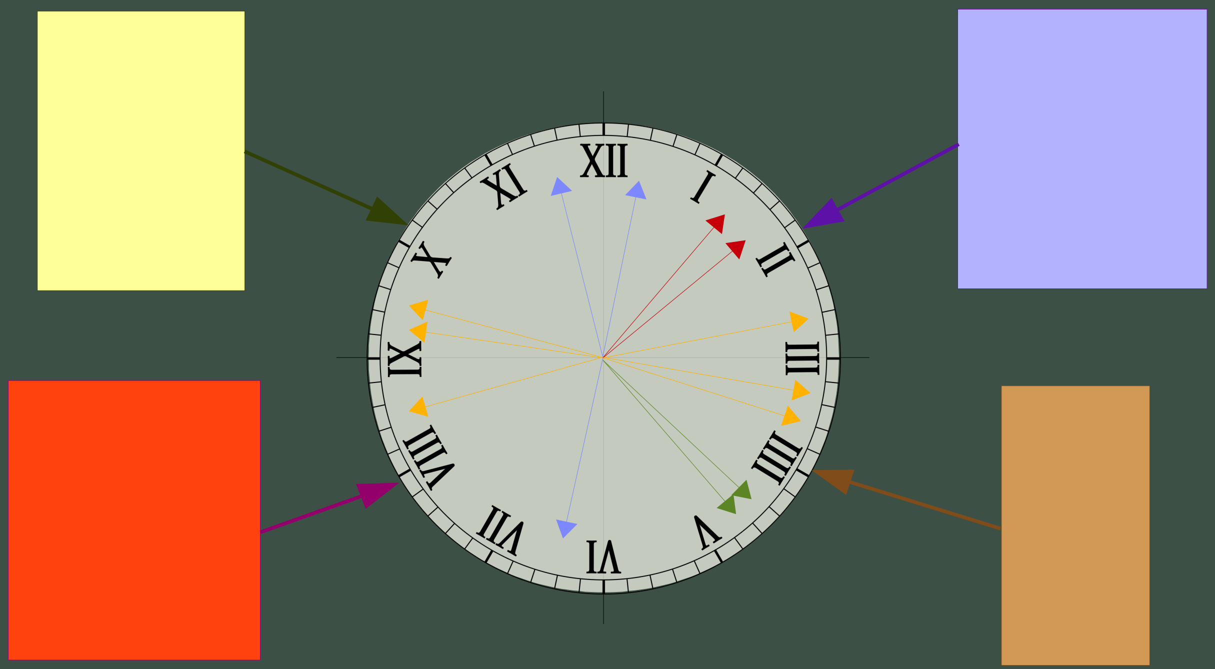
A component-based regularised Cox Regression:

SC-CoxR



X. Bry *IMAG, Univ. Montpellier*

Joint work with:
T. Simac , *S. El Ghachi* and *P. Antoine*

Data and Problem

1. Data

1.1. The Data

A right-censored survival time y, to be modelled through many possibly redundant time-dependent explanatory variables.

