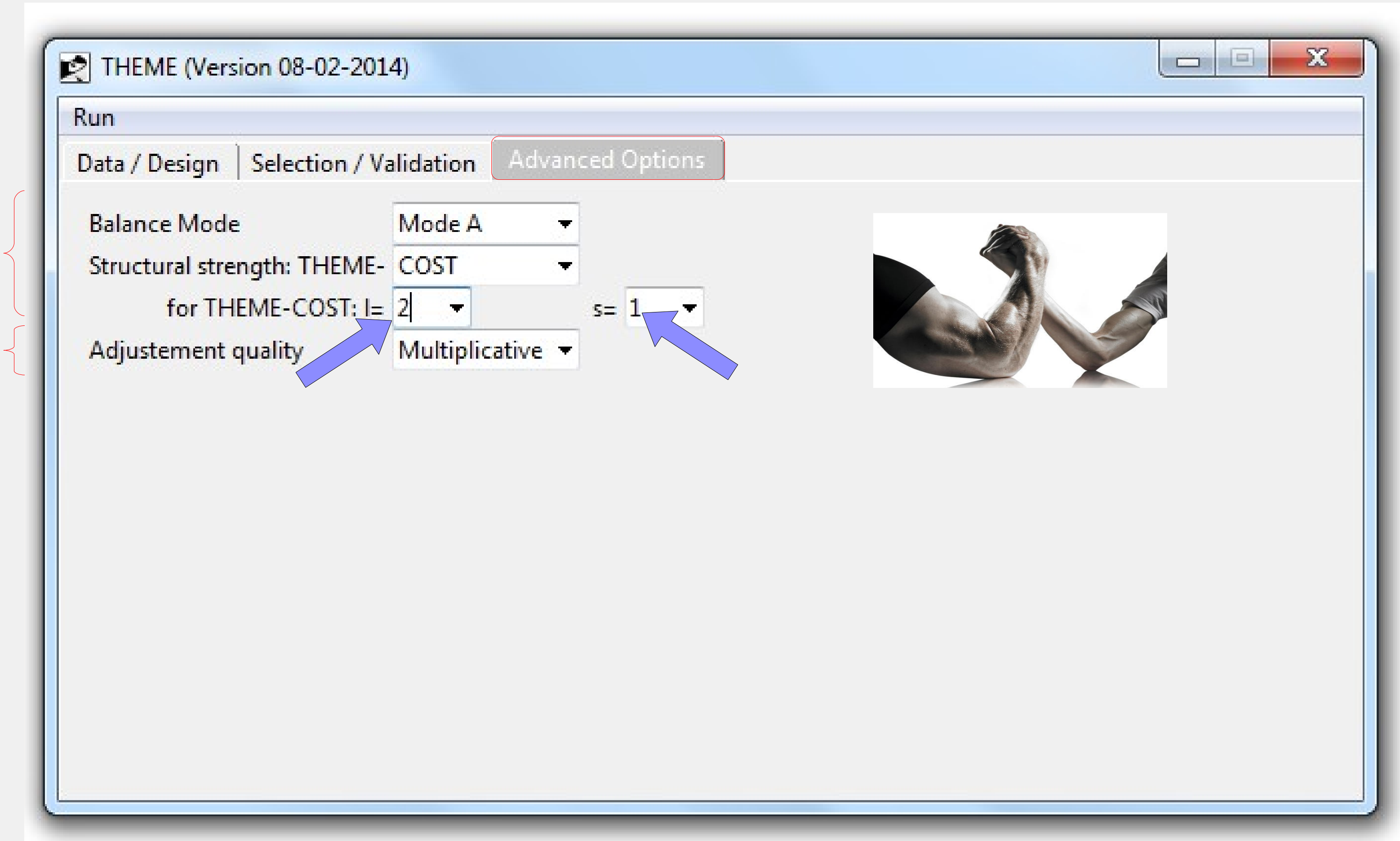
Leave 5 observations out of calibration-sample, and use them for prediction.

How to operate the THEME R-software?

