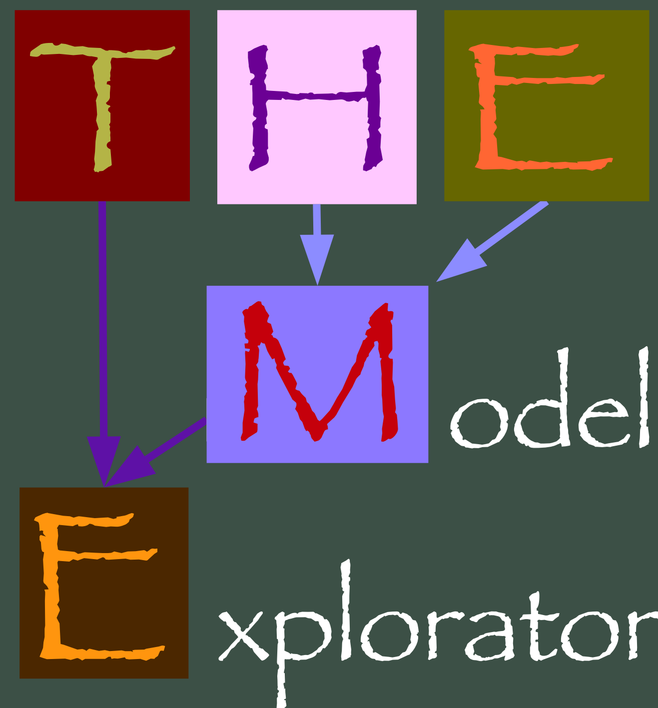
A component-based
Multidimensional Path Modelling technique: THEME
(with a user-friendly R-package)

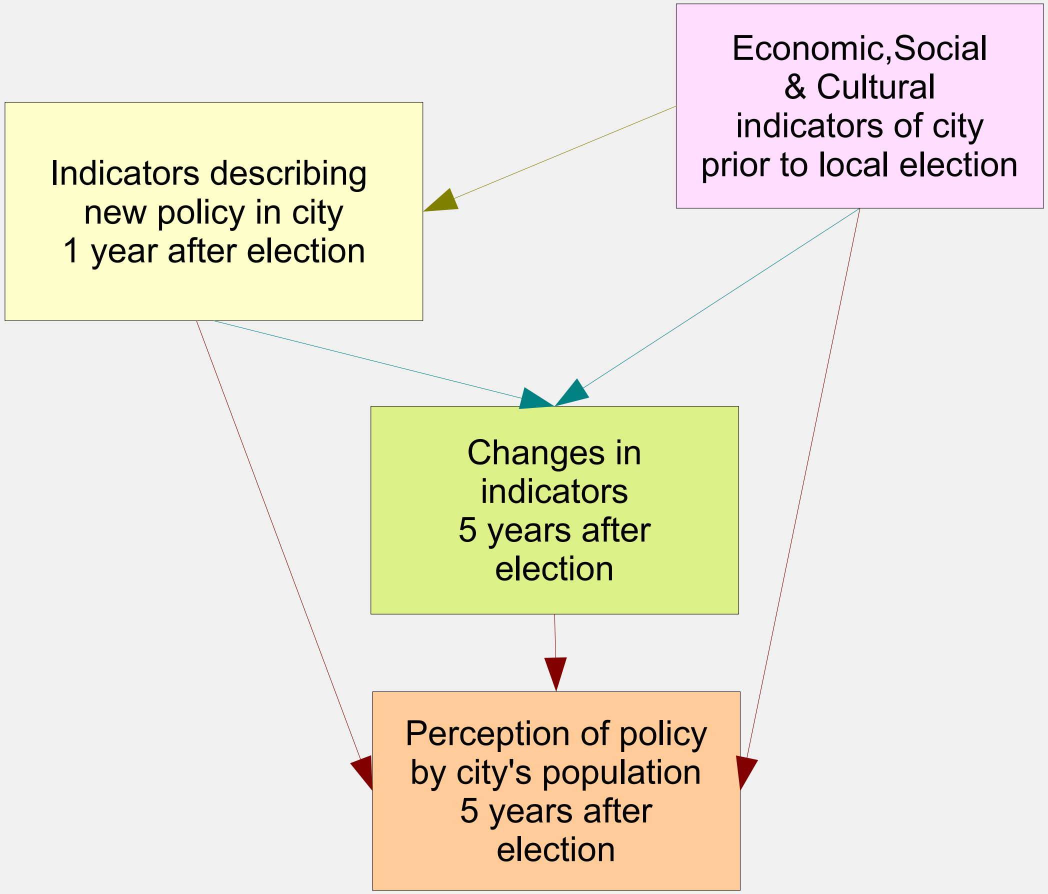


X. Bry *IMAG, Univ. Montpellier*
T. Verron *ITG - SEITA, Centre de recherche*

Data and Problem:

1. Conceptual model of a situation = Thematic Model

Example: *n* cities described through indicators:

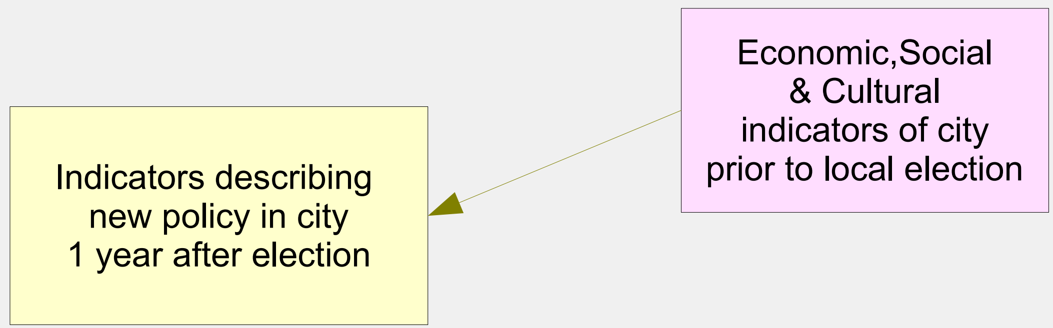


Data and Problem:

1. Conceptual model of a situation = Thematic Model

Example: *n* cities described through indicators:

Thematic equation 1

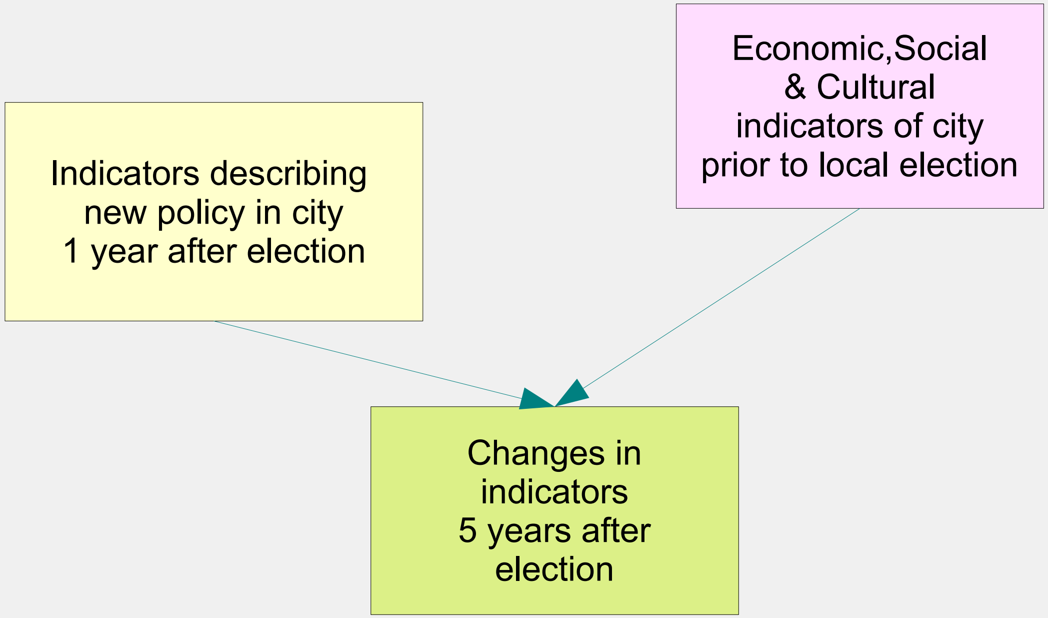


Data and Problem:

1. Conceptual model of a situation = Thematic Model

Example: *n* cities described through indicators:

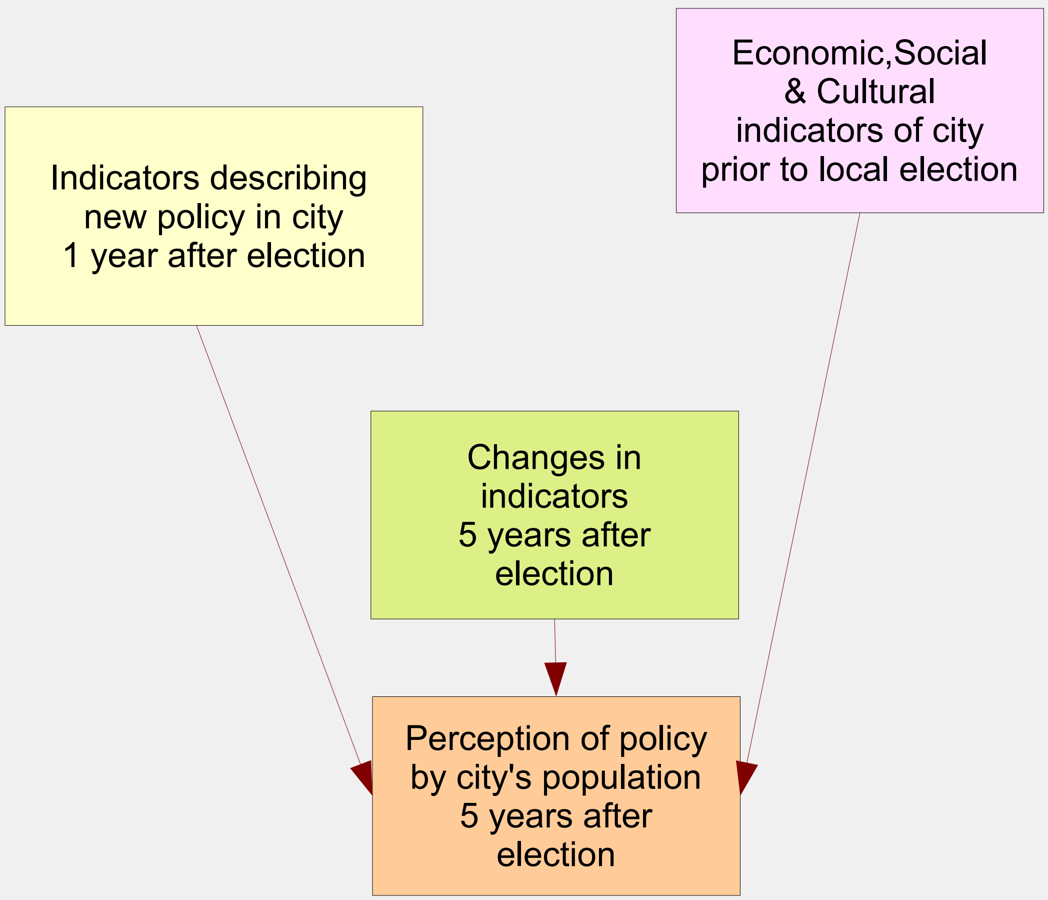
Thematic equation 2



Data and Problem:

1. Conceptual model of a situation = Thematic Model

Example: *n* cities described through indicators:

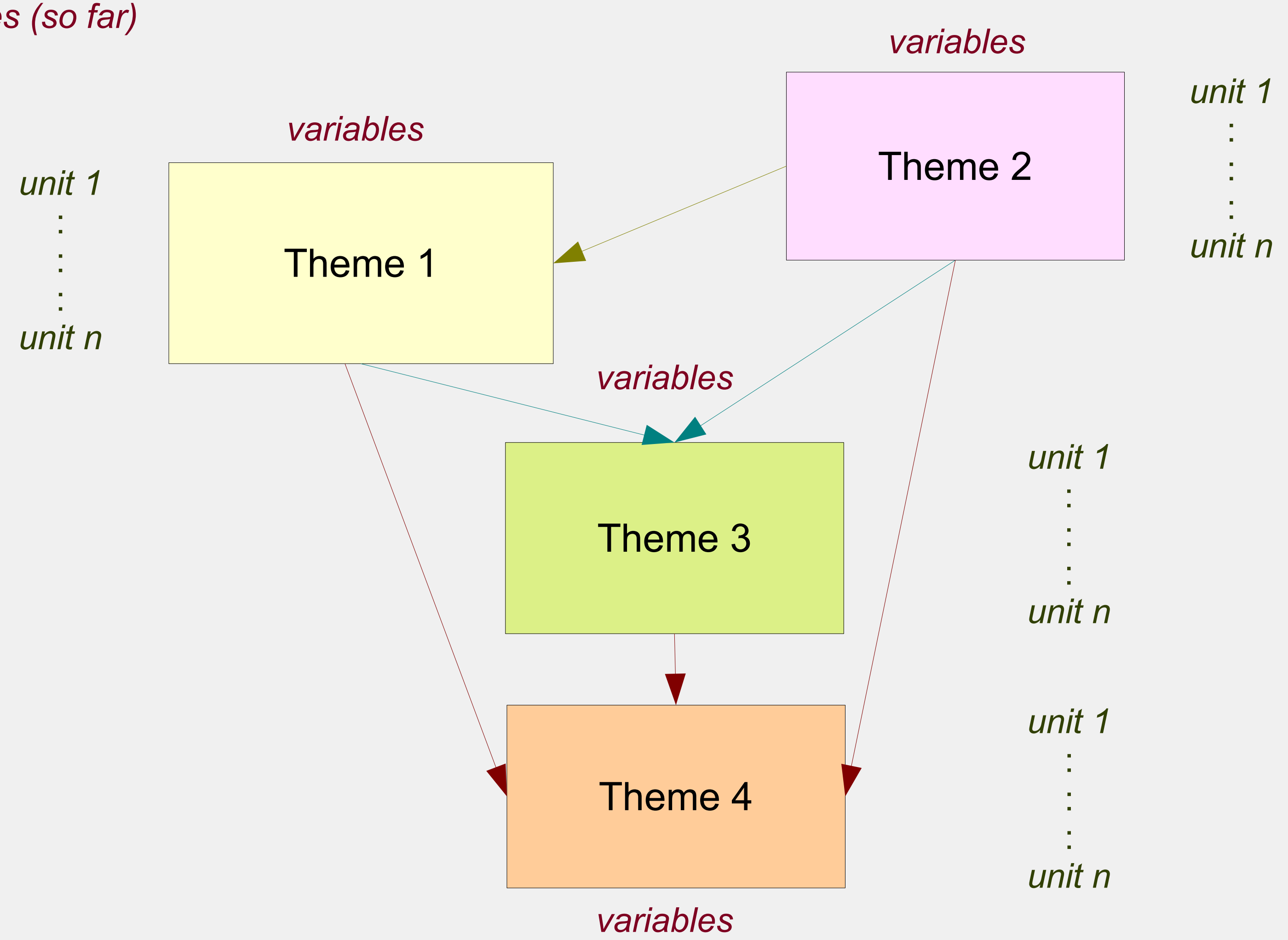


Thematic equation 3

Data and Problem:

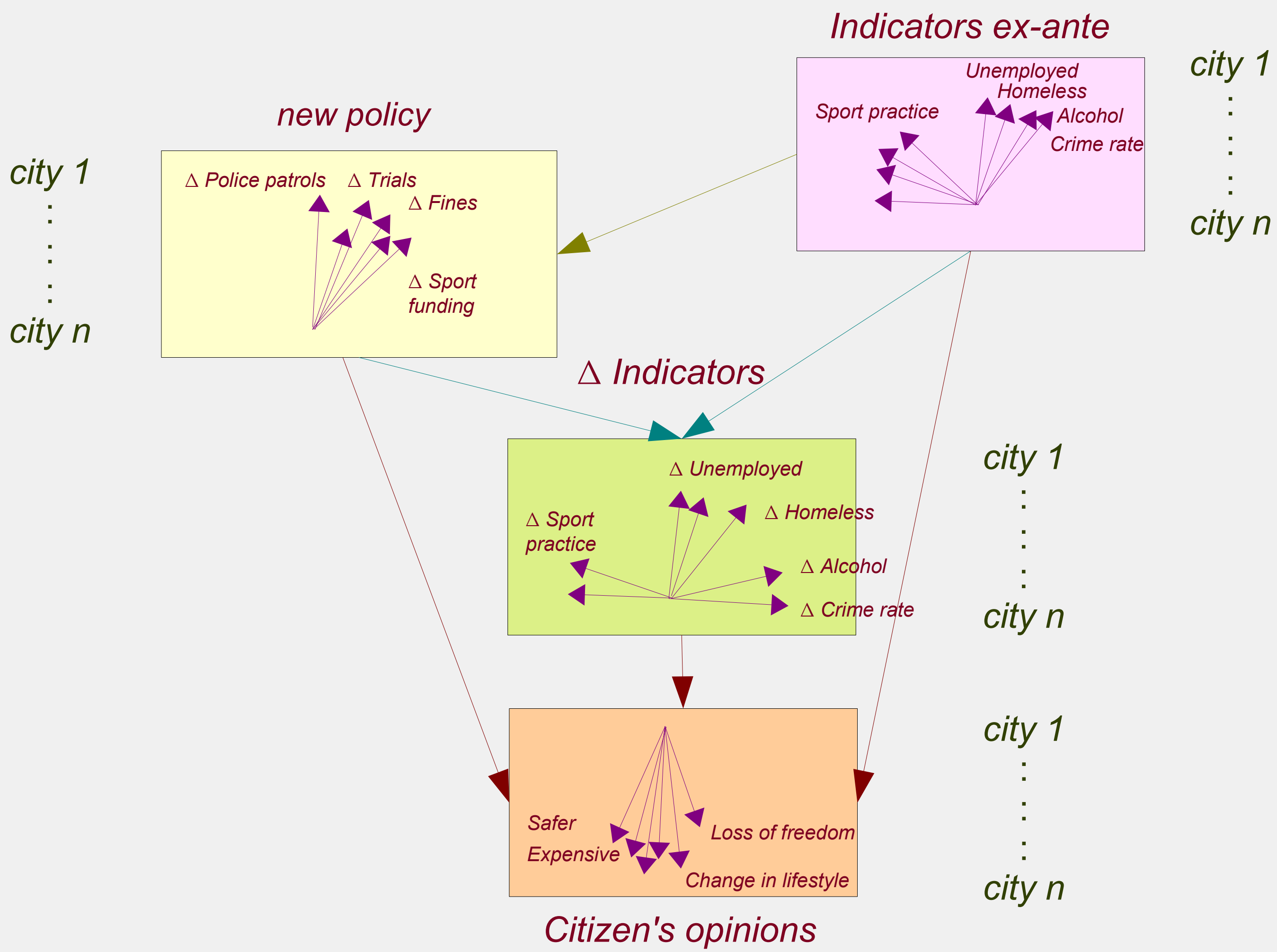
1. Conceptual model of a situation = Thematic Model

Numeric variables (so far)



Data and Problem:

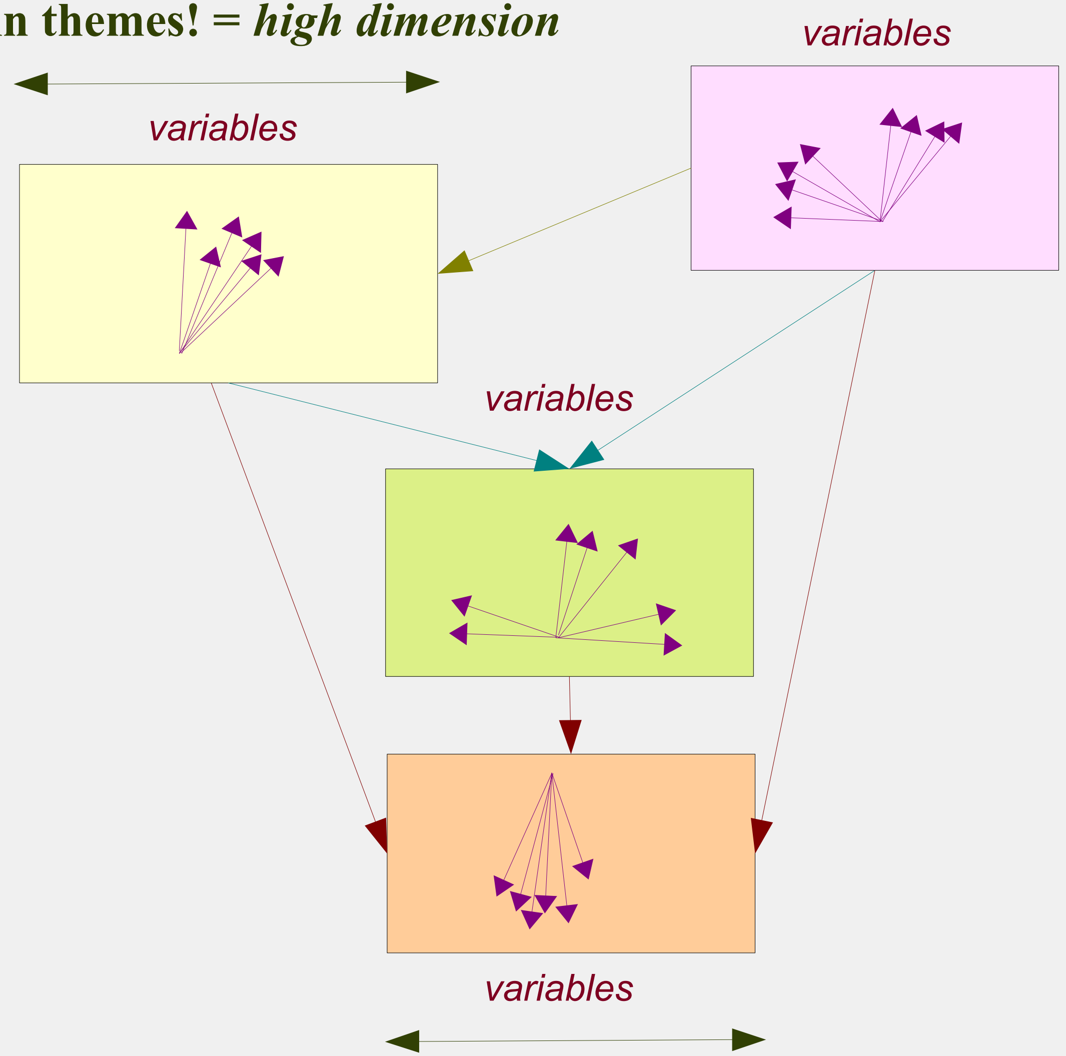
2. The Path-Modelling problem



Data and Problem:

2. The Path-Modelling problem

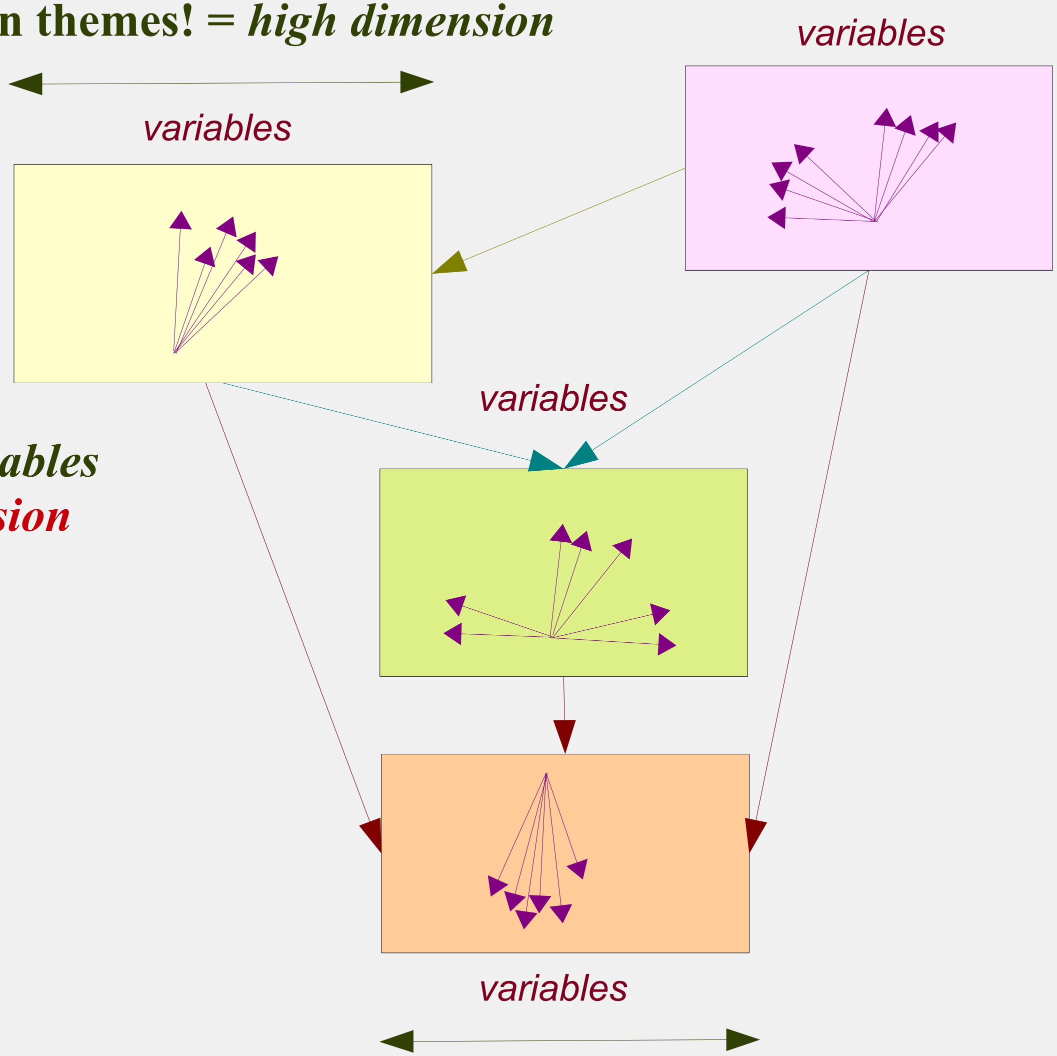
- **Too many variables in themes! =** ***high dimension***



Data and Problem:

2. The Path-Modelling problem

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

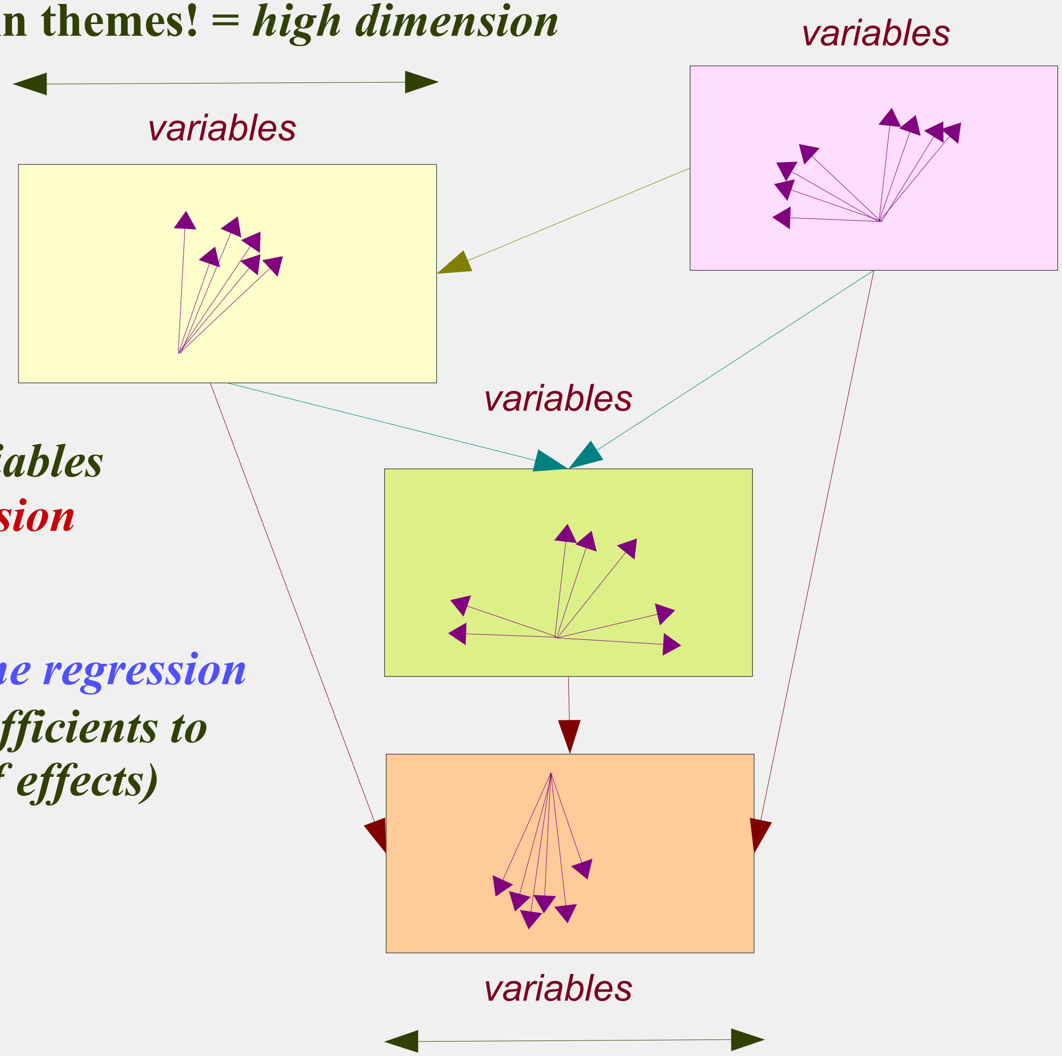


1) *Redundancy of variables*
→ instability of regression model coefficients.

Data and Problem:

2. The Path-Modelling problem

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

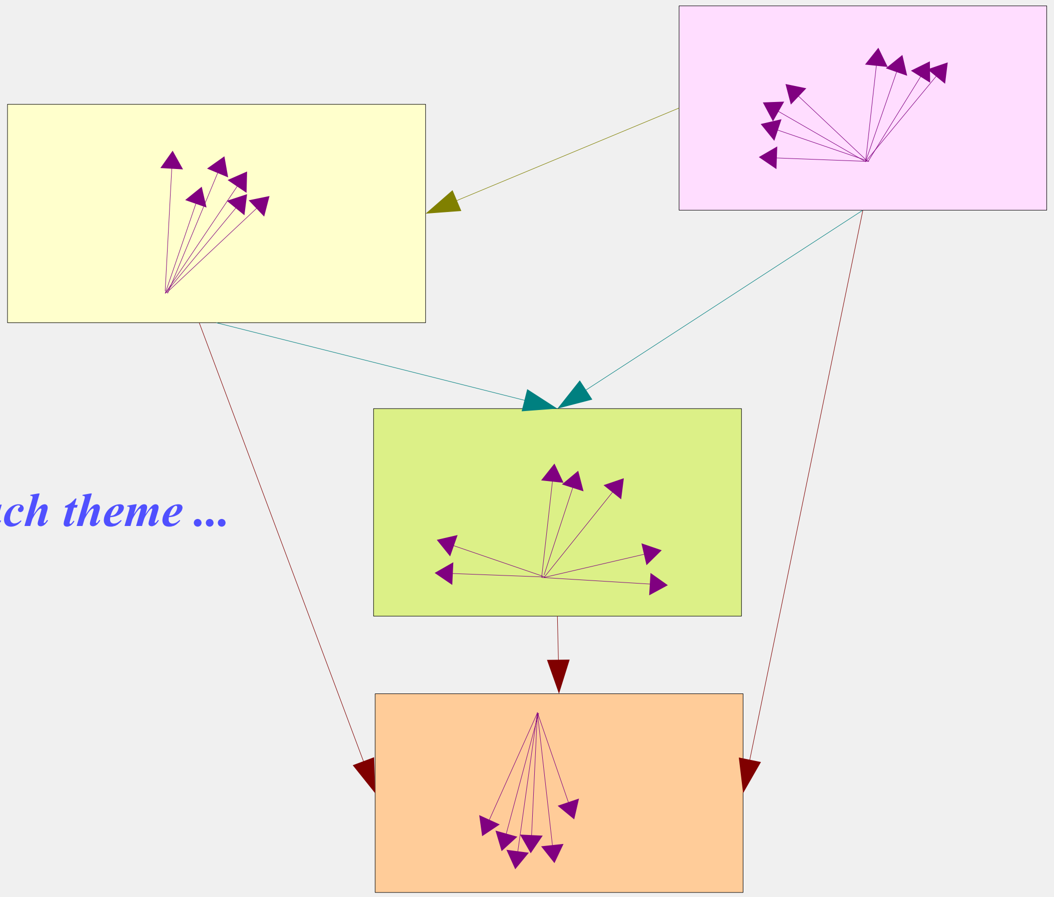


1) *Redundancy of variables*
→ *instability of regression model coefficients.*

⇒ *Regularisation of the regression models (shrinking coefficients to minimise confusion of effects)*

Data and Problem:

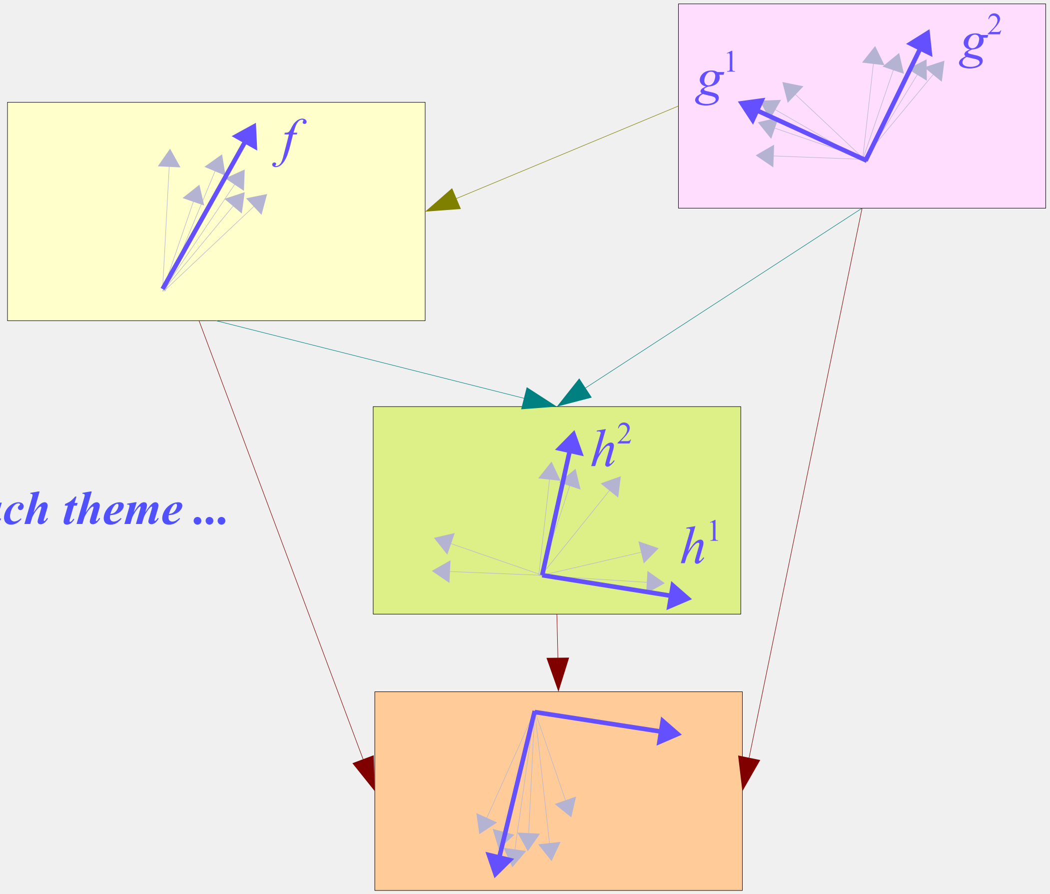
2. The Path-Modelling problem



2) *High dimension*
⇒ *Reduce dimension in each theme ...*

Data and Problem:

2. The Path-Modelling problem

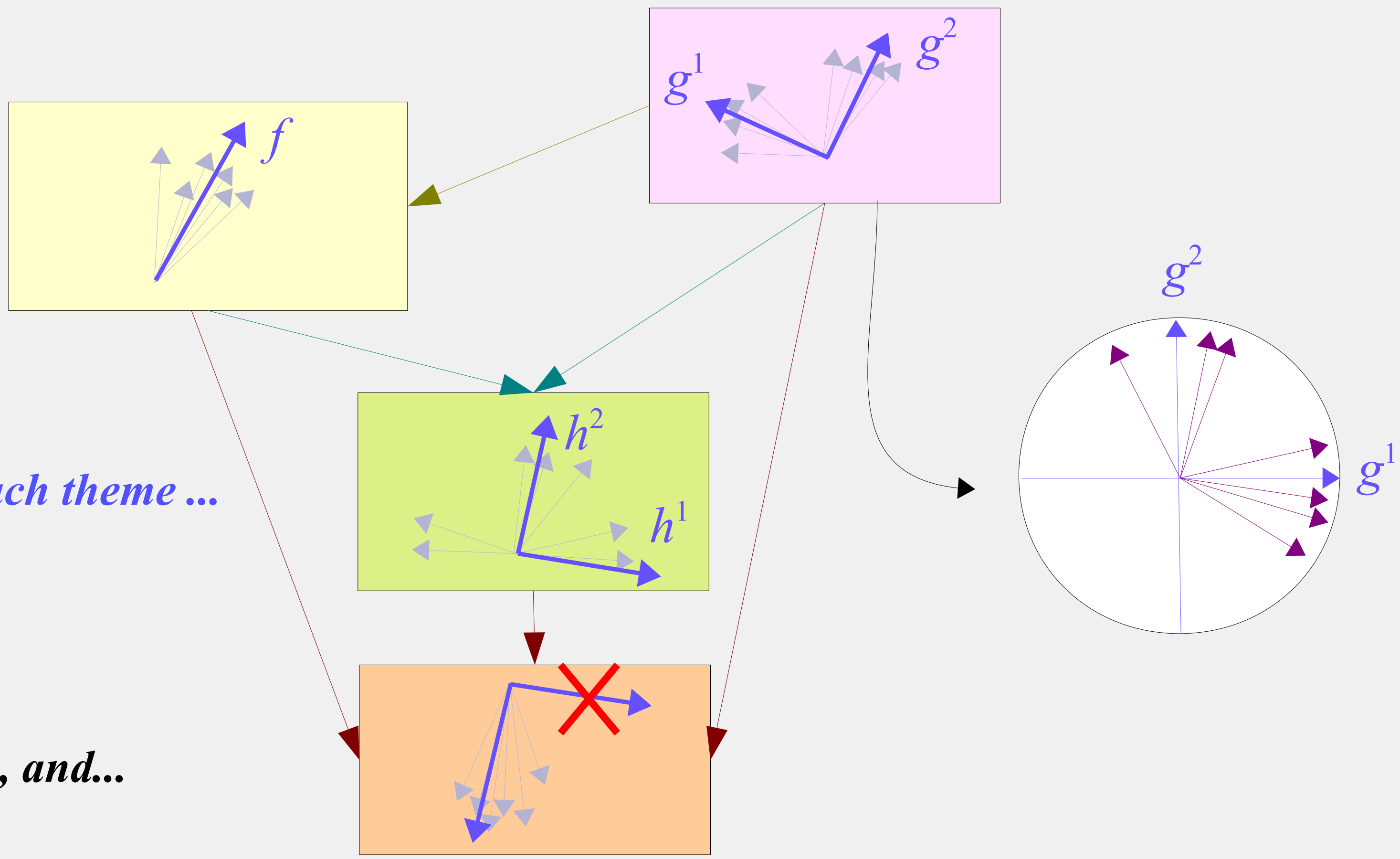


2) ***High dimension***
⇒ ***Reduce dimension in each theme ...***

through a few
Thematic Components

Data and Problem:

2. The Path-Modelling problem



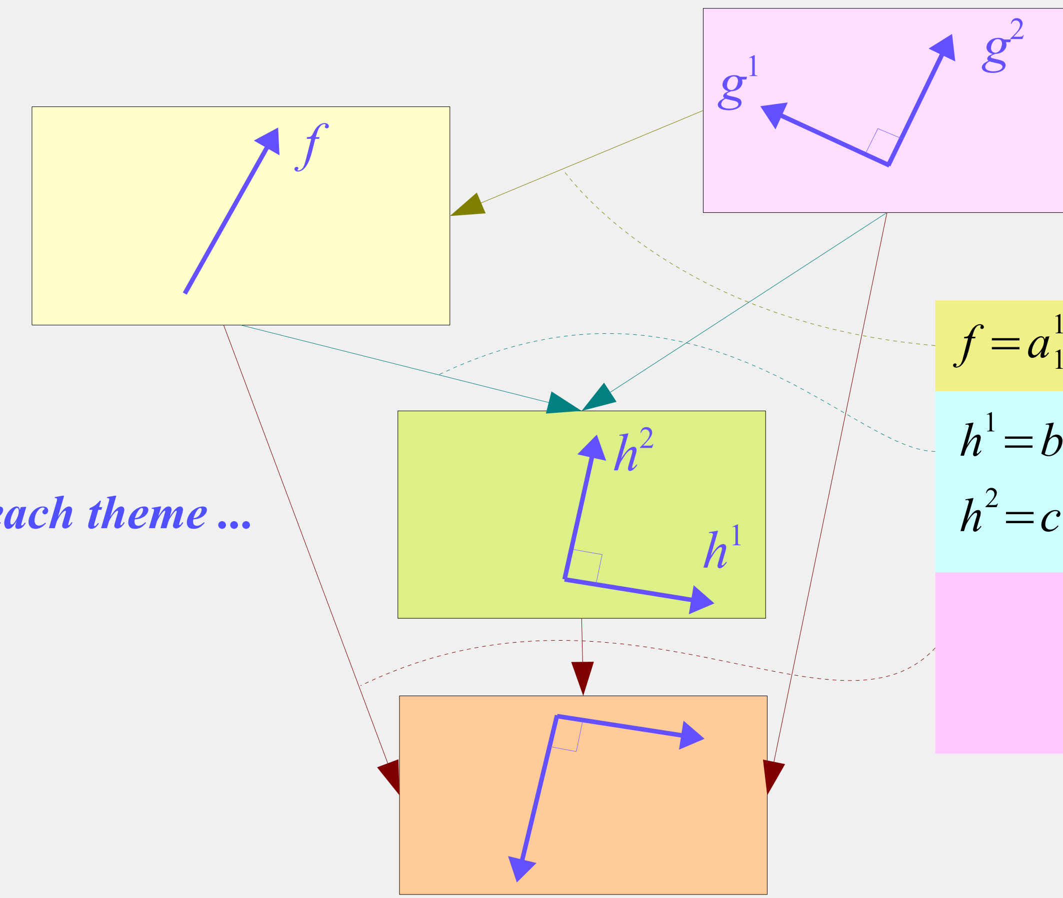
2) ***High dimension***
⇒ *Reduce dimension in each theme ...*

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.