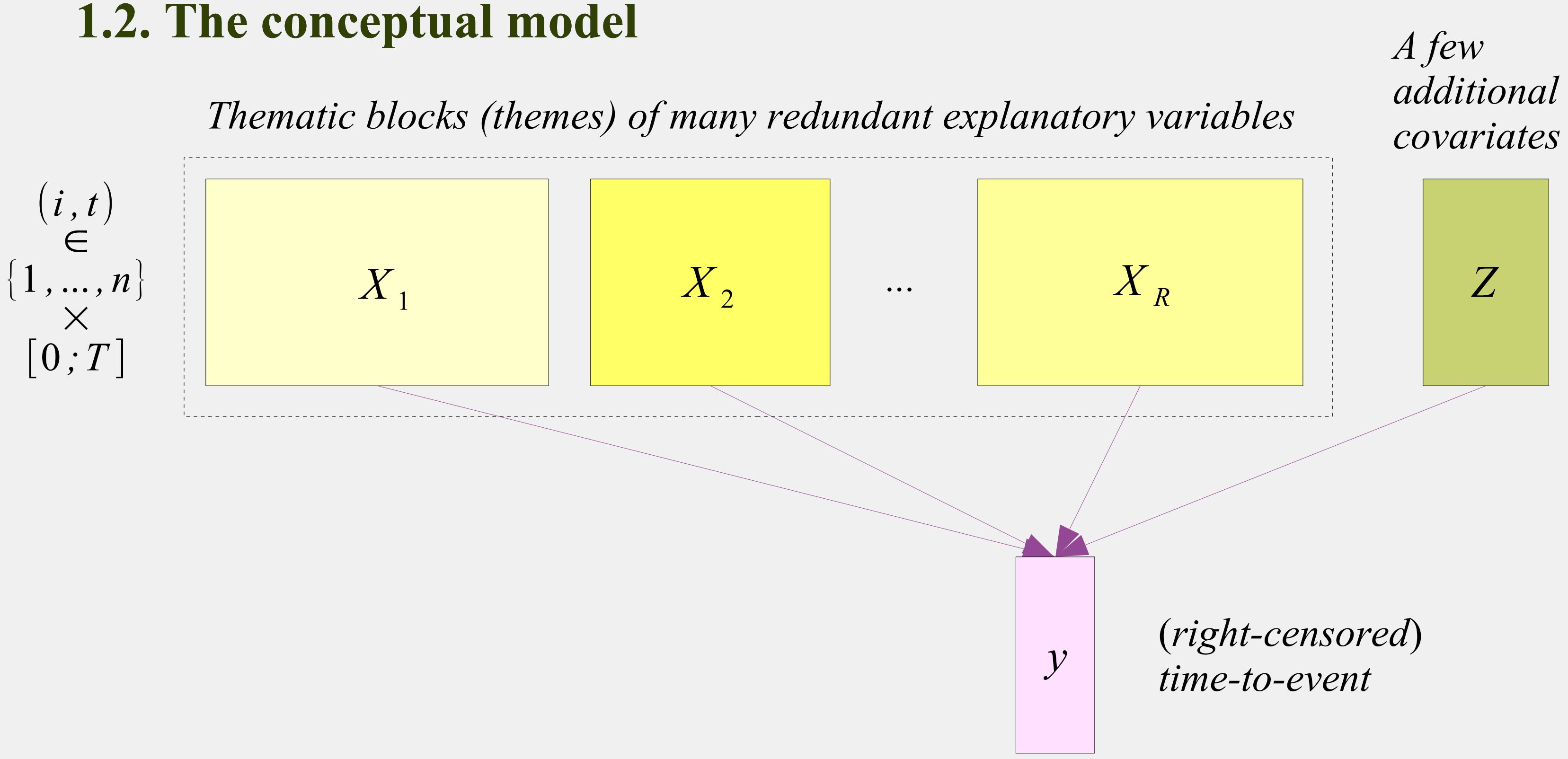
Data and Problem

1. Data

1.1. The Data

A right-censored survival time *y*, to be modelled through many possibly redundant time-dependent explanatory variables.