4. Setting the structural strength and goodness of fit parameters

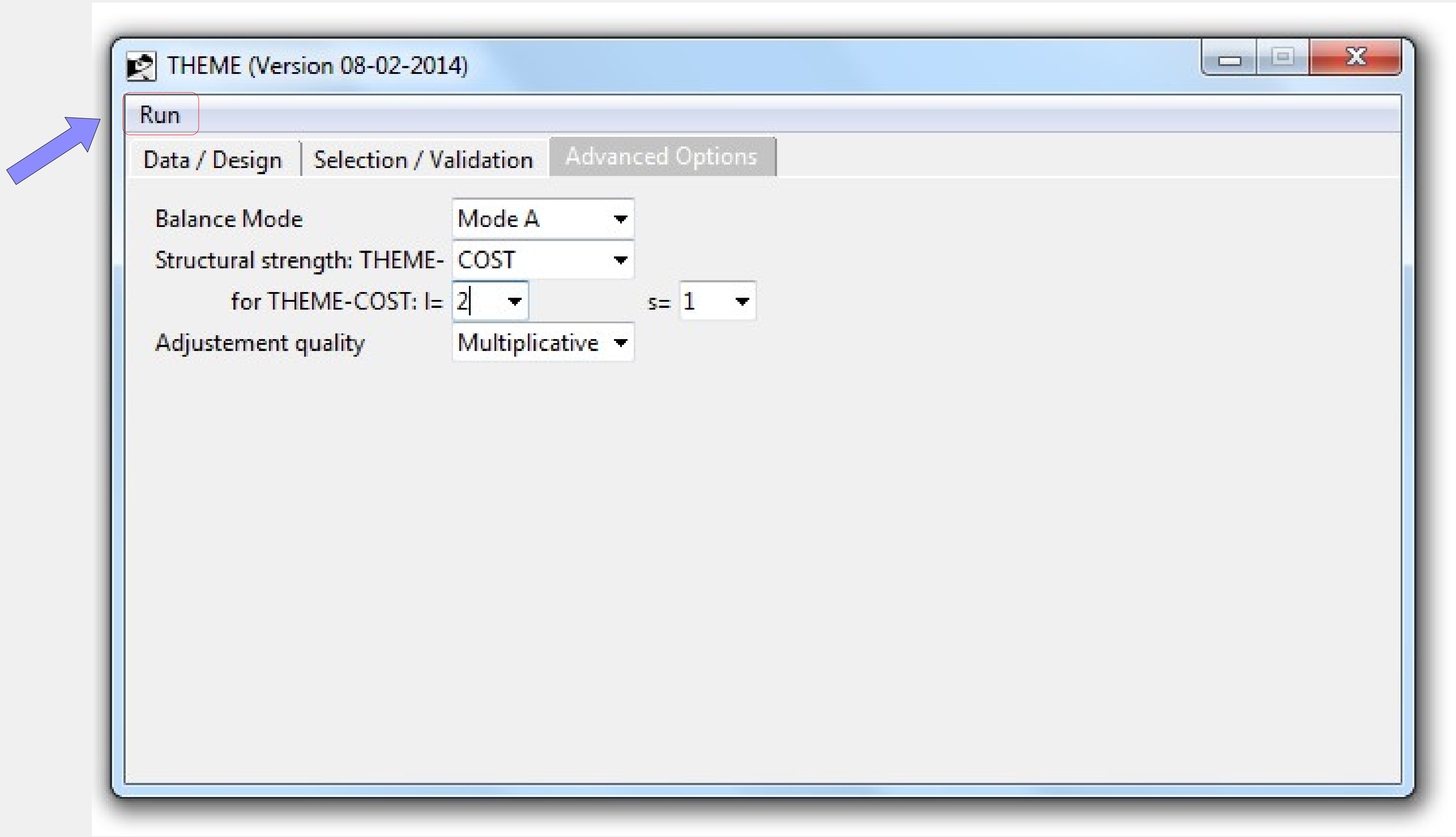
Structural strength parameters

GoF



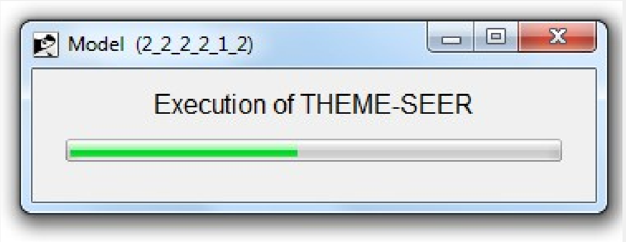
How to operate the THEME R-software?

5. Launching estimation



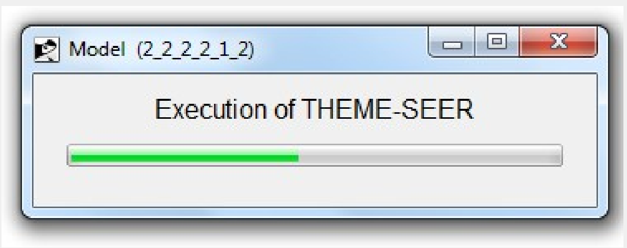
How to operate the THEME R-software?

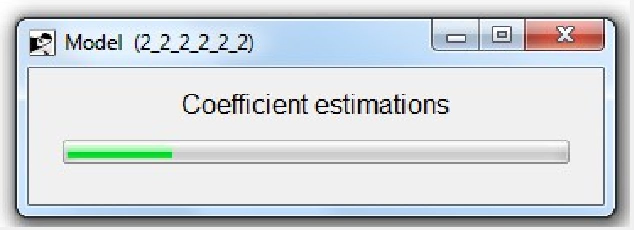
6. Waiting for results



How to operate the THEME R-software?

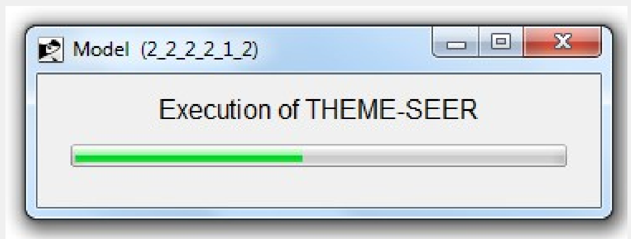
6. Waiting for results

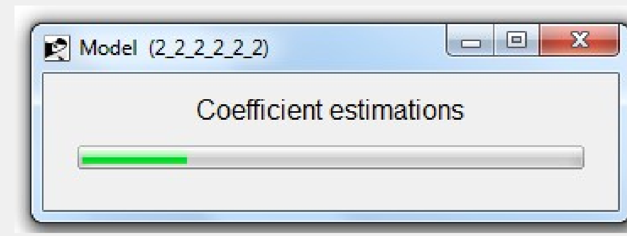


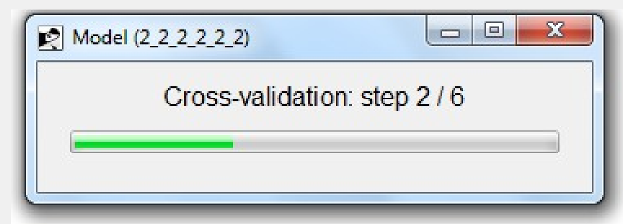


How to operate the THEME R-software?