2) *High dimension*
⇒ *Reduce dimension in each theme ...*

through a few
Thematic Components

... *satisfying the model*.

Data and Problem:

2. The Path-Modelling problem

The advantage of *Thematic Components* :

- ***Thematic*** **: a clear conceptual interpretation**

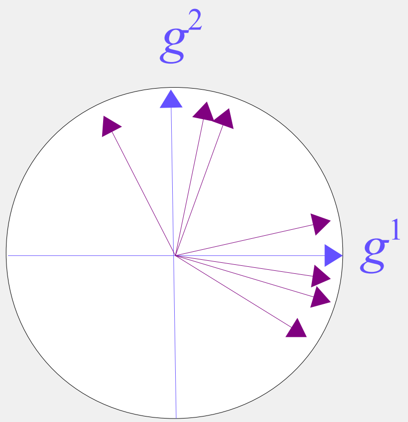
Data and Problem:

2. The Path-Modelling problem

The advantage of *Thematic Components* :

- ***Thematic*** **: a clear conceptual interpretation**
- ***Components*** *= linear combinations of variables estimating a latent variable* **:**

→ closer to the observed variables: easier to interpret substantially



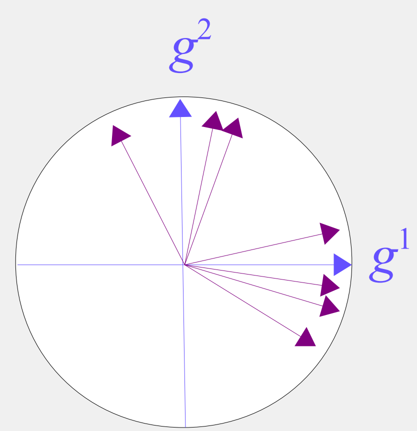
Data and Problem:

2. The Path-Modelling problem

The advantage of ***Thematic Components*** **:**

- ***Thematic*** **: a clear conceptual interpretation**
- ***Components*** *= linear combinations of variables estimating a latent variable* **:**

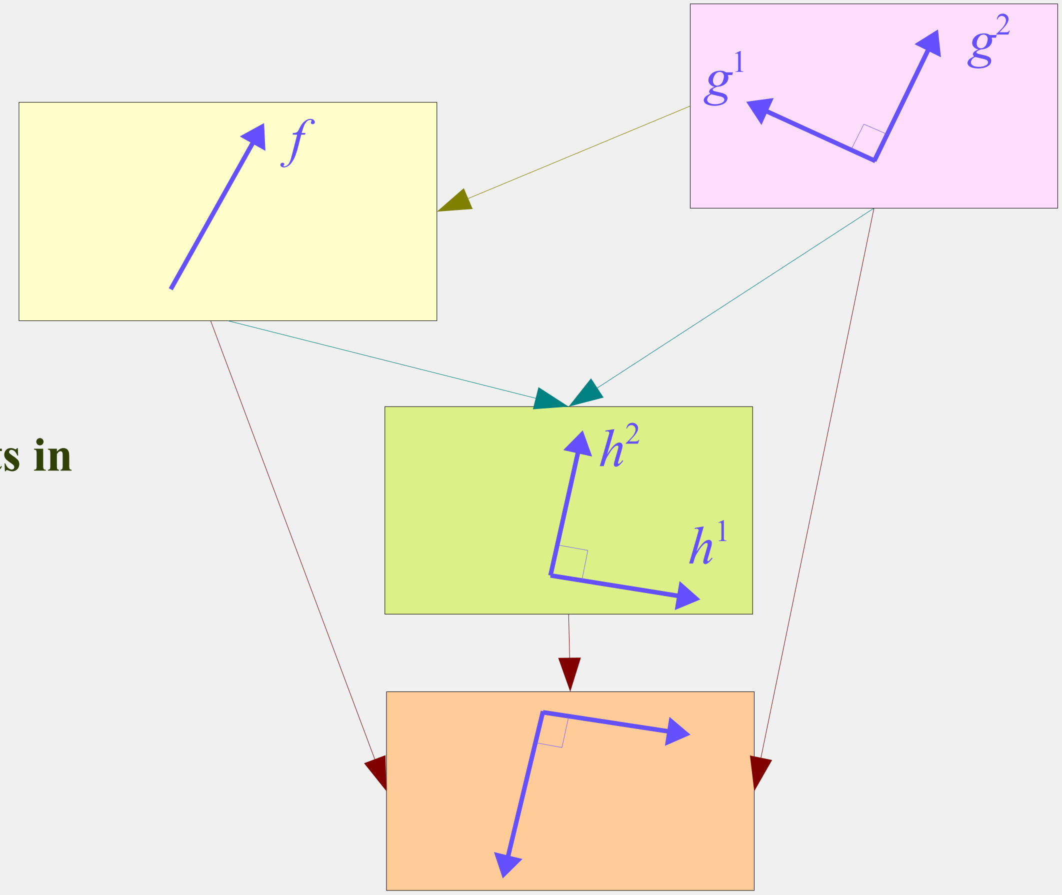
→ closer to the observed variables: easier to interpret substantially



→ possibility to calculate the value of the components from those of the observed variables (contrary to standard SEM estimations)
→ possibility to predict the dependent variables...
... through ***regularised regression models*****!**

Data and Problem:

2. The Path-Modelling problem

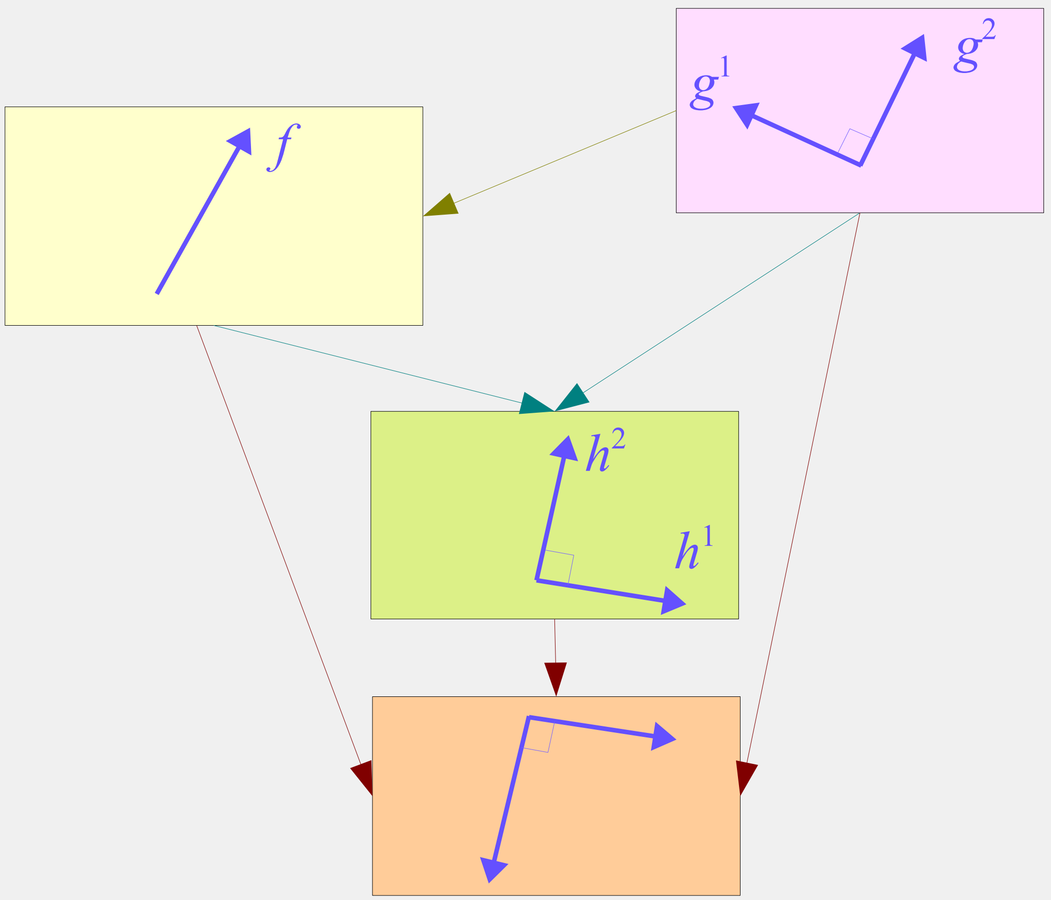


⇒ **How many components in each theme?**

... and which?

Data and Problem:

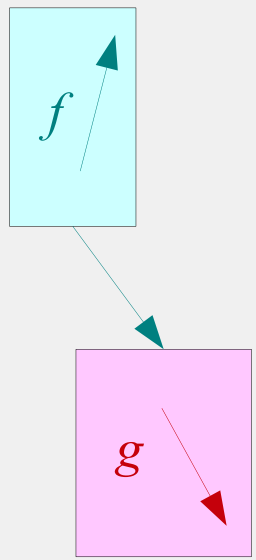
2. The Path-Modelling problem

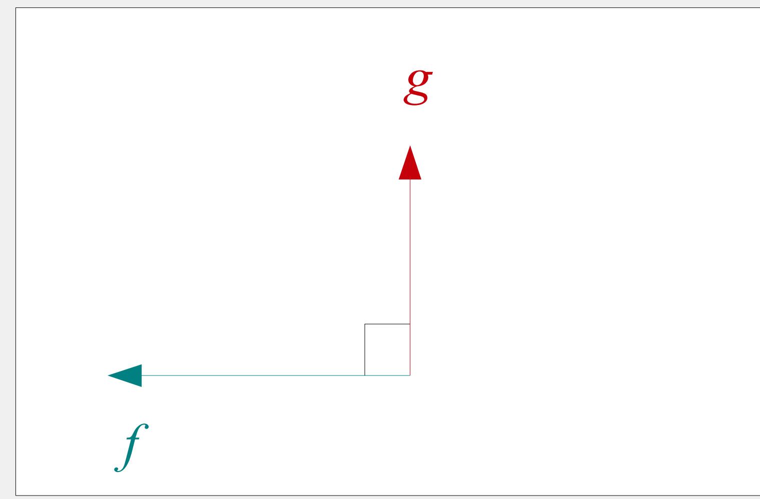


Pb: Every useful component partly depends on all others connected to it, directly or not...

How THEME works

2. The Path-Modelling problem





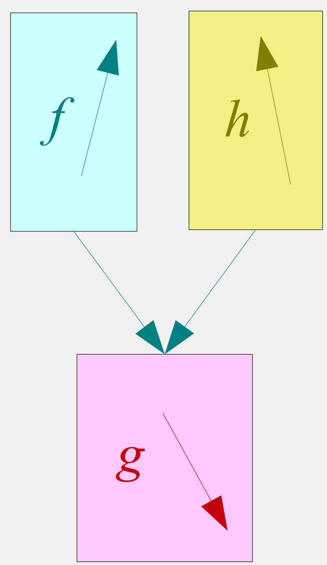
$$\rho^2(g, f) = 0$$

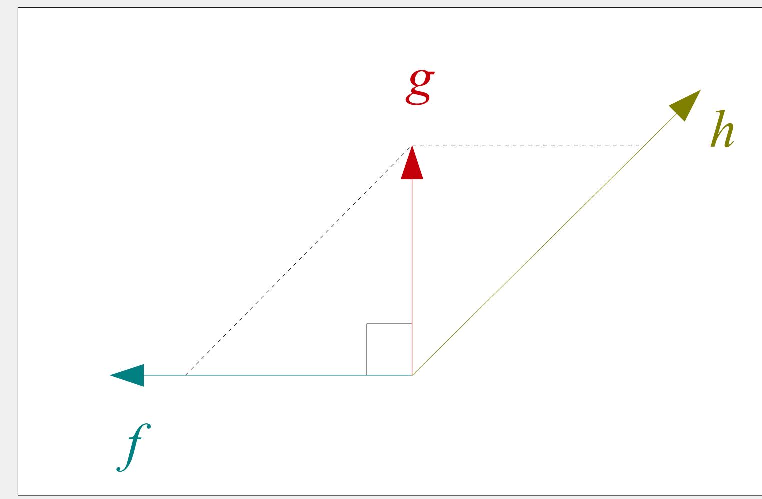
Pb: Every useful component partly depends on *all* others connected to it, directly or not...

No bivariate correlation $\rho(f,g)$

How THEME works

1. Goodness of Fit of the Component Model





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

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

No bivariate correlation $\rho(f,g)$

*Important partial effect of f on g, **conditional on h***

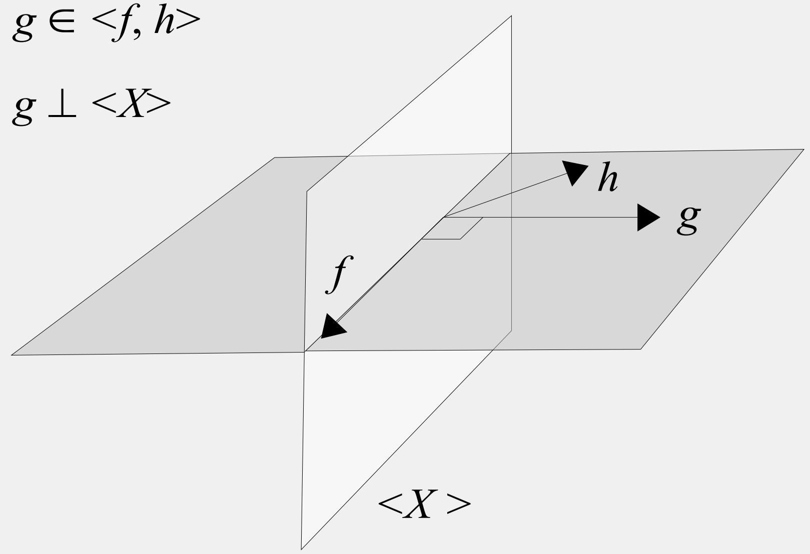
How THEME works

1. Goodness of Fit of the Component Model

A very simple case : $g = af + bh$

$g \in \langle f, h\rangle$

$g \perp \langle X\rangle$



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

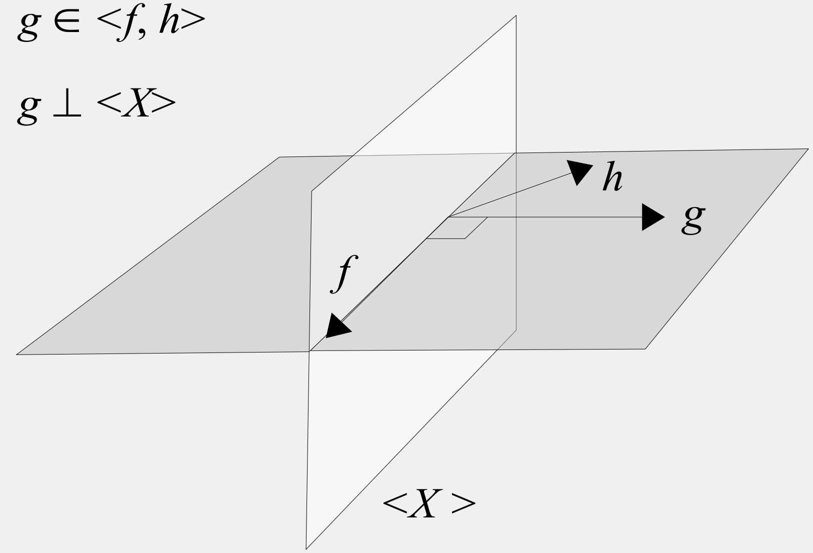
How THEME works

1. Goodness of Fit of the Component Model

A very simple case : $g = af + bh$

$g \in \langle f, h \rangle$

$g \perp \langle X \rangle$



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

A thematic model that should lead to f : $g = \langle X , h \rangle$

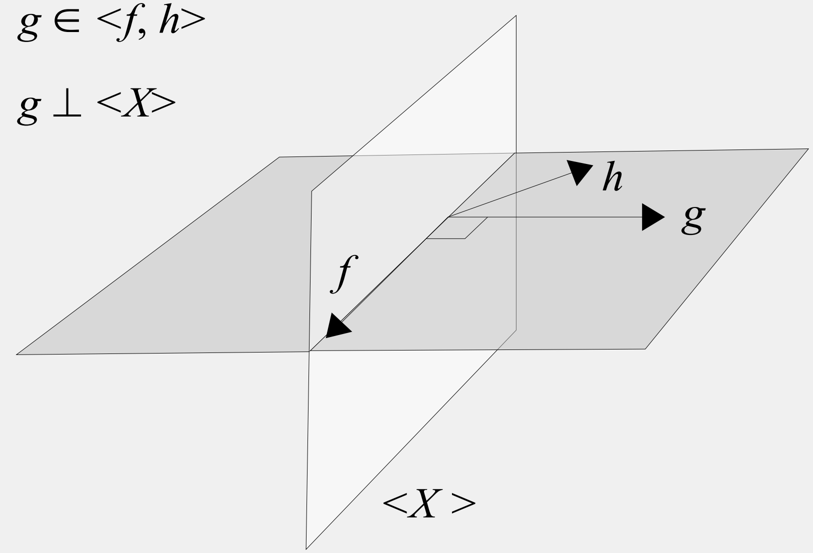
How THEME works

1. Goodness of Fit of the Component Model

A very simple case : $g = af + bh$

$g \in \langle f, h \rangle$

$g \perp \langle X \rangle$



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

Here, f can never be found within $\langle X \rangle$ by means of single bivariate correlation with g.

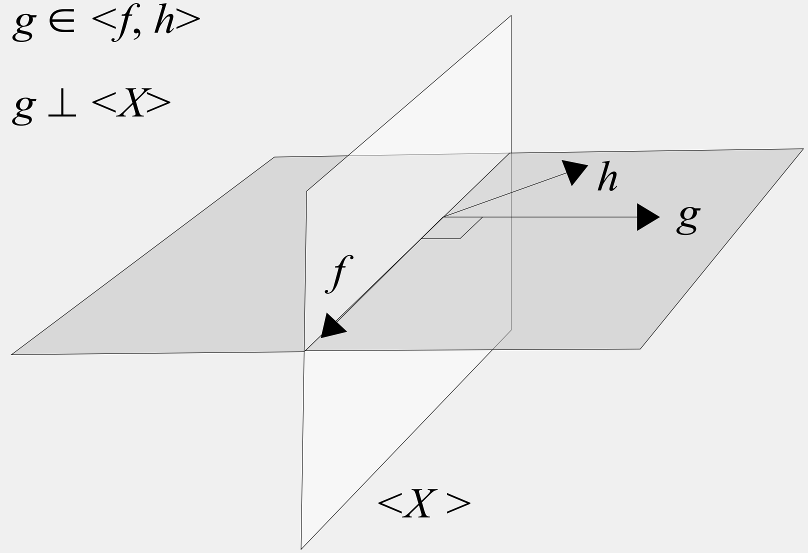
How THEME works

1. Goodness of Fit of the Component Model

A very simple case : $g = af + bh$

$g \in \langle f, h \rangle$

$g \perp \langle X \rangle$



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

*Proper (partial) effects cannot be correctly captured through **bivariate** correlations with the dependent components.*

$\Rightarrow$ *Exeunt: PLS Path-Modeling, Multiblock PLS, RGCCA...*

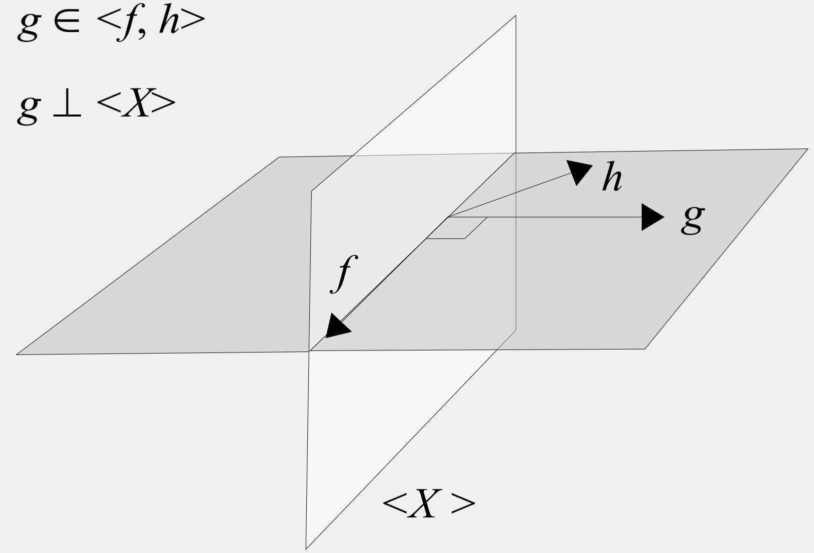
How THEME works

1. Goodness of Fit of the Component Model

A very simple case : $g = af + bh$

$g \in \langle f, h \rangle$

$g \perp \langle X \rangle$



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

*Proper (partial) effects cannot be correctly captured through **bivariate** correlations with the dependent components.*

⇒ *Exeunt: PLS Path-Modeling, Multiblock PLS, RGCCA...*

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

How THEME works

1. Goodness of Fit of the Component Model

... for a single dependent component per equation

- Goodness of fit of one equation:

Equation q : $\mathbf{M}^q : X_{d^q} = \langle \{ X_r ; r \in \mathbf{P}^q \} \rangle$

→ GOF: $\psi_q = g\left(R_q^2\right)$, where g is any positive strictly increasing function

How THEME works

1. Goodness of Fit of the Component Model

... for a single dependent component per equation

- Goodness of fit of one equation:

Equation q : $\mathbf{M}^q : X_{d^q} = \langle \{ X_r ; r \in \mathbf{P}^q \} \rangle$

→ GOF: $\psi_q = g(R_q^2)$, where g is any positive strictly increasing function

- Goodness of fit of the component-model:

$$\psi = g^{-1}\left(\sum_{q=1}^{Q} \varpi_q \psi_q \right)$$

weight of equation q

How THEME works

1. Goodness of Fit of the Component Model

... for a single dependent component per equation

- Goodness of fit of one equation:

Equation q : $\mathbf{M}^q : X_{d^q} = \langle \{ X_r ; r \in \mathbf{P}^q \} \rangle$

→ GOF: $\psi_q = g(R_q^2)$, where g is any positive strictly increasing function

- Goodness of fit of the component-model:

$$\psi = g^{-1}\left(\sum_{q=1}^{Q} \varpi_q \psi_q \right)$$

weight of equation q

Particular cases:

- Arithmetic averaging: $\forall q : \varpi_q = \frac{1}{Q} \; ; \; g = id \quad \Rightarrow \quad \psi = \frac{1}{Q} \sum_q R_q^2$
- Geometric averaging: $\forall q : \varpi_q = \frac{1}{Q} \; ; \; g = \ln \quad \Rightarrow \quad \psi = \left(\prod_q R_q^2 \right)^{\frac{1}{Q}}$

How THEME works

1. Goodness of Fit of the Component Model

... for any number of dependent components per equation

- Goodness of fit of one equation:

Equation q : $\mathbf{M}^q : X_{d^q} = \langle \{ X_r ; r \in \mathbf{P}^q \} \rangle$

$F_{d^q} = [f^1_{d^q}, \ldots, f^{K_{d^q}}_{d^q}]$ = dependent components of eq. q

$F_{\mathbf{P}^q}$ = all explanatory components of eq. q

How THEME works

1. Goodness of Fit of the Component Model

... for any number of dependent components per equation

- Goodness of fit of one equation:

Equation q : $\mathbf{M}^q : X_{d^q} = \langle \{ X_r ; r \in \mathbf{P}^q \} \rangle$

$F_{d^q} = [f^1_{d^q}, \ldots, f^{K_{d^q}}_{d^q}]$ = dependent components of eq. q

$F_{\mathbf{P}^q}$ = all explanatory components of eq. q

→ GOF: $$\psi_q = g\left(h^{-1}\left(\sum_{k=1,\ldots K_{d^q}} h\left(R^2\left(f^k_{d^q} \middle| F_{\mathbf{P}^q} \right) \right) \right) \right)$$

How THEME works

1. Goodness of Fit of the Component Model

... for any number of dependent components per equation

- Goodness of fit of one equation:

Equation q : $\mathbf{M}^q : X_{d^q} = \langle \{ X_r ; r \in \mathbf{P}^q \} \rangle$

$F_{d^q} = [f^1_{d^q}, \ldots, f^{K_{d^q}}_{d^q}]$ = dependent components of eq. q

$F_{\mathbf{P}^q}$ = all explanatory components of eq. q

GOF: $$\psi_q = g\left(h^{-1}\left(\sum_{k=1,\ldots K_{d^q}} h\left(R^2\left(f^k_{d^q} | F_{\mathbf{P}^q} \right) \right) \right) \right)$$

Particular case: when $h = id$: $\psi_q = g\left(tr\left(\Pi_{F_{d^q}} \Pi_{\langle F_{\mathbf{P}^q} \rangle} \right) \right)$

Classic link between sub-spaces

How THEME works

1. Goodness of Fit of the Component Model

... for any number of dependent components per equation

- Goodness of fit of one equation:

Equation q : $\mathbf{M}^q : X_{d^q} = \langle \{ X_r ; r \in \mathbf{P}^q \} \rangle$

$F_{d^q} = [f^1_{d^q}, \ldots, f^{K_{d^q}}_{d^q}]$ = dependent components of eq. q

$F_{\mathbf{P}^q}$ = all explanatory components of eq. q

→ GOF: $$\psi_q = g\left(h^{-1}\left(\sum_{k=1,\ldots K_{d^q}} h\left(R^2\left(f^k_{d^q} \middle| F_{\mathbf{P}^q}\right)\right)\right)\right)$$

Particular case: when $h = id$: $\psi_q = g\left(tr\left(\Pi_{F_{d^q}} \Pi_{\langle F_{\mathbf{P}^q} \rangle}\right)\right)$

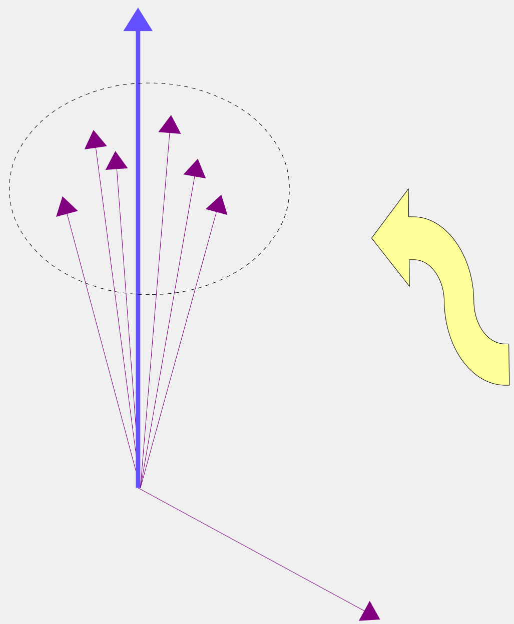
Classic link between sub-spaces

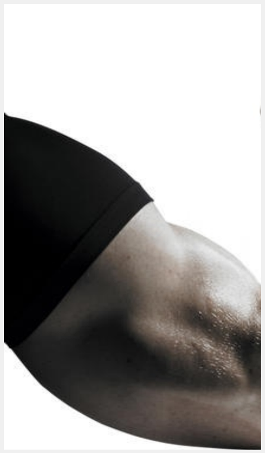
- Goodness of fit of the component-model: $$\psi = g^{-1}\left(\sum_{q=1}^{Q} \varpi_q \psi_q\right)$$

How THEME works

2. Structural relevance of components

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

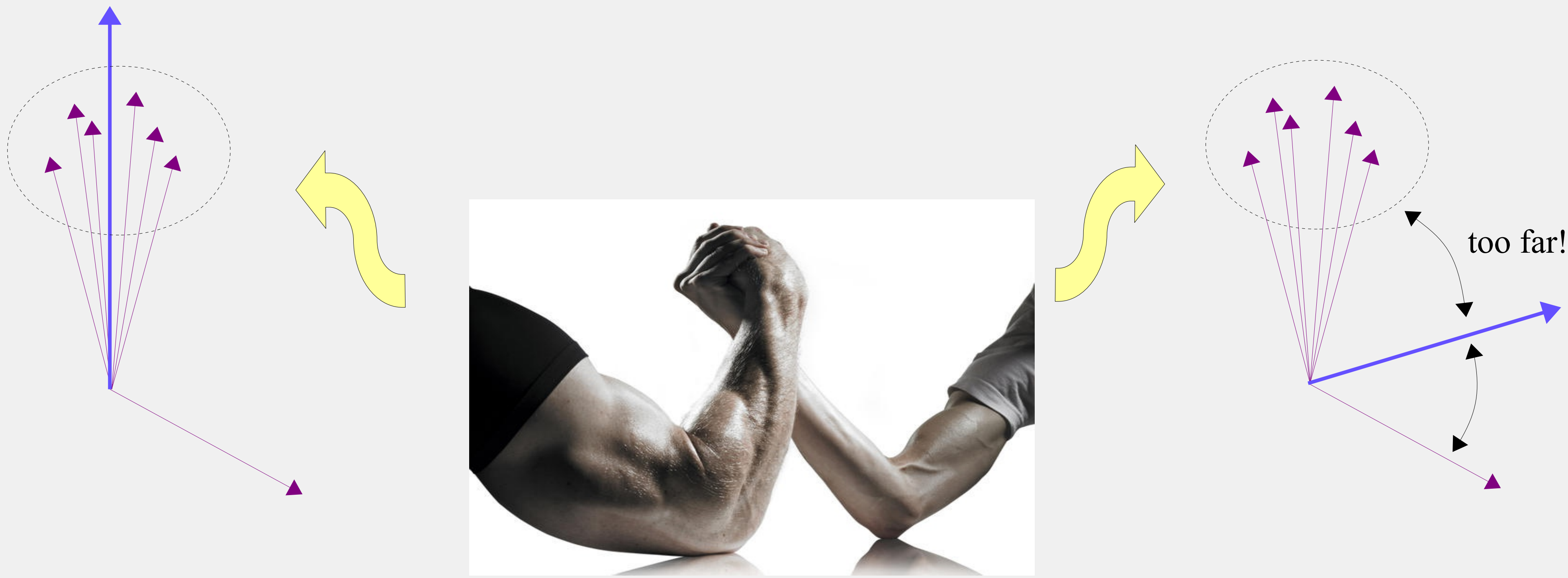




How THEME works

2. Structural relevance of components

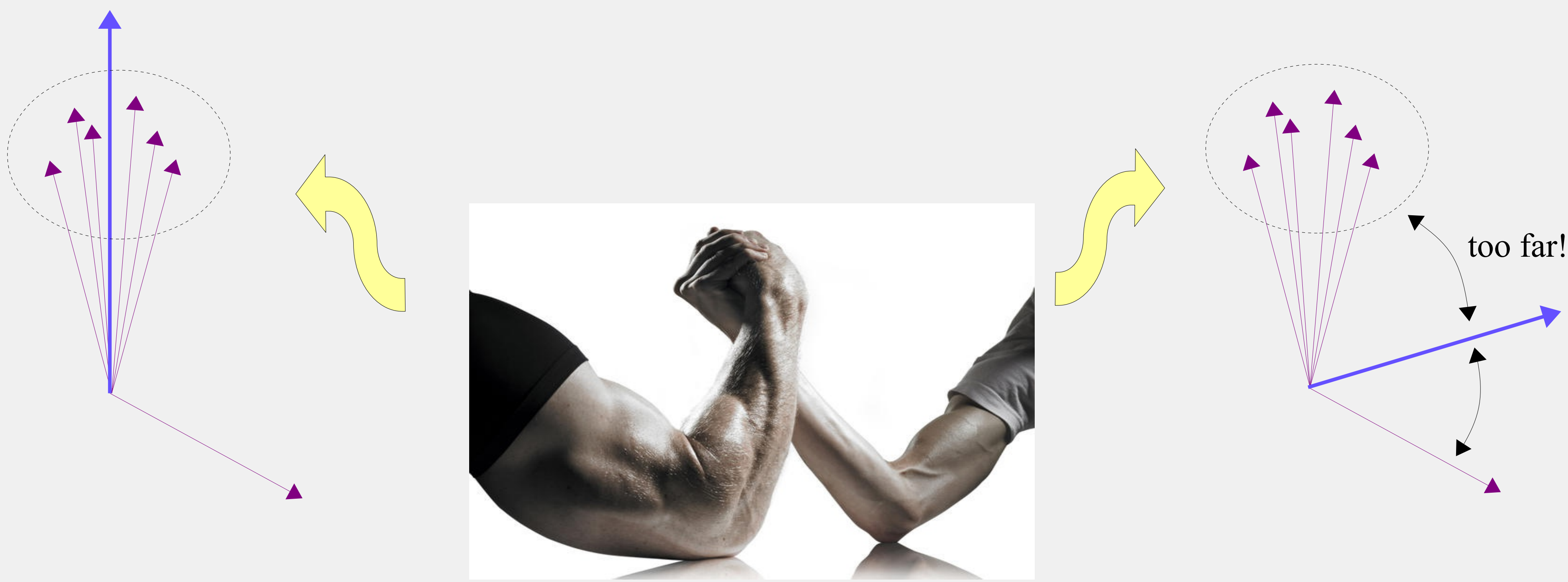
To be interpretable, components must be *structurally strong*, e.g. close to *observed variables bundles*



How THEME works

2. Structural relevance of components

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



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

How THEME works

2. Structural relevance of components

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

- *Component in a theme* X : $f = Xv$
- *Identification / regularisation constraint* : $v' M^{-1} v = 1$

with $M^{-1} = \tau A^{-1} + (1-\tau) X' W X$

where A is such that PCA of (X,A,W) is relevant to X's data

→ • ***The Structural Relevance Indicator:***

$$\phi(v) := \left(\sum_{j=1}^{J} \omega_j (v' N_j v)^l \right)^{\frac{1}{l}} \quad \text{s.t. constraint} \quad v' M^{-1} v = 1$$

weights

N_j's code the directions components should focus on

Structural Relevance of components

2. Structural relevance of components

- **Purpose of N_j's = ?**

$$\phi(v) := \left(\sum_{j=1}^{J} \omega_j (v' N_j v)^l \right)^{\frac{1}{l}}$$

The N_j's are coding directions of concern

Examples:

- Component's variance: $\phi(v) = V(f) = \| X v \|_W^2 = v' (X' W X) v$

$\| v \|^2 = 1 \ \Rightarrow \ M = I$

$\rightarrow$ directions of discrepancy of observations

Structural Relevance of components

2. Structural relevance of components

- **Purpose of N_j's = ?**

$$\phi(v) := \left(\sum_{j=1}^{J} \omega_j (v' N_j v)^l \right)^{\frac{1}{l}}$$

The N_j's are coding directions of concern

Examples:

- Component's variance: $\phi(v) = V(f) = \|X v\|_W^2 = v'(X'WX)v$

$$\|v\|^2 = 1 \;\Rightarrow\; M = I$$

→ directions of discrepancy of observations

- Variable Powered Inertia: $\phi(v) = \left(\sum_{j=1}^{p} \omega_j \rho^{2l}(f, x^j) \right)^{\frac{1}{l}}$

locality parameter

$$= \left(\sum_{j=1}^{p} \omega_j (v' \underbrace{X' W x^j x^{j\,\prime} W X}_{N_j} v)^l \right)^{\frac{1}{l}}$$

$$\|f\|_W^2 = 1 \;\Rightarrow\; M = (X'WX)^{-1}$$

→ directions of observed variables.

Structural Relevance of components

2. Structural relevance of components

- **Purpose of N_j's = ?**

$$\phi(v) := \left(\sum_{j=1}^{J} \omega_j (v' N_j v)^l \right)^{\frac{1}{l}}$$

The N_j's are coding directions of concern

Examples:

Variable Powered Inertia can be extended to:

- Variable Powered Covariance: $\phi(v) = \left(\sum_{j=1}^{p} \omega_j \langle f | x^j \rangle_W^{2l} \right)^{\frac{1}{l}}$

locality parameter

$$= \left(\sum_{j=1}^{p} \omega_j (v' \underbrace{X' W x^j x^{j\prime} W X}_{N_j} v)^l \right)^{\frac{1}{l}}$$

$$M^{-1} = \tau A^{-1} + (1-\tau)(X'WX)$$

Regularisation parameter

Matrix suitable for the PCA of X

Structural Relevance of components

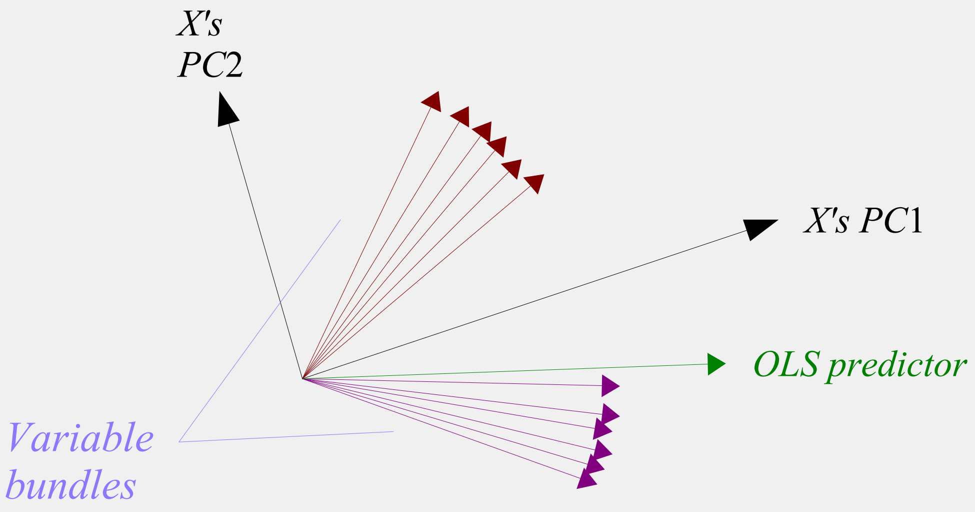
2. Structural relevance of components

- **Purpose of *l* = ?**

$$\phi(v) := \left(\sum_{j=1}^{J} \omega_j (v' N_j v)^l \right)^{\frac{1}{l}}$$

A disturbing phenomenon when regularising regression in the PLS way...

What happens:



Structural Relevance of components

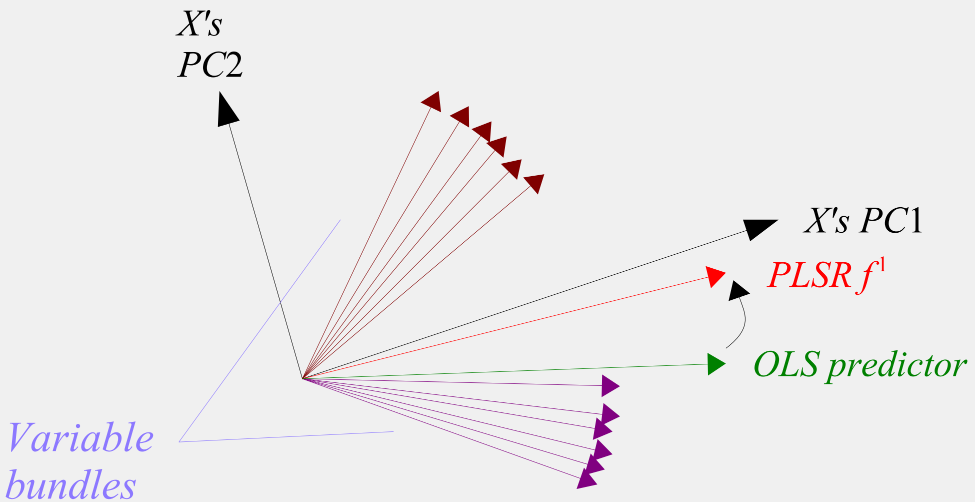
2. Structural relevance of components

- **Purpose of *l* = ?**

$$\phi(v) := \left(\sum_{j=1}^{J} \omega_j (v' N_j v)^l \right)^{\frac{1}{l}}$$

A disturbing phenomenon when regularising regression in the PLS way...

What happens:



A PC is too ***global*** a direction to fit bundle structures.

 We must go beyond component variance.