1.2. The conceptual model



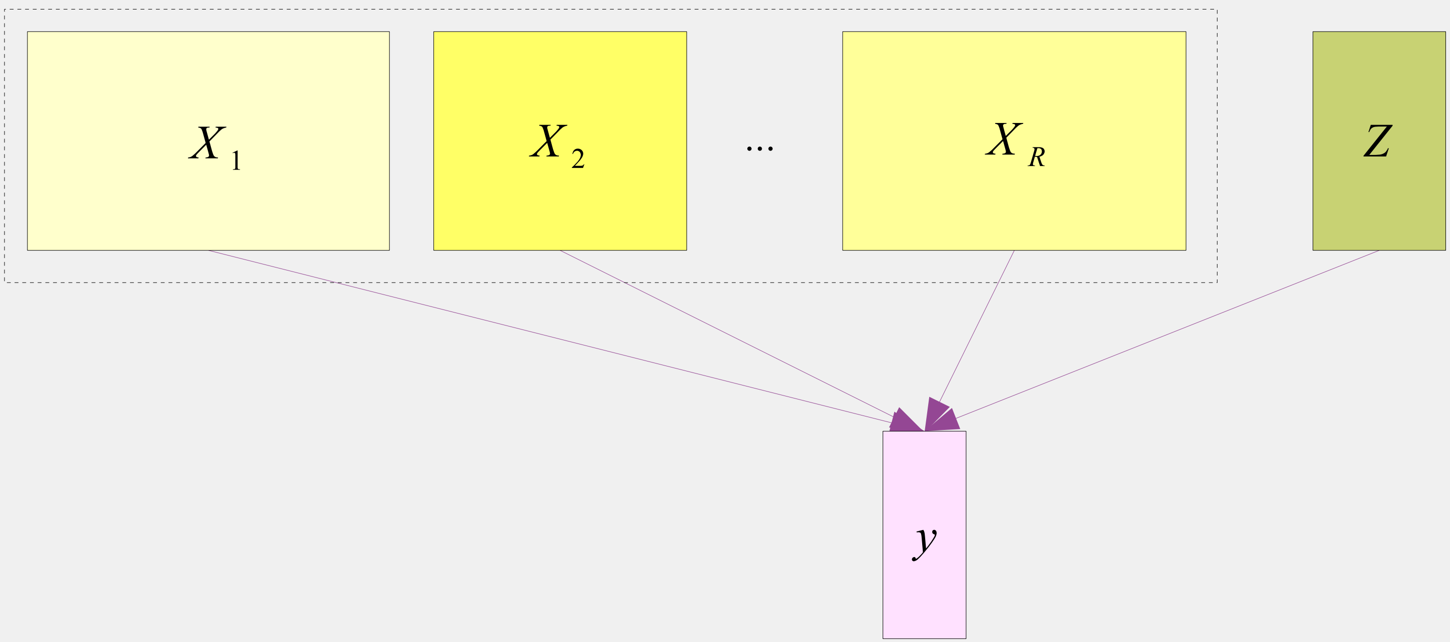
Data and Problem

2. *Problem*

2.1. Dimension reduction

High dimension + redundancy ⇒ dimension reduction in the X's

A few additional covariates



No dimension reduction required in Z

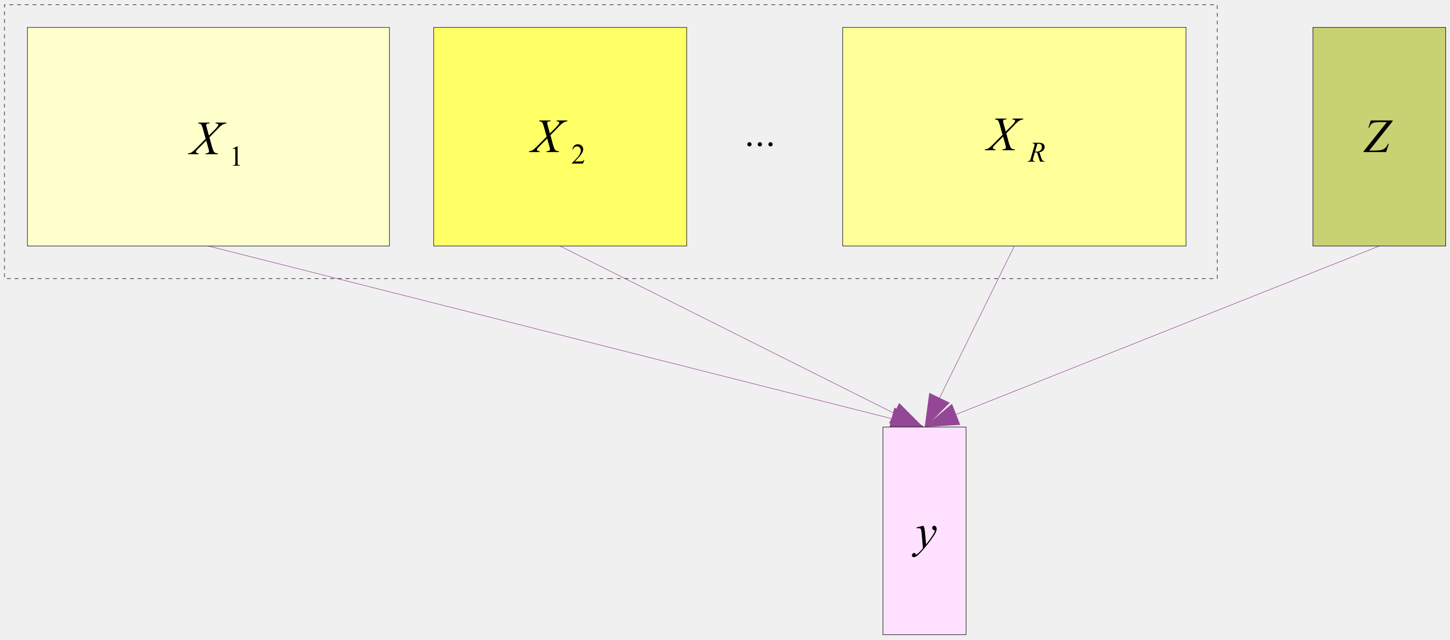
Data and Problem

2. Problem

2.1. Dimension reduction

High dimension + redundancy ⇒ dimension reduction in the X's

A few additional covariates



No dimension reduction required in Z

2.2. Exploratory + explanatory situation

The explanatory dimensions must be **found** AND **easy to interpret**.

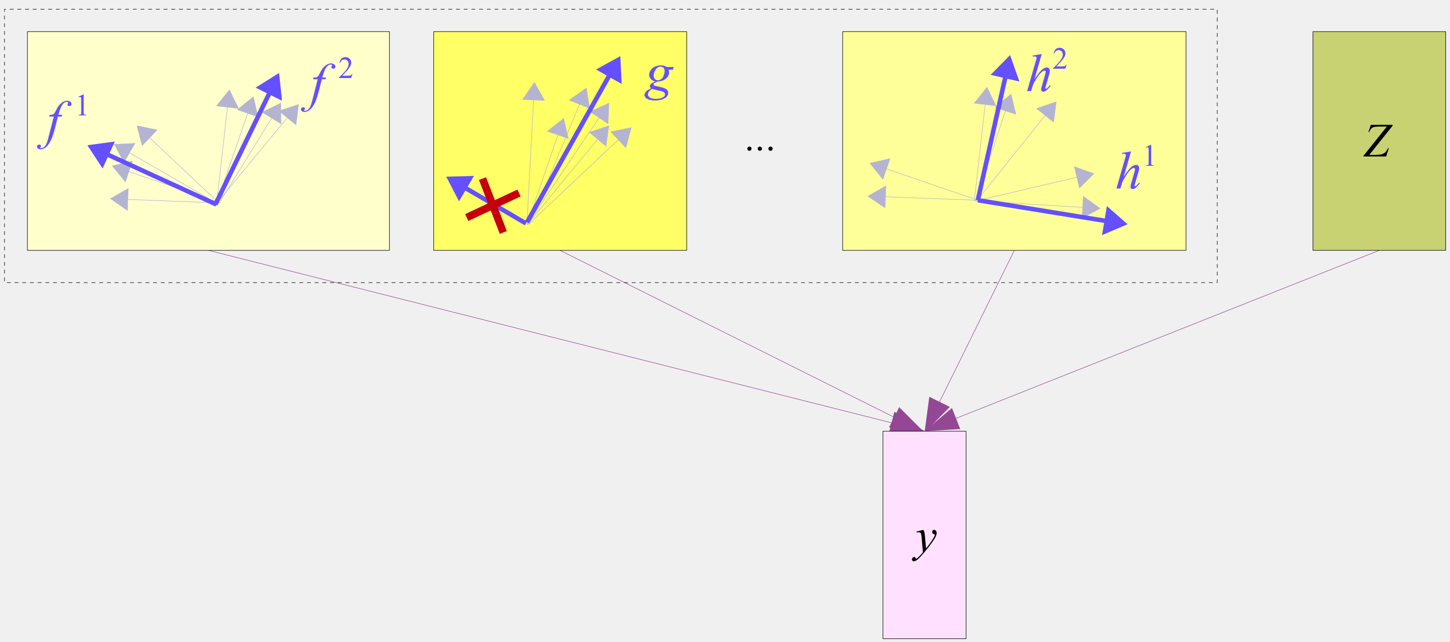
Data and Problem

2. Problem

2.3. How to tackle both issues

We shall look for "**strong**" orthogonal components in each X-theme...