6. Waiting for results

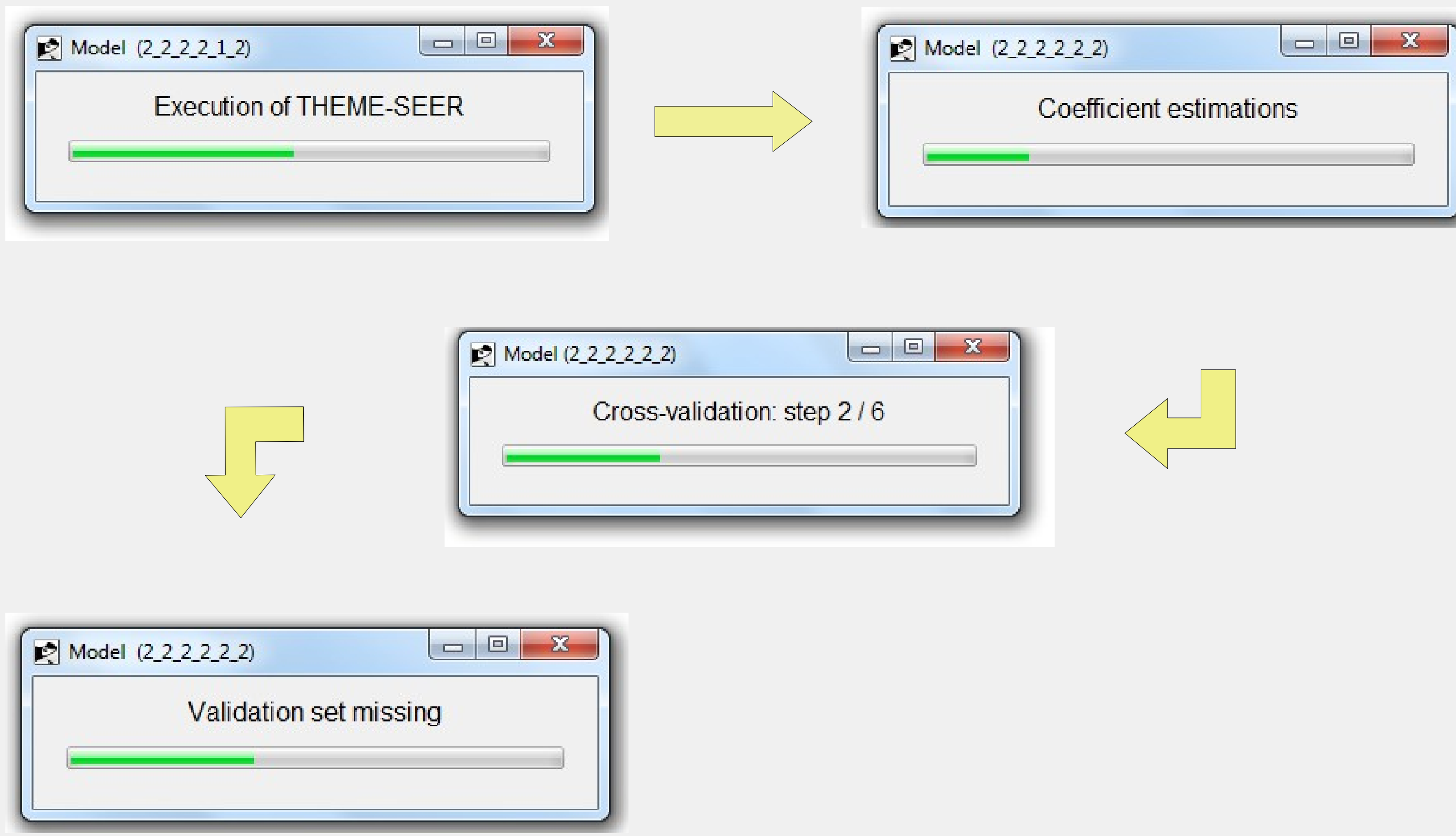






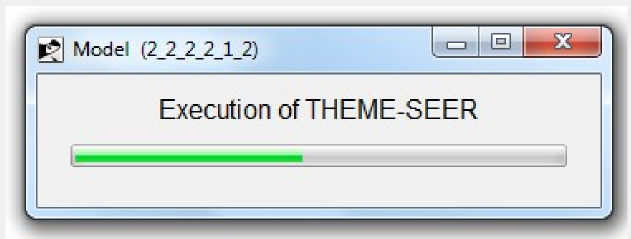
How to operate the THEME R-software?

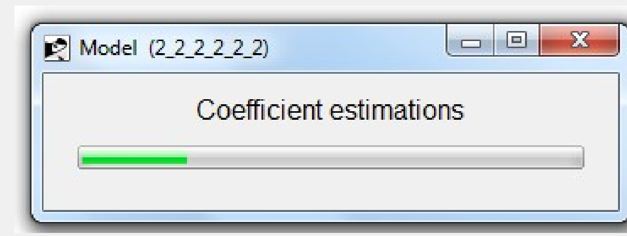
6. Waiting for results

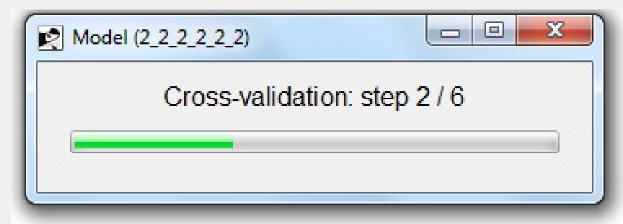


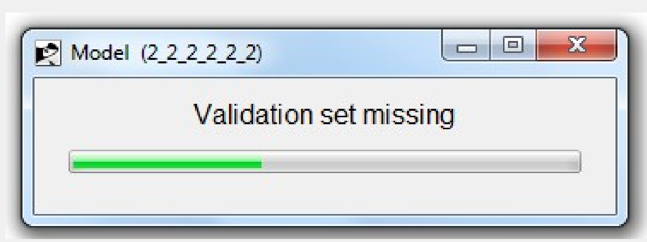
How to operate the THEME R-software?