Structural Relevance of components

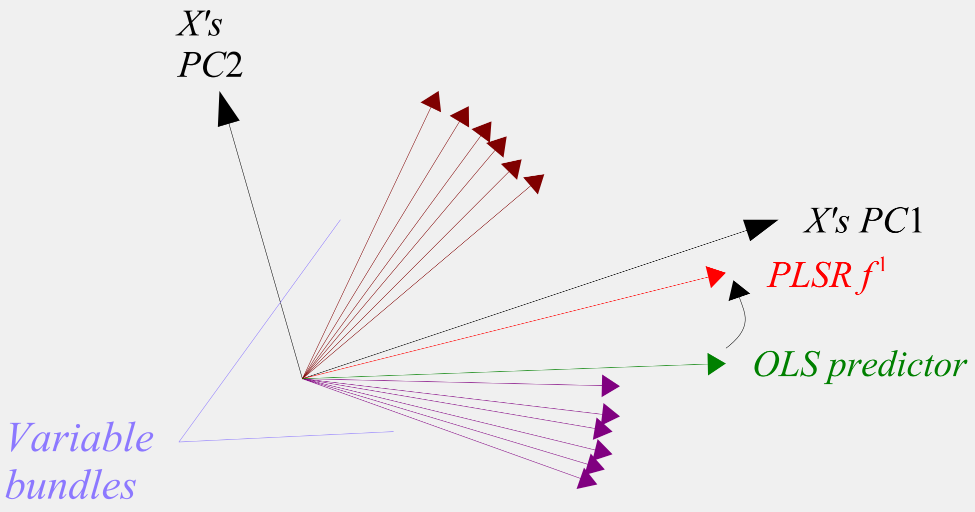
2. Structural relevance of components

- **Purpose of *l* = ?**

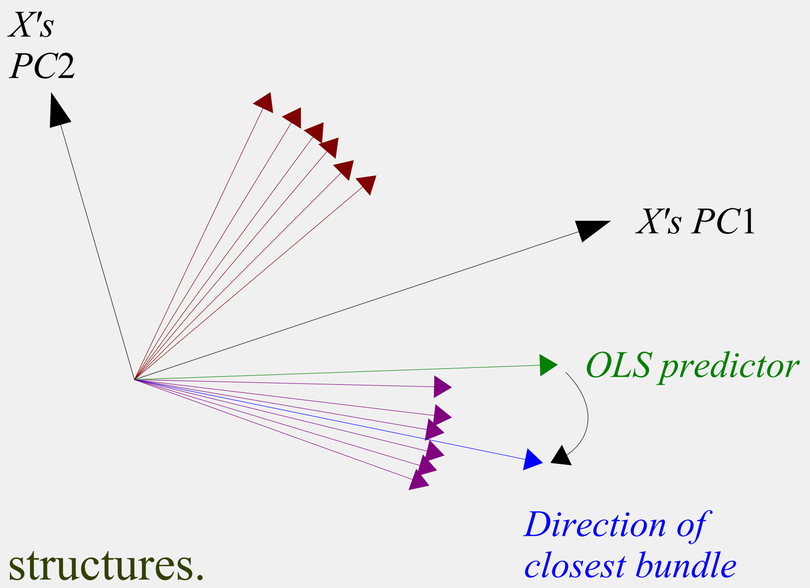
$$\phi(v) := \left(\sum_{j=1}^{J} \omega_j (v' N_j v)^l \right)^{\frac{1}{l}}$$

A disturbing phenomenon when regularising regression in the PLS way...

What happens:



What we would like to happen :



A PC is too ***global*** a direction to fit bundle structures.

□ We must go beyond component variance.

Structural Relevance of components

2. Structural relevance of components

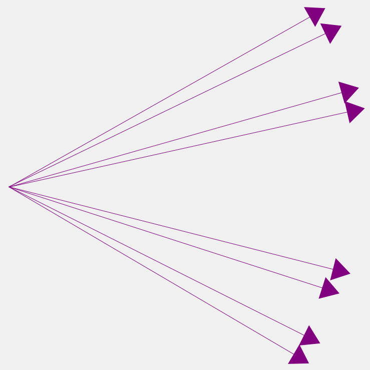
- **Purpose of *l* = ?**

$$\phi(v) := \left(\sum_{j=1}^{J} \omega_j (v' N_j v)^l \right)^{\frac{1}{l}}$$

l : tunes the "locality" of the bundles of directions to focus on

locality = ± the "width" of the bundles of directions considered structurally interesting.

Had this set of directions rather be considered...



How THEME works

2. Structural relevance of components

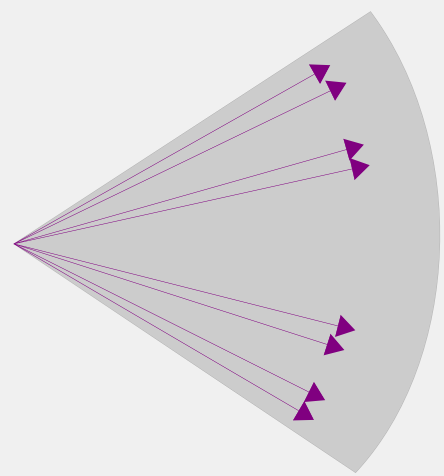
- **Purpose of *l* = ?**

$$\phi(v) := \left(\sum_{j=1}^{J} \omega_j (v' N_j v)^l \right)^{\frac{1}{l}}$$

l : tunes the “locality” of the bundles of directions to focus on

locality = ± the “width” of the bundles of directions considered structurally interesting.

Had this set of directions rather be considered...



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

How THEME works

2. Structural relevance of components

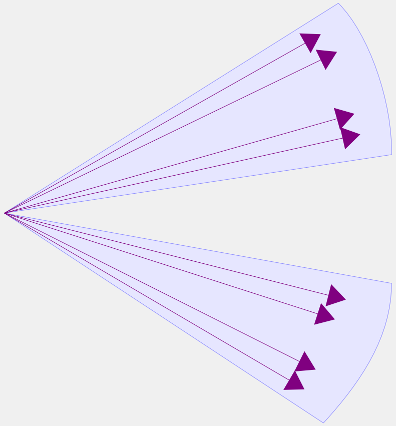
- **Purpose of *l* = ?**

$$\phi(v) := \left(\sum_{j=1}^{J} \omega_j (v' N_j v)^l \right)^{\frac{1}{l}}$$

l : tunes the "locality" of the bundles of directions to focus on

locality = ± the "width" of the bundles of directions considered structurally interesting.

Had this set of directions rather be considered...



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

How THEME works

2. Structural relevance of components

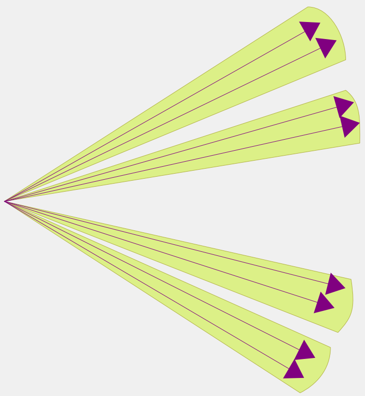
- **Purpose of l = ?**

$$\phi(v) := \left(\sum_{j=1}^{J} \omega_j (v' N_j v)^l \right)^{\frac{1}{l}}$$

l : tunes the "locality" of the bundles of directions to focus on

locality = ± the "width" of the bundles of directions considered structurally interesting.

Had this set of directions rather be considered...



... four bundles? (l ↑↑)

How THEME works

2. Structural relevance of components

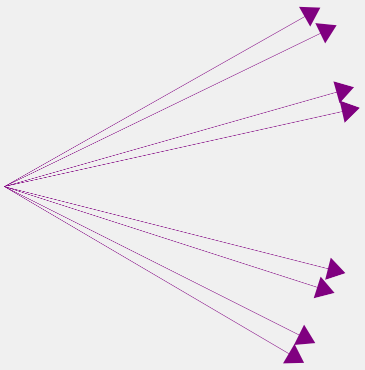
- **Purpose of *l* = ?**

$$\phi(v) := \left(\sum_{j=1}^{J} \omega_j (v' N_j v)^l \right)^{\frac{1}{l}}$$

l : tunes the “locality” of the bundles of directions to focus on

locality = ± the “width” of the bundles of directions considered structurally interesting.

Had this set of directions rather be considered...



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

How THEME works

2. Structural relevance of components

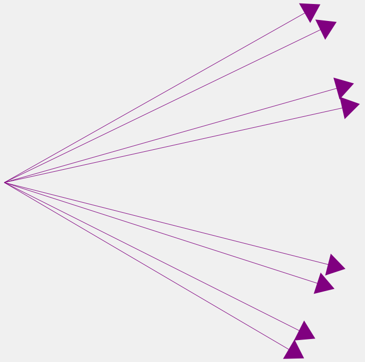
- **Purpose of l = ?**

$$\phi(v) := \left(\sum_{j=1}^{J} \omega_j (v' N_j v)^l \right)^{\frac{1}{l}}$$

l : tunes the "locality" of the bundles of directions to focus on

locality = ± the "width" of the bundles of directions considered structurally interesting.

Had this set of directions rather be considered...



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

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

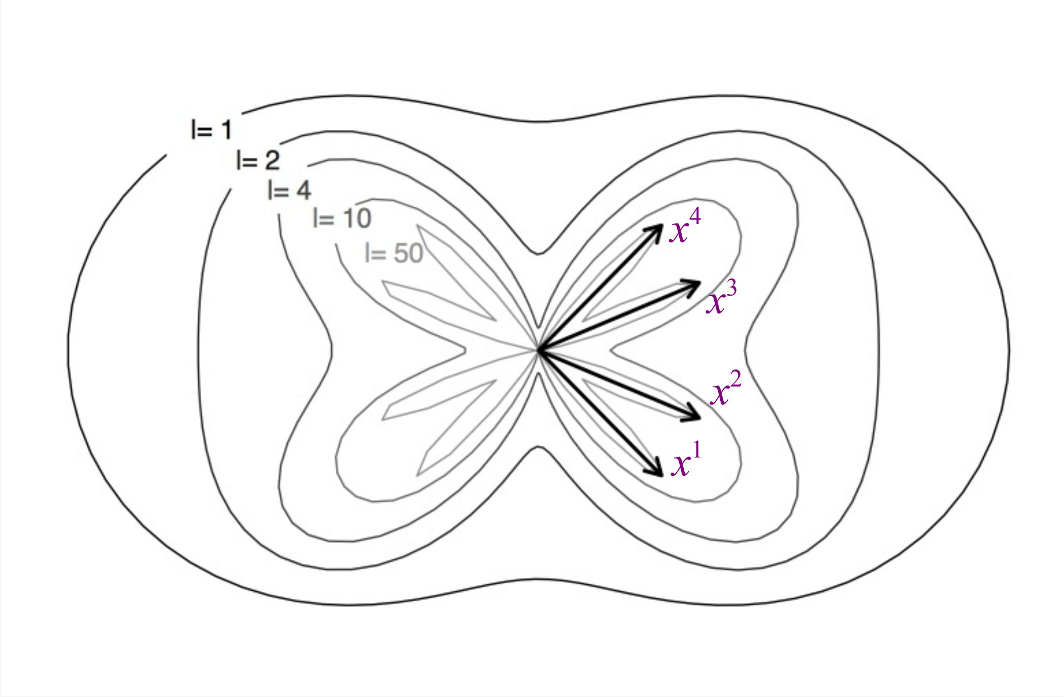
How THEME works

2. Structural relevance of components

l : tunes the "locality" of the bundles of directions to focus on

Example: 4 variables in a plane...

- VPI: $\phi_X^l(v)$ plotted in polar coordinates:



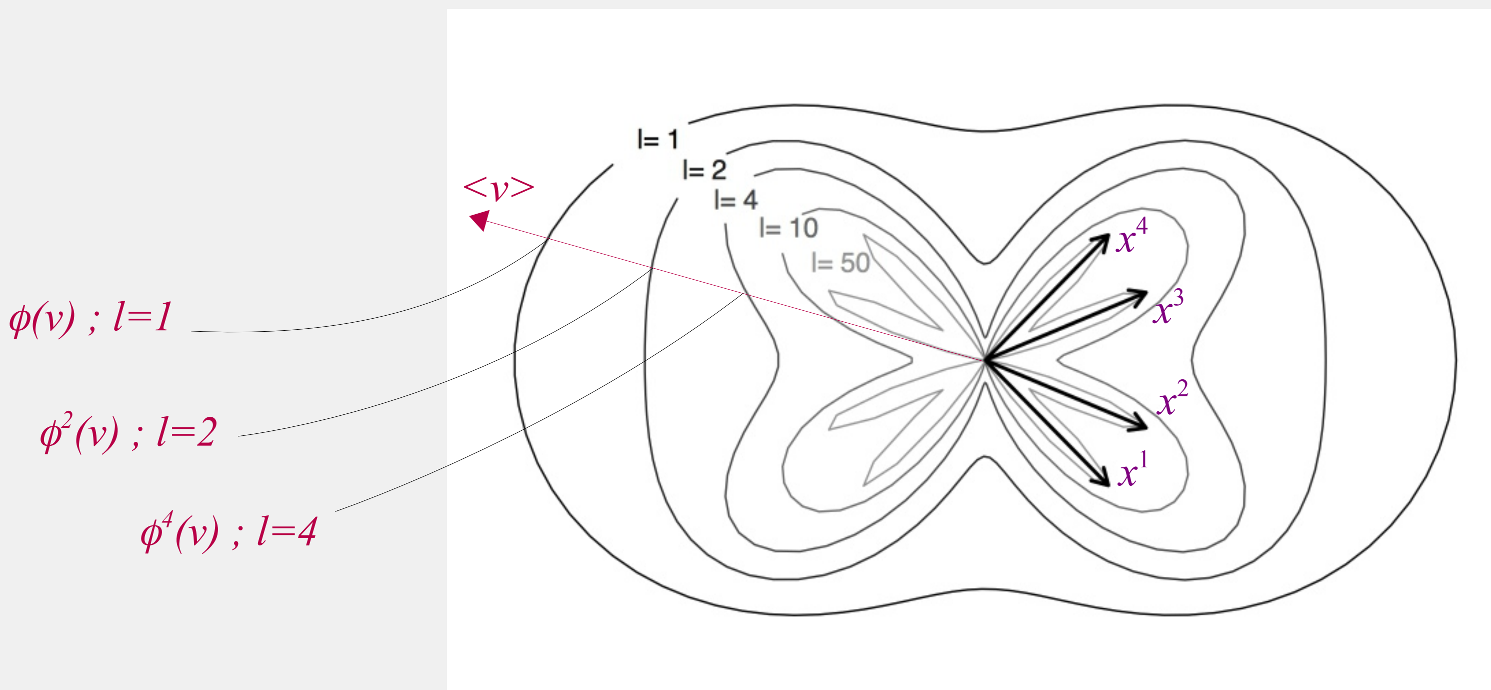
How THEME works

2. Structural relevance of components

l : tunes the "locality" of the bundles of directions to focus on

Example: 4 variables in a plane...

- VPI: $\phi^{l}_{X}(v)$ plotted in polar coordinates:



How THEME works

3. Combining goodness of fit ψ and structural relevance ϕ

The criterion to be maximised by a component $f = Xv$, *given ALL others*, must have the form:

$$\boxed{\psi(v)^{1-s}}\,\boxed{\phi(v)^{s}}$$

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

How THEME works

3. Combining goodness of fit ψ *and structural relevance* ϕ

The criterion to be maximised by a component $f = Xv$, *given ALL others*, must have the form:

$$\boxed{\psi(v)^{1-s}}\,\boxed{\phi(v)}^{s}$$

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

$$\nabla \ln\left(\psi(f)\phi(f)^{s}\right)=0 \quad \Leftrightarrow \quad \frac{\nabla\psi(f)}{\psi(f)} = -\frac{s}{1-s}\frac{\nabla\phi(f)}{\phi(f)}$$

+1% on ϕ is compensated by -s/(1-s)% on ψ

Relative variations compensate at optimum

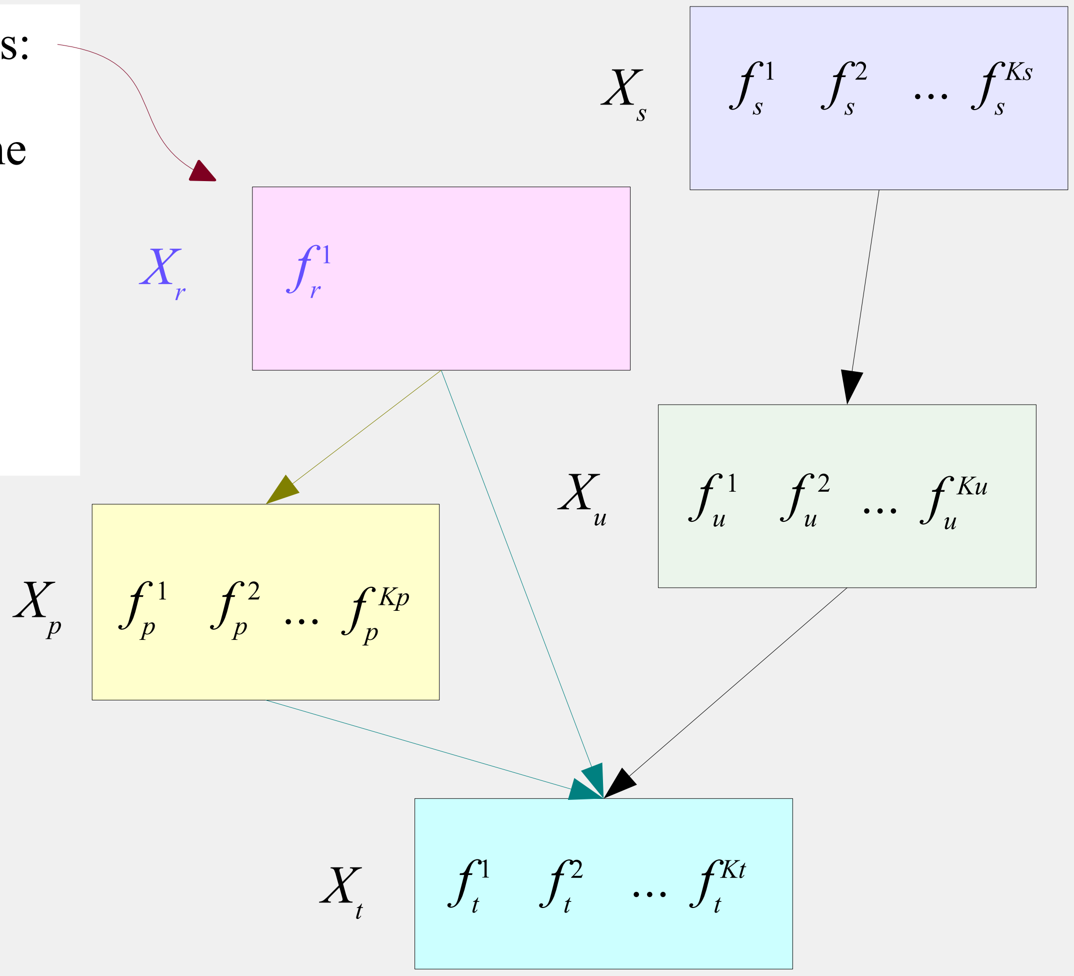
How THEME works

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

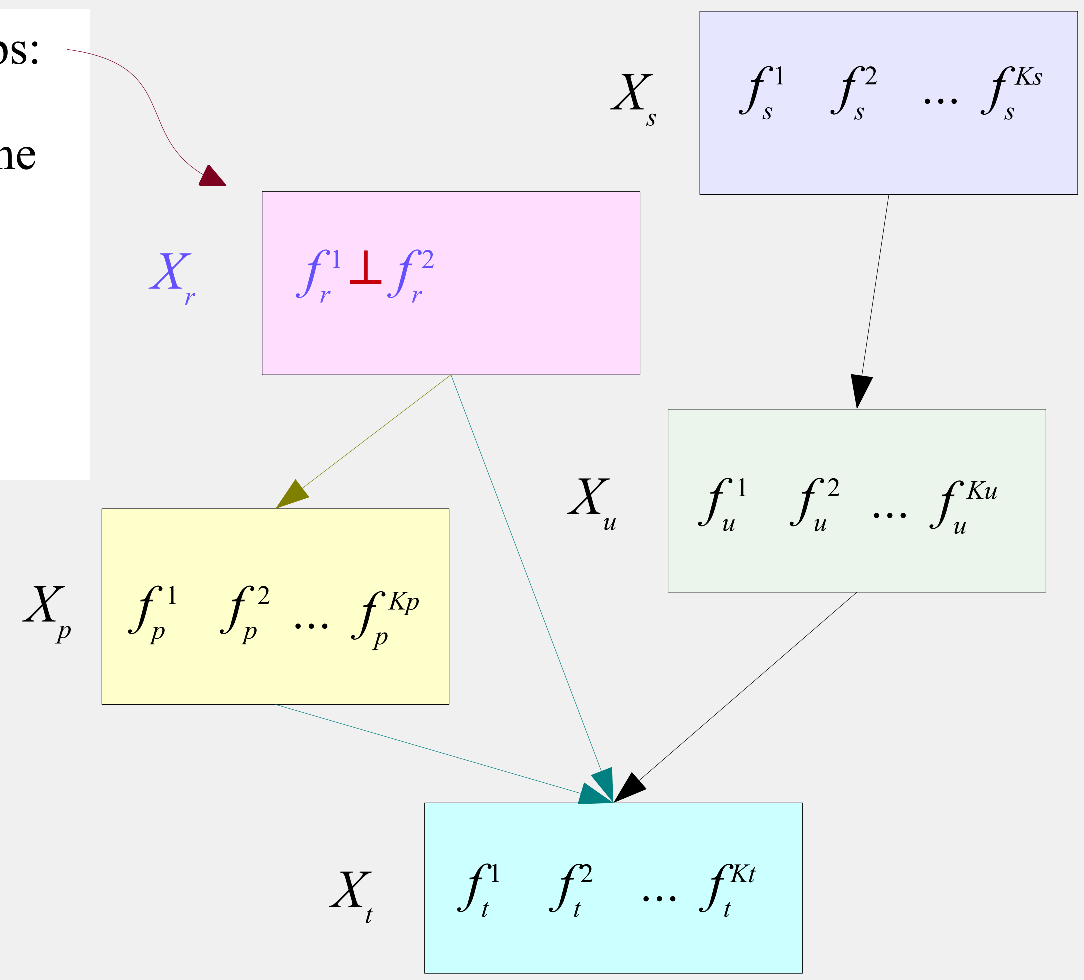
4. Algorithm → component hierarchy

- **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

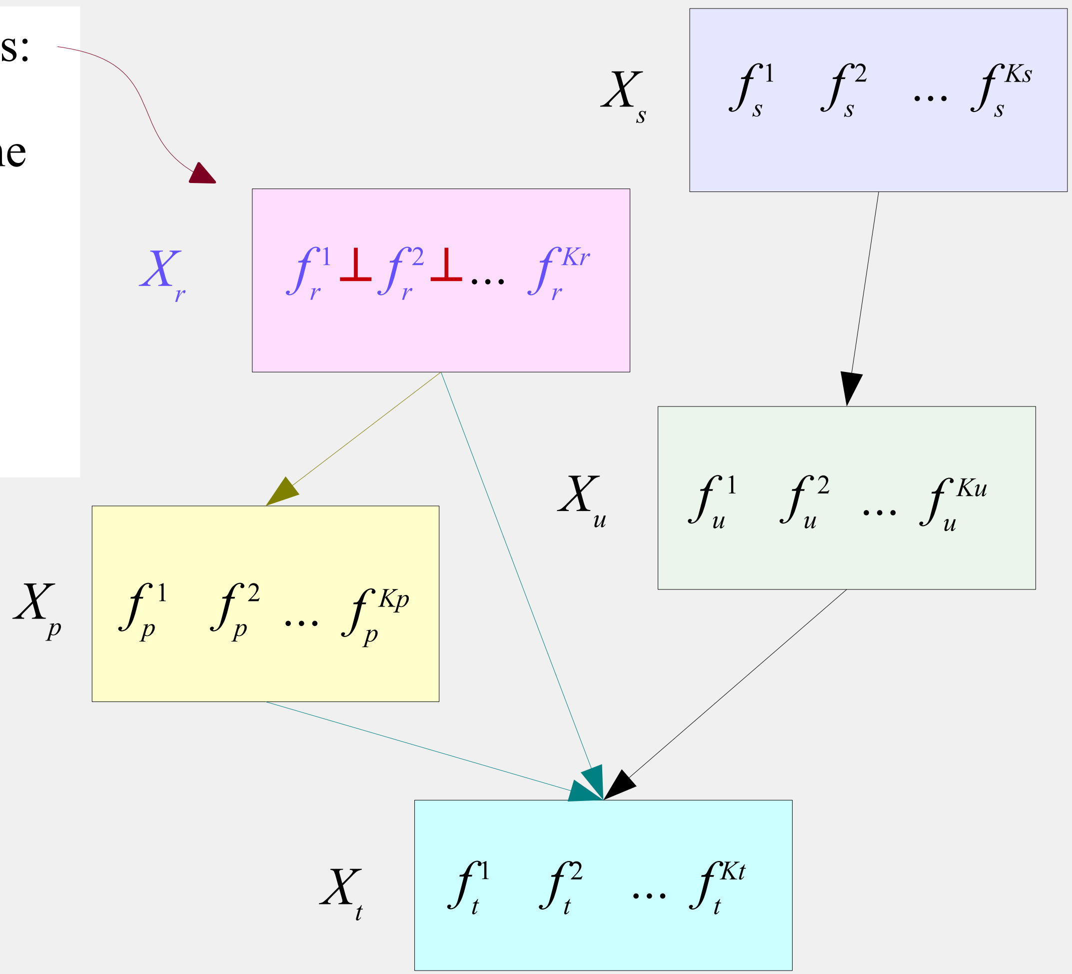
- **The local-nesting (LocNes) principle:**

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

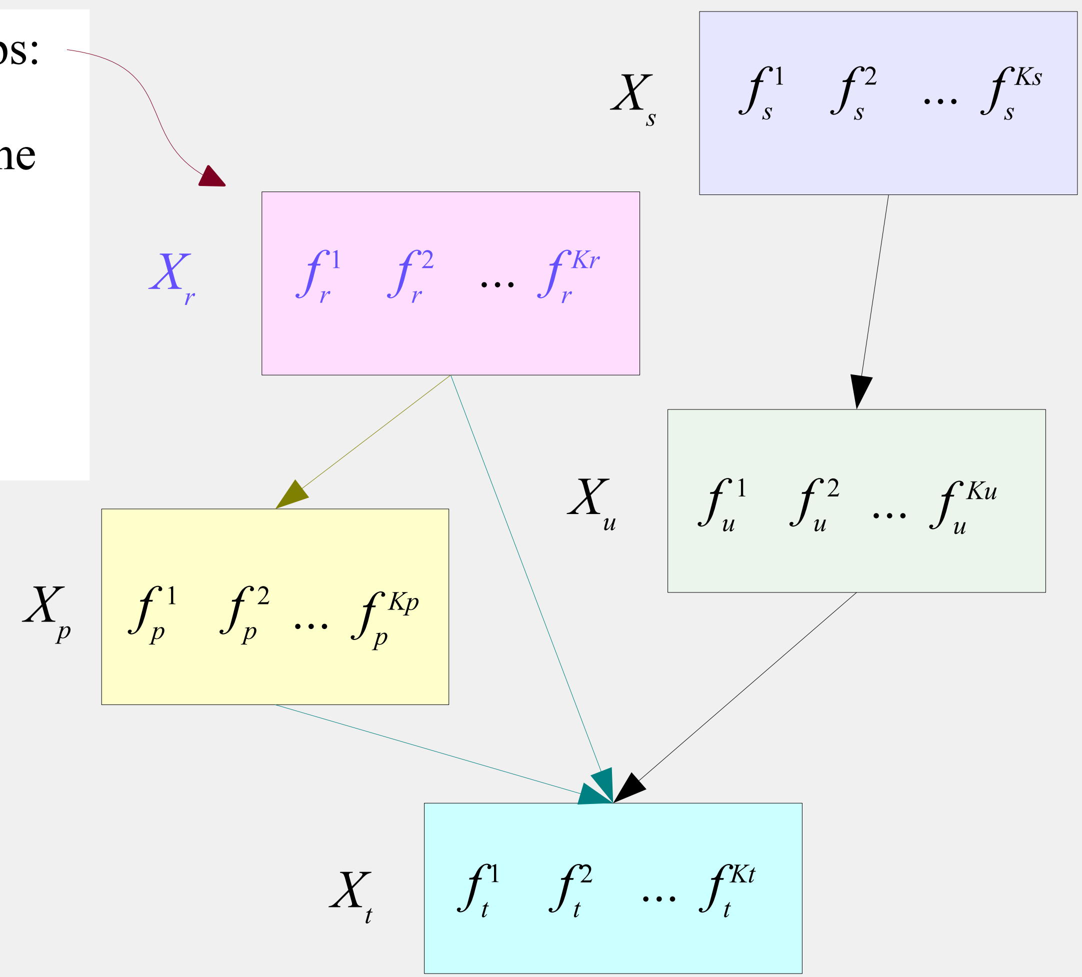
- **The local-nesting (LocNes) principle:**

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.



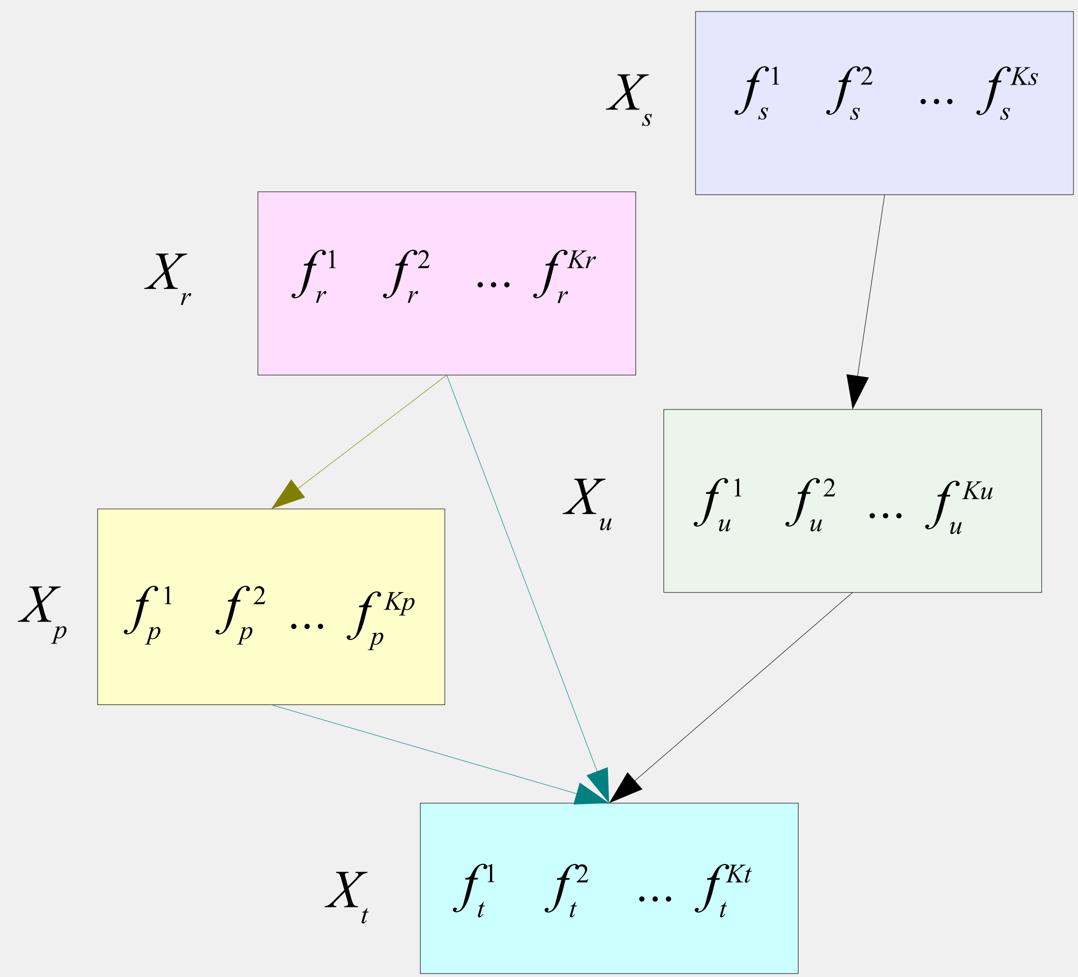
And the algorithm loops over groups X_r until convergence.

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:



How THEME works

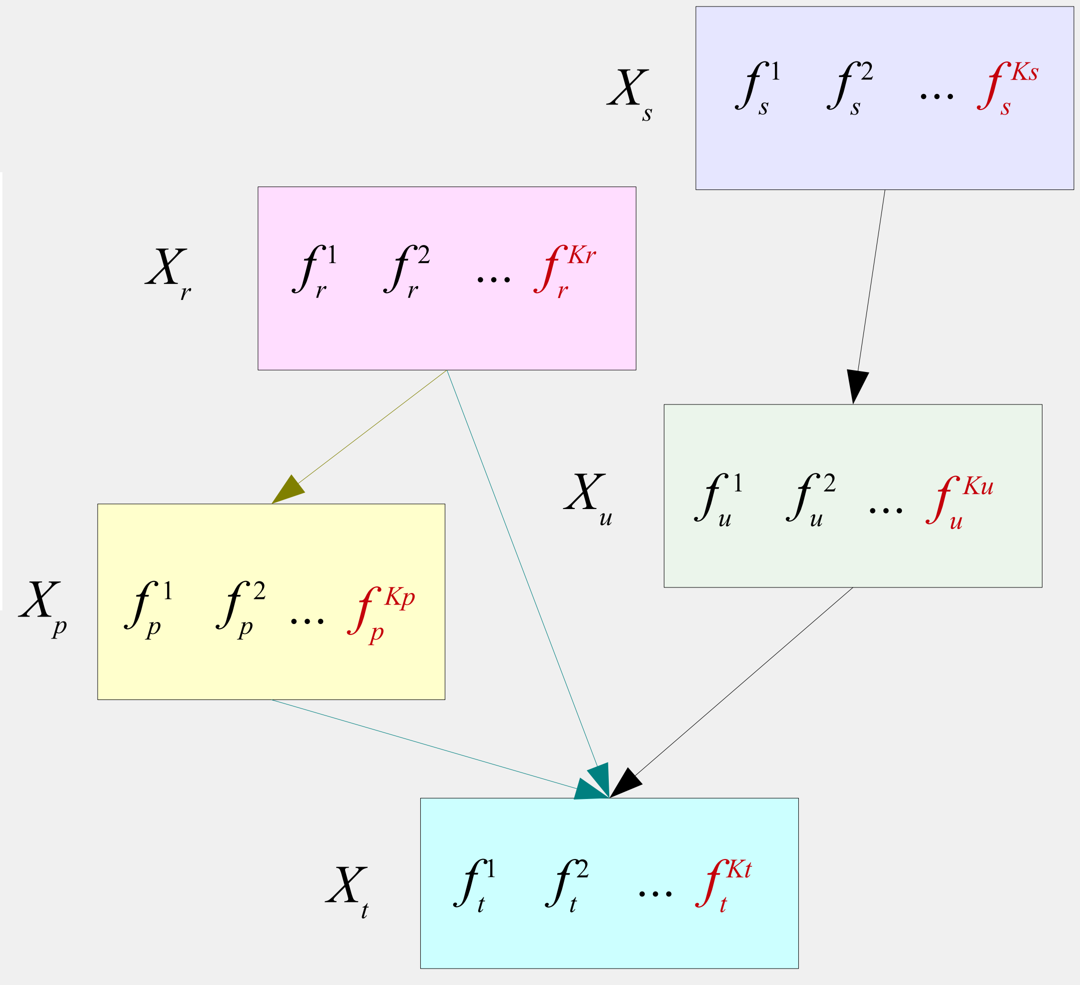
5. Backward component selection

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

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

1) Calculate the model using (too large) K_r components in group X_r.

2) For each group, assess its last component with respect to the overall explanatory / predictive power of the whole model.



How THEME works

5. Backward component selection

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

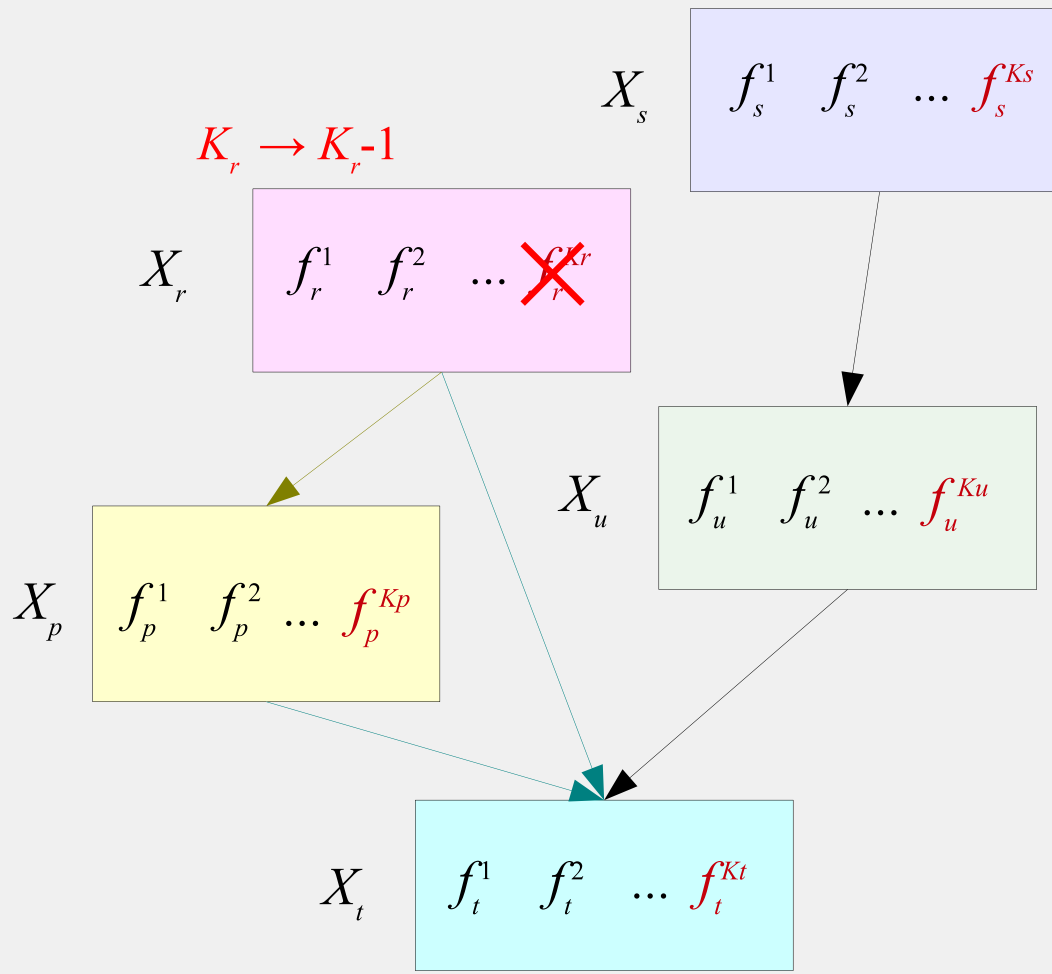
Local nesting makes ***backward selection*** natural and easy:

1) Calculate the model using (too large) K_r components in group X_r.

2) For each group, assess its last component with respect to the overall predictive power of the whole model.

3) Compare these last components between groups with respect to some loss criterion.

4) Discard the least useful last component *if affordable*, and resume (1) with one less component in the corresponding group. Else, stop.