A few additional covariates



No dimension reduction required in Z

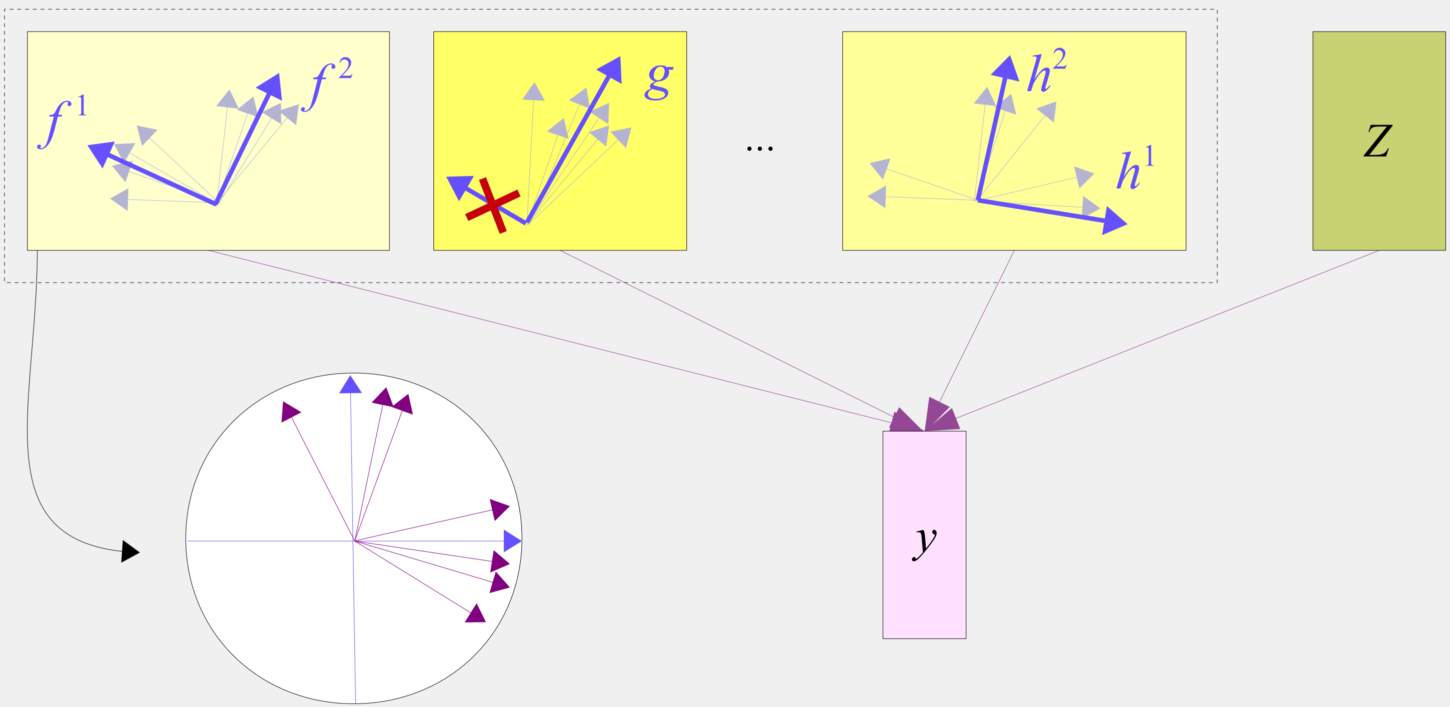
Data and Problem

2. Problem

2.3. How to tackle both issues

We shall look for "**strong**" orthogonal components in each *X*-theme...