6. Waiting for results





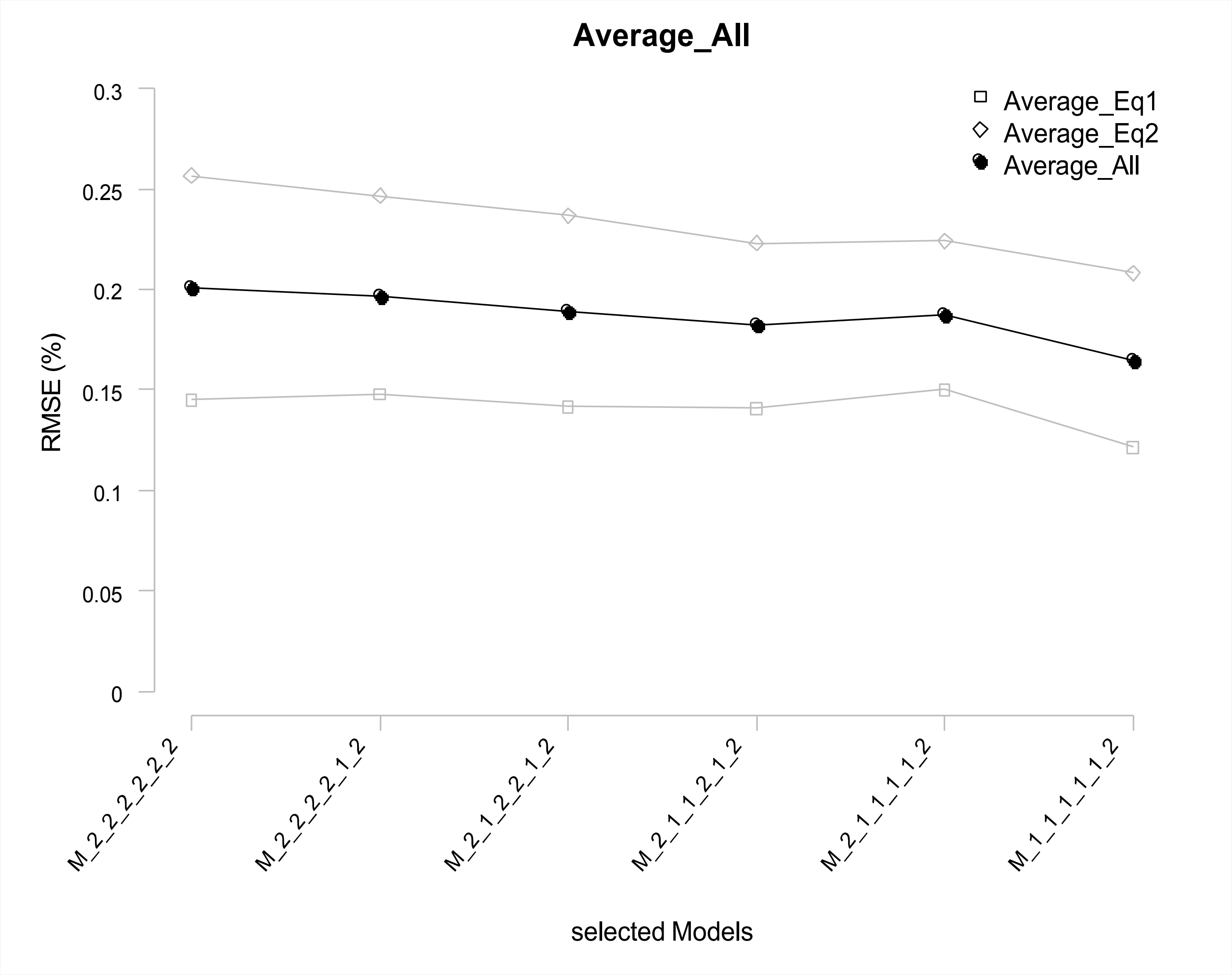




How to operate the THEME R-software?