How THEME works

5. Backward component selection

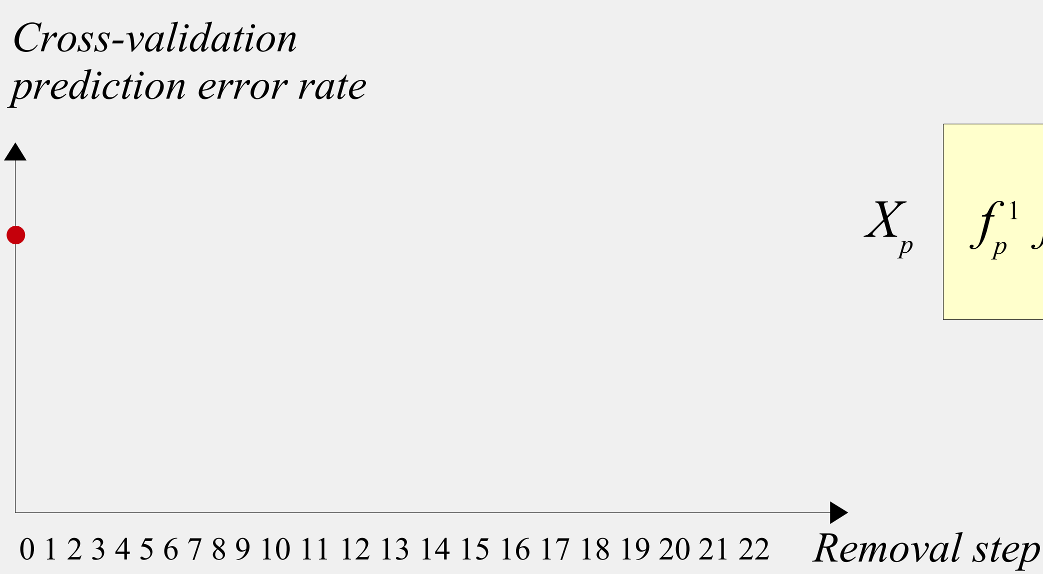
- **How to select the suitable number of components in each group?**

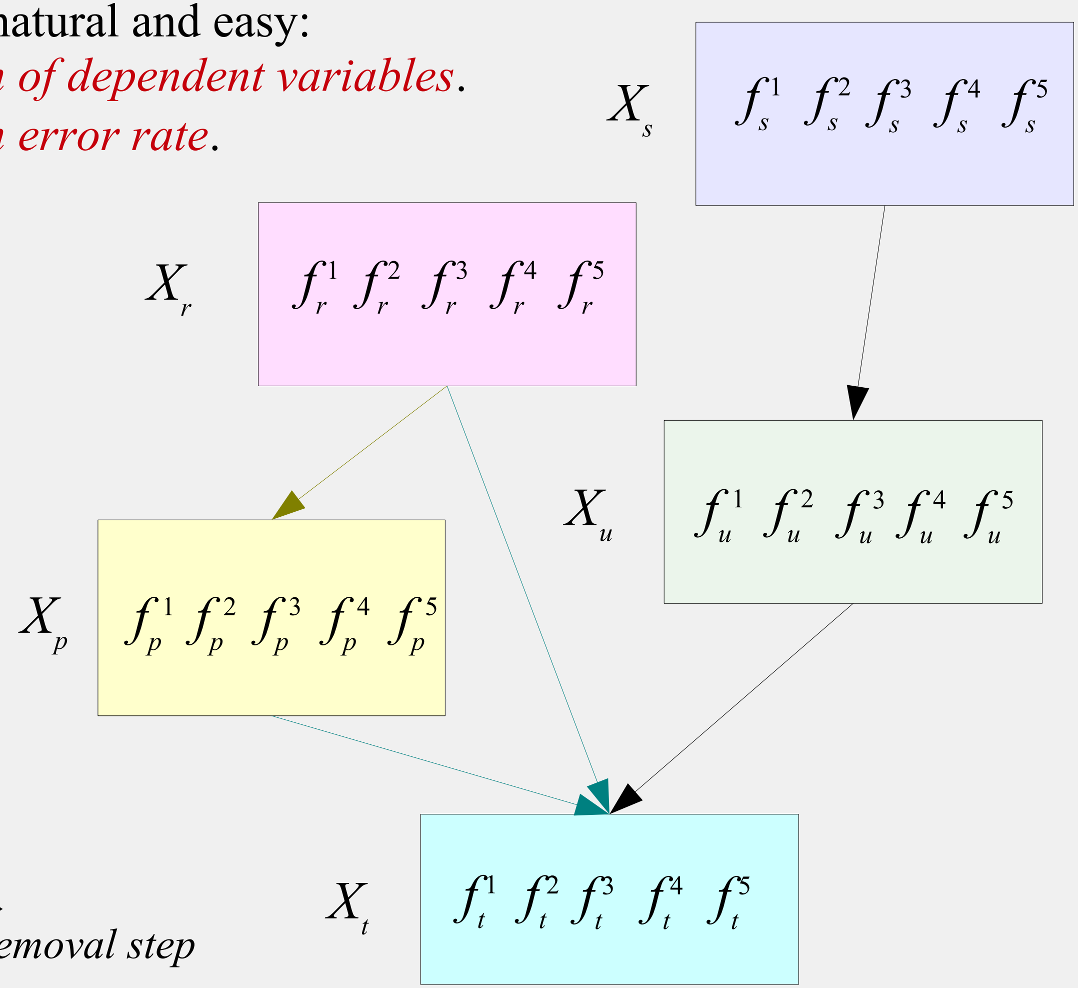
Local nesting makes ***backward selection*** natural and easy:

Explanatory components → *prediction of dependent variables*.

→ *prediction error rate*.

Example:





How THEME works

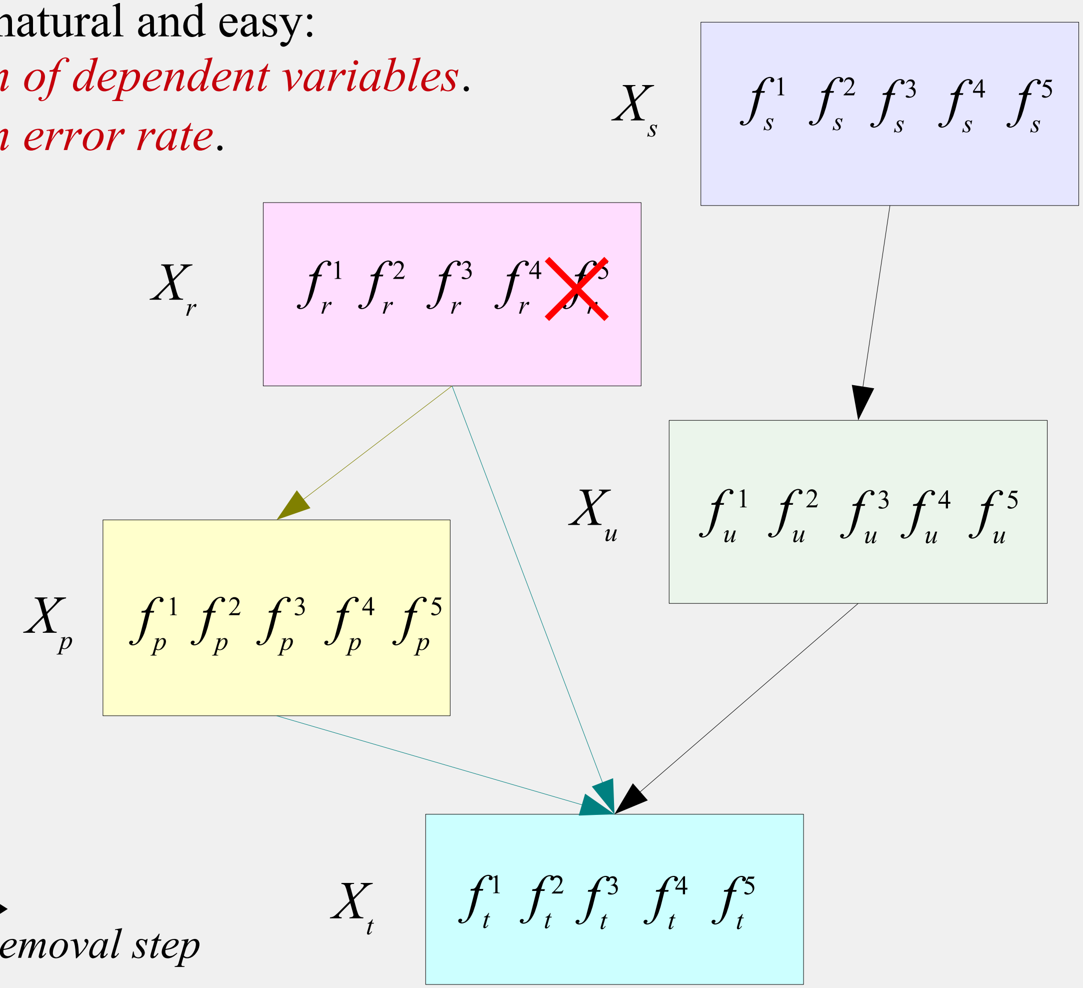
5. Backward component selection

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

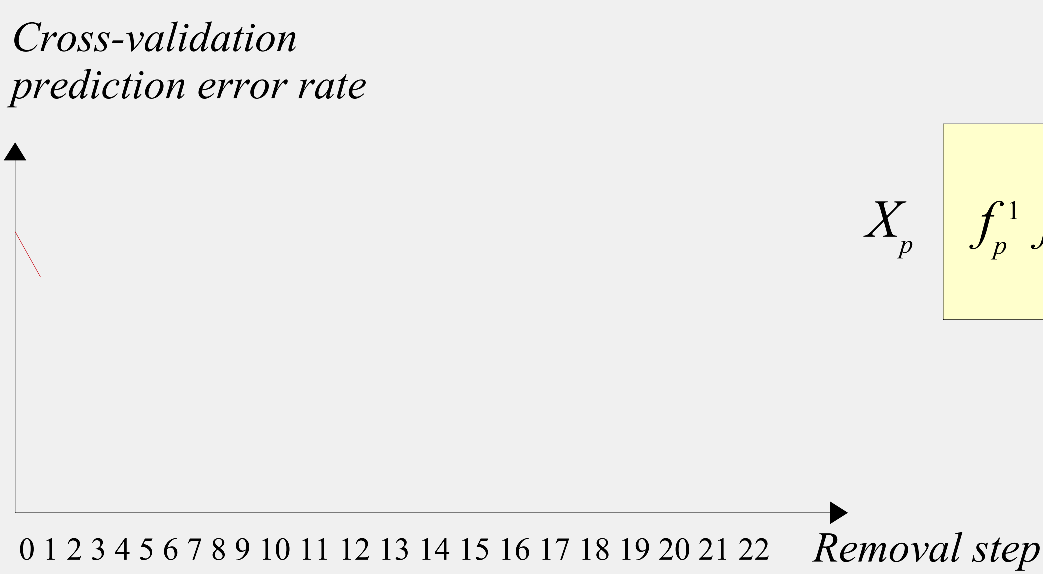
Local nesting makes ***backward selection*** natural and easy:

Explanatory components → *prediction of dependent variables*.

→ *prediction error rate*.



Example:



How THEME works

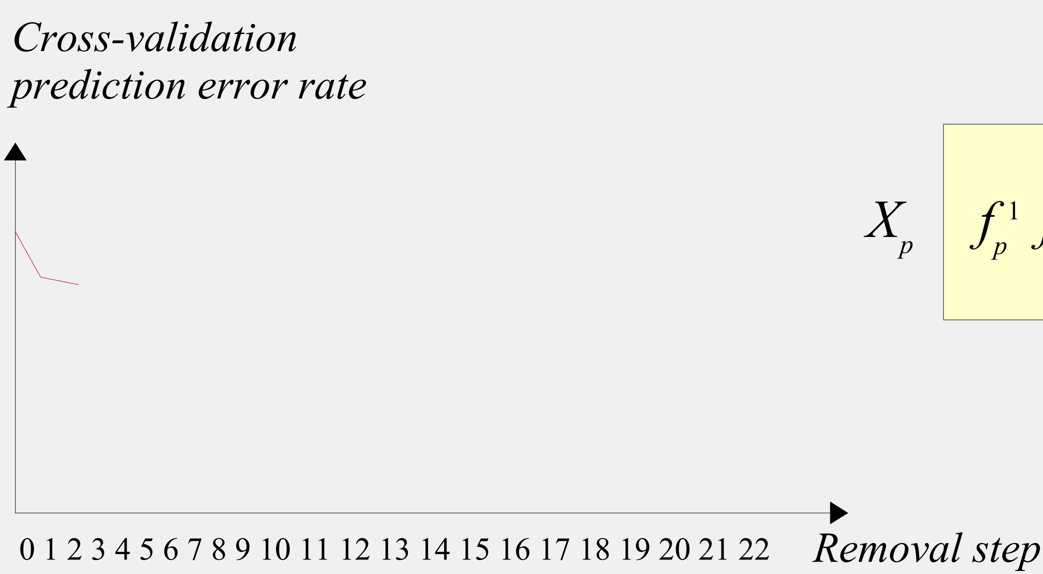
5. Backward component selection

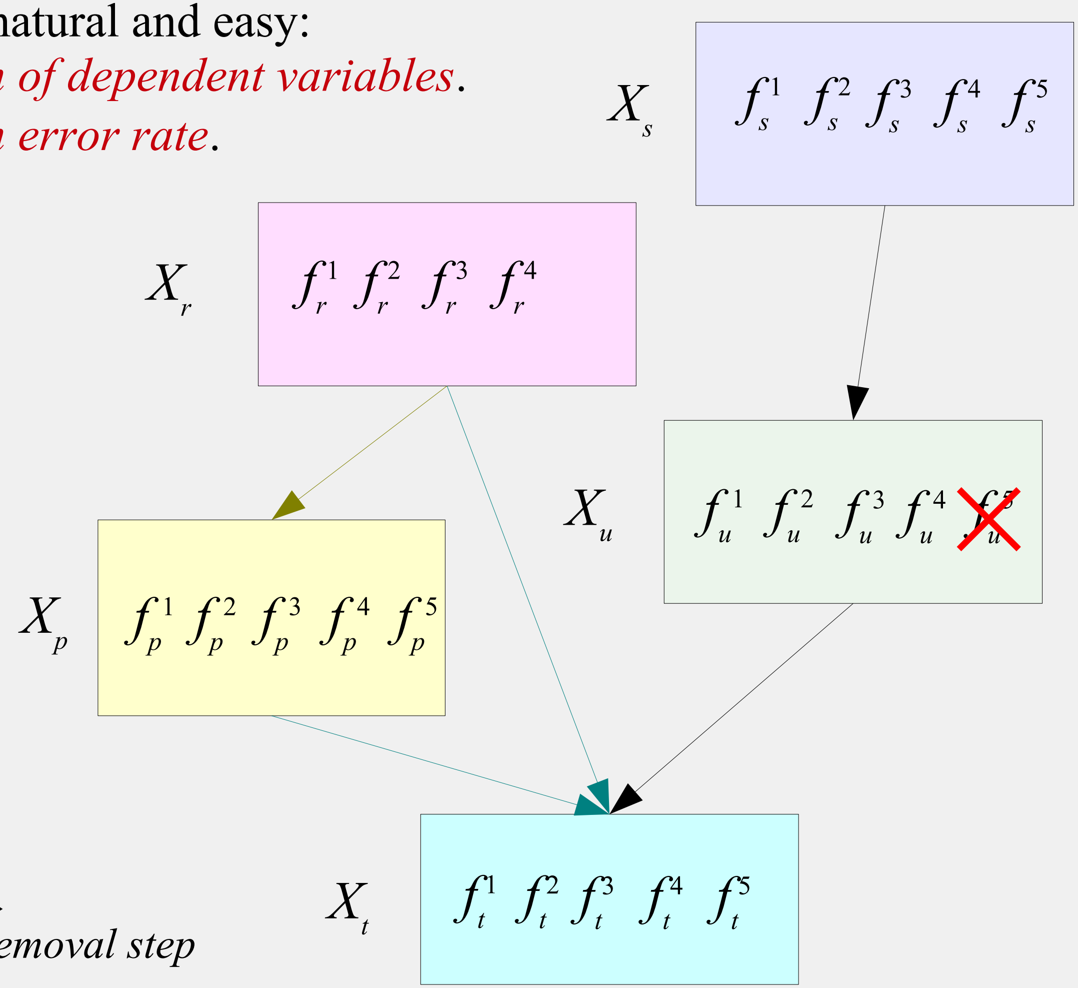
- **How to select the suitable number of components in each group?**

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

Explanatory components → *prediction of dependent variables*.
→ *prediction error rate*.

Example:





How THEME works

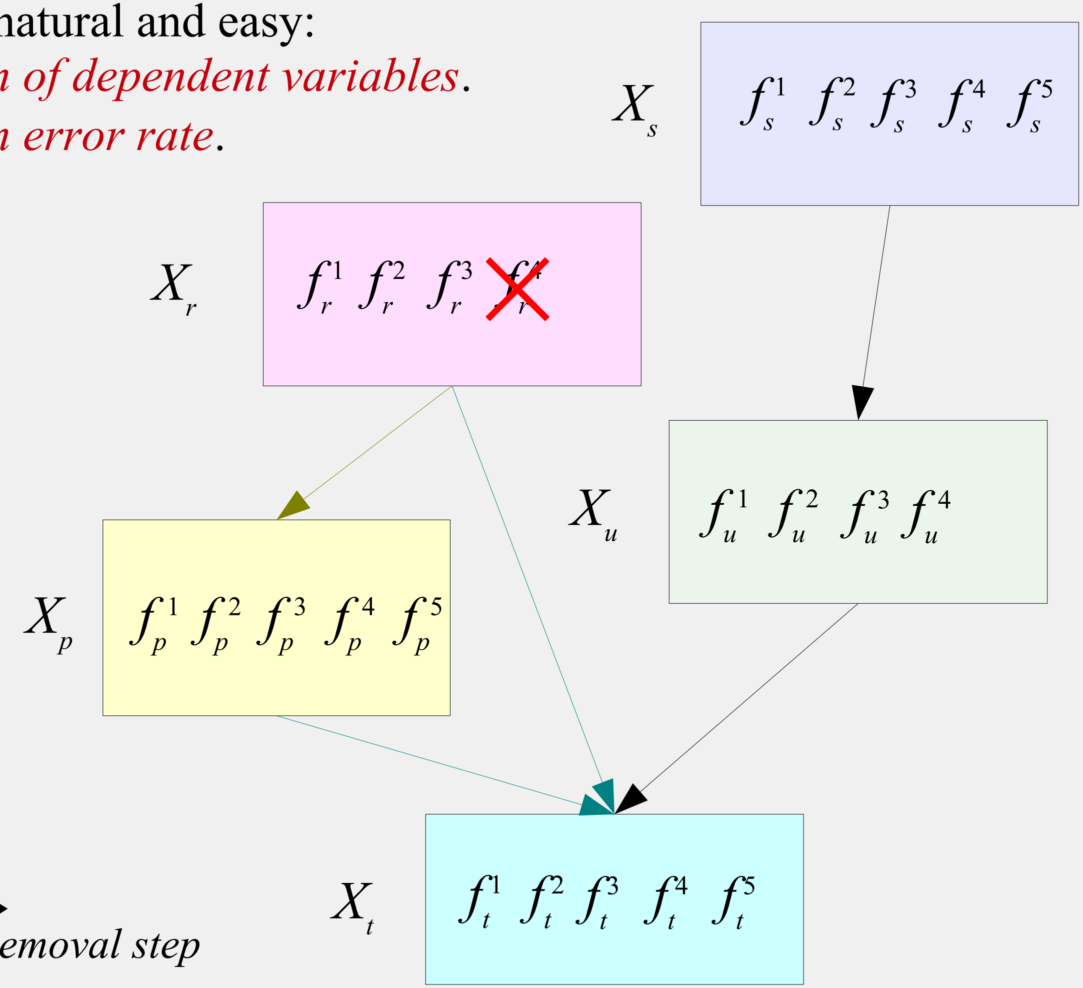
5. Backward component selection

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

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

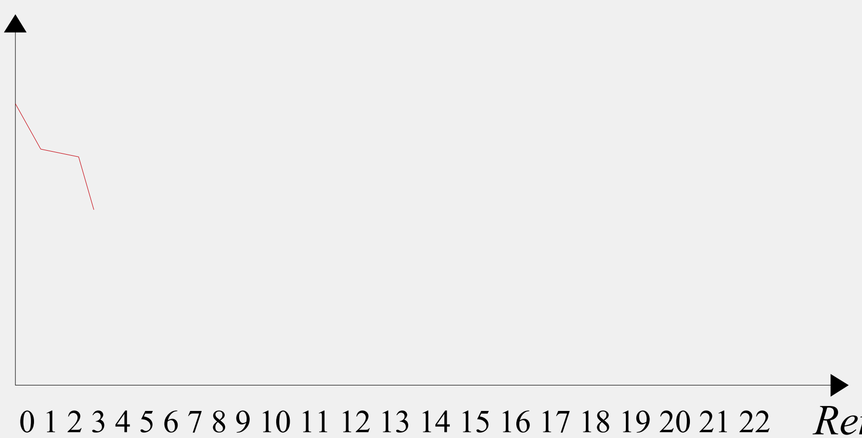
Explanatory components → *prediction of dependent variables*.

→ *prediction error rate*.



Example:

Cross-validation prediction error rate



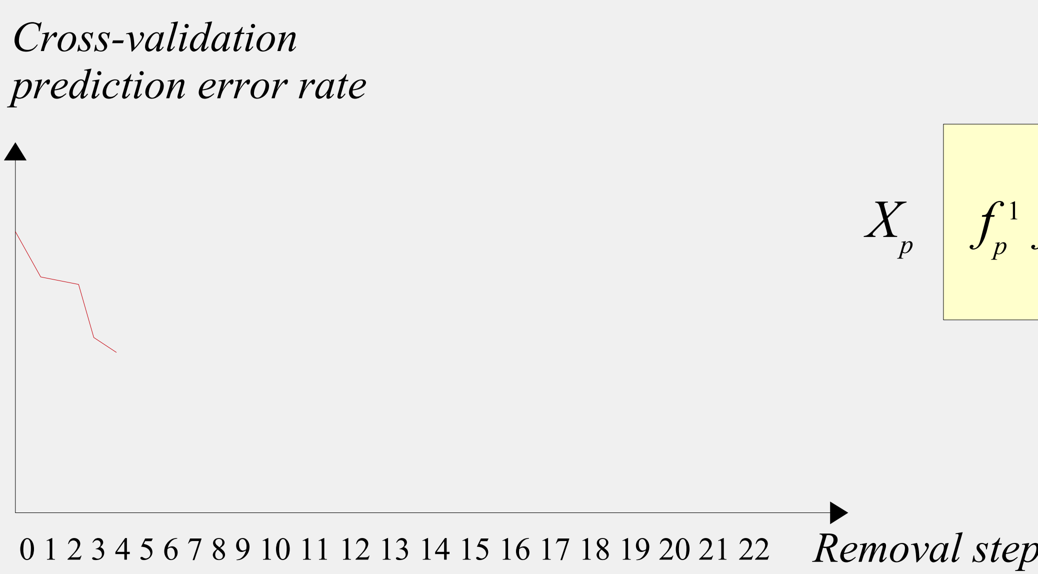
How THEME works

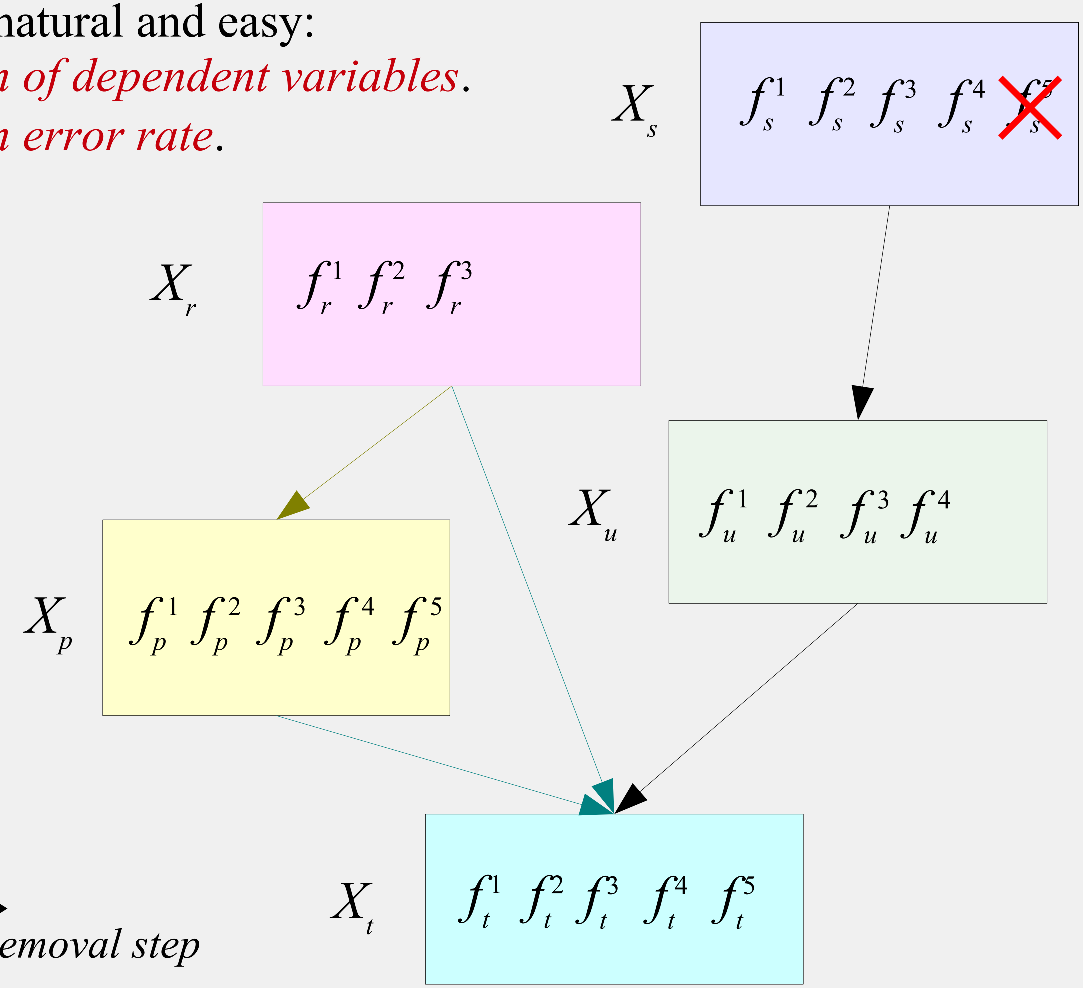
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:
Explanatory components → *prediction of dependent variables*.
→ *prediction error rate*.

Example:



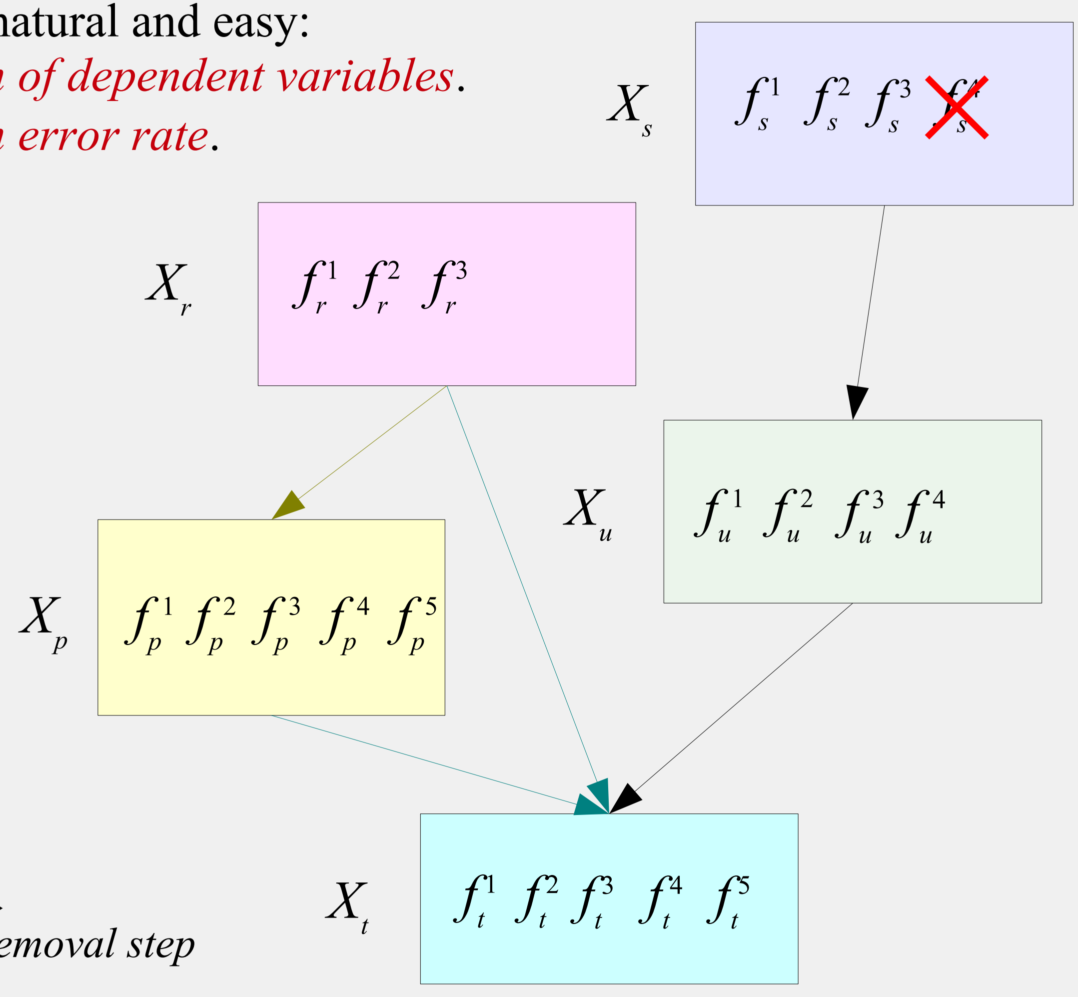


How THEME works

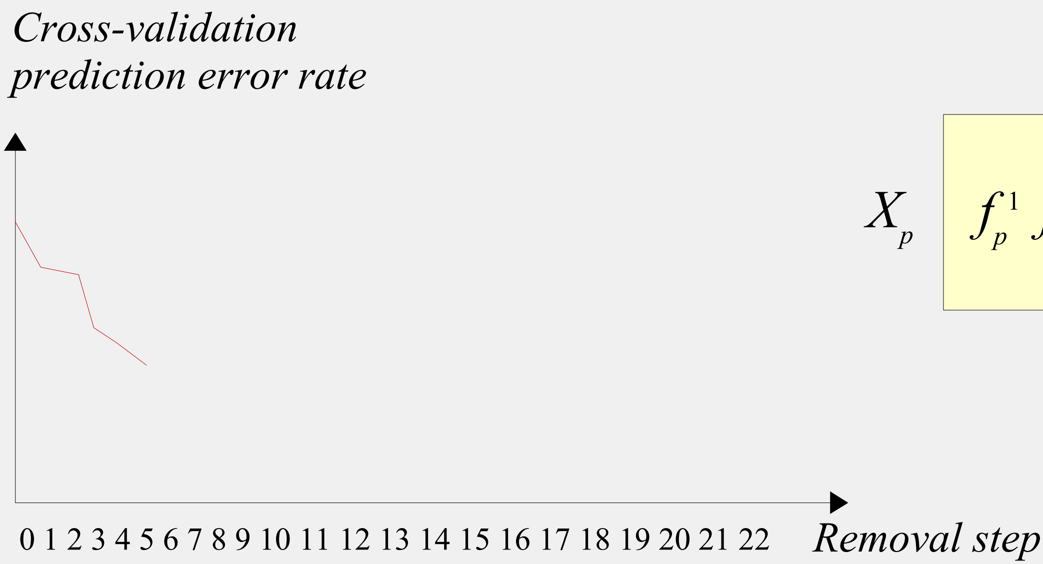
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:
Explanatory components → *prediction of dependent variables*.
→ *prediction error rate*.



Example:

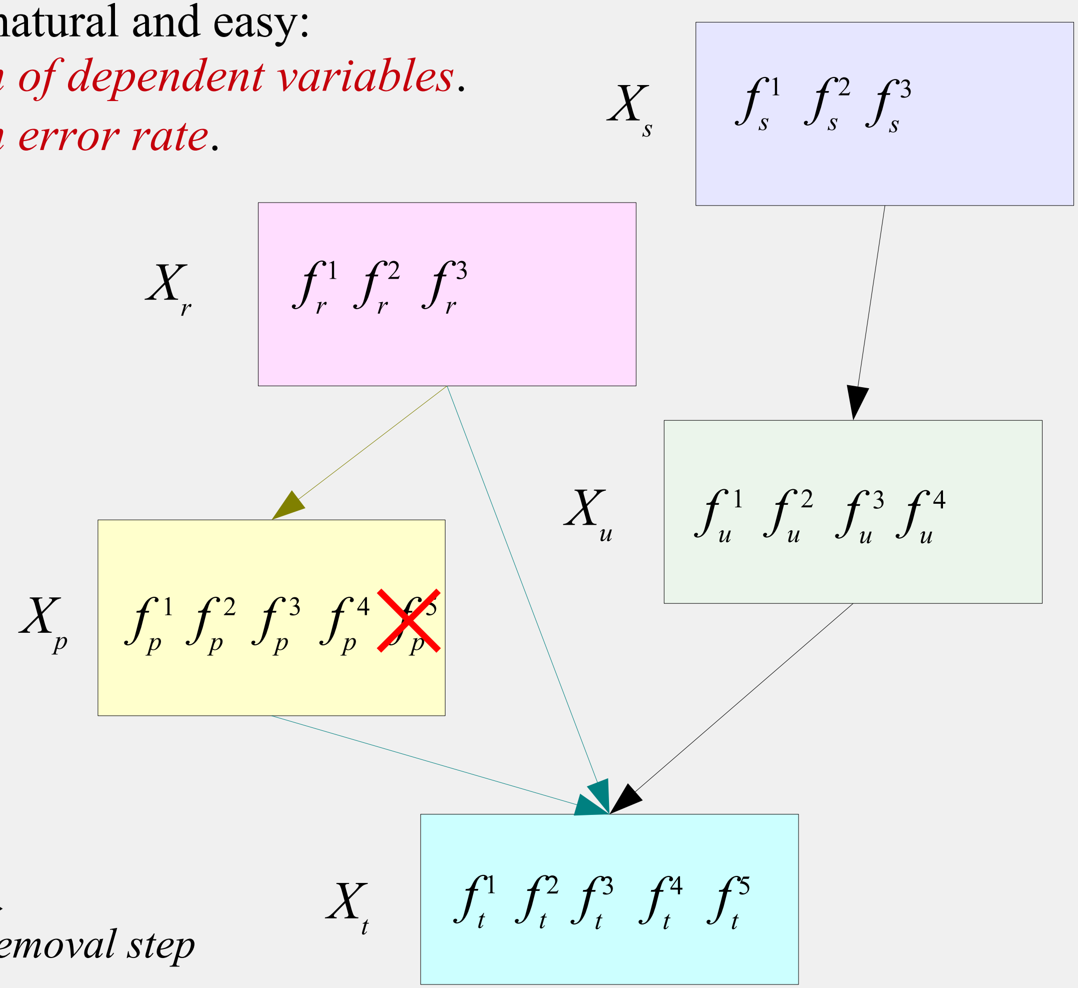


How THEME works

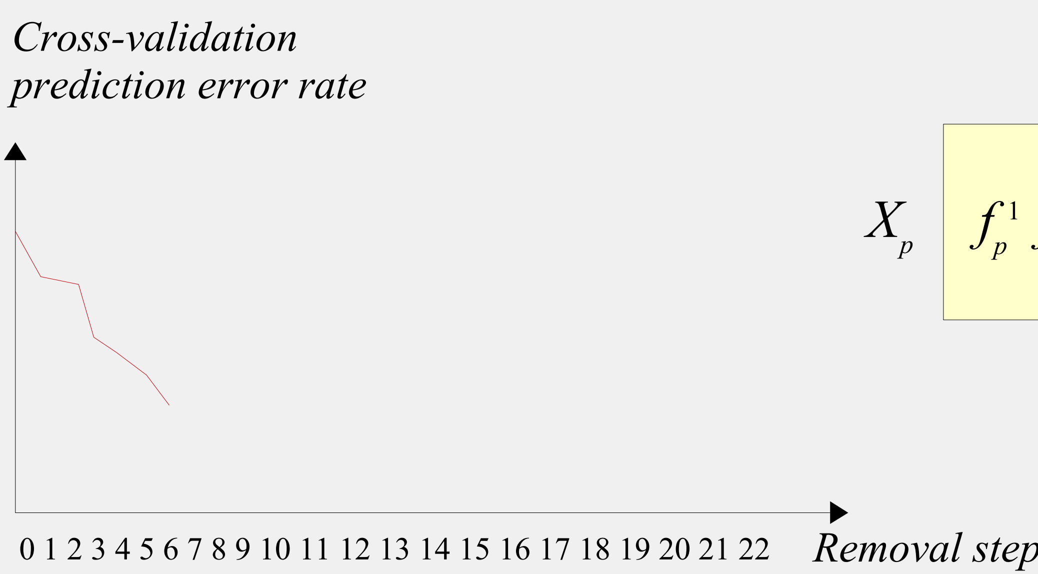
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:
Explanatory components → *prediction of dependent variables*.
→ *prediction error rate*.



Example:

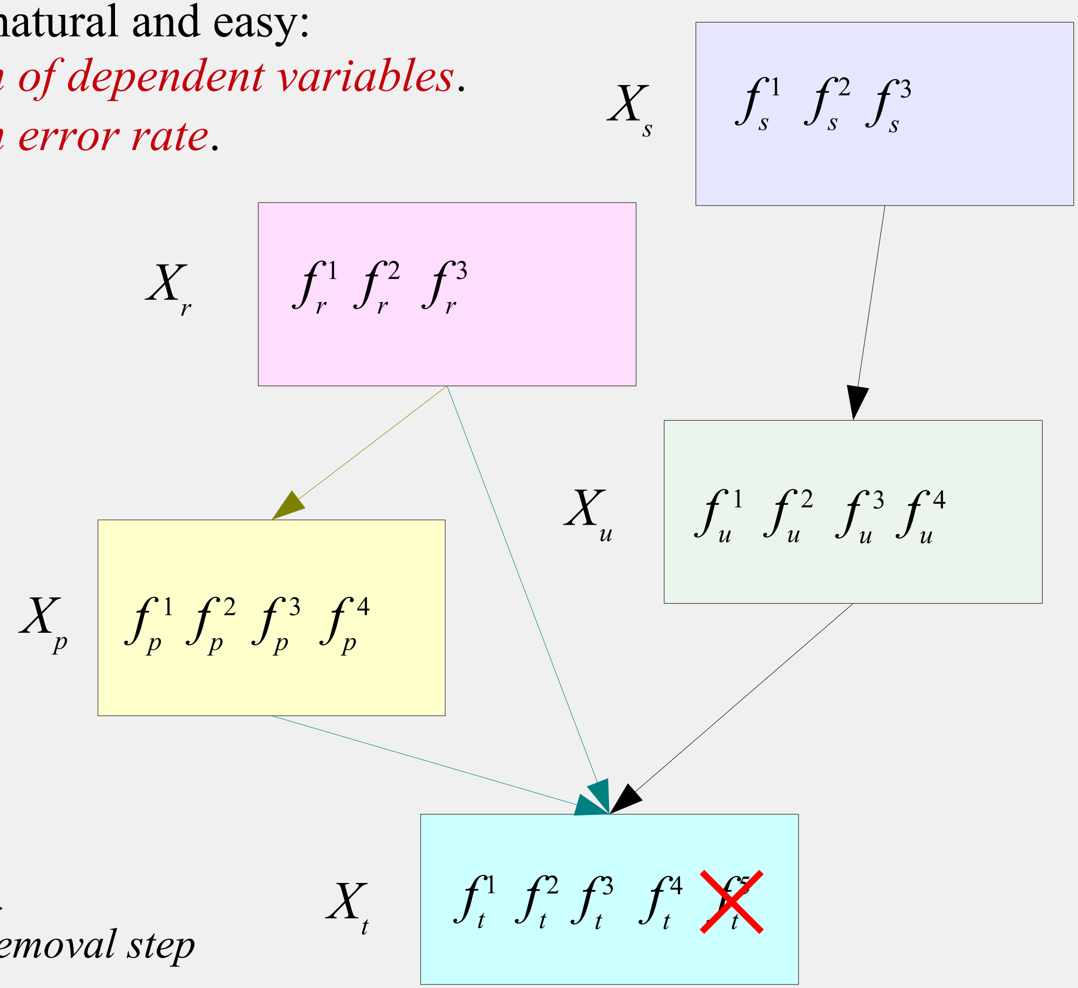


How THEME works

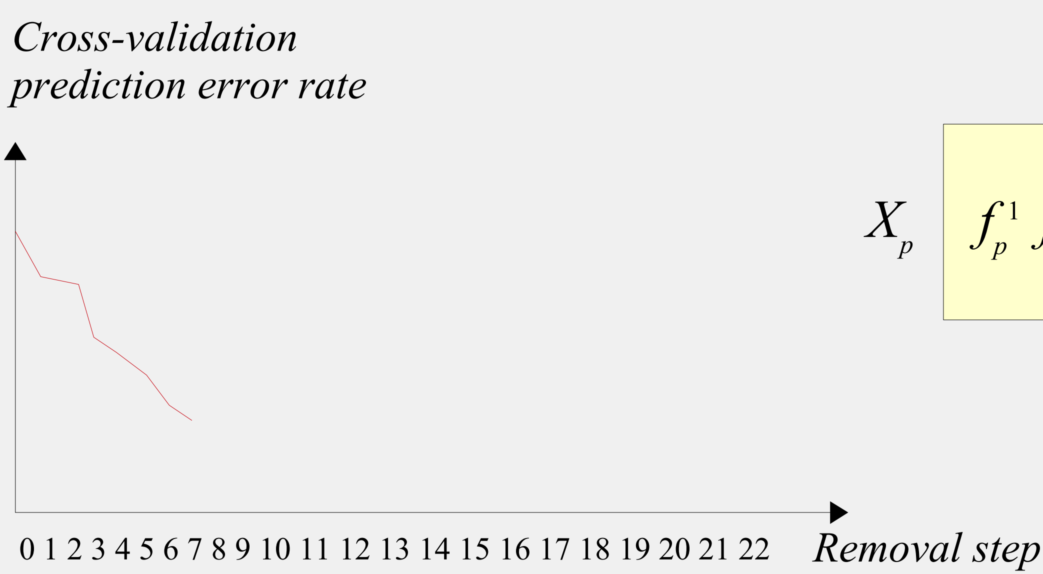
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:
Explanatory components → *prediction of dependent variables*.
→ *prediction error rate*.