A few additional covariates



No dimension reduction required in *Z*

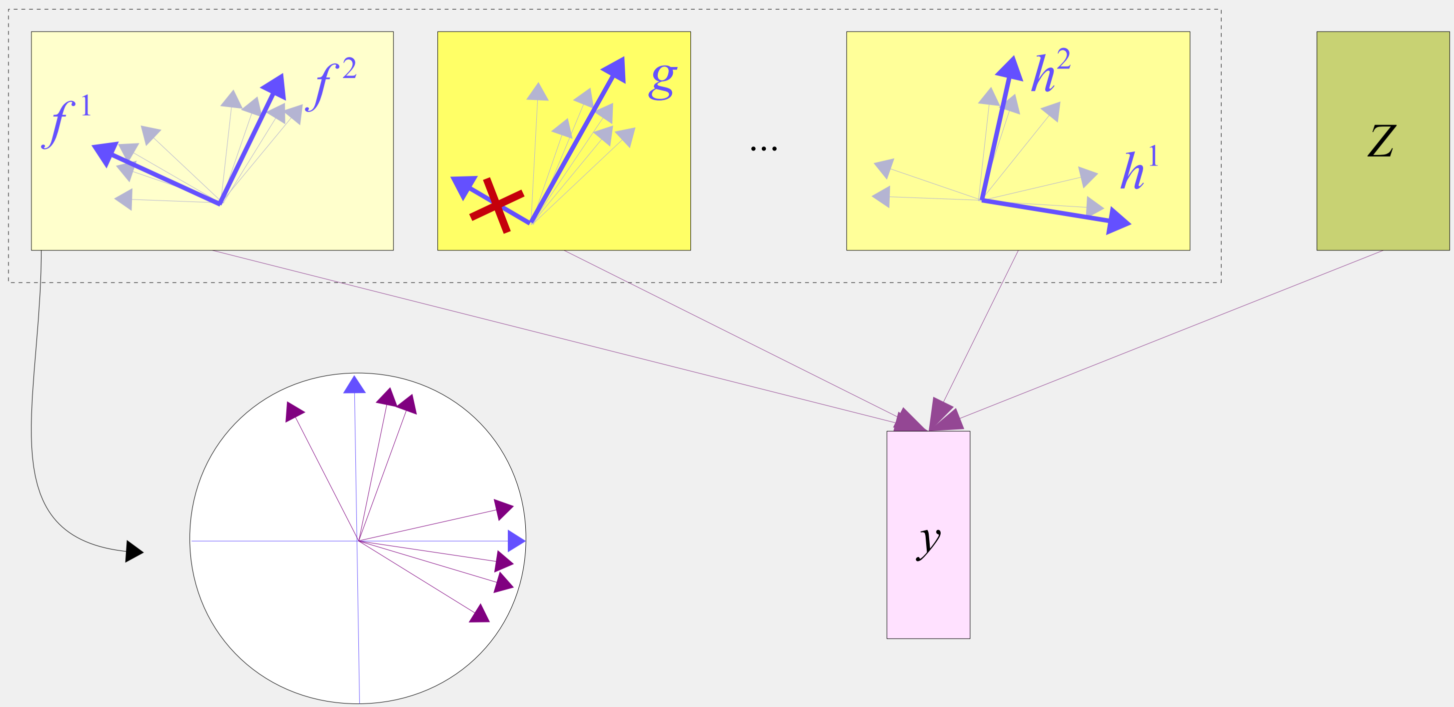
Data and Problem

2. Problem

2.3. How to tackle both issues

We shall look for "**strong**" orthogonal components in each *X*-theme...

A few additional covariates



No dimension reduction required in *Z*

... so as to build a component-based Cox Proportional Hazard Model:

With $f_{(i,t)} := (f^1_{(i,t)}, f^2_{(i,t)}, \dots, g_{(i,t)}, h^1_{(i,t)}, \dots)'$: $h(t\,; x_{(i,t)}, z_{(i,t)}) = h_0(t)\, e^{\delta' f_{(i,t)} + \gamma' z_{(i,t)}}$

Statistical model

1. The classical Cox Proportional hazard Model

Regressor-set $X \rightarrow$ semi-parametric hazard function: $h(t\,;x_{(i,t)}) = h_0(t)\,e^{\beta' x_{(i,t)}}$

Statistical model

1. The classical Cox Proportional hazard Model

Regressor-set $X \rightarrow$ semi-parametric hazard function: $h(t\,;x_{(i,t)}) = h_0(t)\, e^{\beta' x_{(i,t)}}$

2. The component-based Cox-Model

2.1. The single-*X*-theme component Model

Explanatory theme $X \rightarrow$ components $F = [f^1, \ldots, f^k]$, where $f^k = X u^k$

Let $f_{(i,t)} := (f^1_{(i,t)}, \ldots, f^k_{(i,t)})'$

$\rightarrow$ semi-parametric hazard function of the component-model: $h(t\,;x_{(i,t)}, z_{(i,t)}) = h_0(t)\, e^{\alpha' f_{(i,t)} + \gamma' z_{(i,t)}}$

Statistical model

1. The classical Cox Proportional hazard Model

Regressor-set $X \rightarrow$ semi-parametric hazard function: $h(t\,;x_{(i,t)})=h_0(t)\,e^{\beta' x_{(i,t)}}$

2. The component-based Cox-Model

2.1. The single-*X*-theme component Model

Explanatory theme $X \rightarrow$ components $F=[f^1,\dots,f^k]$, where $f^k=X\,u^k$

Let $f_{(i,t)}:=(f^1_{(i,t)},\dots,f^k_{(i,t)})'$

$\rightarrow$ semi-parametric hazard function of the component-model: $h(t\,;x_{(i,t)},z_{(i,t)})=h_0(t)\,e^{\alpha' f_{(i,t)}+\gamma' z_{(i,t)}}$

2.2. The general component Model

Explanatory theme $X_r \rightarrow$ components $F_r=[f^1_r,\dots,f^{k_r}_r]$, where $f^k_r=X_r\,u^k_r$

Let $f_{r(i,t)}:=(f^1_{r(i,t)},\dots,f^{k_r}_{r(i,t)})'$

$\rightarrow$ semi-param. hazard function of the component-model: $h(t\,;x_{(i,t)},z_{(i,t)})=h_0(t)\,e^{\sum_{r=1}^{R}\alpha_r' f_{r(i,t)}+\gamma' z_{(i,t)}}$