7. Reaping results

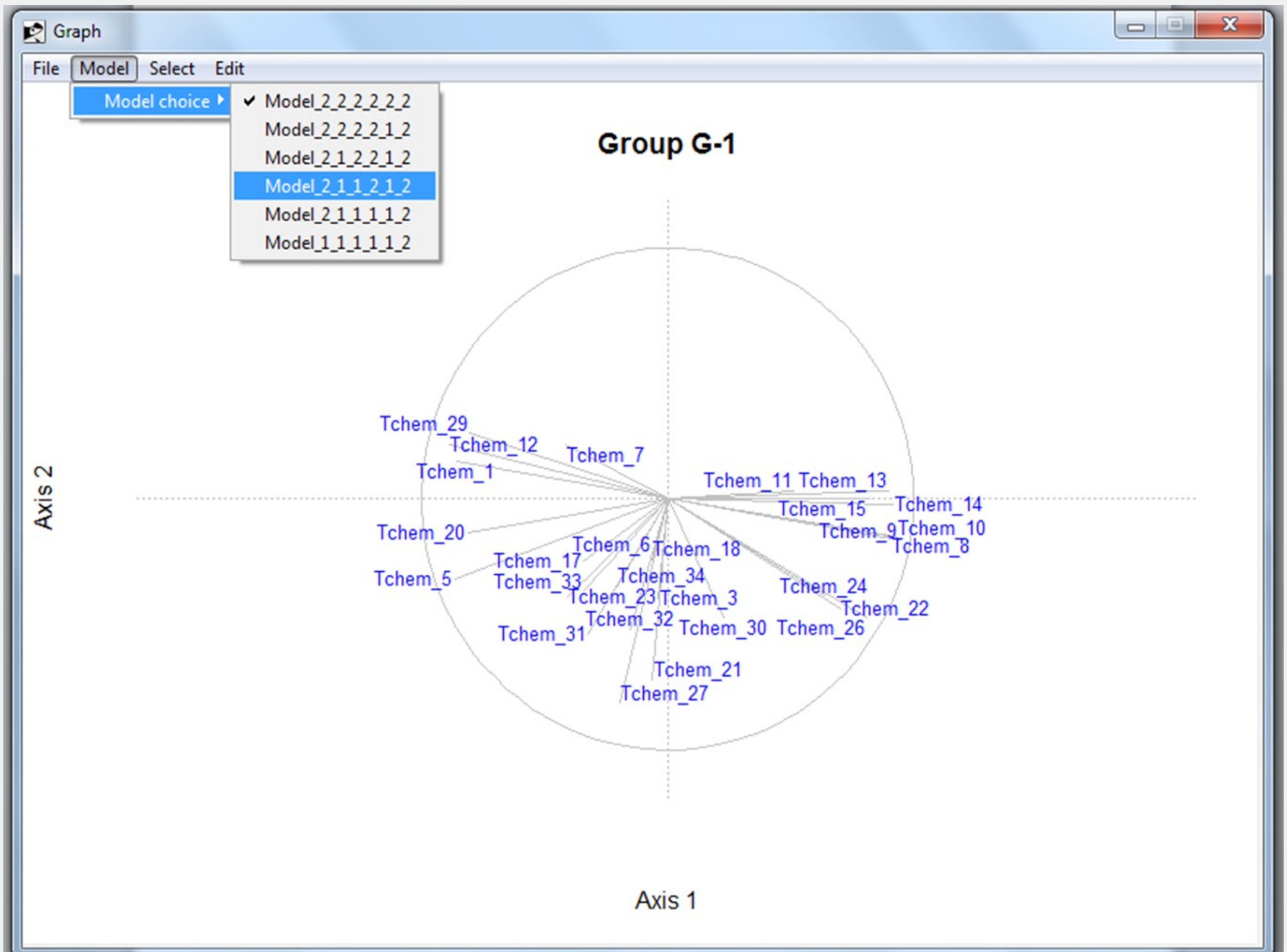
Model-selection



How to operate the THEME R-software?

7. Reaping results

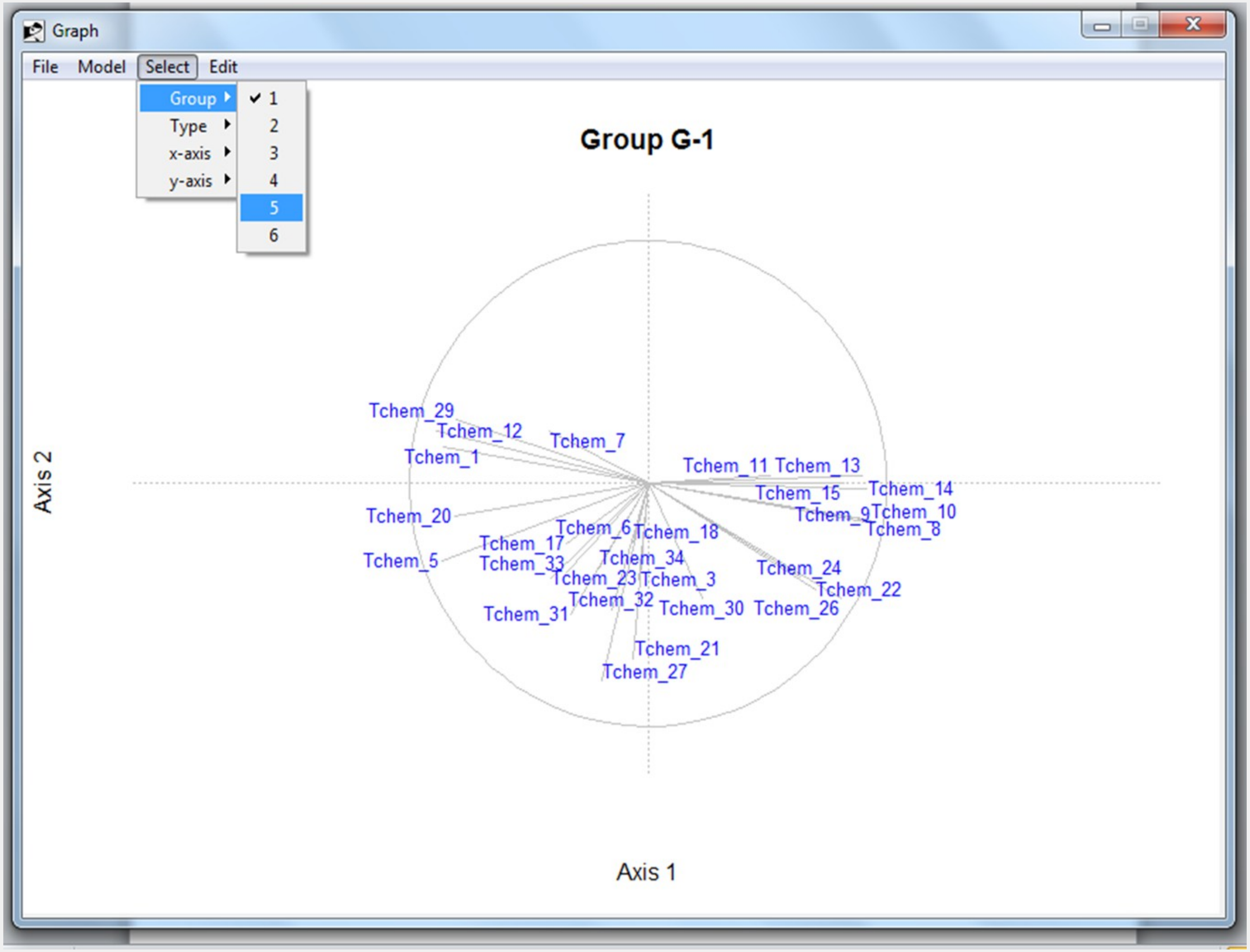
Model-selection* → *Graphing variables



How to operate the THEME R-software?