Example:

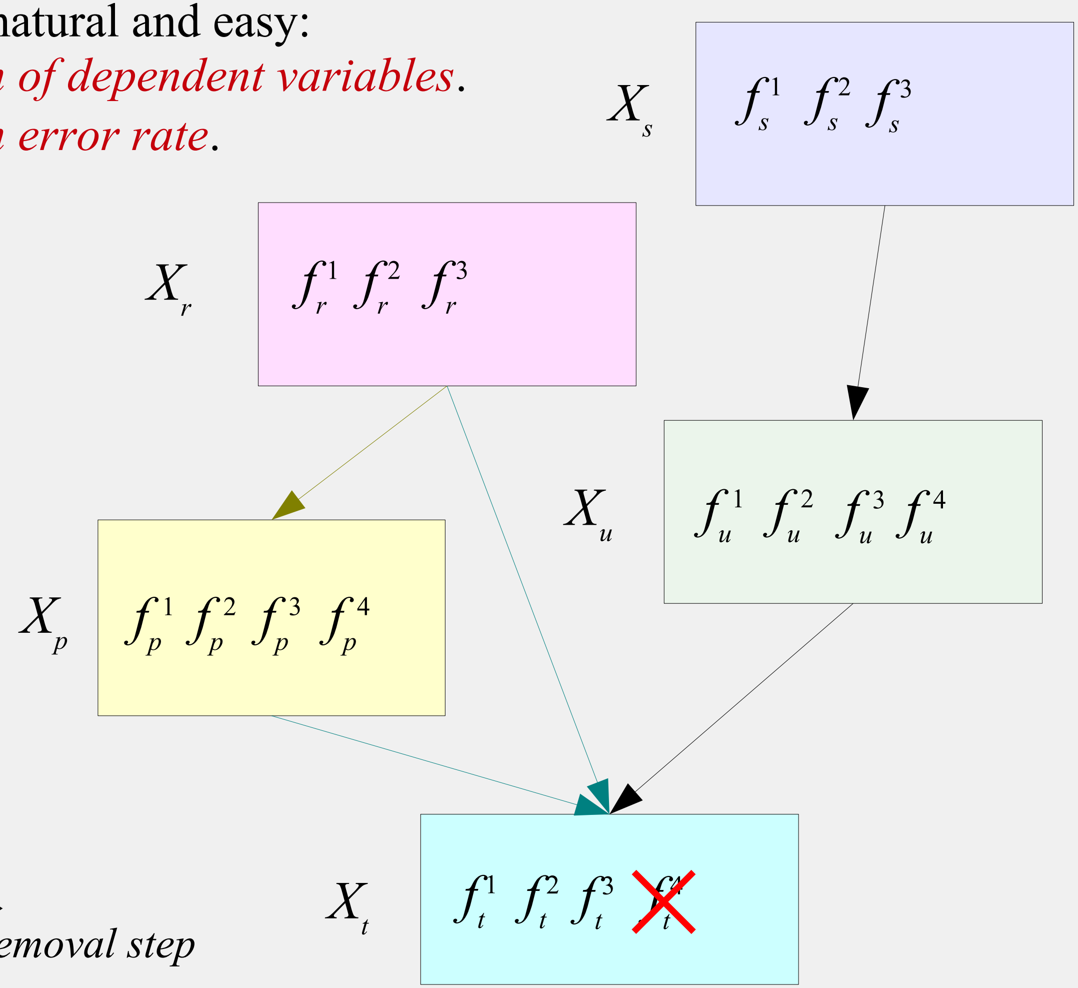


How THEME works

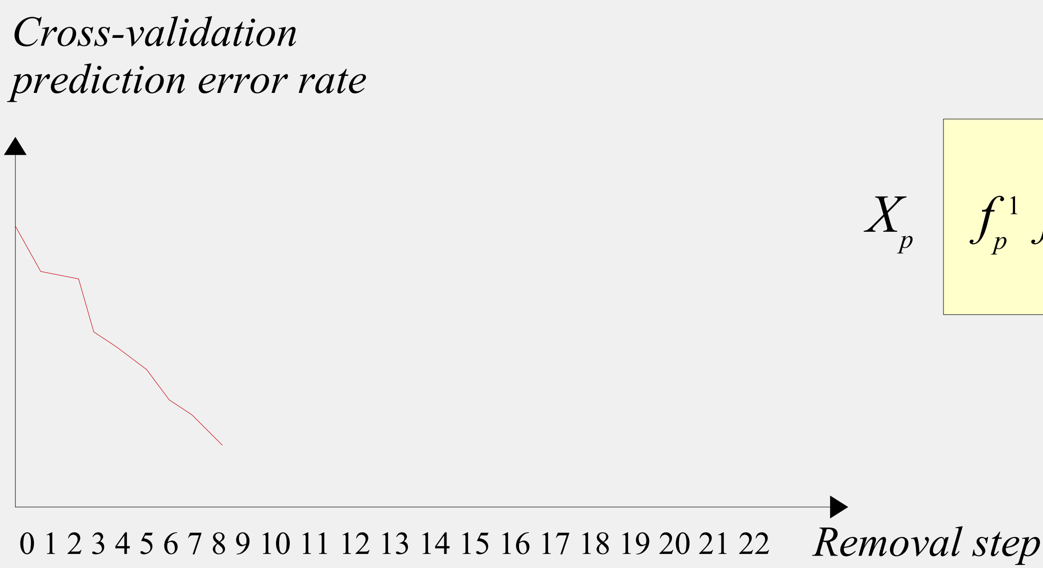
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:
Explanatory components → *prediction of dependent variables*.
→ *prediction error rate*.



Example:



How THEME works

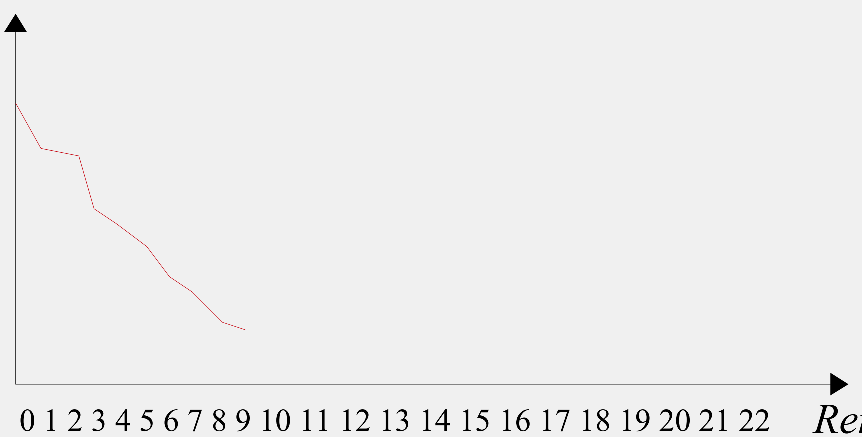
5. Backward component selection

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

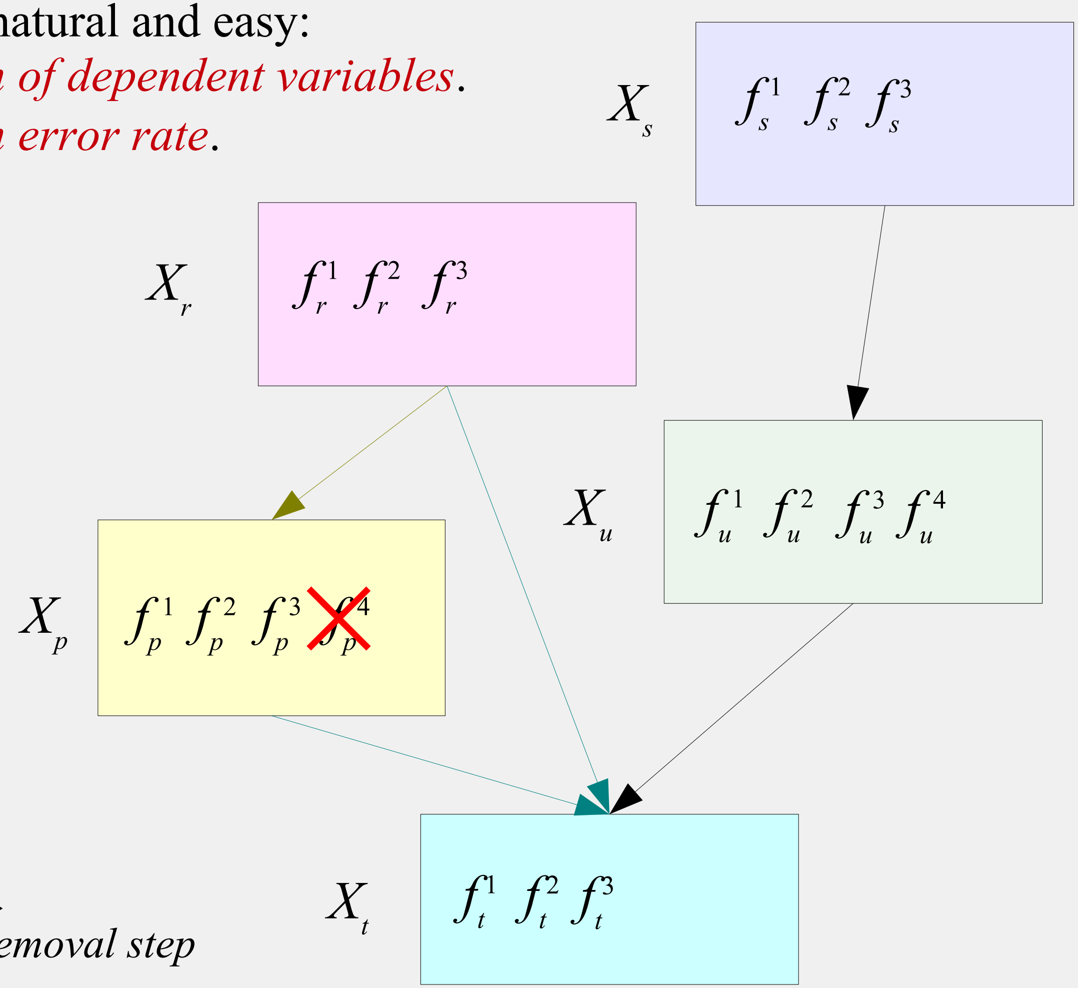
Local nesting makes ***backward selection*** natural and easy:
Explanatory components → *prediction of dependent variables*.
→ *prediction error rate*.

Example:

Cross-validation prediction error rate



Removal step



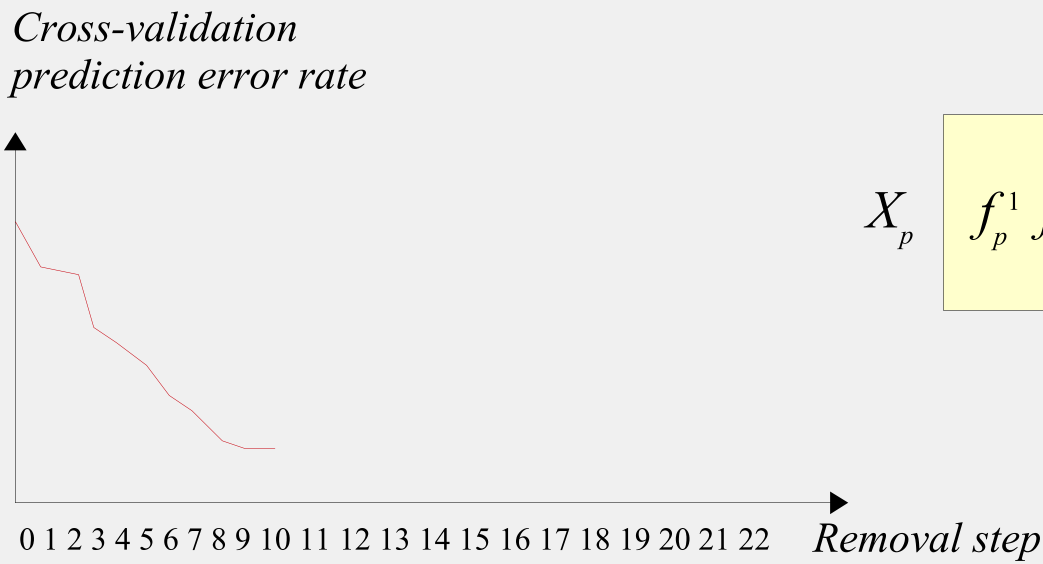
How THEME works

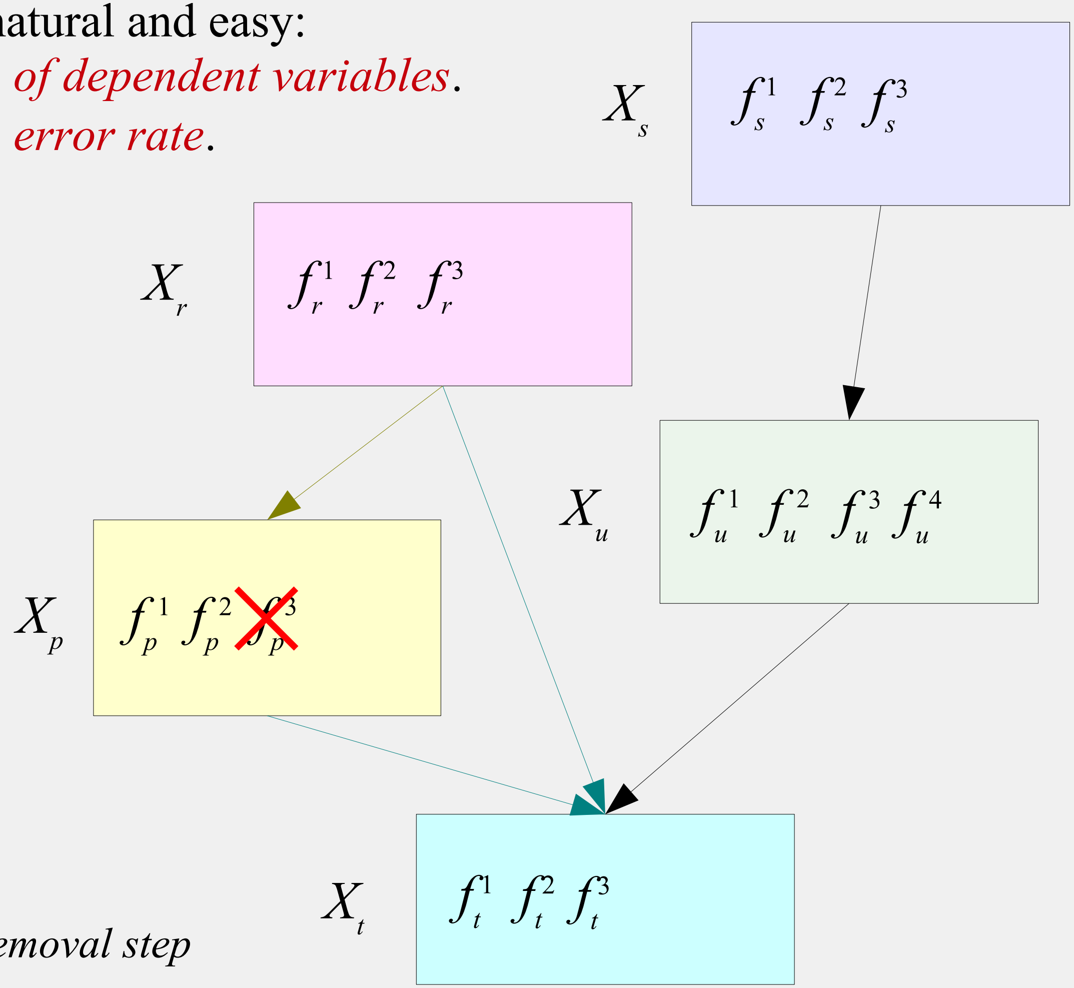
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:
Explanatory components → *prediction of dependent variables*.
→ *prediction error rate*.

Example:





How THEME works

5. Backward component selection

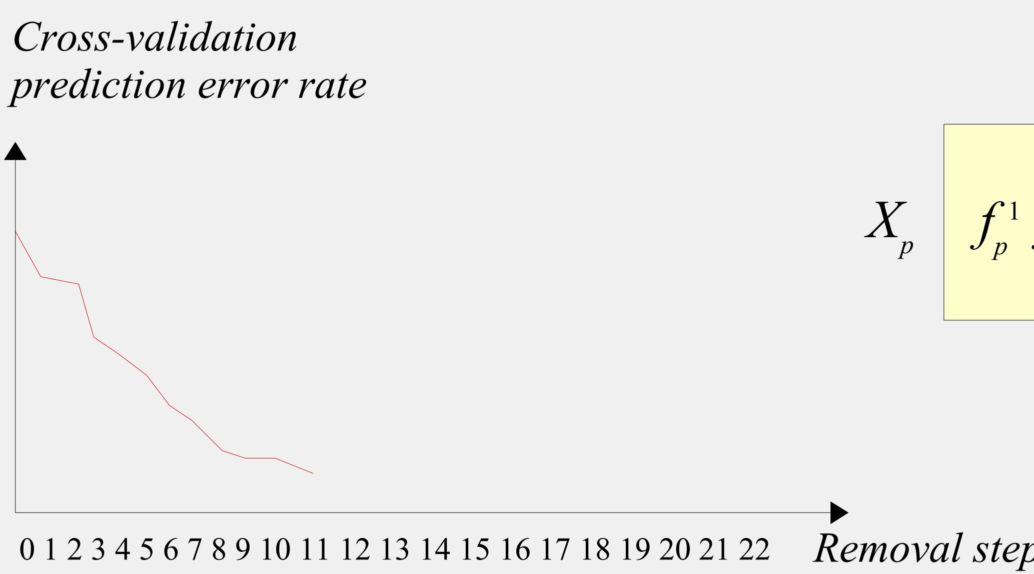
- **How to select the suitable number of components in each group?**

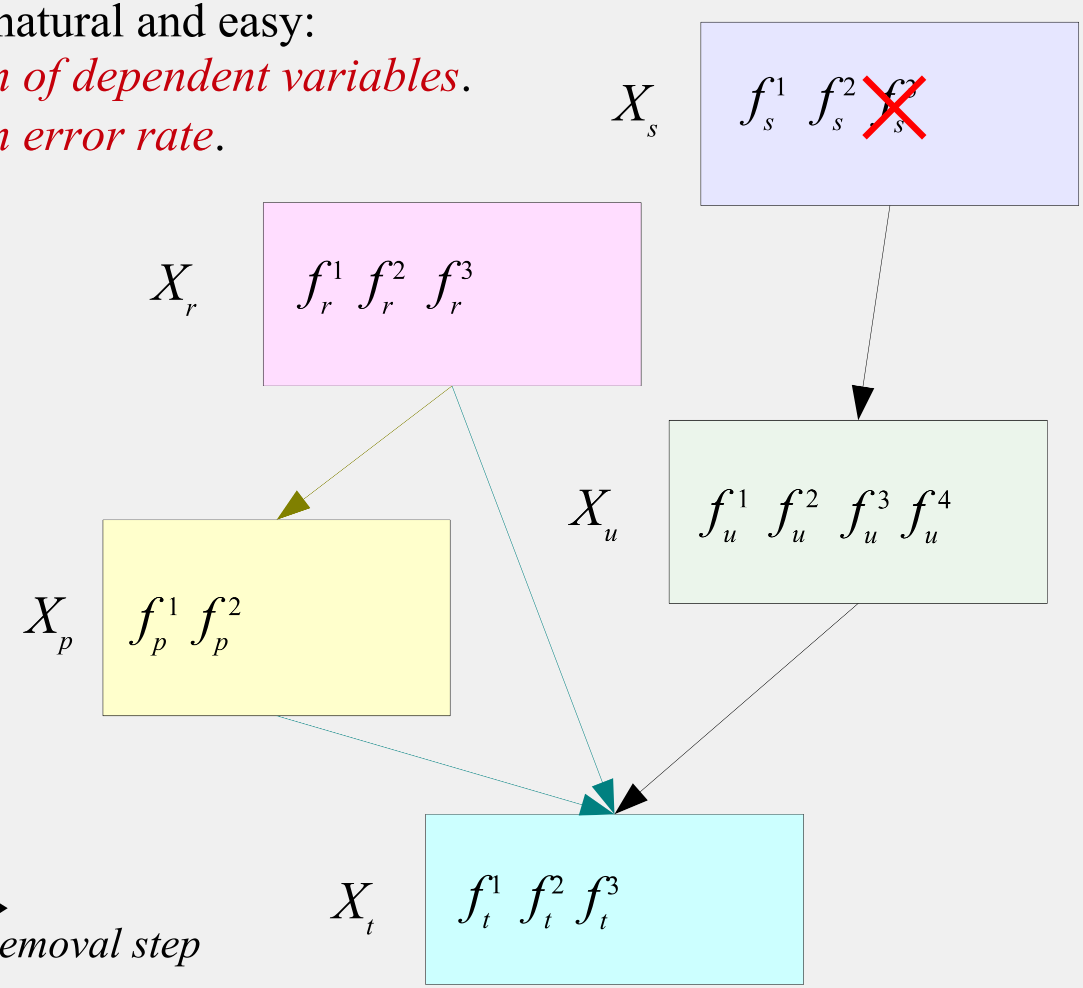
Local nesting makes ***backward selection*** natural and easy:

Explanatory components → *prediction of dependent variables*.

→ *prediction error rate*.

Example:





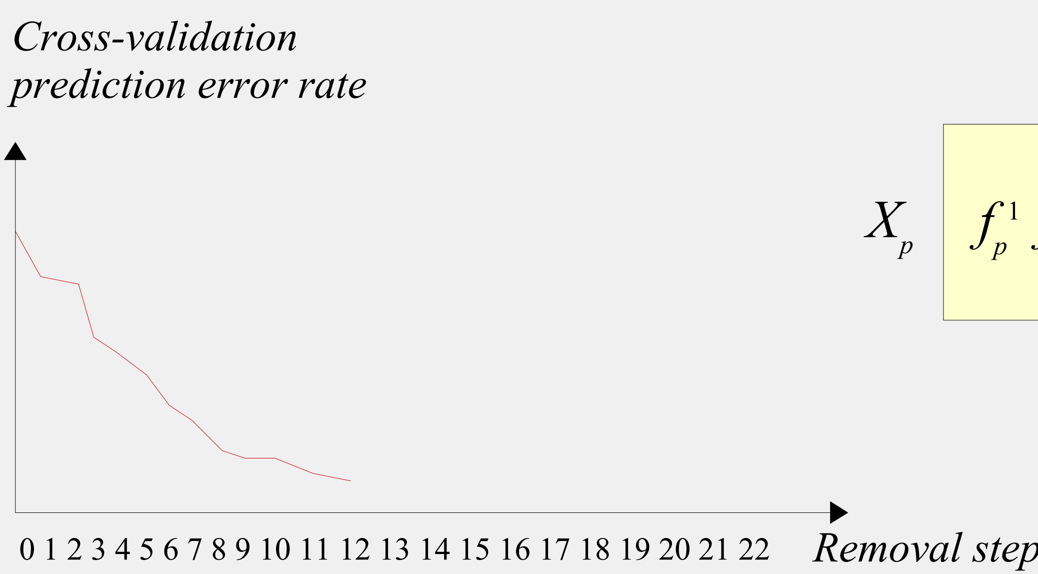
How THEME works

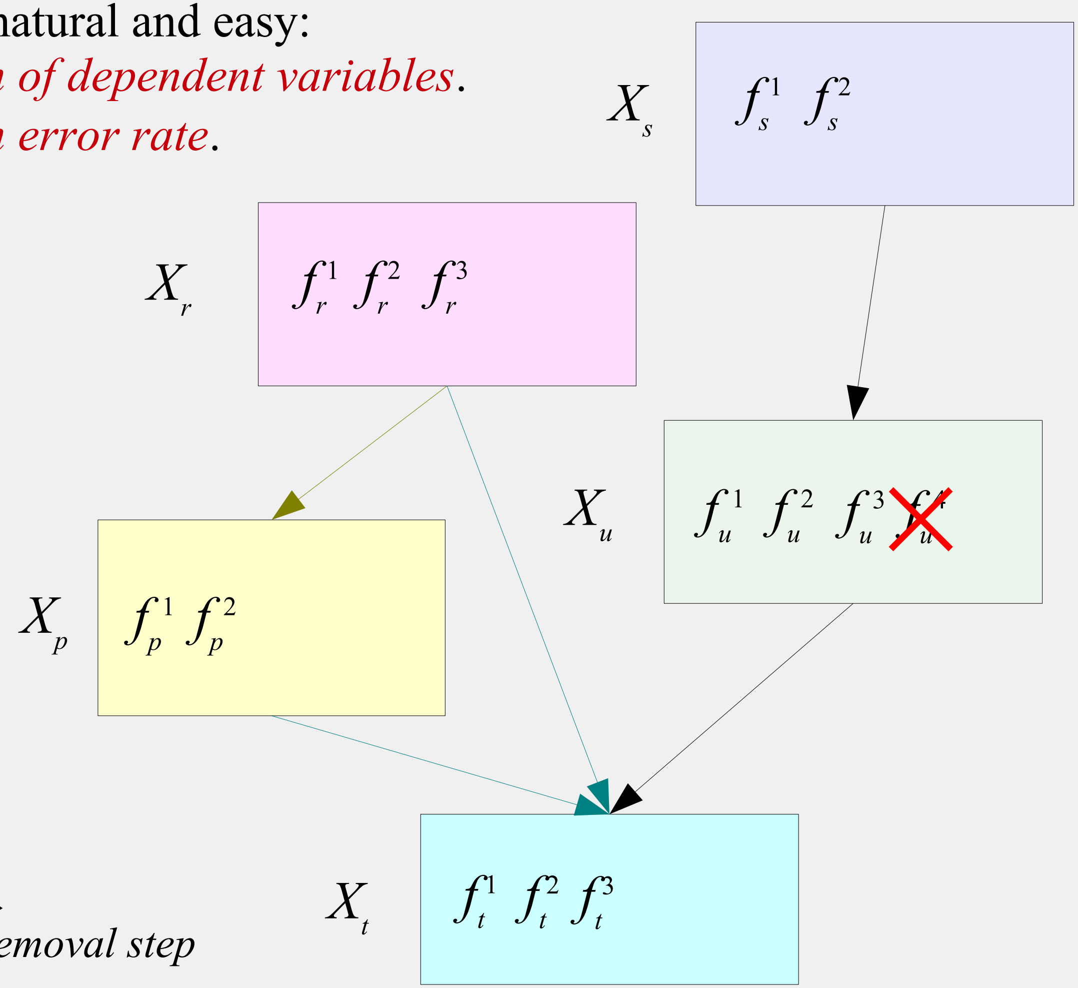
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:
Explanatory components → *prediction of dependent variables*.
→ *prediction error rate*.

Example:





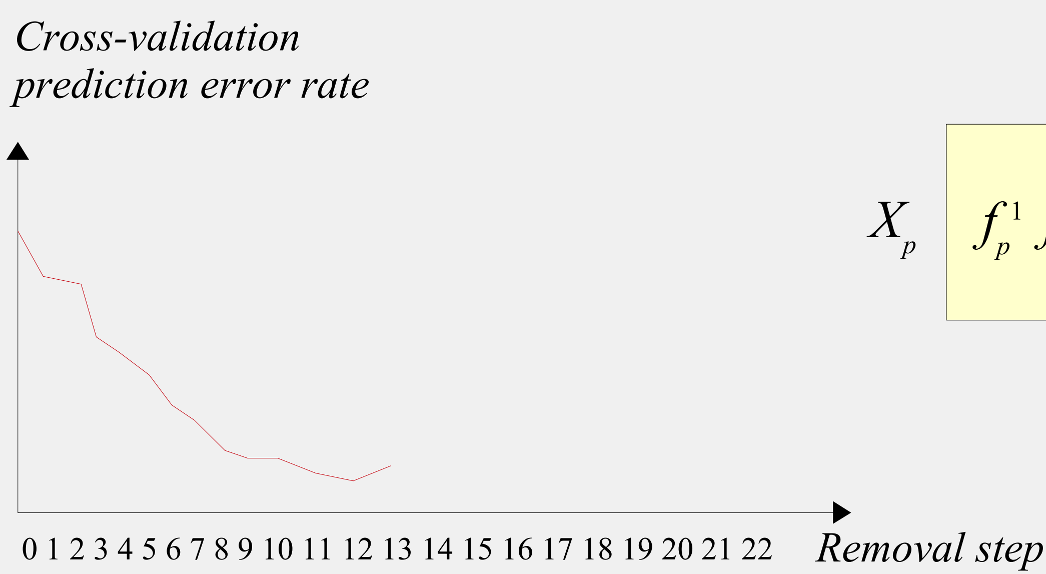
How THEME works

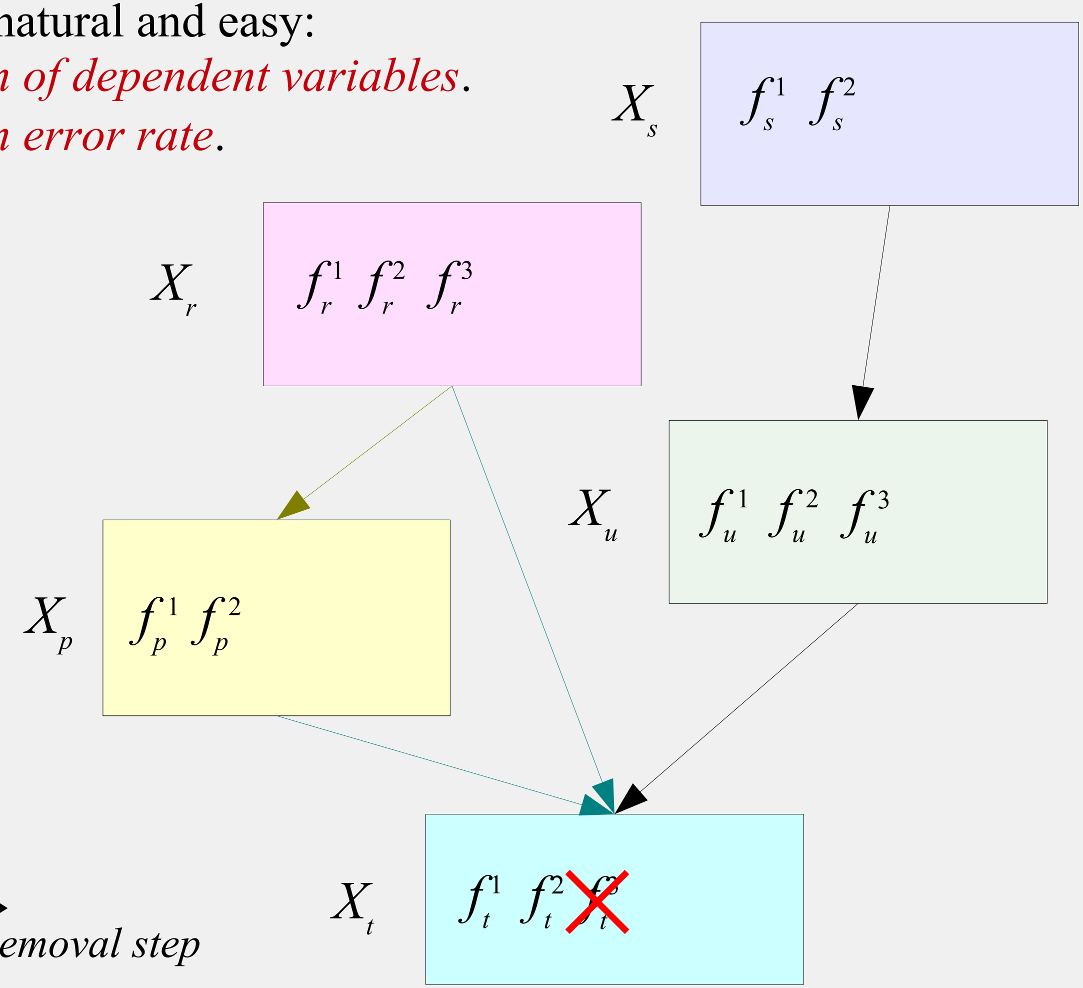
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:
Explanatory components → *prediction of dependent variables*.
→ *prediction error rate*.

Example:





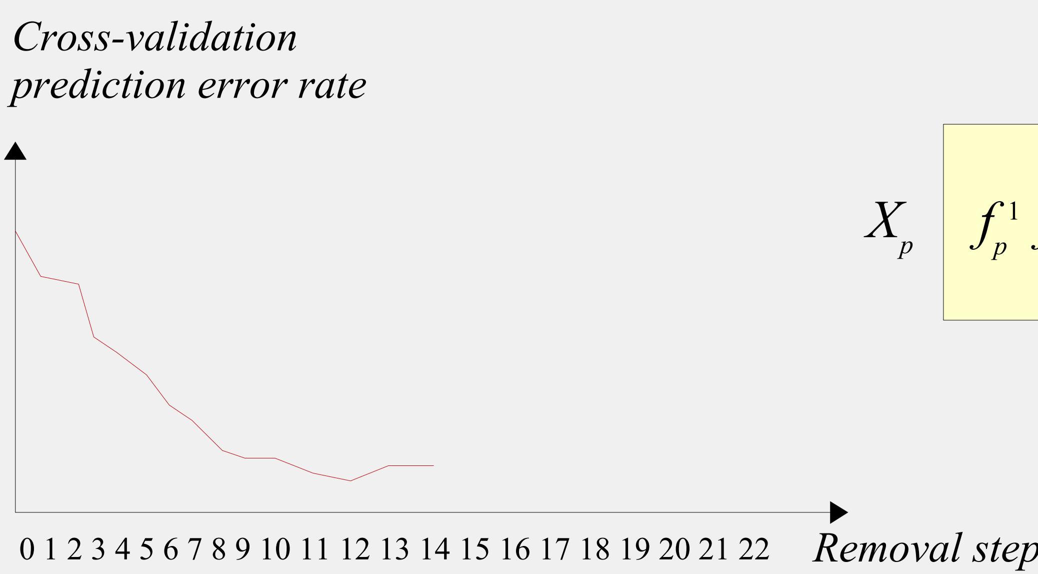
How THEME works

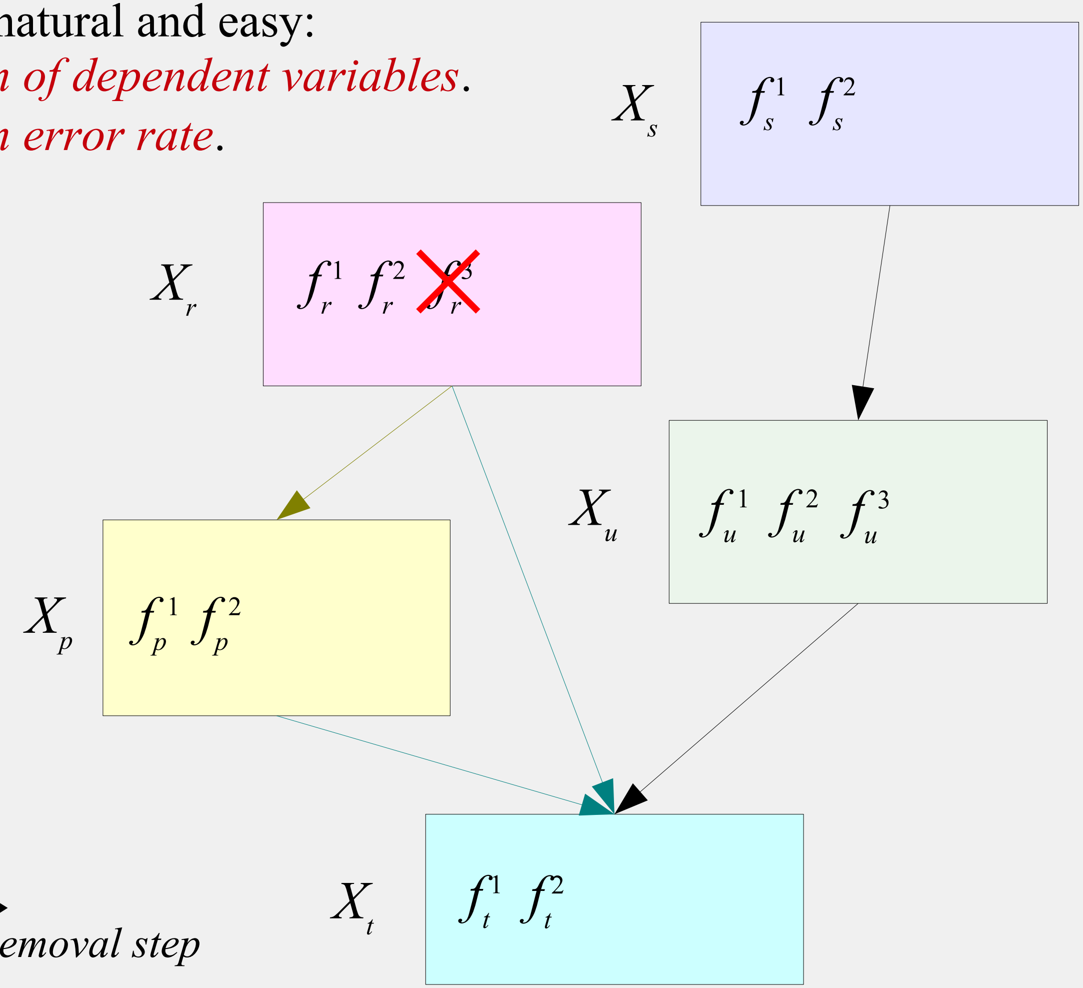
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:
Explanatory components → *prediction of dependent variables*.
→ *prediction error rate*.

Example:



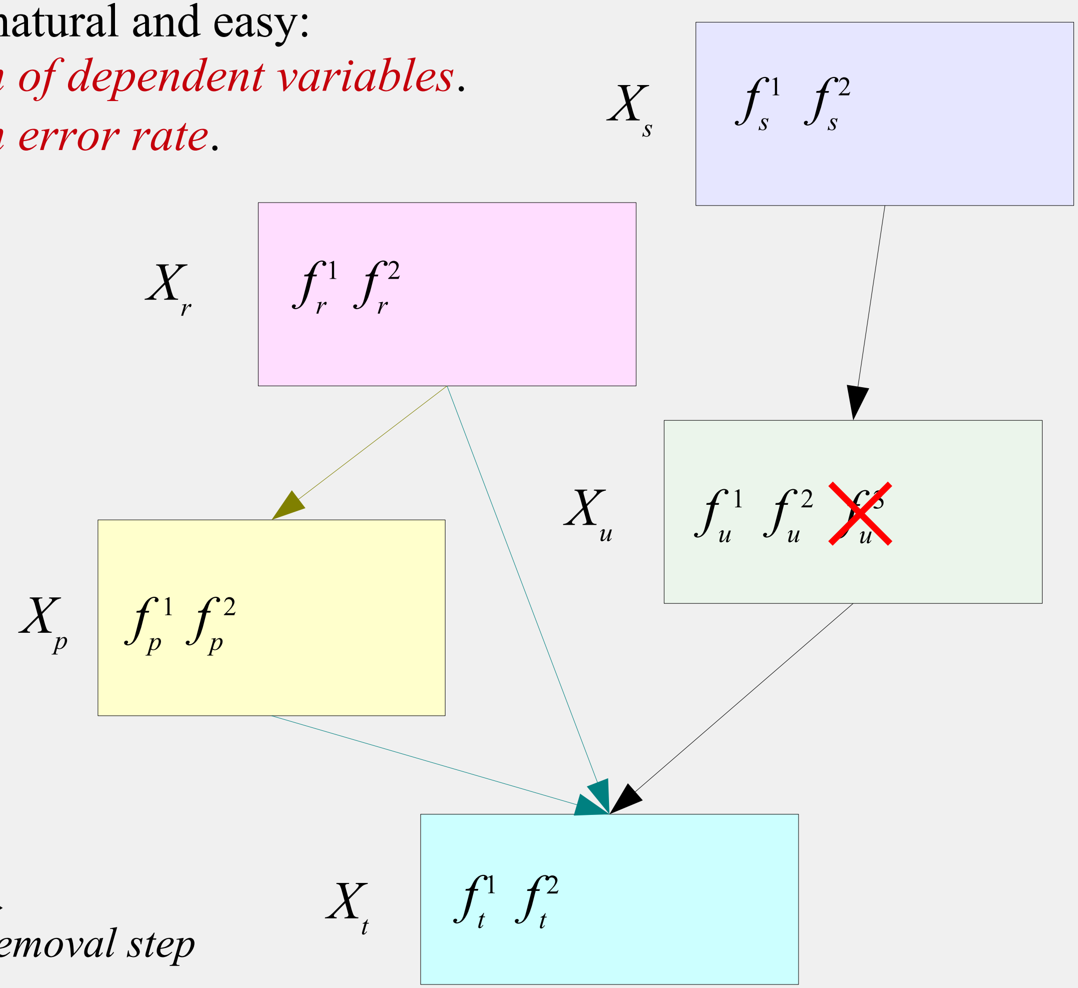


How THEME works

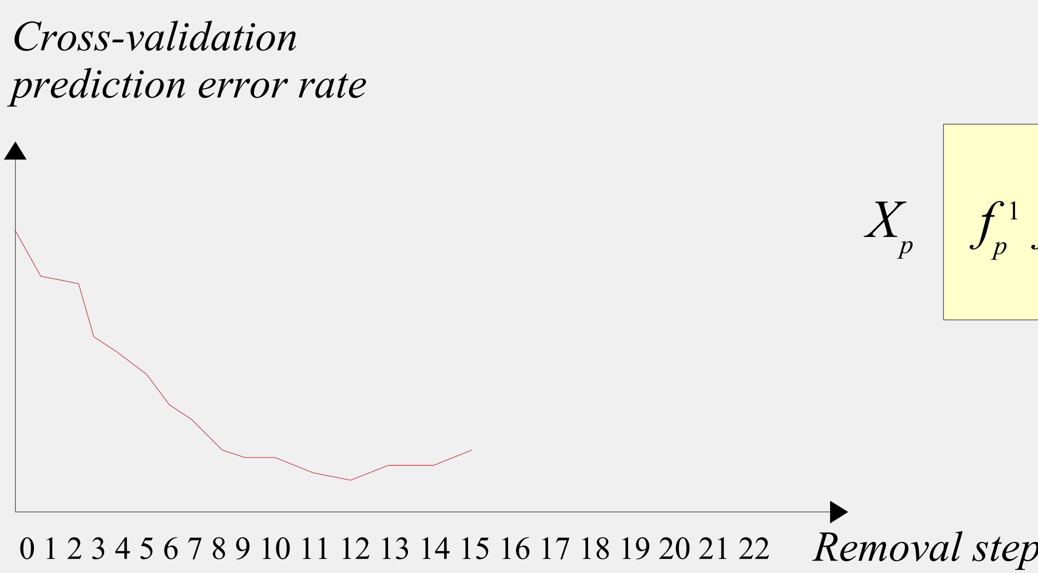
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:
Explanatory components → *prediction of dependent variables*.
→ *prediction error rate*.