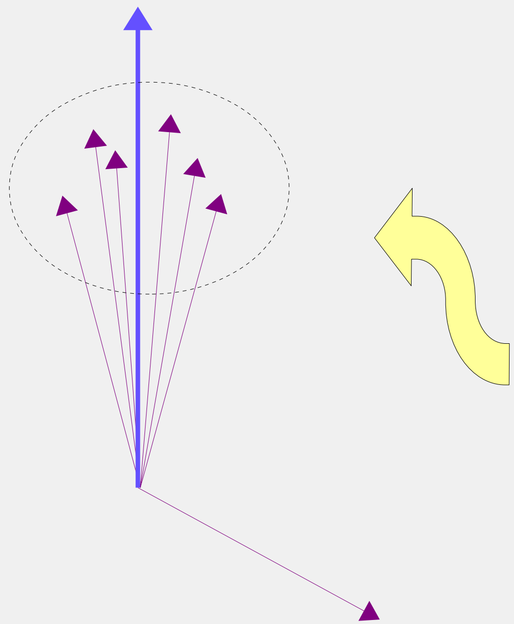
Structural Relevance of components

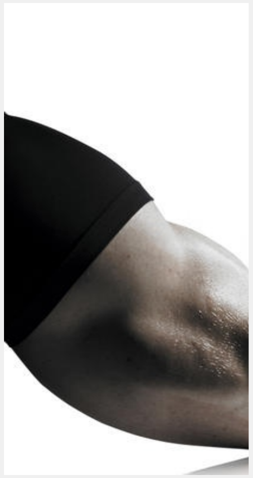
1. The notion of Structural Relevance

Components must capture *interpretable* variable structures

⇒ Components must be *structurally relevant*, i.e.:

- close to *bundles of observed variables*





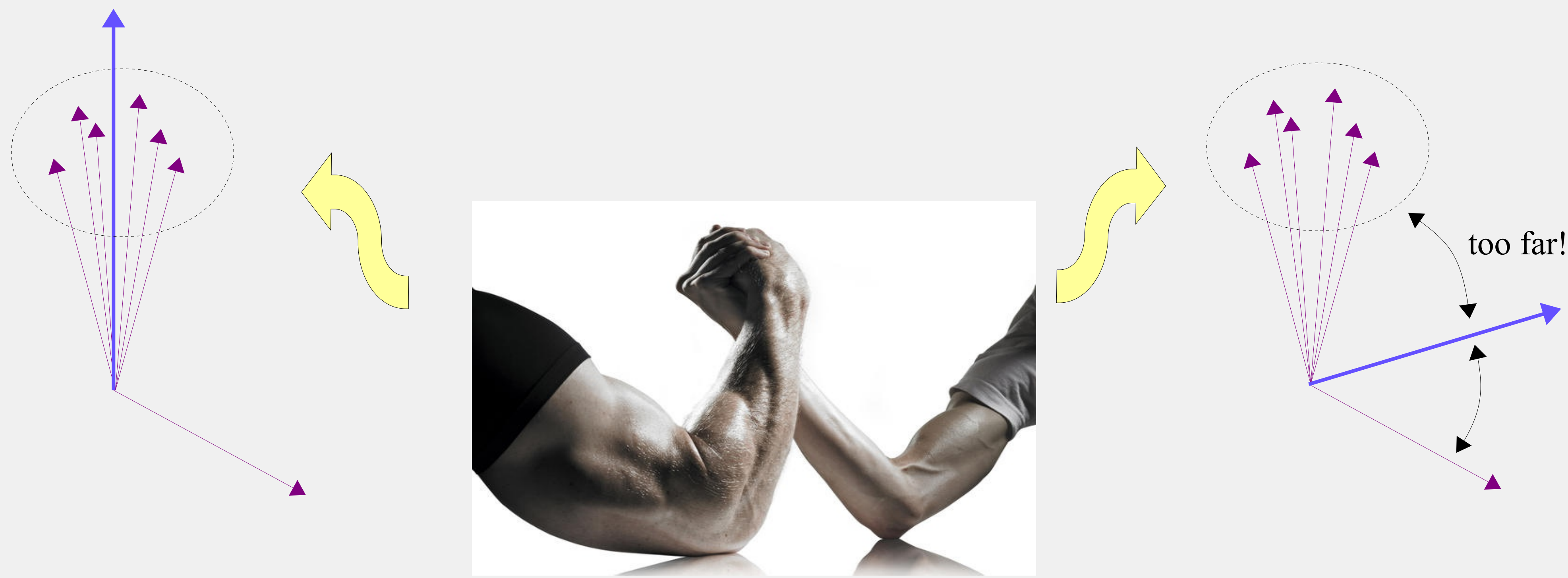
Structural Relevance of components

1. The notion of Structural Relevance

Components must capture *interpretable* variable structures

⇒ Components must be *structurally relevant*, i.e.:

- close to *bundles of observed variables*



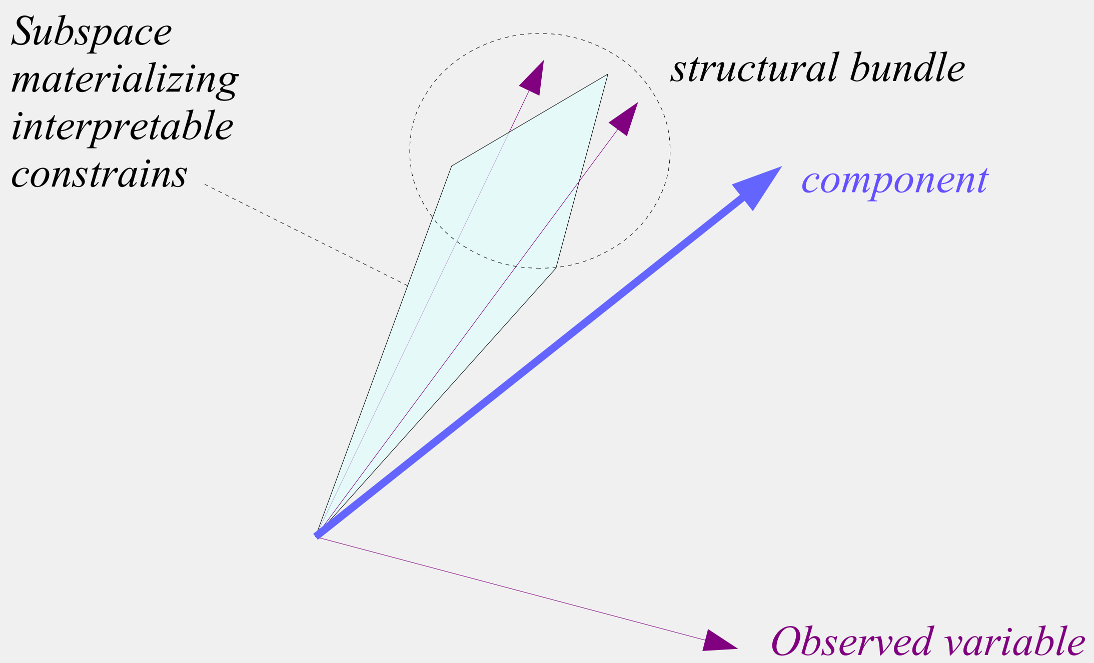
Structural Relevance of components

1. The notion of structural relevance

Components must capture *interpretable* variable structures

⇒ Components must be *structurally relevant*, i.e.:

- or close to *bundles of interpretable subspaces (e.g. embodying theory-based constraints)*



Structural Relevance of components

2. The expression of Structural Relevance

- *Component in a theme* X : $f = Xu$

Structural Relevance of components

2. The expression of Structural Relevance

- *Component in a theme X :* $f = Xu$

- *Identification / regularisation constraint :* $u' M^{-1} u = 1$

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, and $\tau \in [0,1]$ is a parameter tuning regularisation:

  - $\tau = 0$ means no regularisation;
  - $\tau = 1$ means PLS-strong regularisation.

Structural Relevance of components

2. The expression of Structural Relevance

- *Component in a theme X :* $f = Xu$

- *Identification / regularisation constraint :* $u' M^{-1} u = 1$

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, and $\tau \in [0,1]$ is a parameter tuning regularisation:

  - $\tau = 0$ means no regularisation;
  - $\tau = 1$ means PLS-strong regularisation.

- ***The Structural Relevance Indicator:***

$$\phi_{\mathbf{N},\mathbf{\Omega},l}(u) := \left(\sum_{j=1}^{J} \omega_j (u' N_j u)^l \right)^{\frac{1}{l}} \quad \text{s.t. constraint} \quad u' M^{-1} u = 1$$

weights

N_j's code the directions components should focus on

Structural Relevance of components

2. The expression of Structural Relevance

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

$$\phi_{\mathbf{N},\mathbf{\Omega},l}(u) := \left(\sum_{j=1}^{J} \omega_j (u' N_j u)^l \right)^{\frac{1}{l}}$$

The N_j's are coding *directions of concern*

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

  $(W = \text{matrix of line-weights})$ $\|u\|^2 = 1 \Rightarrow M = I$

  $\rightarrow$ directions of discrepancy of observations

Structural Relevance of components

2. The expression of Structural Relevance

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

$$\phi_{\mathbf{N},\boldsymbol{\Omega},l}(u) := \left(\sum_{j=1}^{J} \omega_j (u' N_j u)^l \right)^{\frac{1}{l}}$$

The N_j's are coding *directions of concern*

Examples:

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

  $(W = \text{matrix of line-weights})$ $\quad \|u\|^2 = 1 \;\Rightarrow\; M = I$

  → directions of discrepancy of observations

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

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

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

→ directions of observed variables.

Structural Relevance of components

2. The expression of Structural Relevance

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

$$\phi_{\mathbf{N},\mathbf{\Omega},l}(u) := \left(\sum_{j=1}^{J} \omega_j (u' N_j u)^l \right)^{\frac{1}{l}}$$

The N_j's are coding *directions of concern*

Examples:

Variable Powered Inertia can be extended to:

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

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

$M^{-1} = \tau A^{-1} + (1-\tau)(X' W X)$ where A = suitable metric matrix for X's PCA

Regularisation parameter:

$\tau = 0$: no regularisation.

$\tau = 1$: PLS-strong regularisation.

Structural Relevance of components

2. The expression of Structural Relevance

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

$$\phi_{\mathbf{N},\mathbf{\Omega},l}(u) := \left(\sum_{j=1}^{J} \omega_j (u' N_j u)^l \right)^{\frac{1}{l}}$$

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

locality = ± the "narrowness" of the bundles of directions of structural interest.

Structural Relevance of components

2. The expression of Structural Relevance

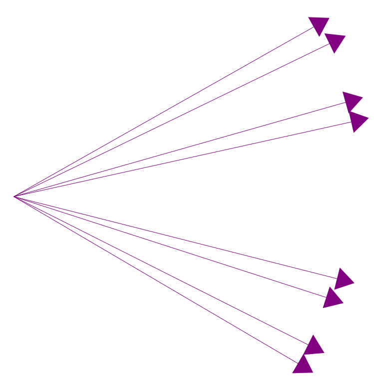
- **Purpose of *l* = ?**

$$\phi_{\mathbf{N},\boldsymbol{\Omega},l}(u) := \left(\sum_{j=1}^{J} \omega_j (u' N_j u)^l \right)^{\frac{1}{l}}$$

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

locality = ± the "narrowness" of the bundles of directions of structural interest.