7. Reaping results

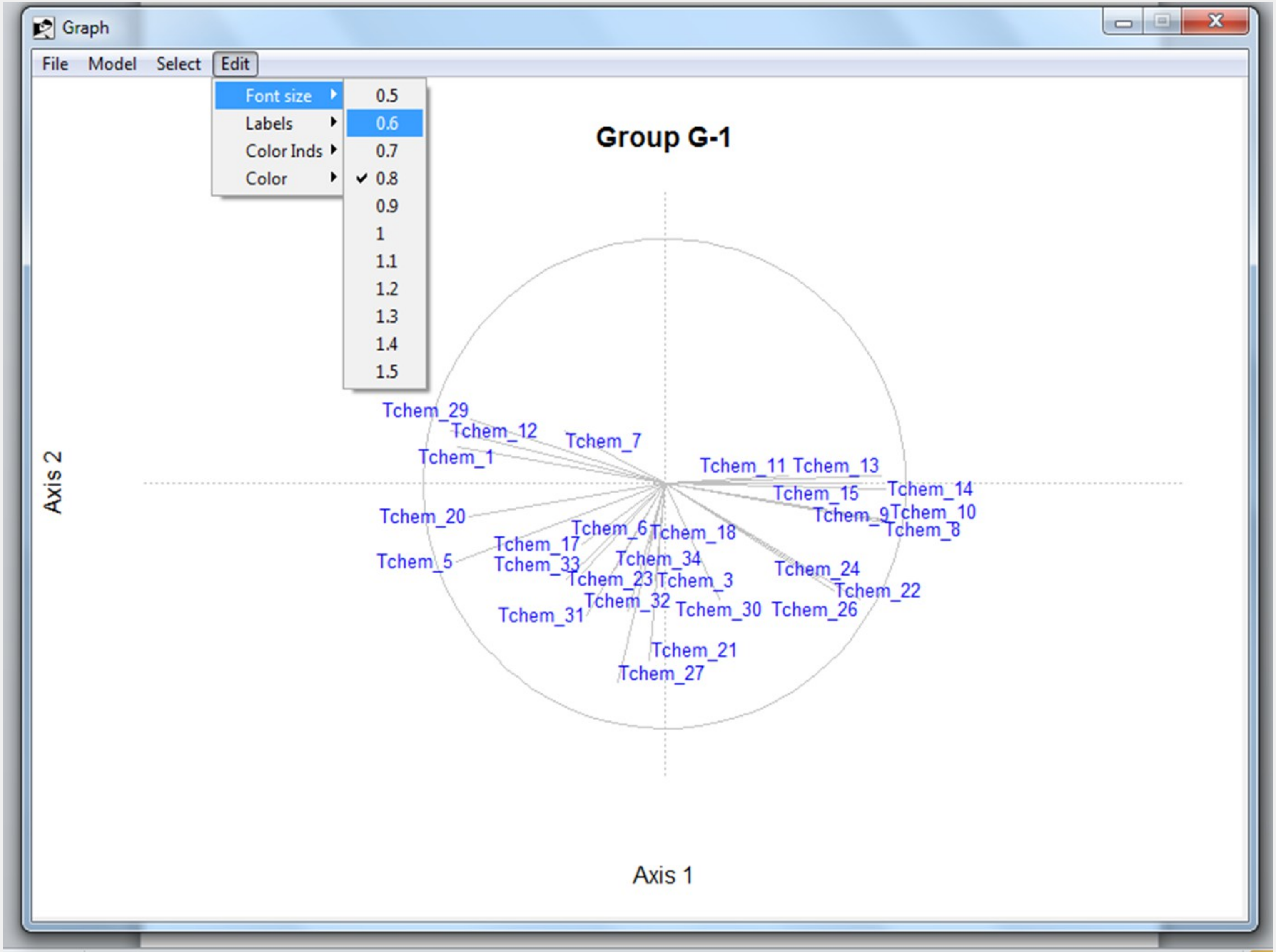
Graphing variables



How to operate the THEME R-software?