Example:



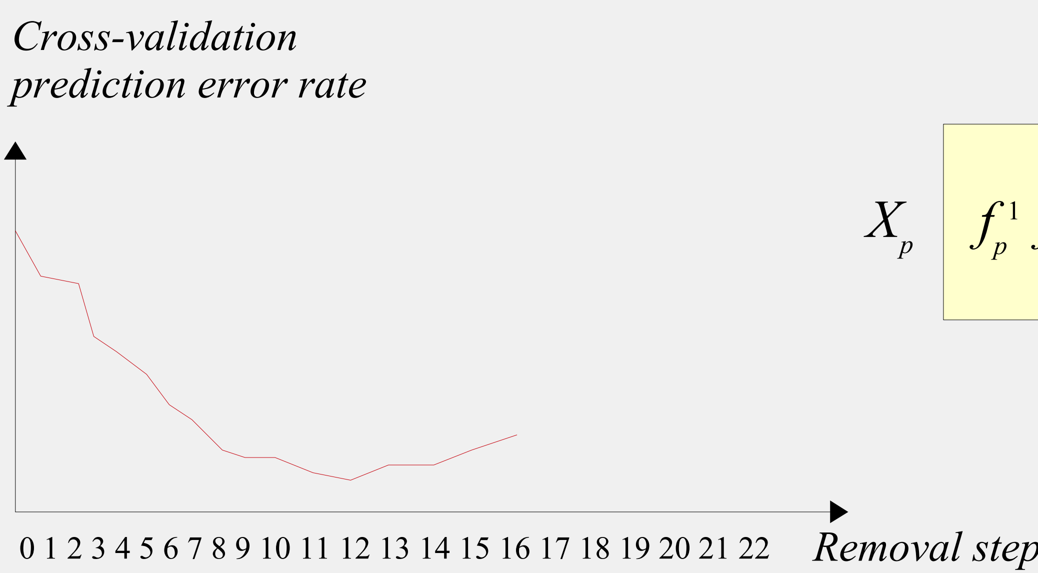
How THEME works

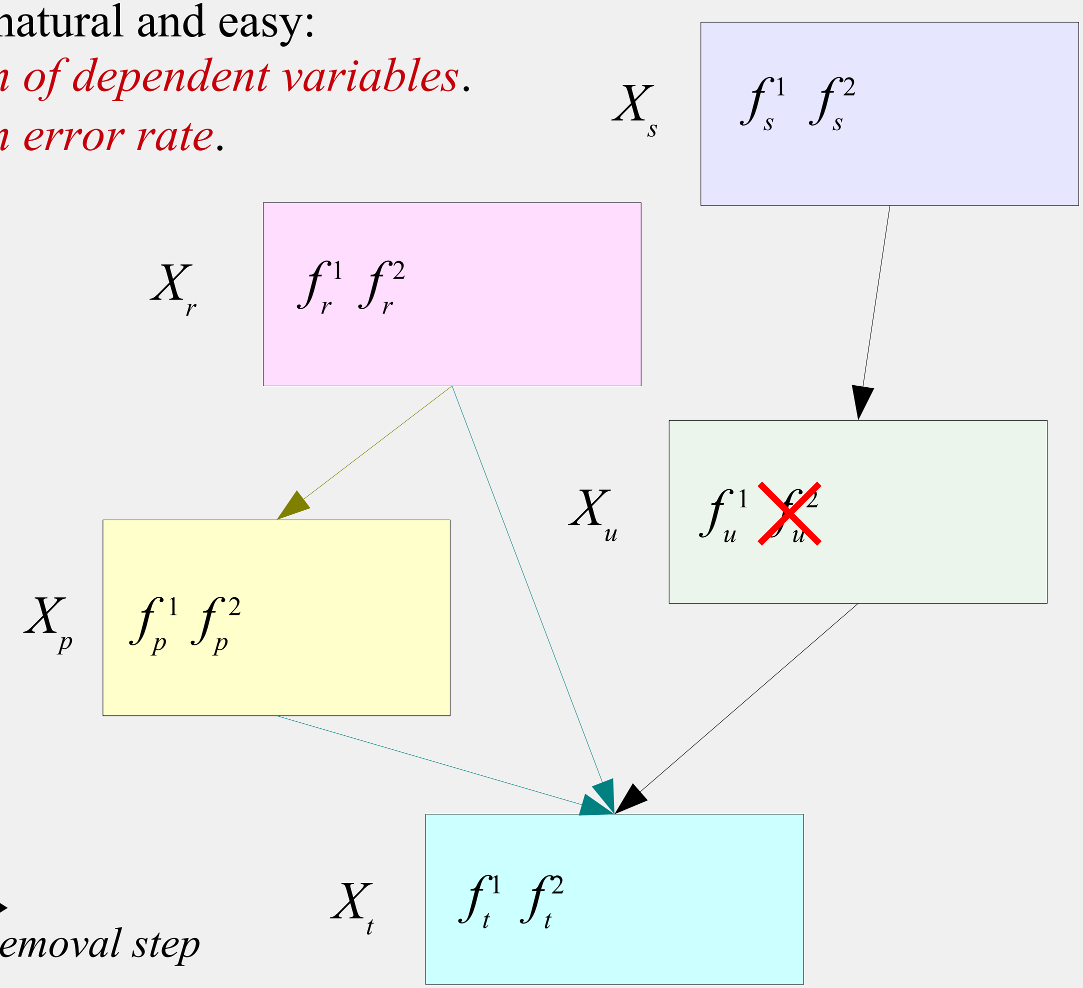
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:
Explanatory components → *prediction of dependent variables*.
→ *prediction error rate*.

Example:



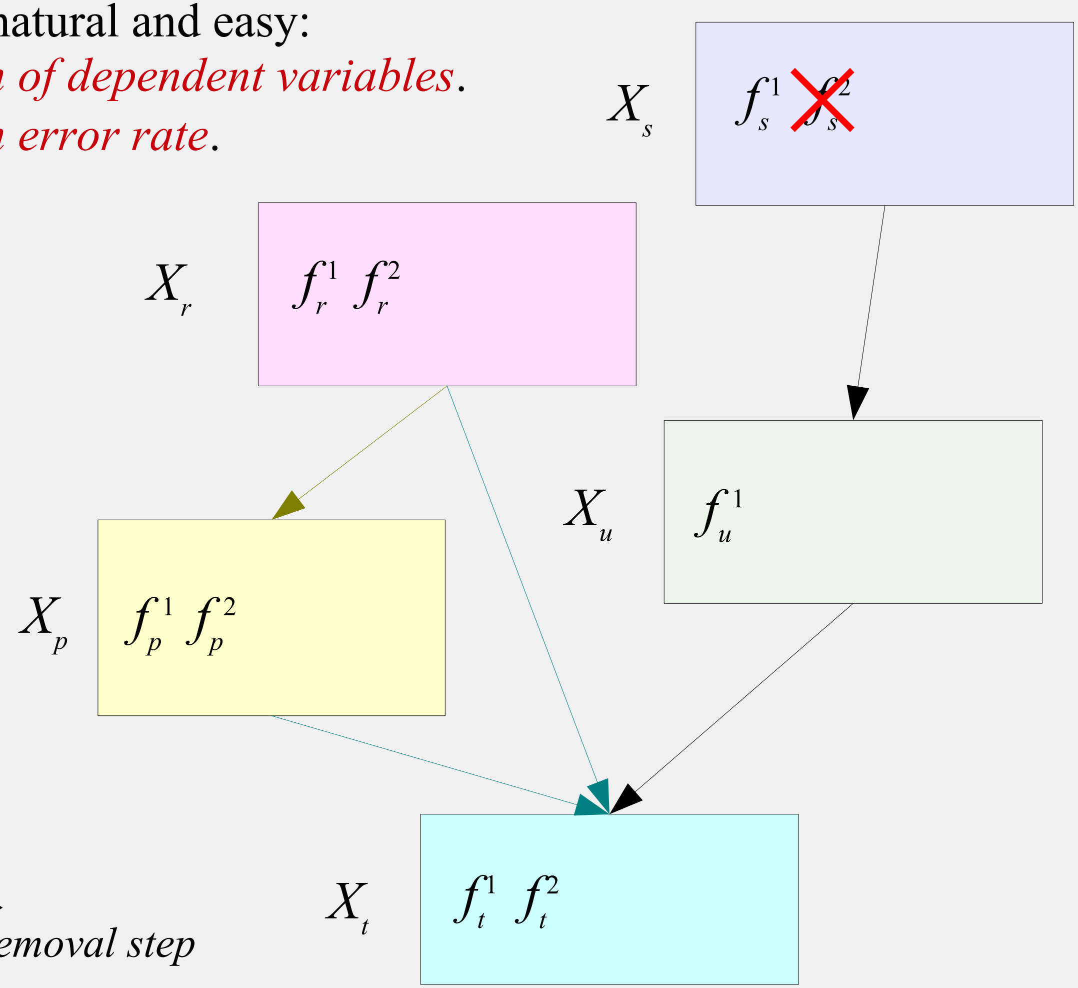


How THEME works

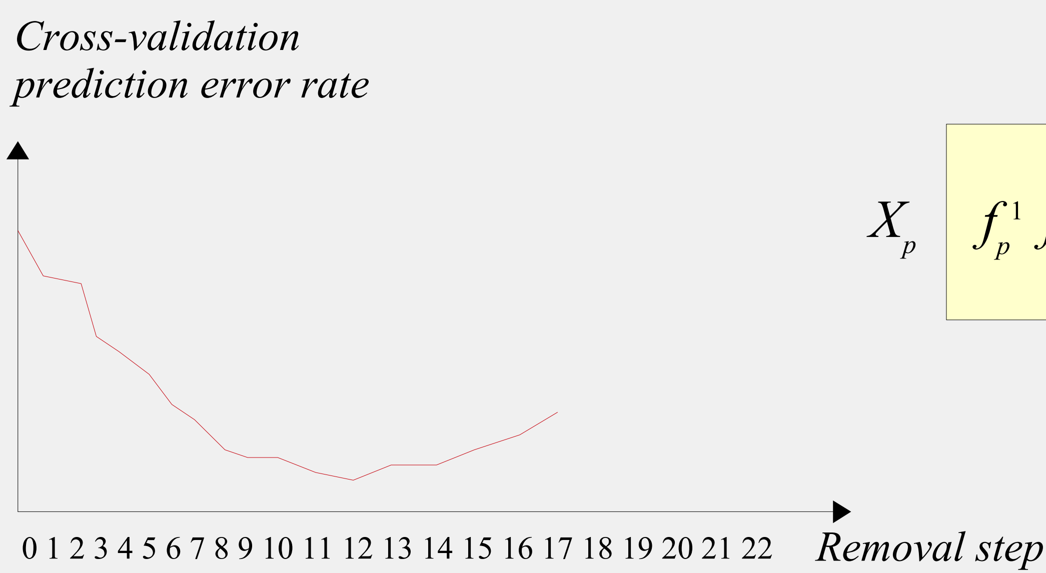
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:
Explanatory components → *prediction of dependent variables*.
→ *prediction error rate*.



Example:



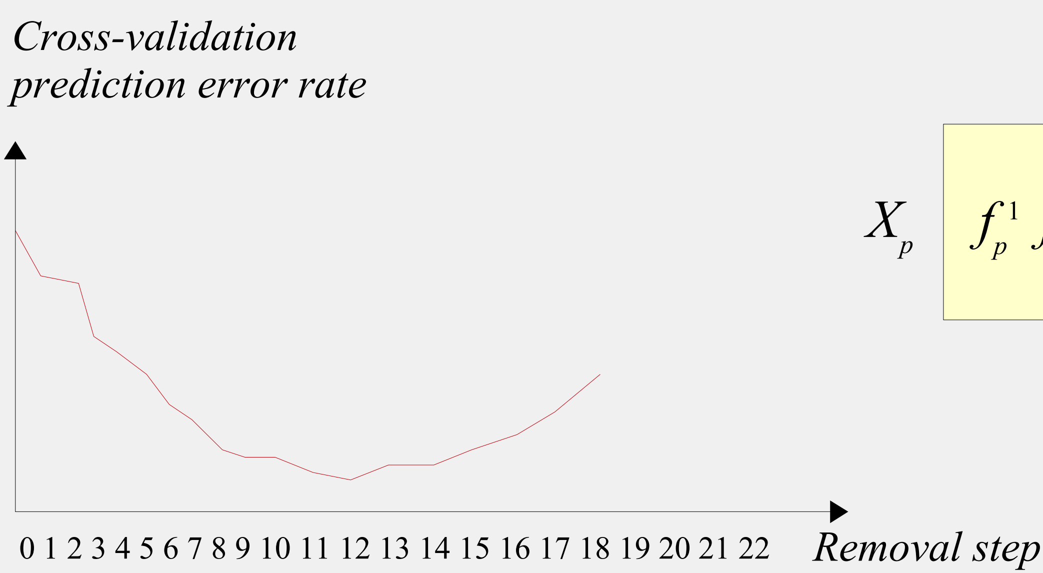
How THEME works

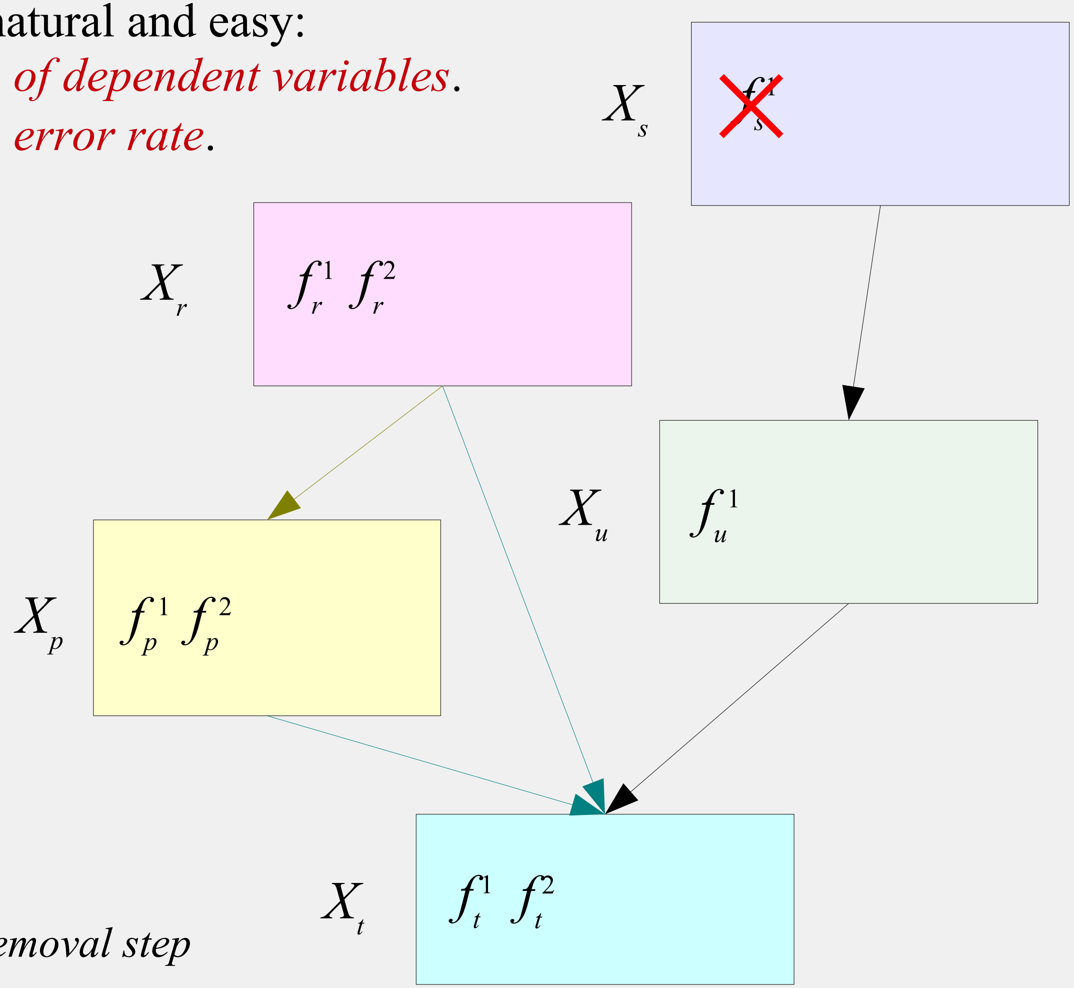
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:
Explanatory components → *prediction of dependent variables*.
→ *prediction error rate*.

Example:



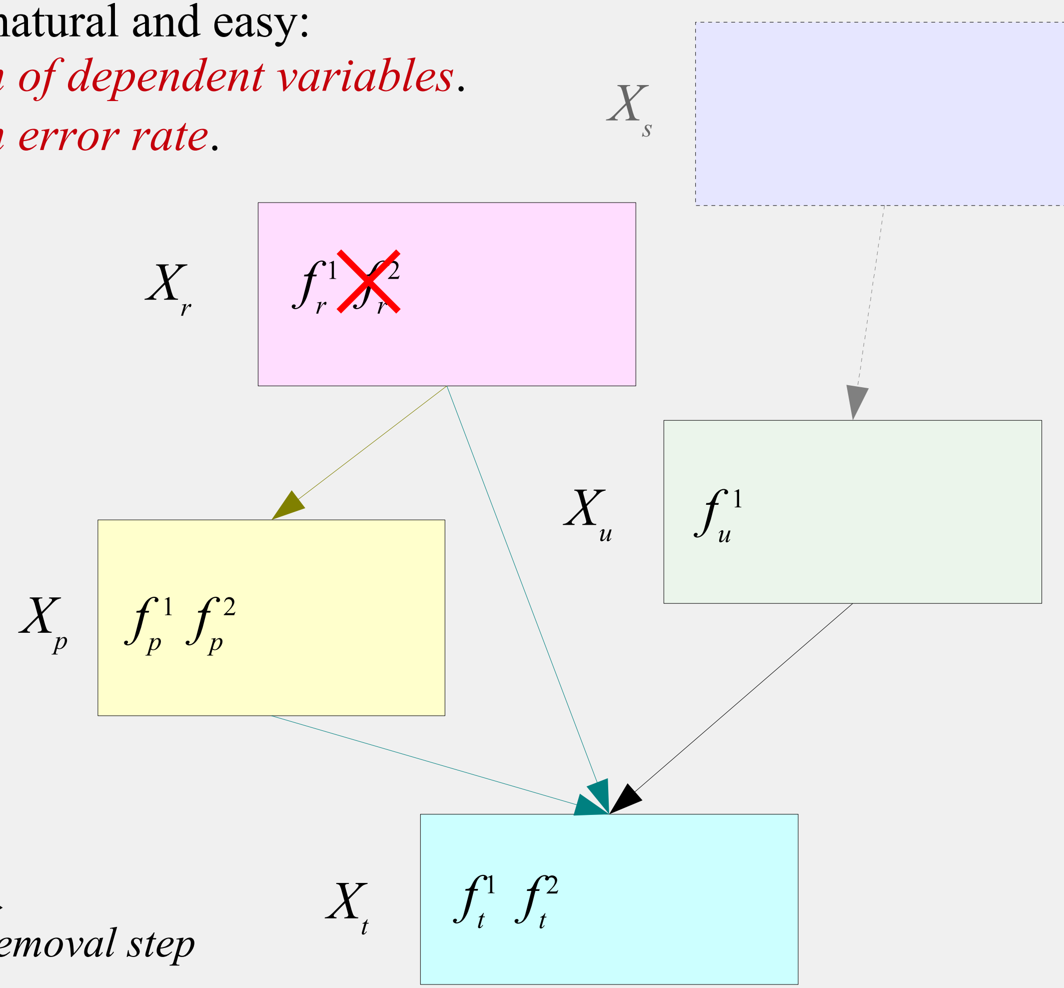


How THEME works

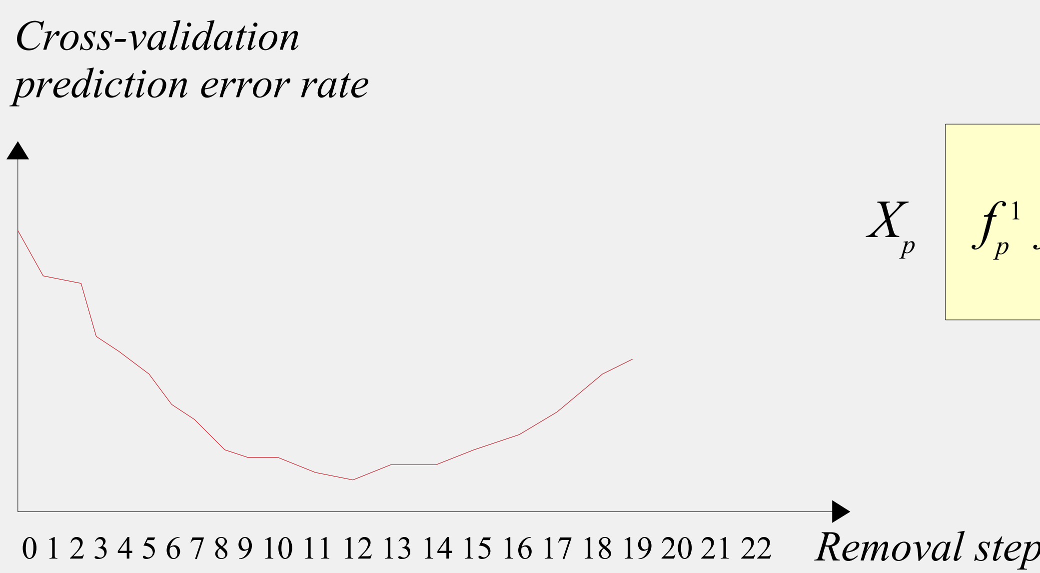
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:
Explanatory components → *prediction of dependent variables*.
→ *prediction error rate*.



Example:



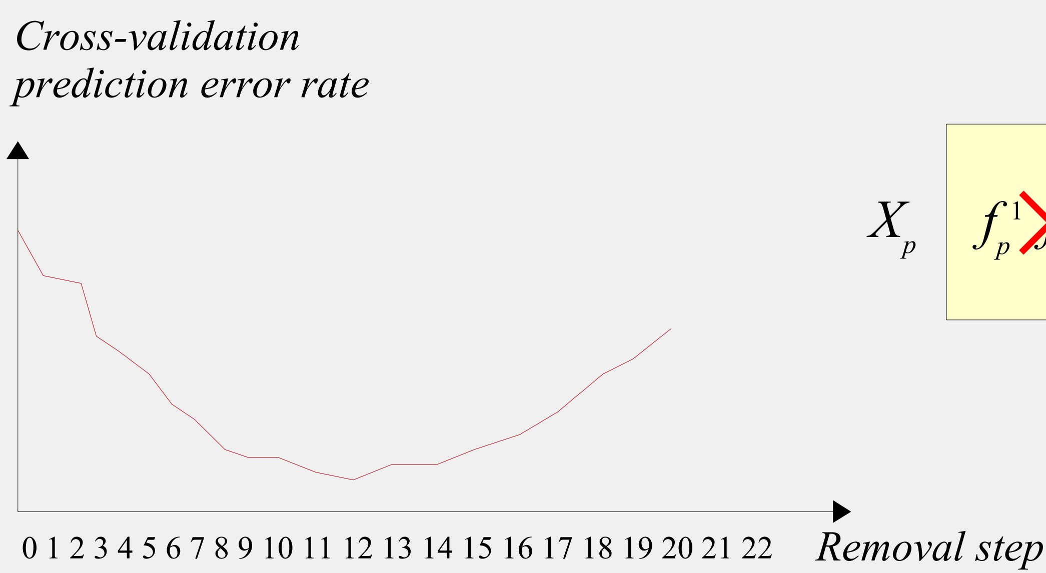
How THEME works

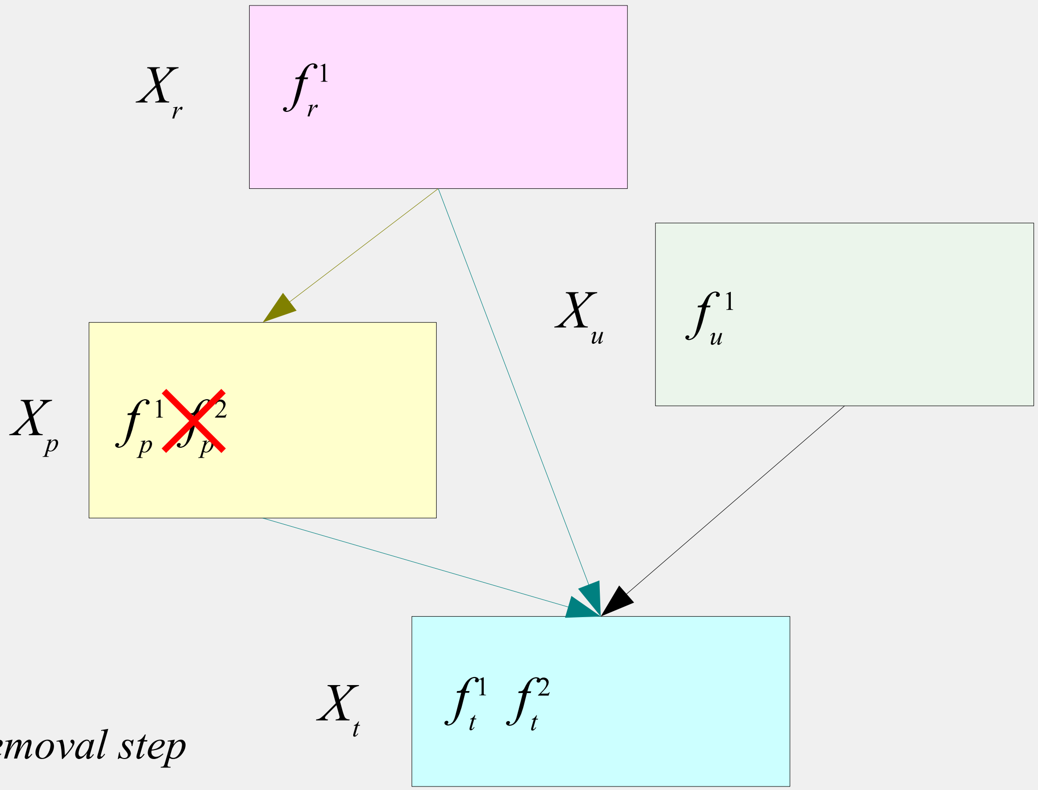
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:
Explanatory components → *prediction of dependent variables*.
→ *prediction error rate*.

Example:





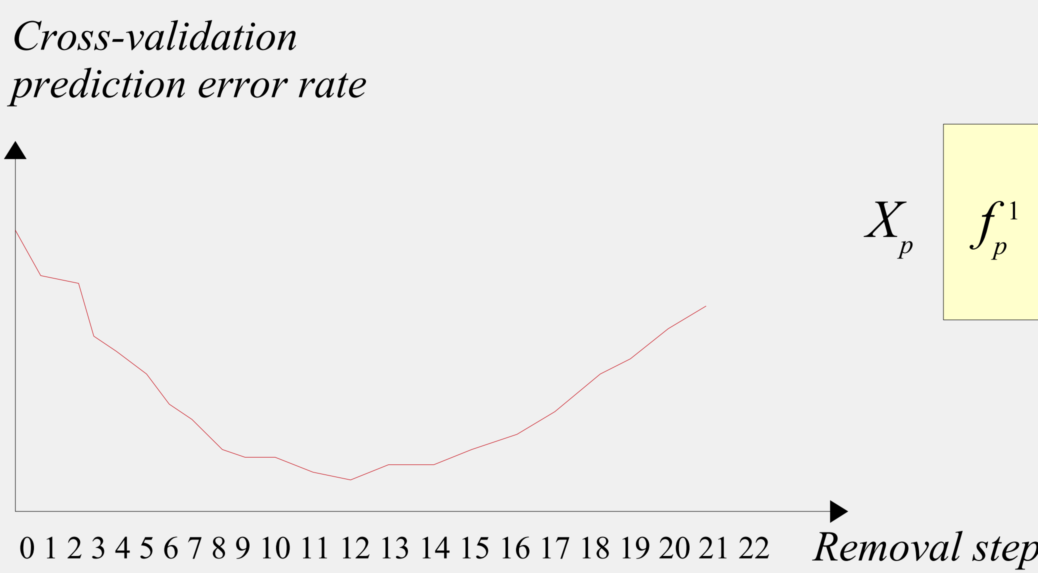
How THEME works

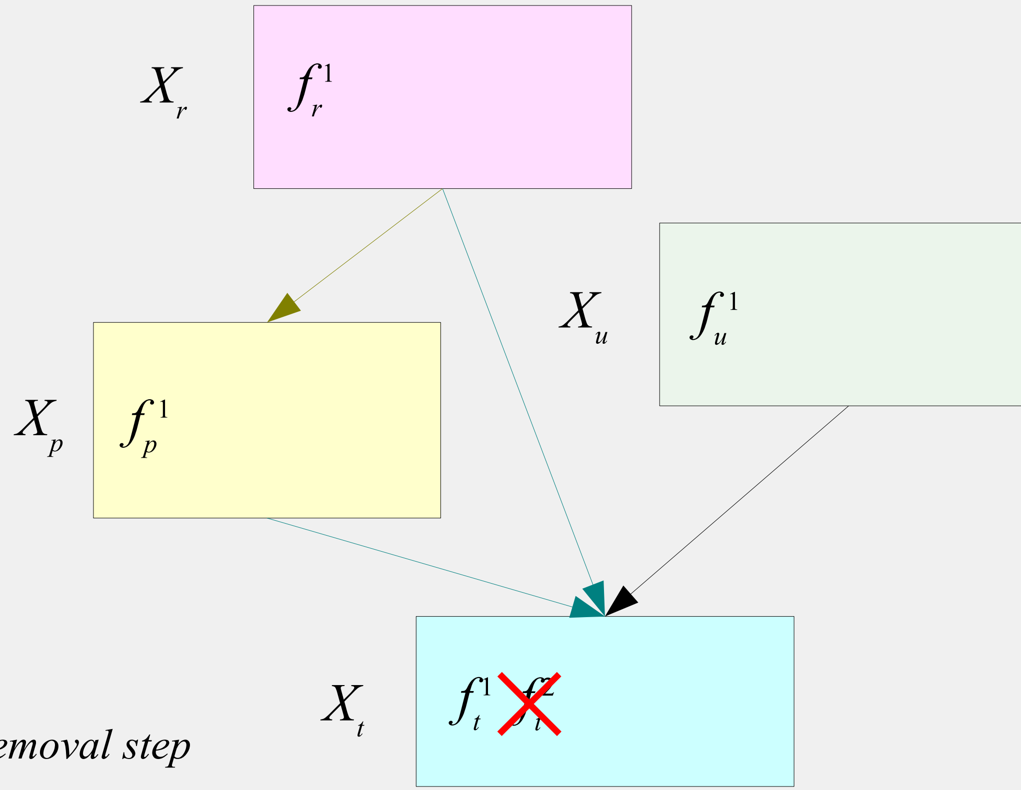
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:
Explanatory components → *prediction of dependent variables*.
→ *prediction error rate*.

Example:





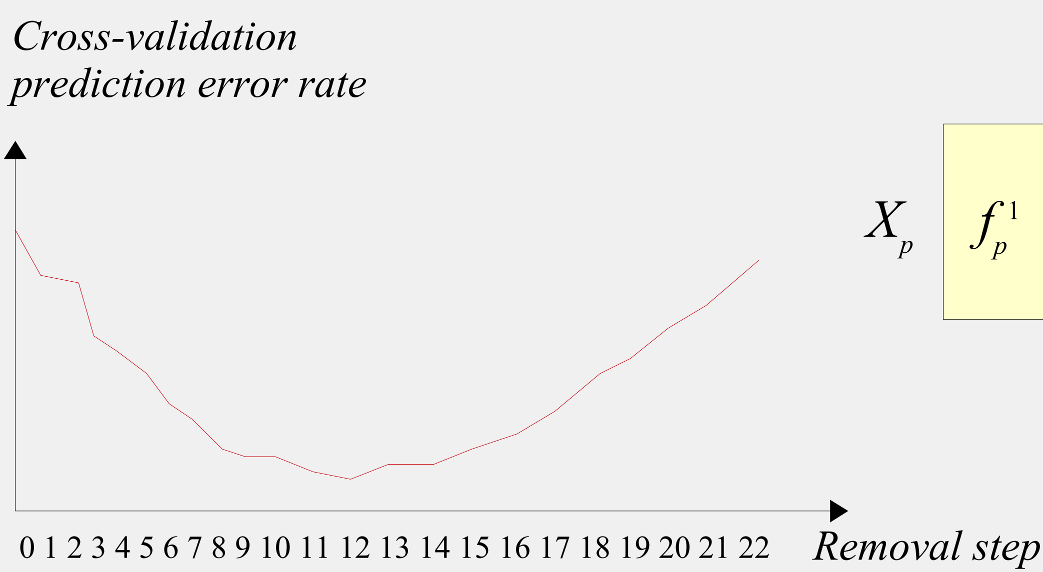
How THEME works

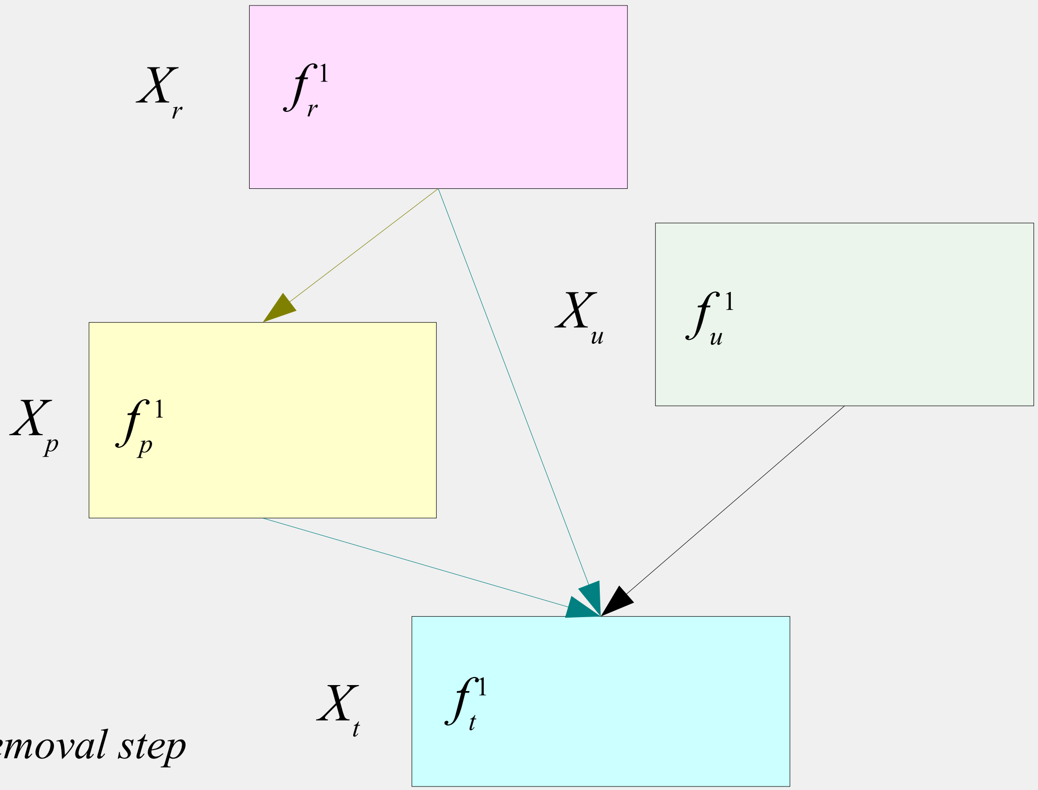
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:
Explanatory components → *prediction of dependent variables*.
→ *prediction error rate*.

Example:





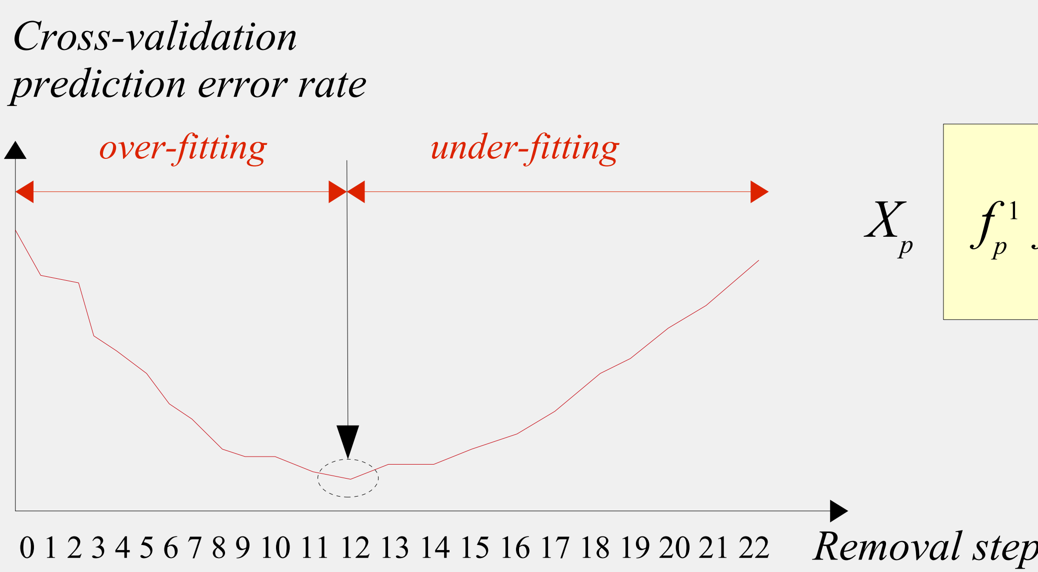
How THEME works

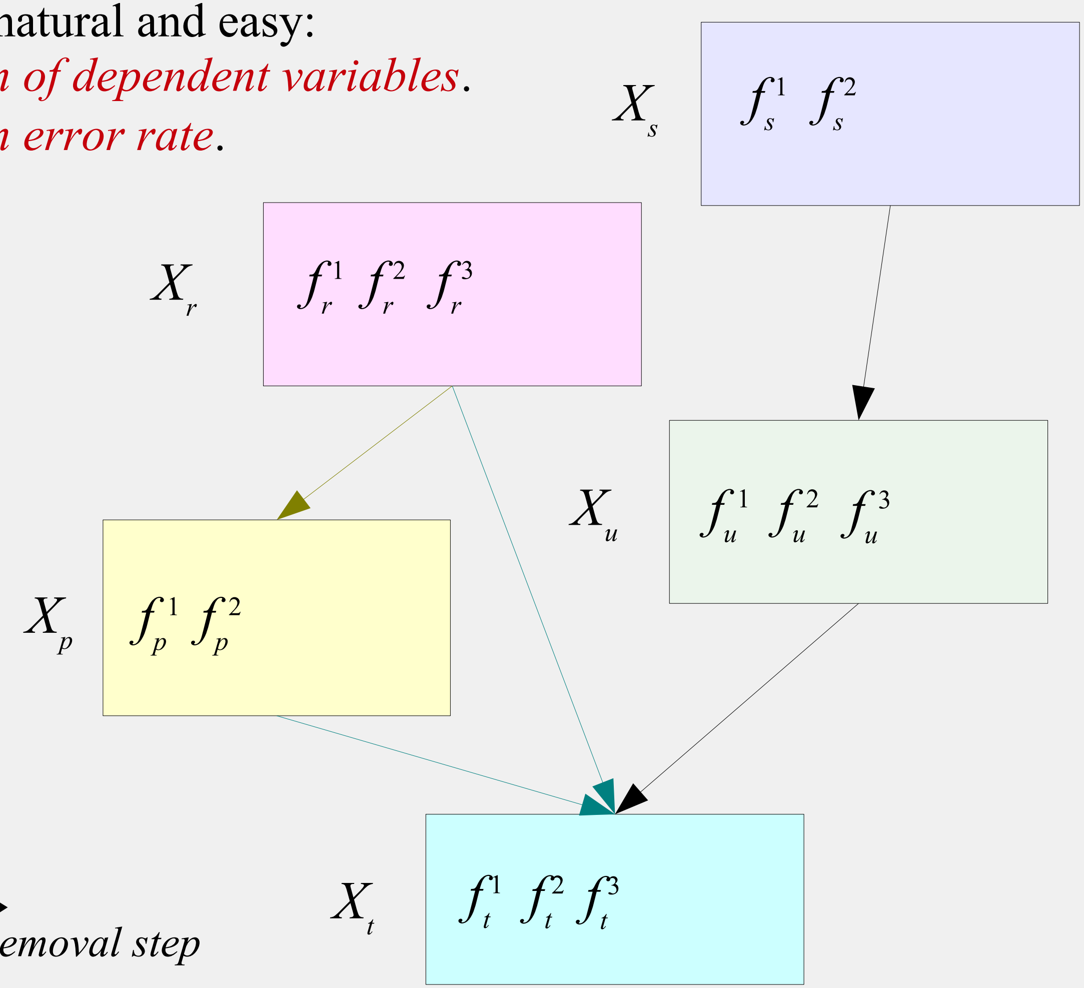
5. Backward component selection

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

Local nesting makes ***backward selection*** natural and easy:
Explanatory components → *prediction of dependent variables*.
→ *prediction error rate*.

Example:





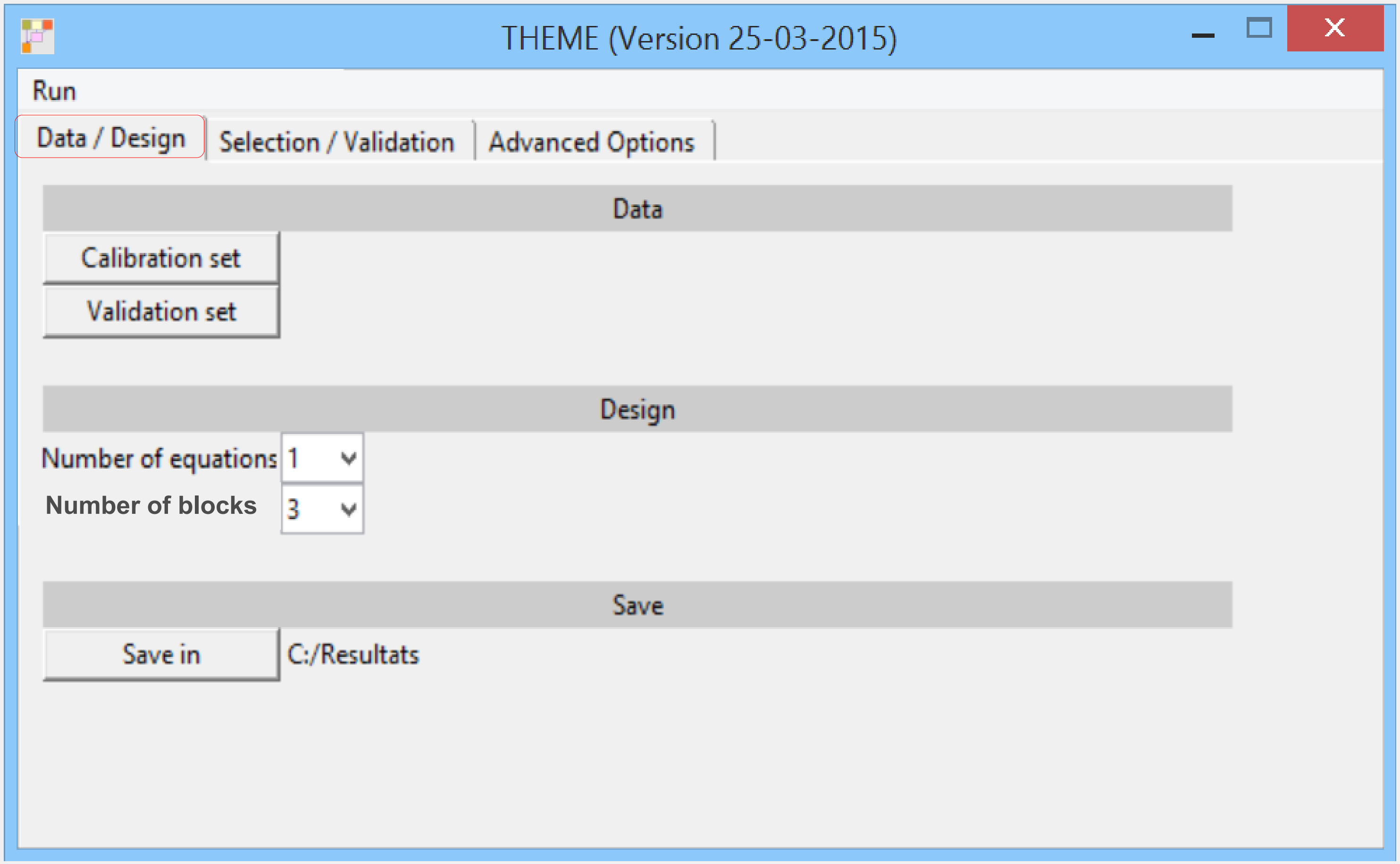
How to operate the THEME R-software?

1. The main window

Sample selection

Model design

Output

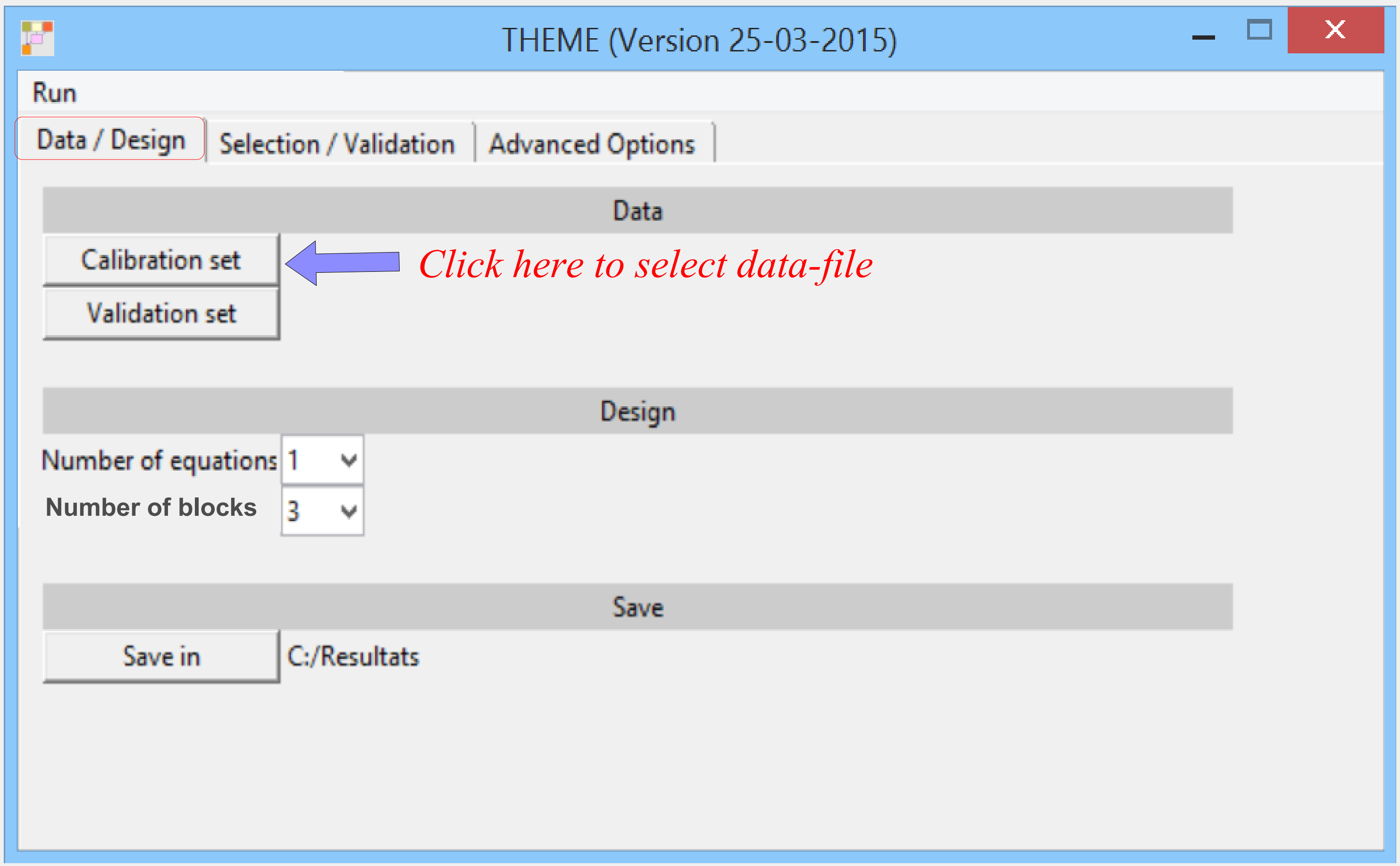


How to operate the THEME R-software?

1. The main window

Data Input

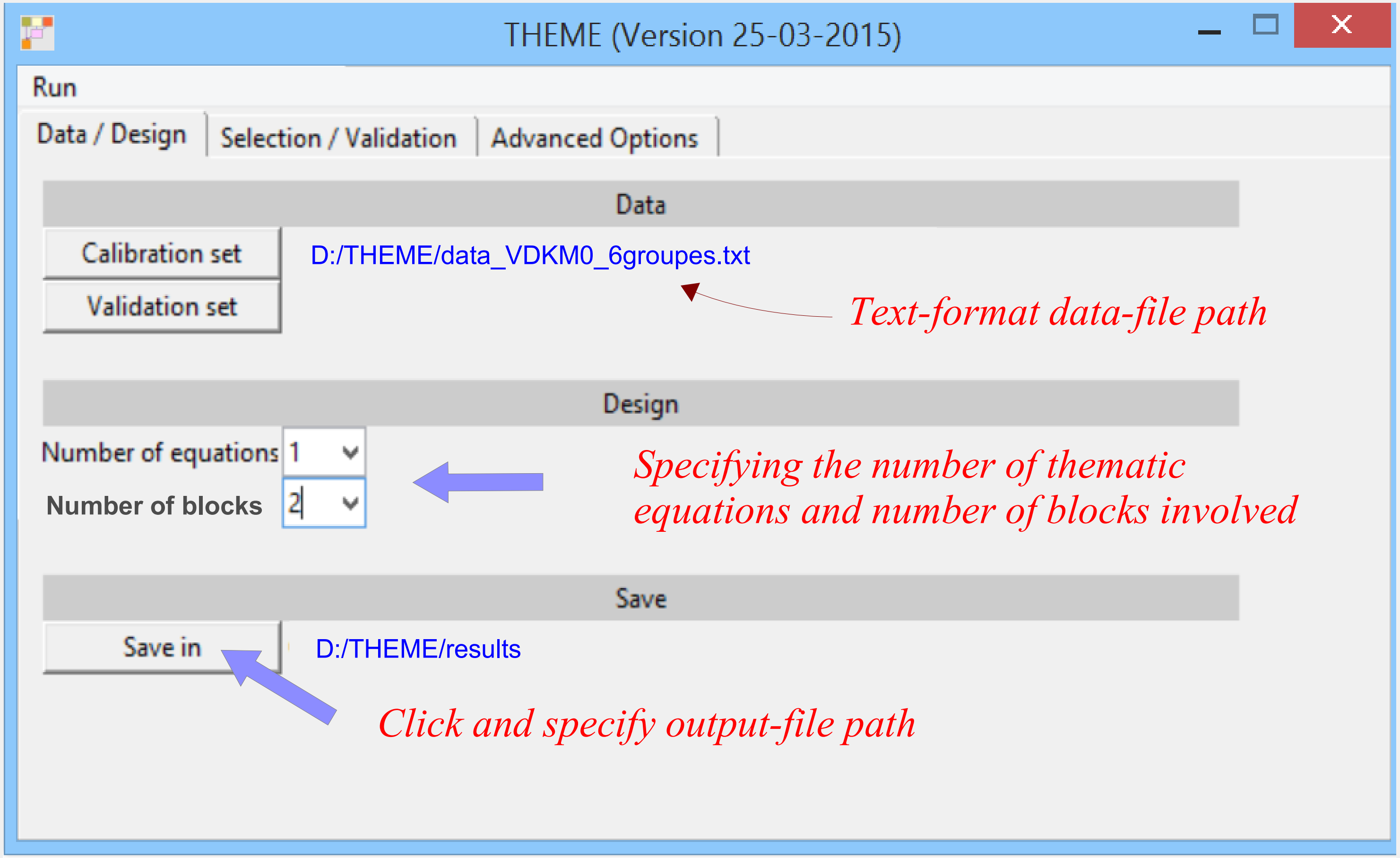
Output



Click here to select data-file

How to operate the THEME R-software?

1. The main window



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*

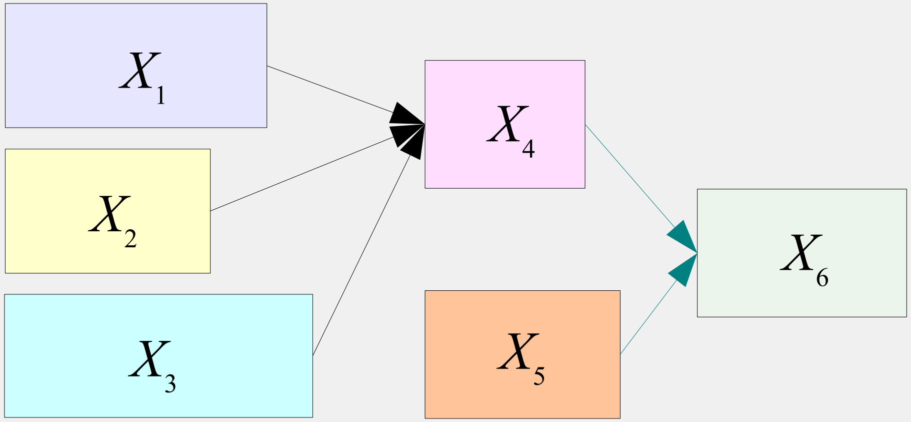
- ***Data file*** = ASCII-file with tab separator: data_VDKM0_6groupes.txt

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:***



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

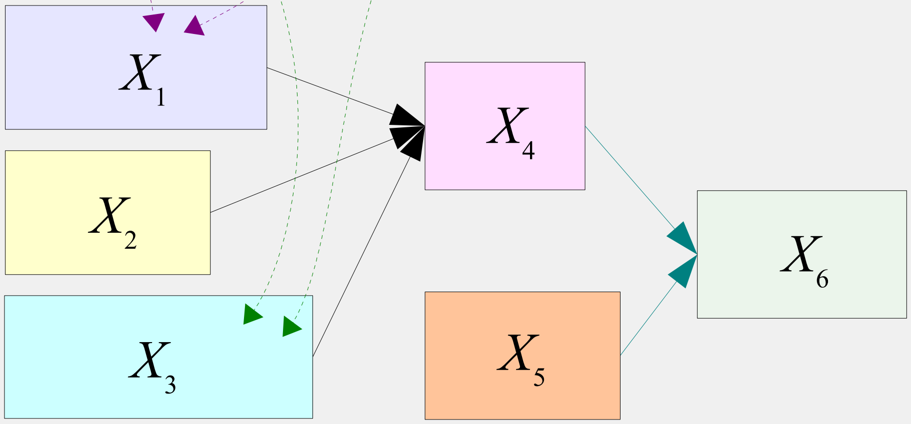
Variables

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*)

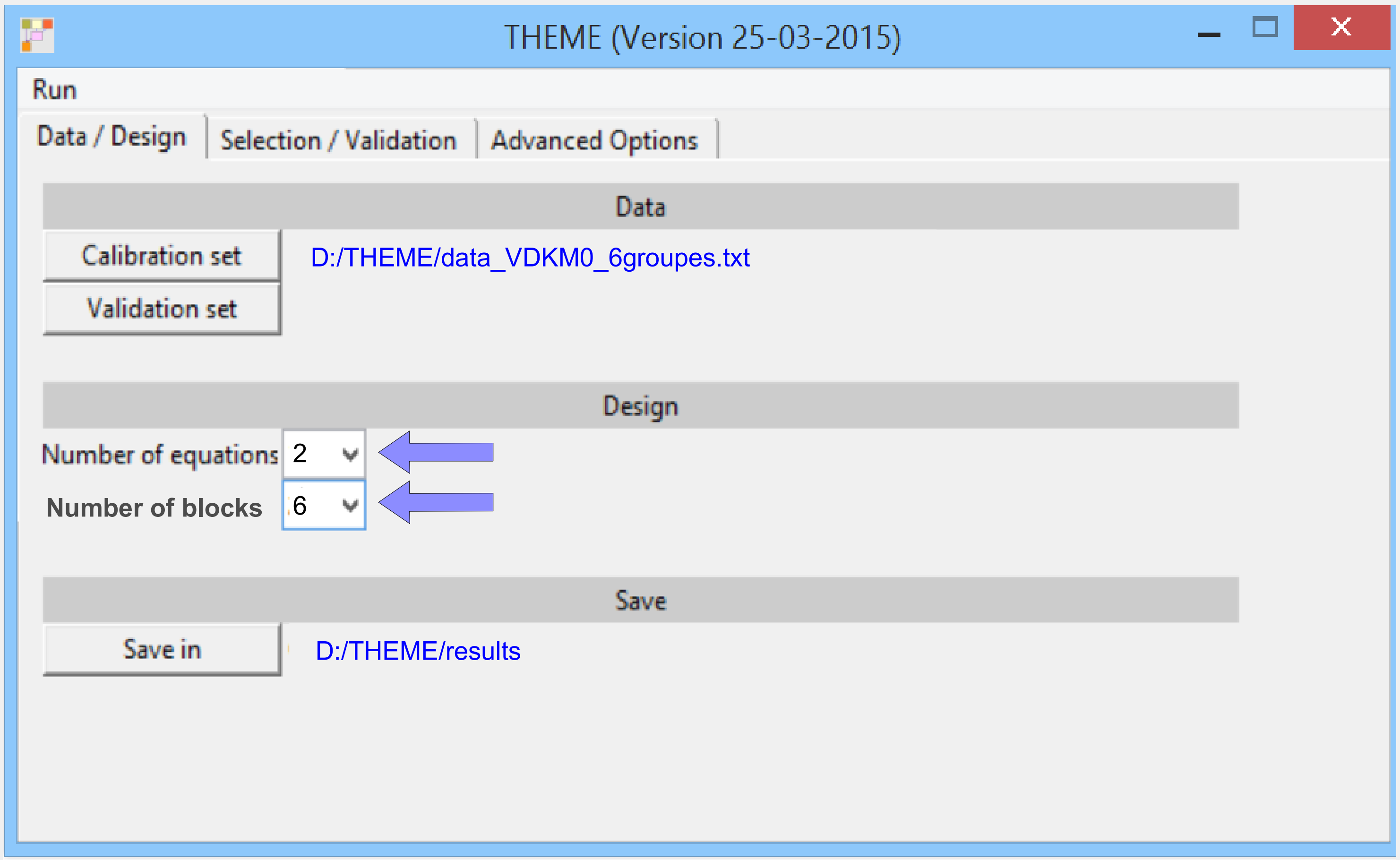
- ***Design of the thematic model:***



How to operate the THEME R-software?

2. From raw data to Thematic Model

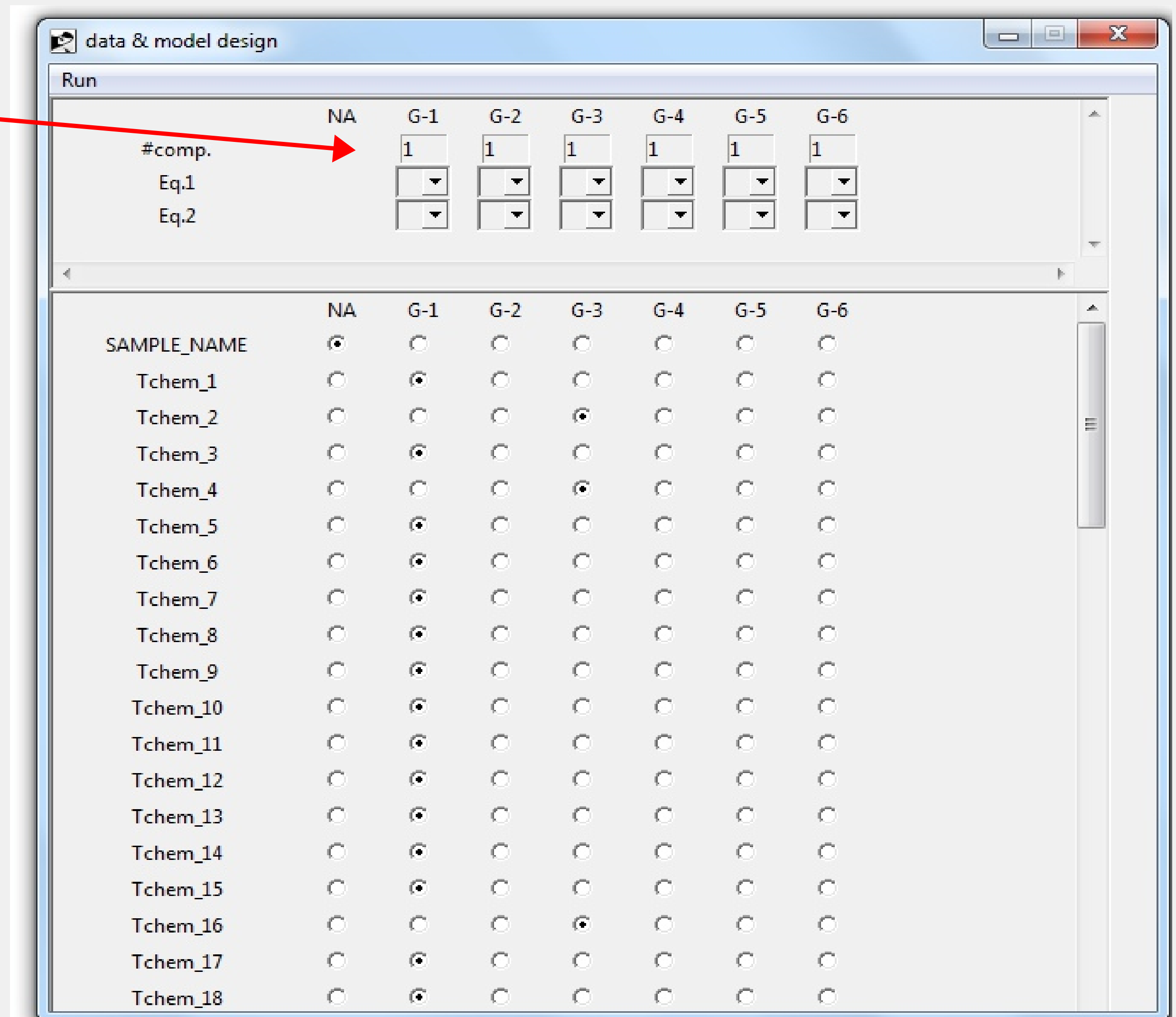
Model design



How to operate the THEME R-software?

2. From raw data to Thematic Model

Number of components in groups

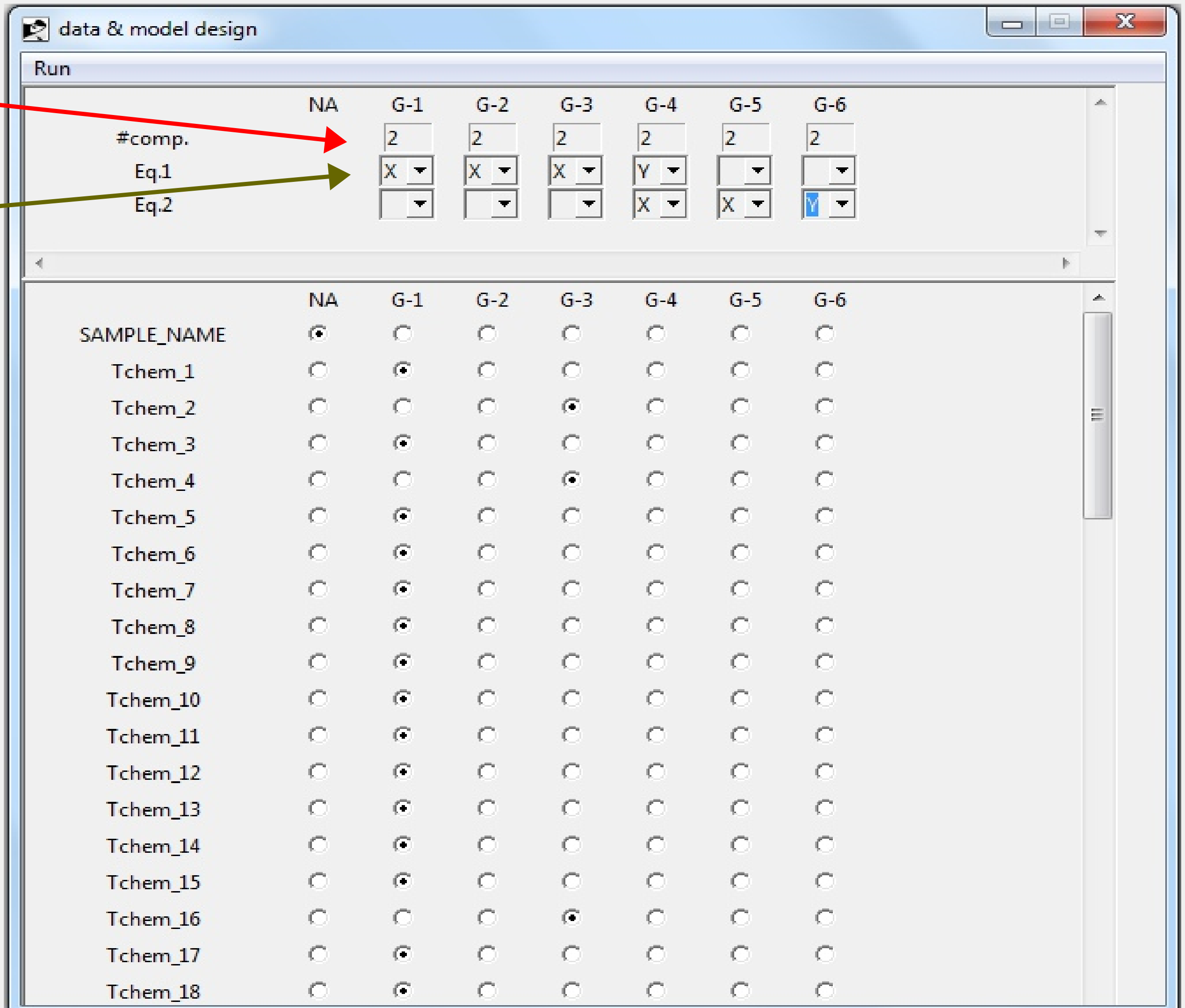


How to operate the THEME R-software?

2. From raw data to Thematic Model

Number of components in groups

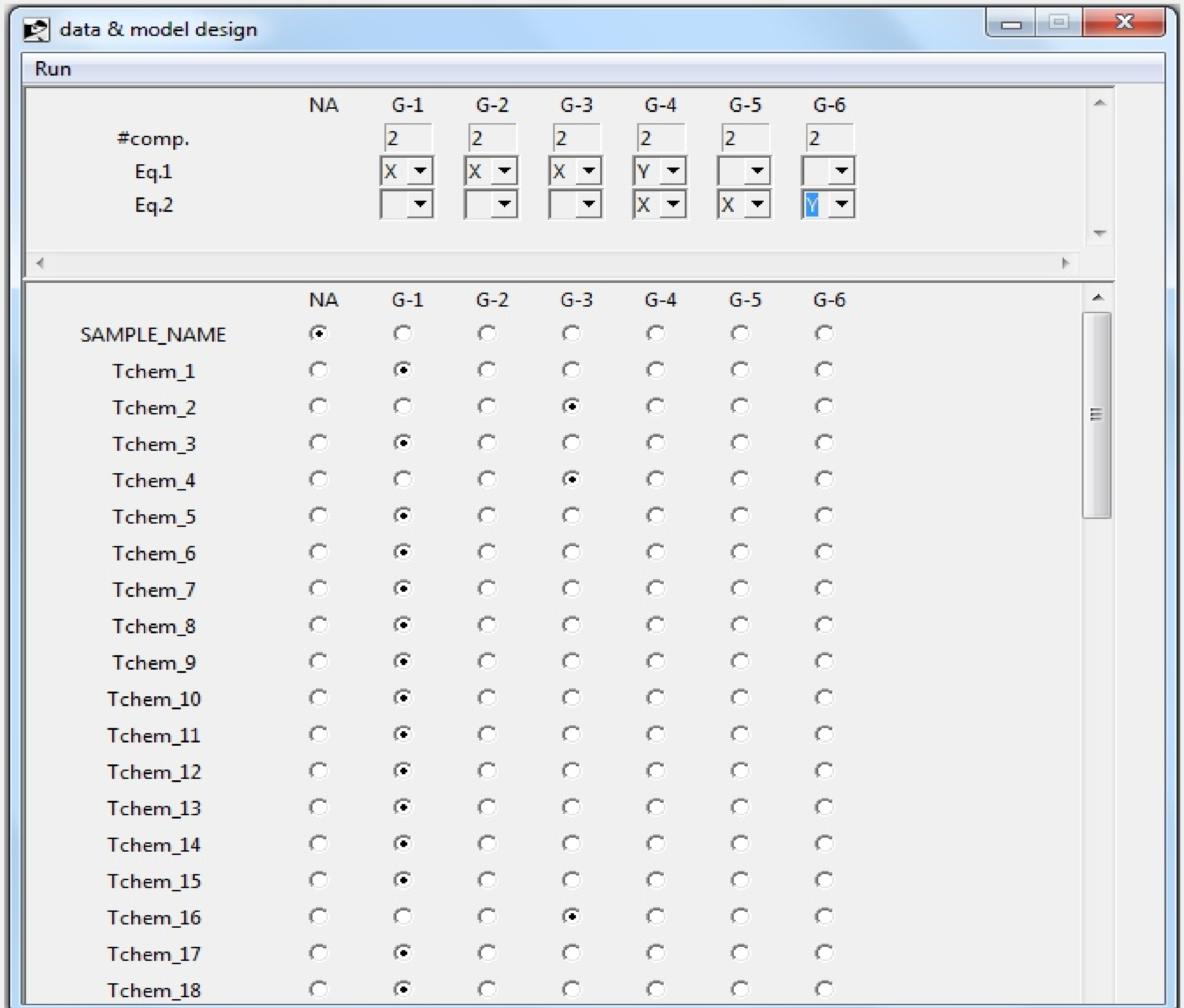
Role of groups in equation (explanatory=X, dependent =Y)



How to operate the THEME R-software?

2. From raw data to Thematic Model

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

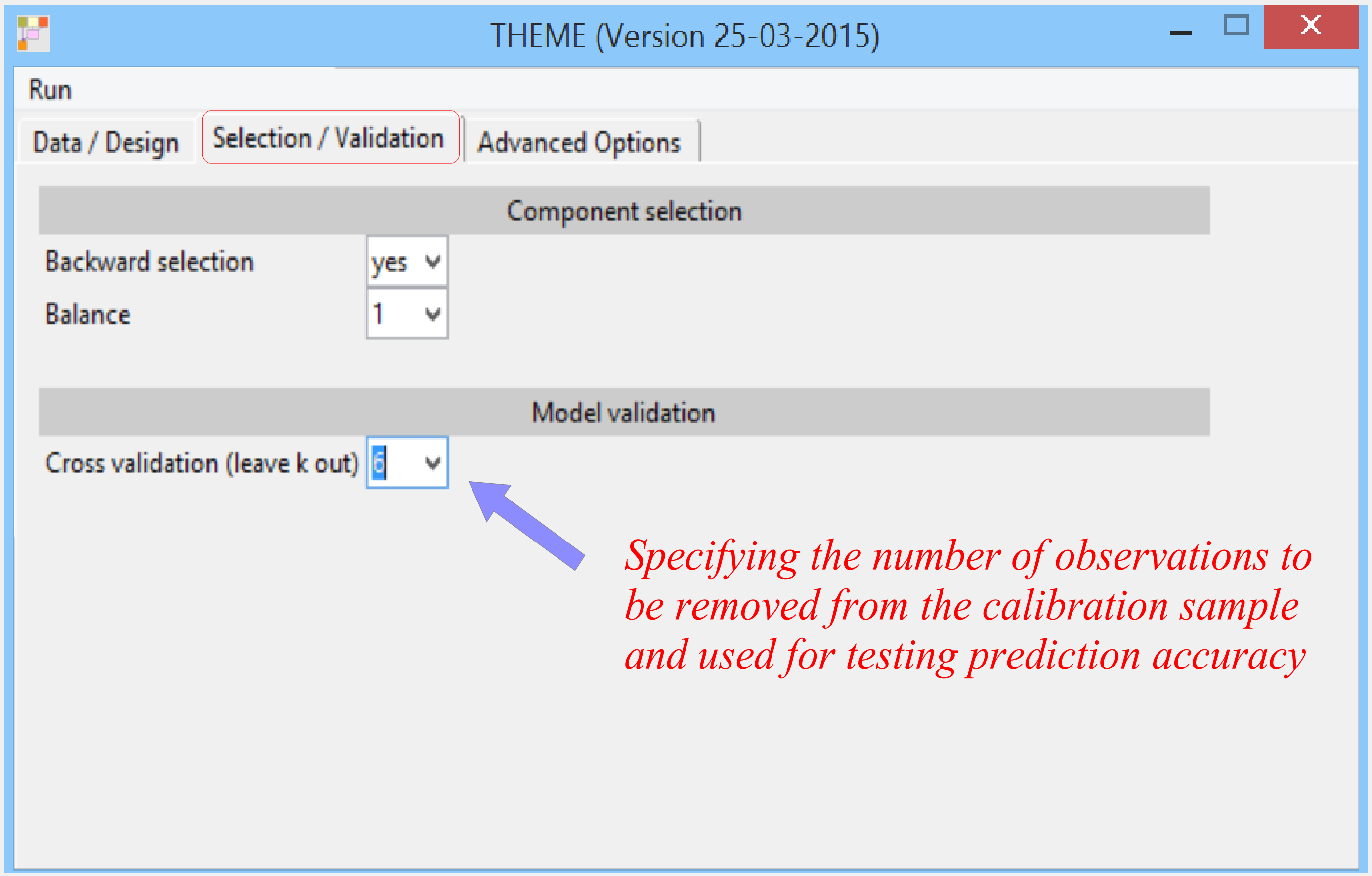


How to operate the THEME R-software?

3. Setting the selection & validation parameters

Component selection

Model cross-validation

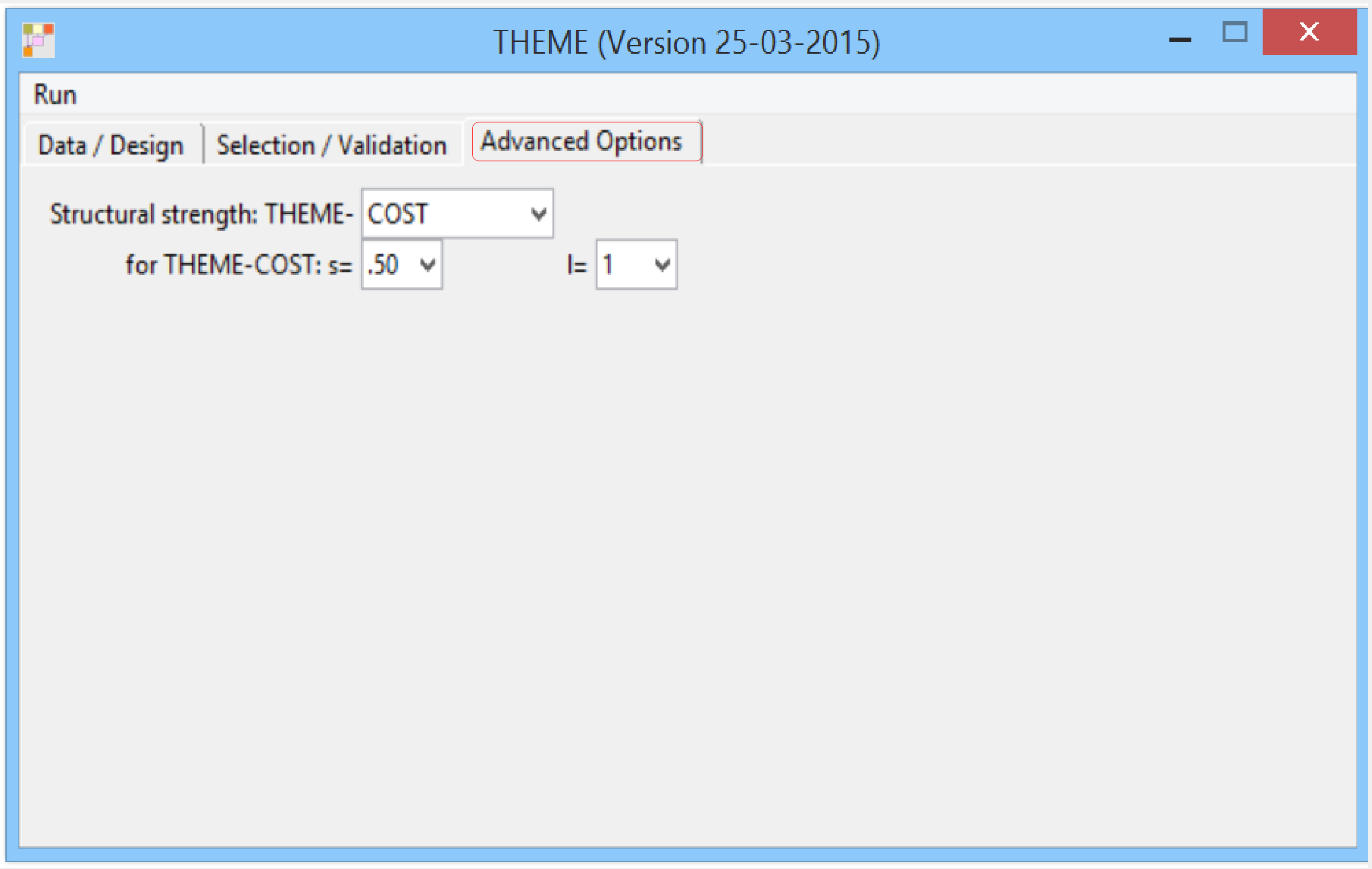


Specifying the number of observations to be removed from the calibration sample and used for testing prediction accuracy

How to operate the THEME R-software?

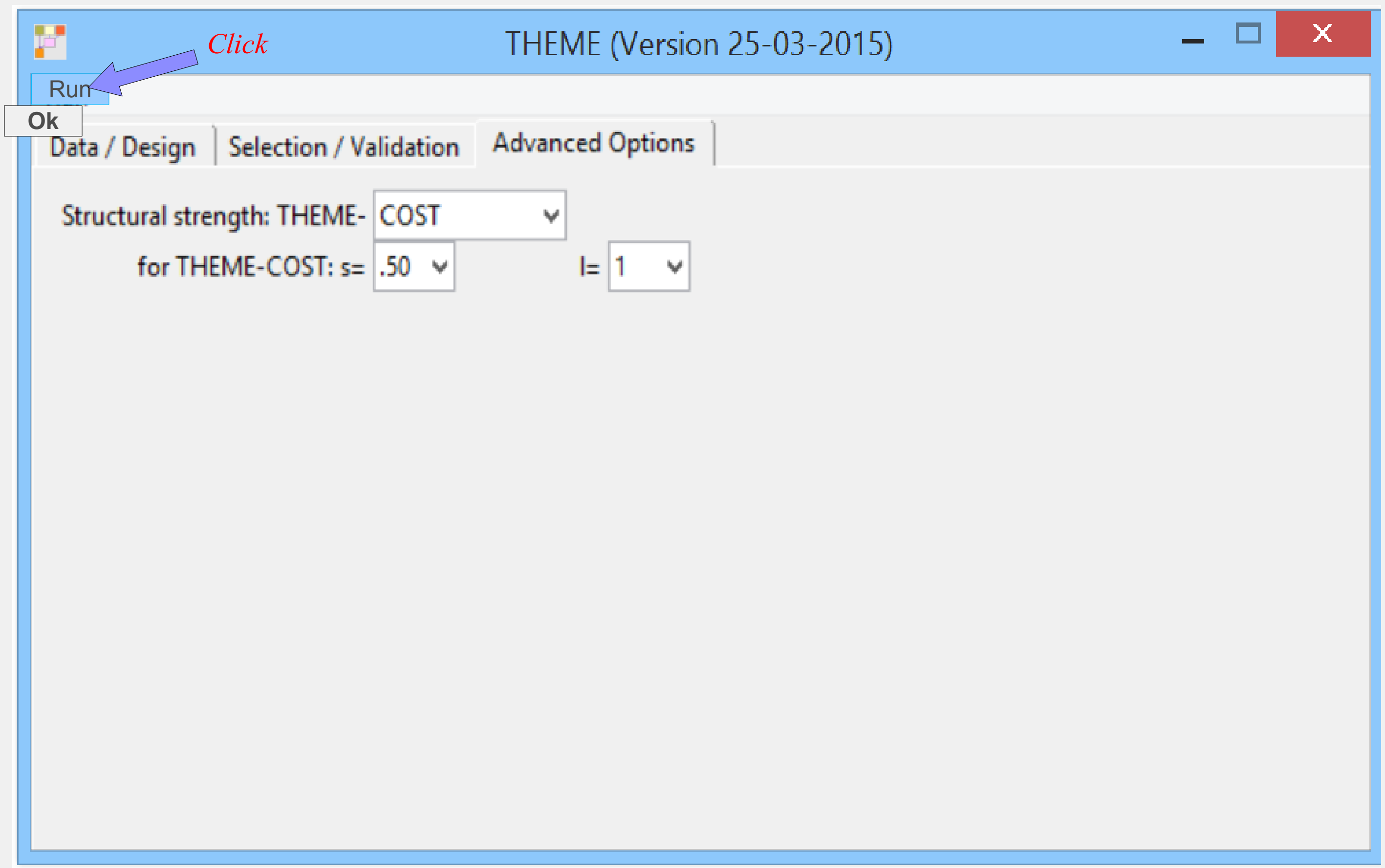
4. Setting the structural strength parameters

Structural strength parameters



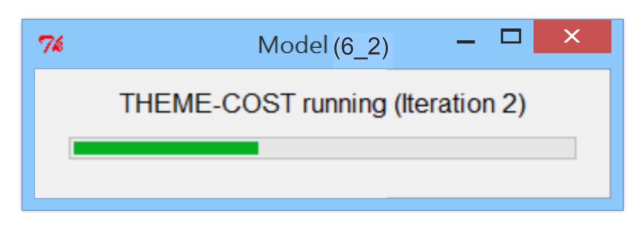
How to operate the THEME R-software?

5. Launching estimation



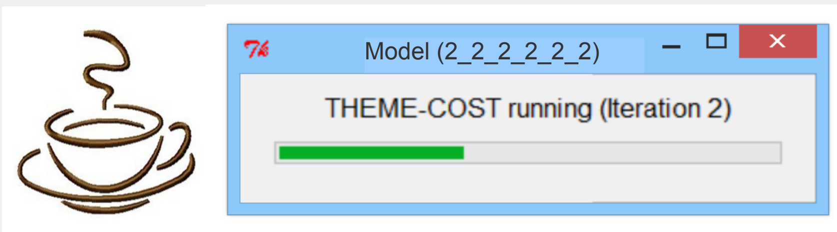
How to operate the THEME R-software?

6. Waiting for results

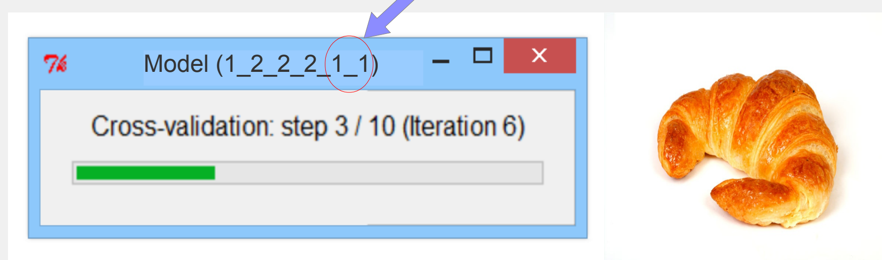


How to operate the THEME R-software?