7. Reaping results

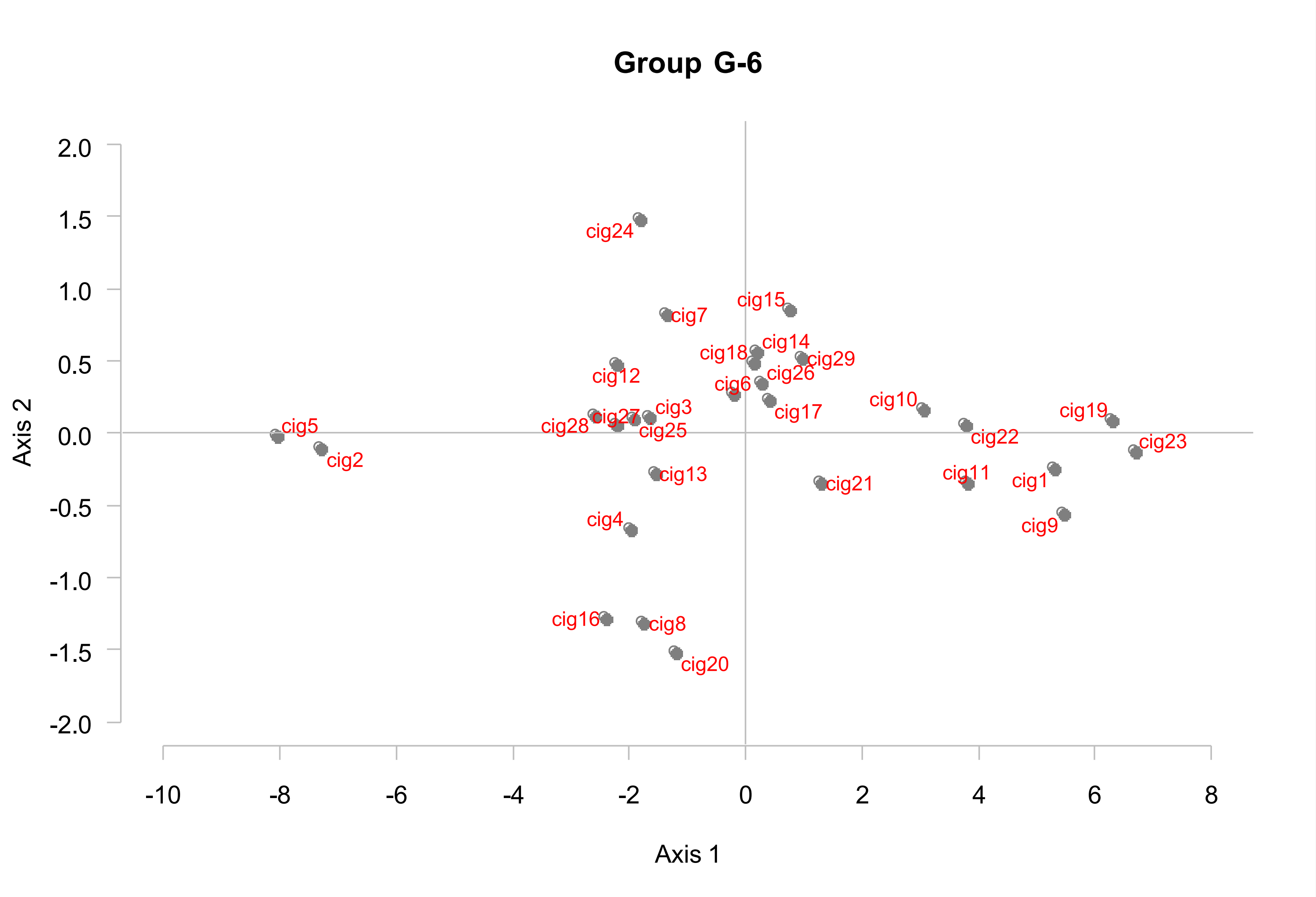
Graphing variables



How to operate the THEME R-software?

7. Reaping results

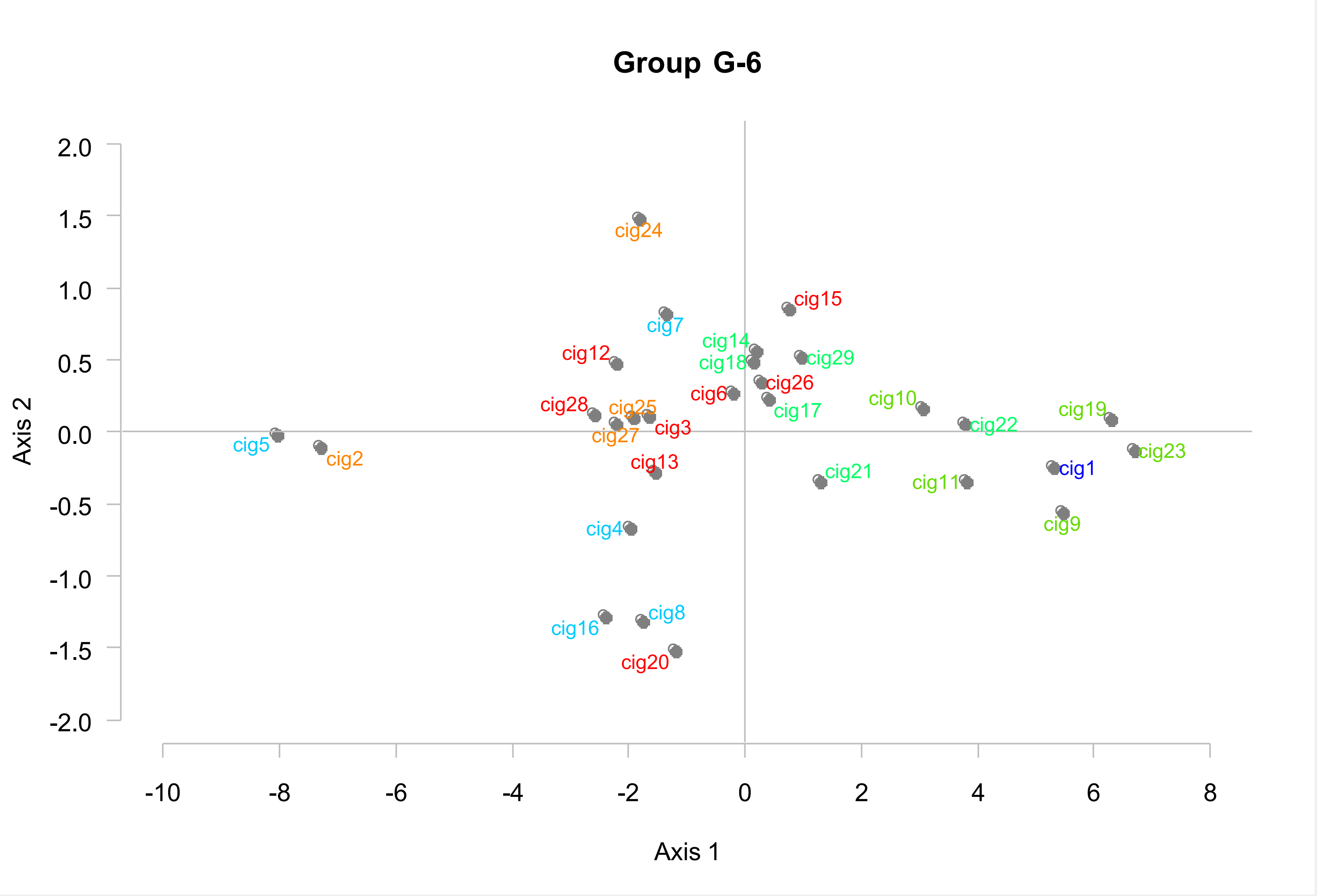
Graphing observations



How to operate the THEME R-software?