6. Waiting for results



Number of components decreasing



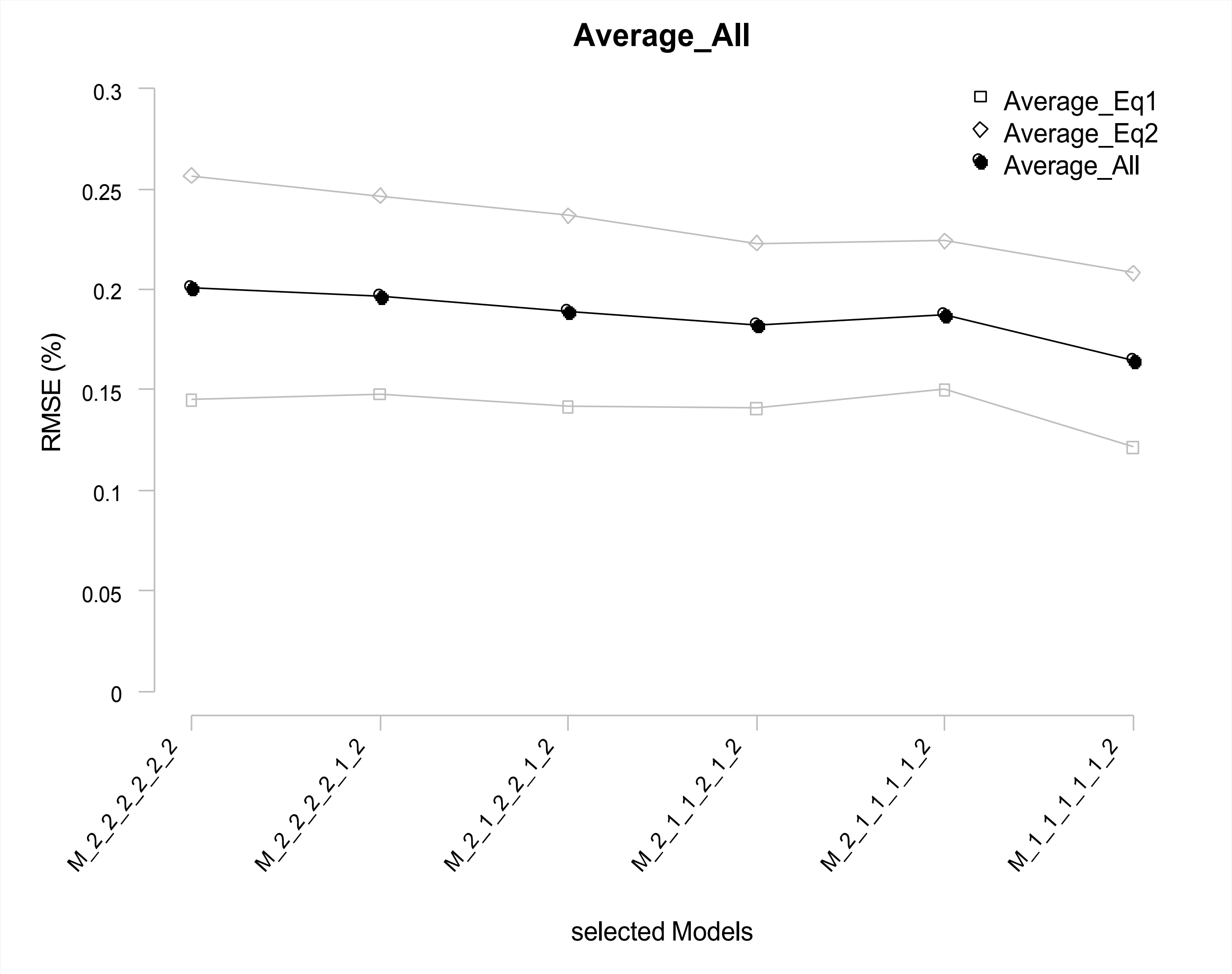
How to operate the THEME R-software?

6. Waiting for results

How to operate the THEME R-software?

7. Reaping results

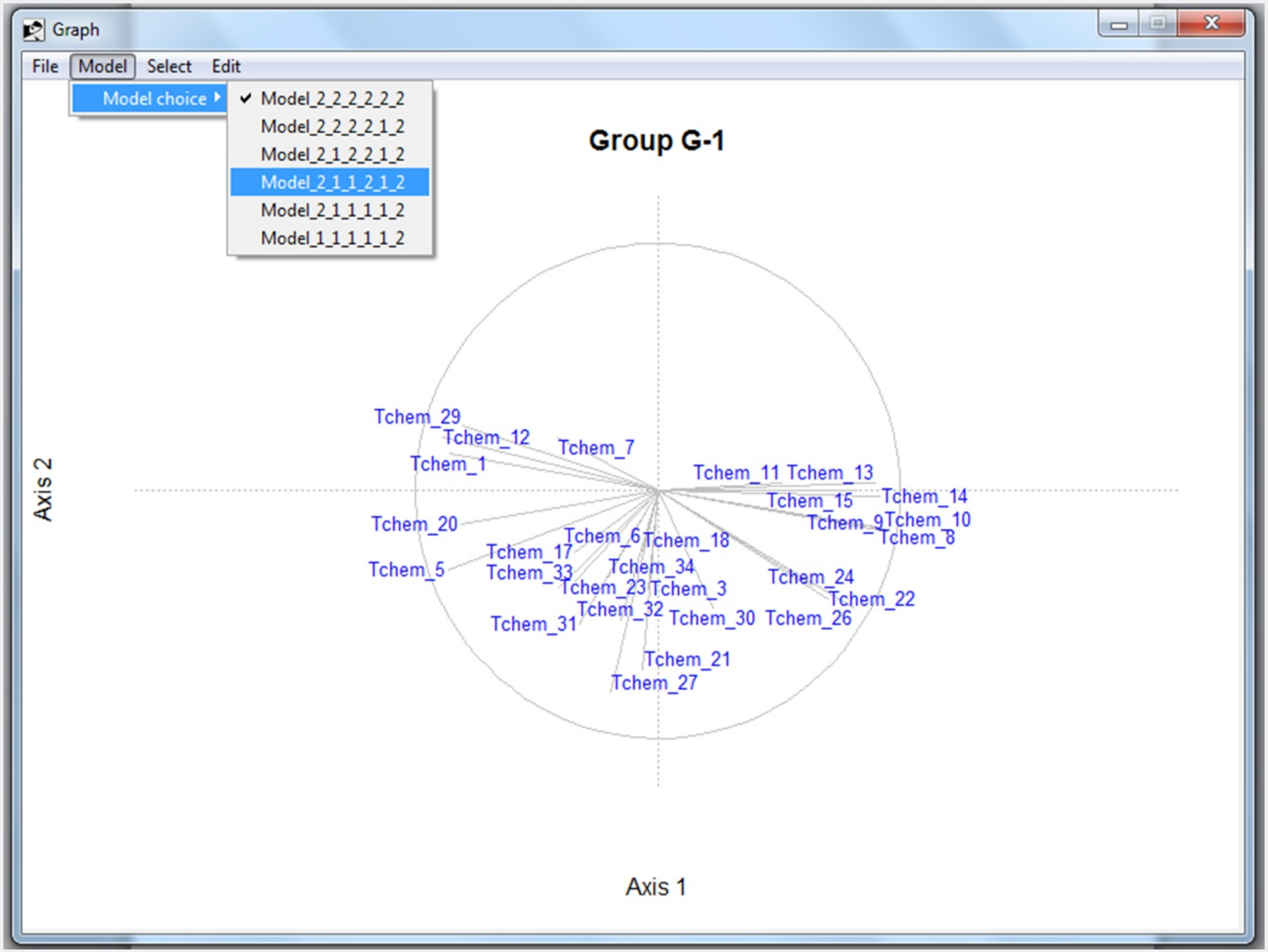
Model-selection



How to operate the THEME R-software?

7. Reaping results

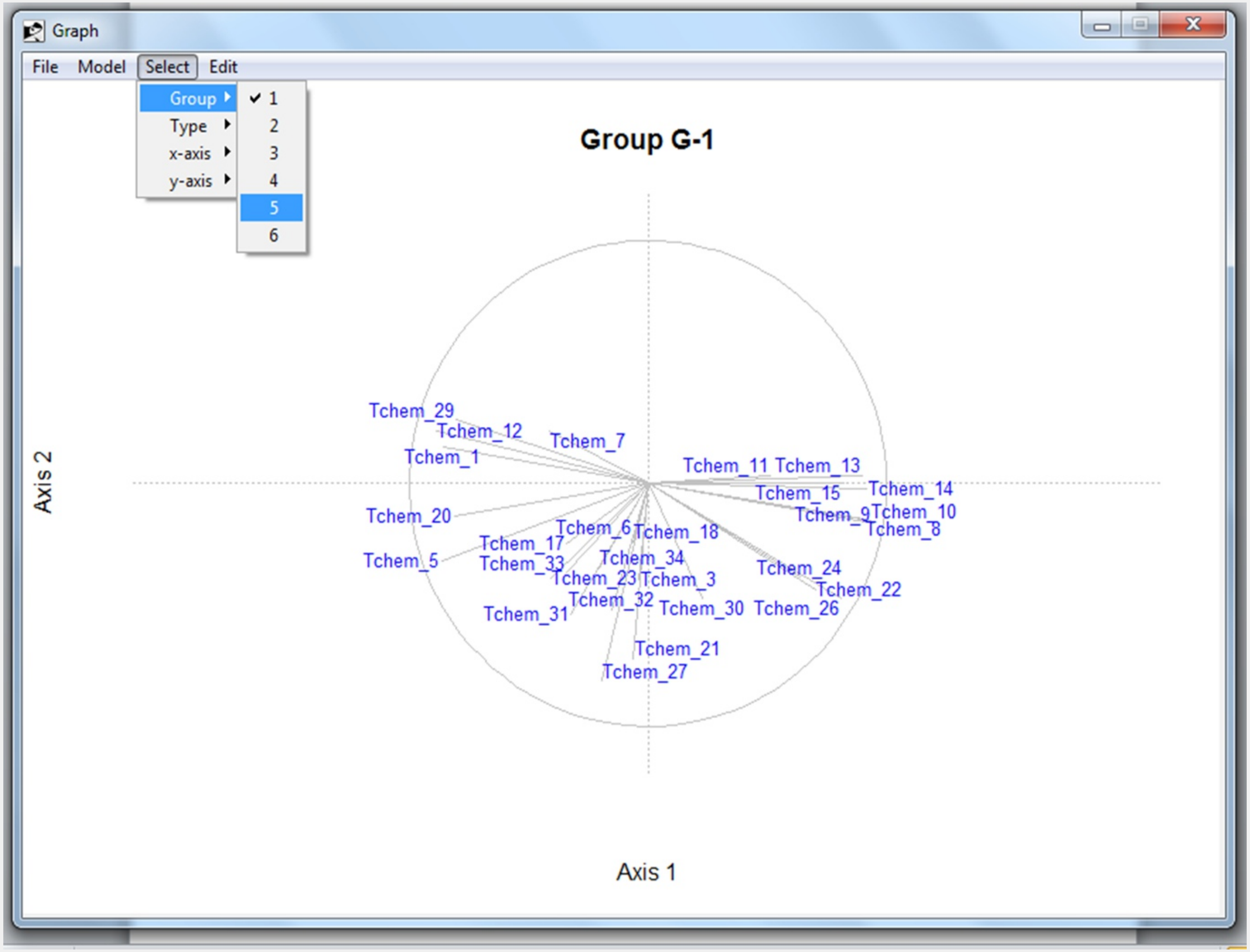
Model-selection* → *Graphing variables



How to operate the THEME R-software?

7. Reaping results

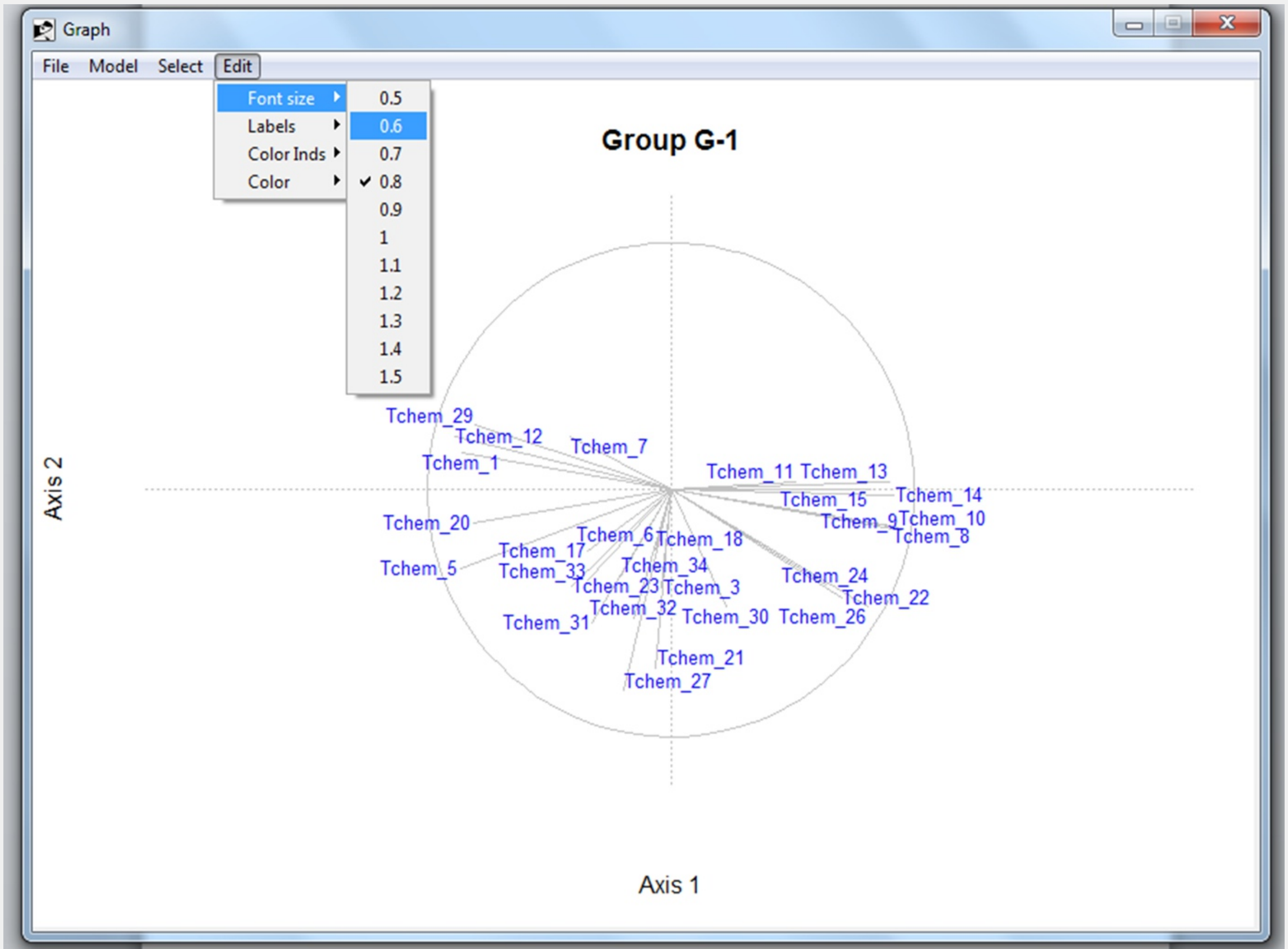
Graphing variables



How to operate the THEME R-software?

7. Reaping results

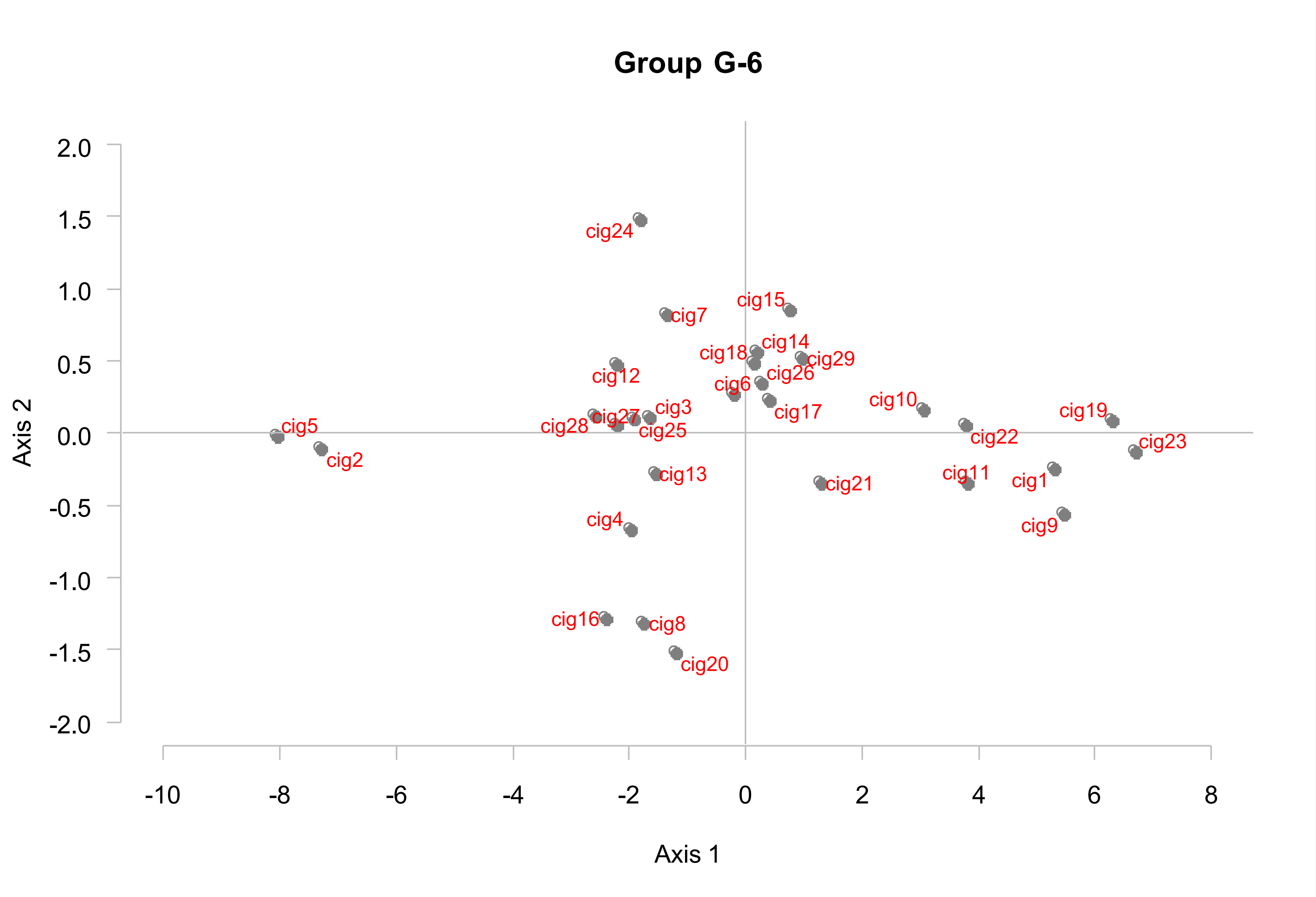
Graphing variables



How to operate the THEME R-software?

7. Reaping results

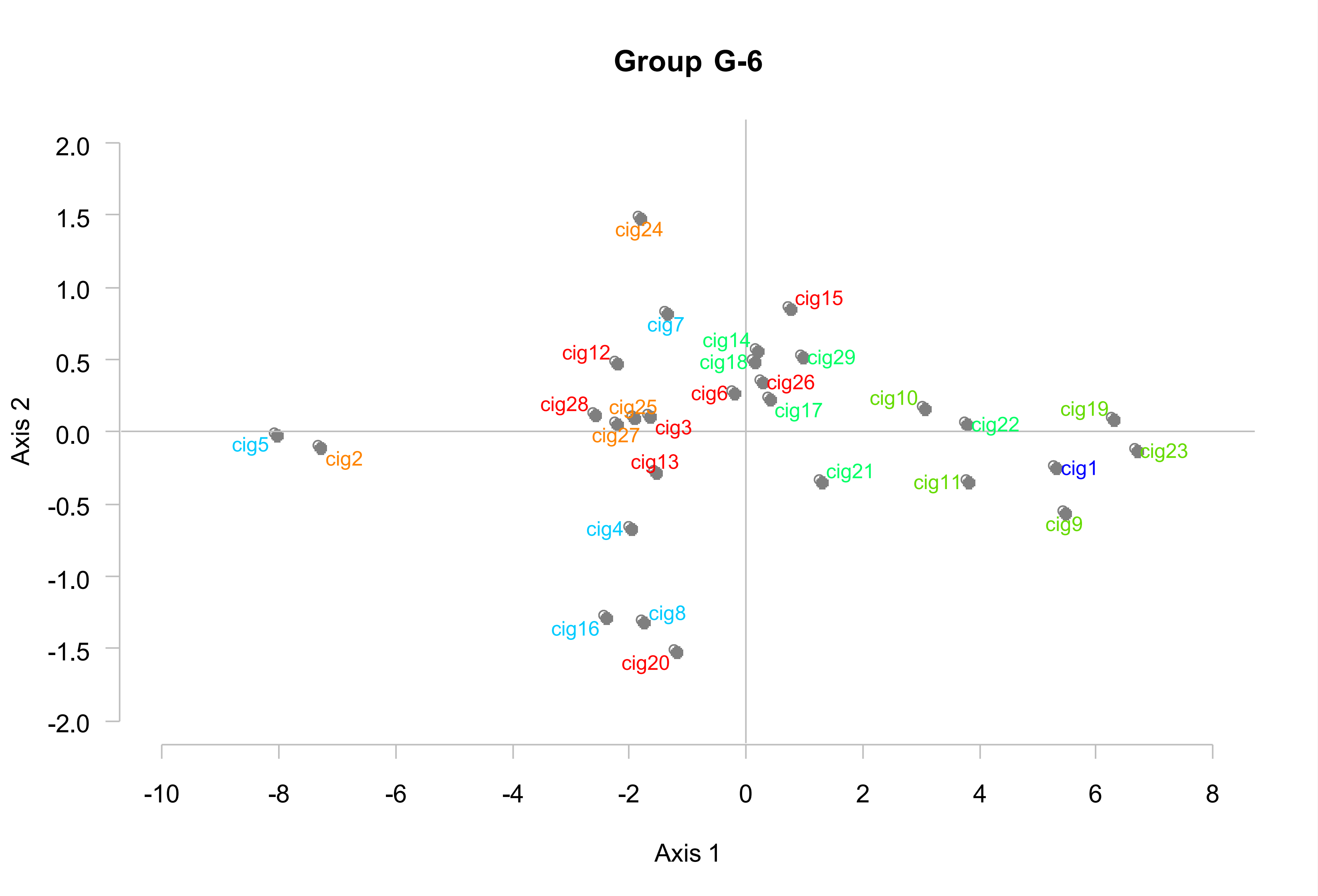
Graphing observations



How to operate the THEME R-software?

7. Reaping results

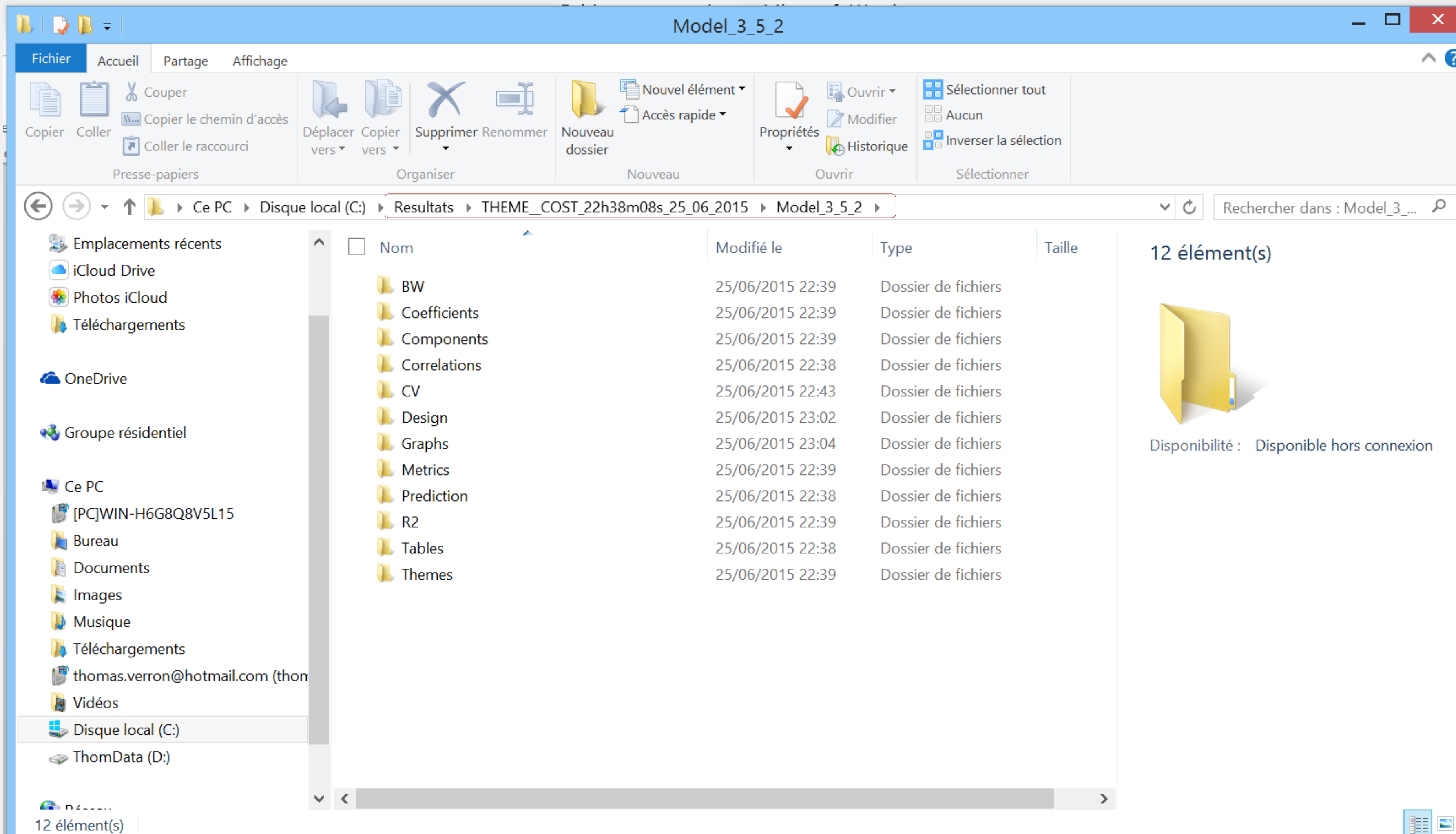
Graphing observations



How to operate the THEME R-software?

7. *Reaping results*

Getting ALL the results in sub-folders:



How to operate the THEME R-software?

7. Reaping results

Equation 1

.	*	**	***

Dependent variables

		NFDPM	Nicotine	CO	Acetaldehyde	Acrolein	Formaldehyde	BaP	NNK	NNN
Group 1	F1	0,03	-0,09	0,24	0,13	0,21	0,28	0,02	-0,40	-0,32
	F2	-0,22	-0,64	0,34	0,26	0,48	0,00	-0,53	-0,21	0,06
Group 2	F1	-0,19	-0,28	0,09	-0,06	-0,06	-0,10	-0,27	-0,47	-0,07
Group 3	F1	0,30	0,40	0,16	0,13	-0,03	0,17	0,41	0,19	0,05
	F2	0,06	0,06	-0,12	0,02	0,02	0,03	0,15	-0,18	0,38
Group 4	F1	-0,67	-1,02	0,10	-0,12	0,11	-0,09	-0,74	-0,95	-0,46
	F2	0,17	0,10	0,24	0,22	0,10	0,18	0,23	0,25	-0,34

Equation 2

		NFDPM	Nicotine	CO	Acetaldehyde	Acrolein	Formaldehyde	BaP	NNK	NNN
Group 7	F1	-0,13	-0,13	-0,08	-0,11	-0,10	-0,04	-0,22	-0,38	0,13
	F2	-0,12	-0,20	0,01	0,02	0,02	0,17	-0,07	-0,37	-0,48
	F3	0,06	0,22	-0,15	0,06	0,13	0,18	0,12	0,14	-0,60
Group 5	F1	0,50	0,43	0,60	0,50	0,51	0,51	0,33	-0,04	0,61
	F2	-0,01	-0,05	-0,04	0,08	0,08	0,25	0,00	0,01	-0,57

How to operate the THEME R-software?

7. Reaping results

Equation 1 . * ** *** *Dependent variables*

		NFDPM	Nicotine	CO	Acetaldehyde	Acrolein	Formaldehyde	BaP	NNK	NNN
Group 1	F1	0,03	-0,09	0,24	0,13	0,21	0,28	0,02	-0,40	-0,32
	F2	-0,22	-0,64	0,34	0,26	0,48	0,00	-0,53	-0,21	0,06
Group 2	F1	-0,19	-0,28	0,09	-0,06	-0,06	-0,10	-0,27	-0,47	-0,07
Group 3	F1	0,30	0,40	0,16	0,13	-0,03	0,17	0,41	0,19	0,05
	F2	0,06	0,06	-0,12	0,02	0,02	0,03	0,15	-0,18	0,38
Group 4	F1	-0,67	-1,02	0,10	-0,12	0,11	-0,09	-0,74	-0,95	-0,46
	F2	0,17	0,10	0,24	0,22	0,10	0,18	0,23	0,25	-0,34

Equation 2

		NFDPM	Nicotine	CO	Acetaldehyde	Acrolein	Formaldehyde	BaP	NNK	NNN
Group 7	F1	-0,13	-0,13	-0,08	-0,11	-0,10	-0,04	-0,22	-0,38	0,13
	F2	-0,12	-0,20	0,01	0,02	0,02	0,17	-0,07	-0,37	-0,48
	F3	0,06	0,22	-0,15	0,06	0,13	0,18	0,12	0,14	-0,60
Group 5	F1	0,50	0,43	0,60	0,50	0,51	0,51	0,33	-0,04	0,61
	F2	-0,01	-0,05	-0,04	0,08	0,08	0,25	0,00	0,01	-0,57

Equation 1 . * ** *** *Dependent variables*

		NFDPM	Nicotine	CO	Acetaldehyde	Acrolein	Formaldehyde	BaP	NNK	NNN
Group 1	F1	0,03	-0,09	0,24	0,13	0,21	0,28	0,02	-0,40	-0,32
	F2	-0,22	-0,64	0,34	0,26	0,48	0,00	-0,53	-0,21	0,06
	C_TO	0,99	0,25	-1,02	-37,51	-6,55	-1,60	1,71	6,58	1,14
	Mal_TO	-0,63	-0,18	0,88	30,60	5,23	3,03	-1,14	-7,07	-7,91
	N_TO	0,19	0,13	-1,11	-33,58	-5,43	-8,17	0,51	12,52	28,28
	PP_TO	0,92	0,16	-0,07	-9,60	-2,10	6,05	1,42	-4,67	-25,84
	MV_TO	0,00	0,00	0,00	0,14	0,02	0,00	-0,01	-0,02	0,01
	Asp_TO	2,50	0,84	-4,91	-162,08	-27,19	-24,08	4,80	45,33	74,36
	Cit_TO	-0,25	-0,01	-0,34	-7,62	-1,04	-4,74	-0,32	5,68	18,09
	NO3_TO	-2,53	-0,53	1,31	58,58	10,86	-7,11	-4,13	-0,82	37,11
	Alka_TO	1,67	0,46	-2,15	-75,58	-13,00	-6,33	2,98	16,32	14,87
	GFS_TO	0,05	0,00	0,09	2,14	0,31	1,10	0,05	-1,36	-4,13
	NH3_TO	-4,76	-0,64	-1,70	-10,28	1,39	-49,24	-6,92	49,83	197,75
	NAB_TO	-3,71	0,52	-12,94	-342,77	-51,81	-138,27	-3,06	181,82	510,65
	NAT_TO	-0,29	-0,01	-0,38	-8,56	-1,17	-5,36	-0,37	6,42	20,47
	NNK_TO	-2,83	-0,56	1,05	53,45	10,24	-11,55	-4,53	4,23	54,31
	NNN_TO	-0,06	0,02	-0,32	-8,88	-1,37	-3,20	-0,02	4,33	11,66
Group 2	F1	-0,19	-0,28	0,09	-0,06	-0,06	-0,10	-0,27	-0,47	-0,07
	Cit_PA	-1,80	-0,22	0,48	-16,81	-1,59	-4,72	-1,84	-20,54	-10,52
	PO4_PA	8,20	0,98	-2,18	76,34	7,22	21,43	8,34	93,27	47,77
	Acet_PA	-2,09	-0,25	0,56	-19,45	-1,84	-5,46	-2,12	-23,76	-12,17
	CaCO3_PA	-0,38	-0,05	0,10	-3,58	-0,34	-1,00	-0,39	-4,37	-2,24
	PERM1_SOD	-0,02	0,00	0,00	-0,16	-0,02	-0,04	-0,02	-0,19	-0,10
Group 3	F1	0,30	0,40	0,16	0,13	-0,03	0,17	0,41	0,19	0,05
	F2	0,06	0,06	-0,12	0,02	0,02	0,03	0,15	-0,18	0,38
	Mg_Ca_pc	0,06	0,01	0,01	0,79	-0,01	0,18	0,07	0,11	0,65
	Cl_TO	4,00	0,41	-1,05	46,02	1,11	10,30	5,15	-14,62	170,04
	PO4_TO	-2,85	-0,42	-6,40	-47,85	5,96	-10,45	0,17	-73,63	383,50
	K_pc_TO	4,34	0,47	0,63	53,63	-0,46	11,93	4,63	4,98	59,30
	Hg_TO	0,21	0,02	0,09	2,69	-0,08	0,60	0,19	0,97	-1,60
	Pb_TO	0,80	0,09	0,49	10,71	-0,44	2,37	0,65	5,34	-15,58
	Cd_TO	1,43	0,16	0,26	17,83	-0,20	3,97	1,50	2,28	15,76
	NO3_TO.1	2,70	0,31	1,00	34,67	-0,86	7,69	2,56	10,21	-5,76
Group 4	F1	-0,67	-1,02	0,10	-0,12	0,11	-0,09	-0,74	-0,95	-0,46
	F2	0,17	0,10	0,24	0,22	0,10	0,18	0,23	0,25	-0,34
	FDENSC	0,16	0,02	0,00	1,34	-0,04	0,19	0,13	1,06	1,01
	HC_BIN	-0,01	0,01	-0,09	-3,26	-0,20	-0,47	-0,03	-0,11	4,16
	PDEF	-0,07	-0,01	0,01	-0,36	0,03	-0,05	-0,06	-0,48	-0,80

Equation 2

		NFDPM	Nicotine	CO	Acetaldehyde	Acrolein	Formaldehyde	BaP	NNK	NNN
Group 7	F1	-0,13	-0,13	-0,08	-0,11	-0,10	-0,04	-0,22	-0,38	0,13
	F2	-0,12	-0,20	0,01	0,02	0,02	0,17	-0,07	-0,37	-0,48
	F3	0,06	0,22	-0,15	0,06	0,13	0,18	0,12	0,14	-0,60
	TAR	0,05	0,01	0,01	2,17	0,24	0,07	0,05	0,55	-0,77
	NICO	0,78	0,13	-0,47	32,32	4,77	2,55	0,79	8,61	-25,48
	CO	0,00	-0,01	0,12	0,87	-0,16	-0,24	0,00	0,02	2,37
	Acetal_MS	0,00	0,00	0,00	0,03	0,00	0,00	0,00	0,01	0,04
	Acro_MS	0,00	0,00	0,02	0,32	-0,01	-0,03	0,00	0,04	0,32
	Fo_MS	0,00	0,00	0,00	0,46	0,05	0,05	0,01	0,02	-0,44
	BaP_MS	0,07	0,01	-0,03	3,70	0,50	0,32	0,08	0,73	-3,05
	NNK_MS	0,01	0,00	0,00	0,05	0,00	-0,07	0,01	0,16	0,52
	NNN_MS	0,00	0,00	0,00	-0,07	-0,01	-0,03	0,00	0,04	0,24
Group 5	F1	0,50	0,43	0,60	0,50	0,51	0,51	0,33	-0,04	0,61
	F2	-0,01	-0,05	-0,04	0,08	0,08	0,25	0,00	0,01	-0,57
	FV	-0,06	0,00	-0,07	-3,45	-0,36	-0,25	-0,03	0,02	-0,92
	PD	0,05	0,00	0,05	4,65	0,49	0,58	0,02	0,00	-1,45
	PDFNE	-0,09	-0,01	-0,11	-4,60	-0,48	-0,26	-0,04	0,03	-2,02

How to get the THEME R-software?

Beta-version available for free, with quick guide.

Conclusion

Some differences with SEM estimation

***SEM* :**	***THEME* :**

Conclusion

Some differences with SEM estimation

SEM :	***THEME :***
Constraints on Latent Variables (LV): ➢ Distribution asumption (classically normal). ➢ One LV per theme. ➢ Regression-models relating LV's. ➢ Regression-models relating each LV to *all* observed variables in its theme (even the least useful for modelling!).	➢ Components belong to the observed space of their theme. No distribution assumption. ➢ Possibly several LV's per theme. Number found out. Hence full exploitation of theme dimensionality. ➢ Regression-models relating LV's. ➢ Bonus to flexible structural relevance (SR), apt to relate a component to the useful variables mostly.

Conclusion

Some differences with SEM estimation

SEM :	*THEME* :
Constraints on Latent Variables (LV): ➢ Distribution asumption (classically normal). ➢ One LV per theme. ➢ Regression-models relating LV's. ➢ Regression-models relating each LV to *all* observed variables in its theme (even the least useful for modelling!).	➢ Components belong to the observed space of their theme. No distribution assumption. ➢ Possibly several LV's per theme. Number found out. Hence full exploitation of theme dimensionality. ➢ Regression-models relating LV's. ➢ Bonus to flexible structural relevance (SR), apt to relate a component to the useful variables mostly.
Estimation strategy: ➢ Plain likelihood maximization (MLE). No tuning parameters. Influence of theme-size. ➢ Estimation of *regression coefficients* based on the MLE of a constrained covariance matrix of the observed variables, or the EM algorithm. ➢ Possible estimation of expected values of LV's conditional on the data (through EM only) ➢ Problems with small samples.	➢ Tunable maximisation of trade-off criterion between GOF and SR. ➢ Calculation of component-values. ➢ Estimation of regression coefficients of all dependent variables (latent or observed). ➢ Cross-validation and assessment of model(s). ➢ Small samples OK.

Conclusion

Some differences with SEM estimation

SEM :	*THEME* :
Constraints on Latent Variables (LV):	
➢ Distribution asumption (classically normal). ➢ One LV per theme. ➢ Regression-models relating LV's. ➢ Regression-models relating each LV to *all* observed variables in its theme (even the least useful for modelling!).	➢ Components belong to the observed space of their theme. No distribution assumption. ➢ Possibly several LV's per theme. Number found out. Hence full exploitation of theme dimensionality. ➢ Regression-models relating LV's. ➢ Bonus to flexible structural relevance (SR), apt to relate a component to the useful variables mostly.
Estimation strategy:	
➢ Plain likelihood maximization (MLE). No tuning parameters. Influence of theme-size. ➢ Estimation of *regression coefficients* based on the MLE of a constrained covariance matrix of the observed variables, or the EM algorithm. ➢ Possible estimation of expected values of LV's conditional on the data (through EM only) ➢ Problems with small samples.	➢ Tunable maximisation of trade-off criterion between GOF and SR. ➢ Calculation of component-values. ➢ Estimation of regression coefficients of all dependent variables (latent or observed). ➢ Cross-validation and assessment of model(s). ➢ Small samples OK.
Prediction:	
➢ No prediction/regularisation performed.	➢ Regularised prediction of all dependent variables.

Conclusion

Some differences with SEM estimation

SEM :	*THEME :*
Constraints on Latent Variables (LV):	
➢ Distribution asumption (classically normal). ➢ One LV per theme. ➢ Regression-models relating LV's. ➢ Regression-models relating each LV to *all* observed variables in its theme (even the least useful for modelling!).	➢ Components belong to the observed space of their theme. No distribution assumption. ➢ Possibly several LV's per theme. Number found out. Hence full exploitation of theme dimensionality. ➢ Regression-models relating LV's. ➢ Bonus to flexible structural relevance (SR), apt to relate a component to the useful variables mostly.
Estimation strategy:	
➢ Plain likelihood maximization (MLE). No tuning parameters. Influence of theme-size. ➢ Estimation of *regression coefficients* based on the MLE of a constrained covariance matrix of the observed variables, or the EM algorithm. ➢ Possible estimation of expected values of LV's conditional on the data (through EM only) ➢ Problems with small samples.	➢ Tunable maximisation of trade-off criterion between GOF and SR. ➢ Calculation of component-values. ➢ Estimation of regression coefficients of all dependent variables (latent or observed). ➢ Cross-validation and assessment of model(s). ➢ Small samples OK.
Prediction:	
➢ No prediction/regularisation performed.	➢ Regularised prediction of all dependent variables.
Exploratory power:	
➢ Weak.	➢ High.

THE END

Thank you, all

Bry X., Verron T. (2015) : *THEME: THEmatic Model Exploration through Multiple Co-Structure maximization*, Journal of Chemometrics, vol.29, 12; pp.637-647

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.