7. Reaping results

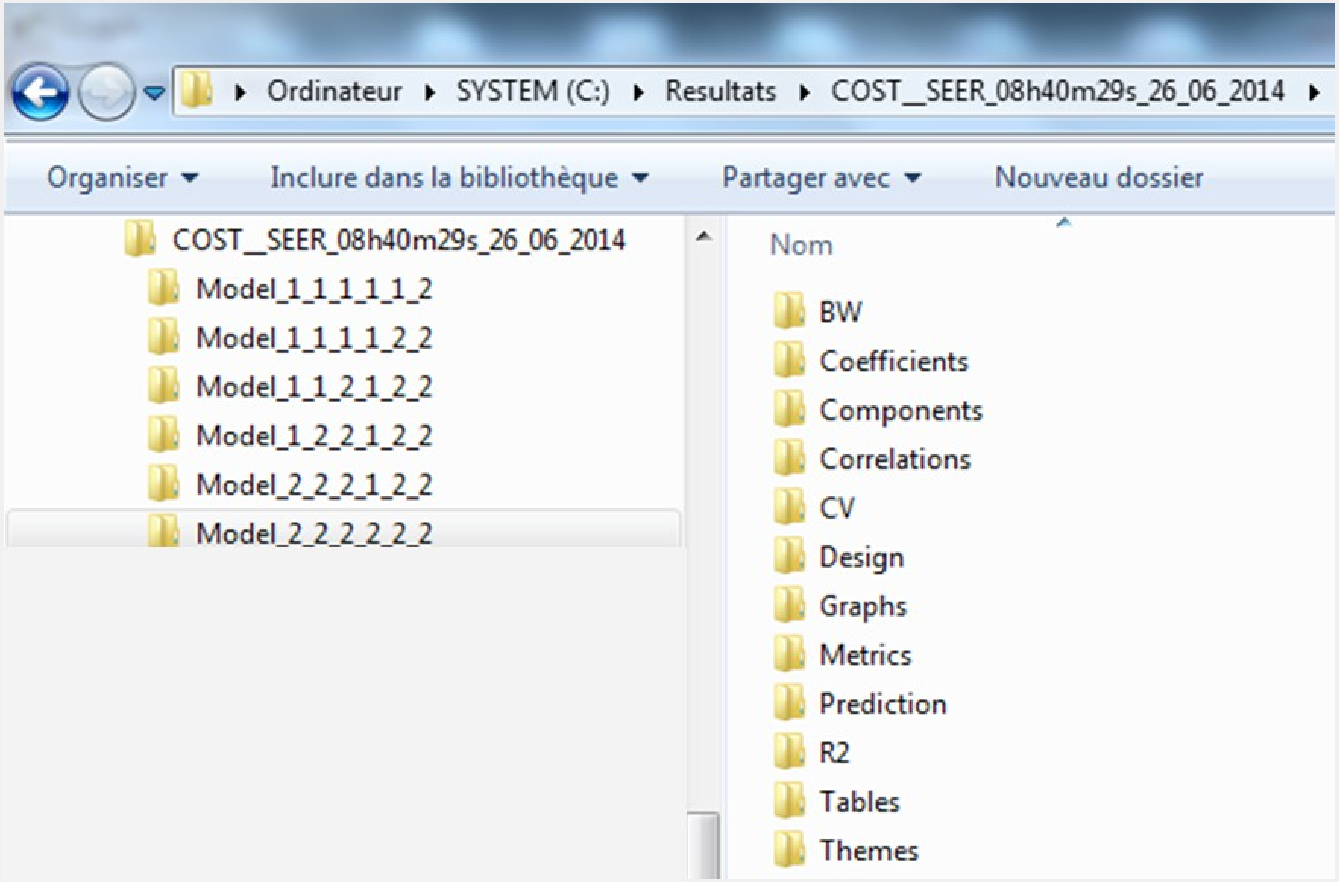
Graphing observations



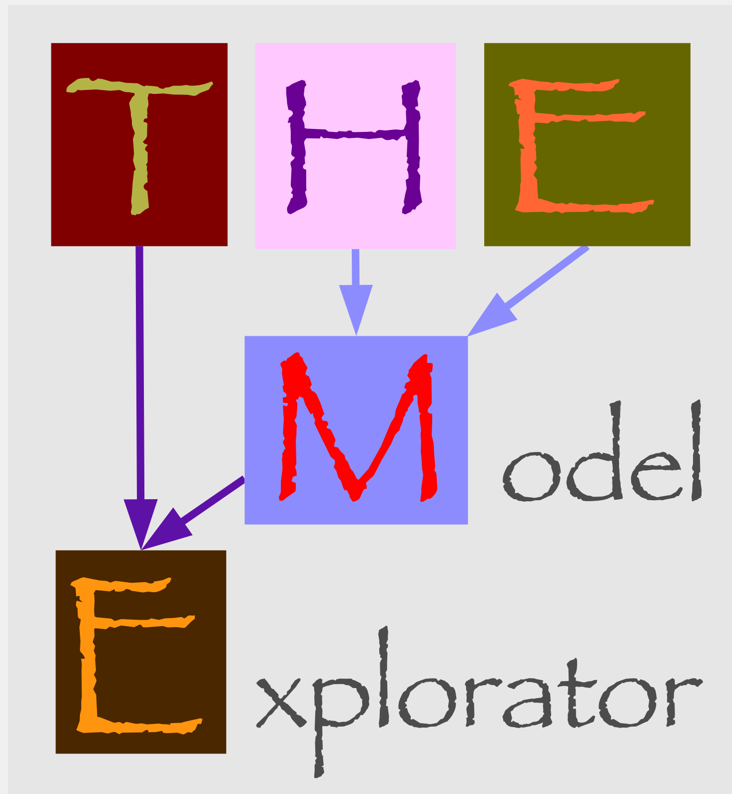
How to operate the THEME R-software?

7. Reaping results

Getting ALL the results as an object



How to get the THEME R-software?



Help-files yet to be written...

Soon available on the CRAN

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.