Would this set of directions rather be considered...



Structural Relevance of components

2. The expression of Structural Relevance

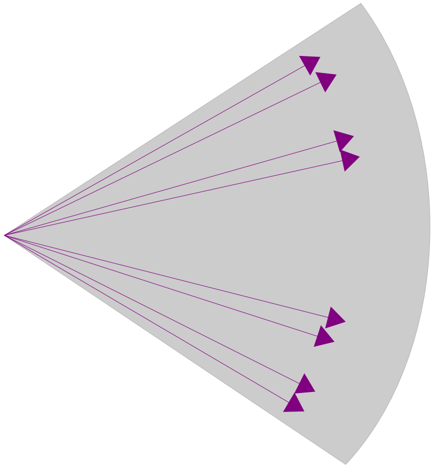
- **Purpose of *l* = ?**

$$\phi_{\mathbf{N},\mathbf{\Omega},l}(u) := \left(\sum_{j=1}^{J} \omega_j (u' N_j u)^l \right)^{\frac{1}{l}}$$

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

locality = ± the "narrowness" of the bundles of directions of structural interest.

Would this set of directions rather be considered...



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

Structural Relevance of components

2. The expression of Structural Relevance

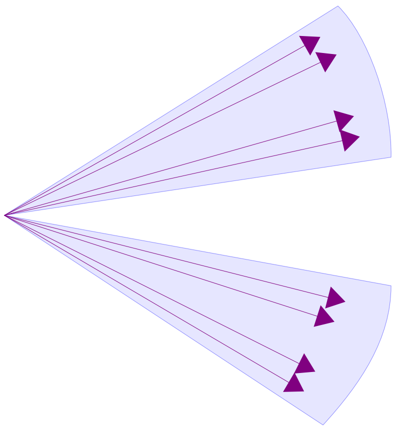
- **Purpose of *l* = ?**

$$\phi_{\mathbf{N},\mathbf{\Omega},l}(u):=\left(\sum_{j=1}^{J}\omega_j(u'N_ju)^l\right)^{\frac{1}{l}}$$

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

locality = ± the "narrowness" of the bundles of directions of structural interest.

Would this set of directions rather be considered...



... two bundles? (*l* ↑)

Structural Relevance of components

2. The expression of Structural Relevance

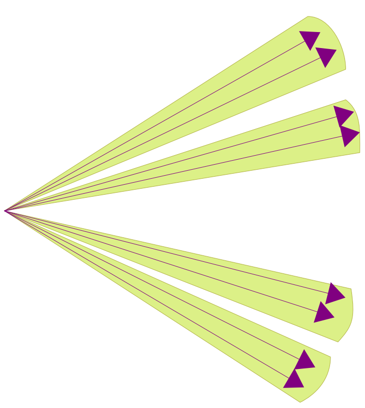
- **Purpose of *l* = ?**

$$\phi_{\mathbf{N},\mathbf{\Omega},l}(u) := \left(\sum_{j=1}^{J} \omega_j \left(u' N_j u \right)^l \right)^{\frac{1}{l}}$$

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

locality = ± the "narrowness" of the bundles of directions of structural interest.

Would this set of directions rather be considered...



... four bundles? (l ↑↑)

Structural Relevance of components

2. The expression of Structural Relevance

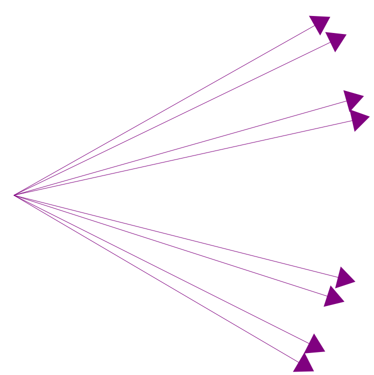
- **Purpose of l = ?**

$$\phi_{\mathbf{N},\mathbf{\Omega},l}(u) := \left(\sum_{j=1}^{J} \omega_j \left(u' N_j u\right)^l \right)^{\frac{1}{l}}$$

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

locality = ± the "narrowness" of the bundles of directions of structural interest.

Would this set of directions rather be considered...



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

Structural Relevance of components

2. The expression of Structural Relevance

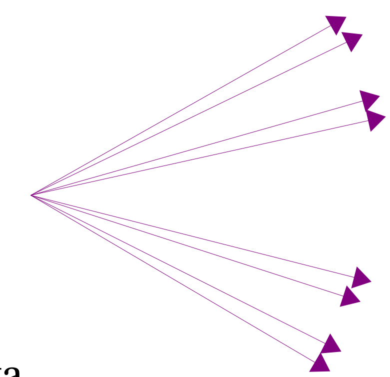
- **Purpose of *l* = ?**

$$\phi_{\mathbf{N},\mathbf{\Omega},l}(u) := \left(\sum_{j=1}^{J} \omega_j (u' N_j u)^l \right)^{\frac{1}{l}}$$

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

locality = ± the "narrowness" of the bundles of directions of structural interest.

Would this set of directions rather be considered...



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

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

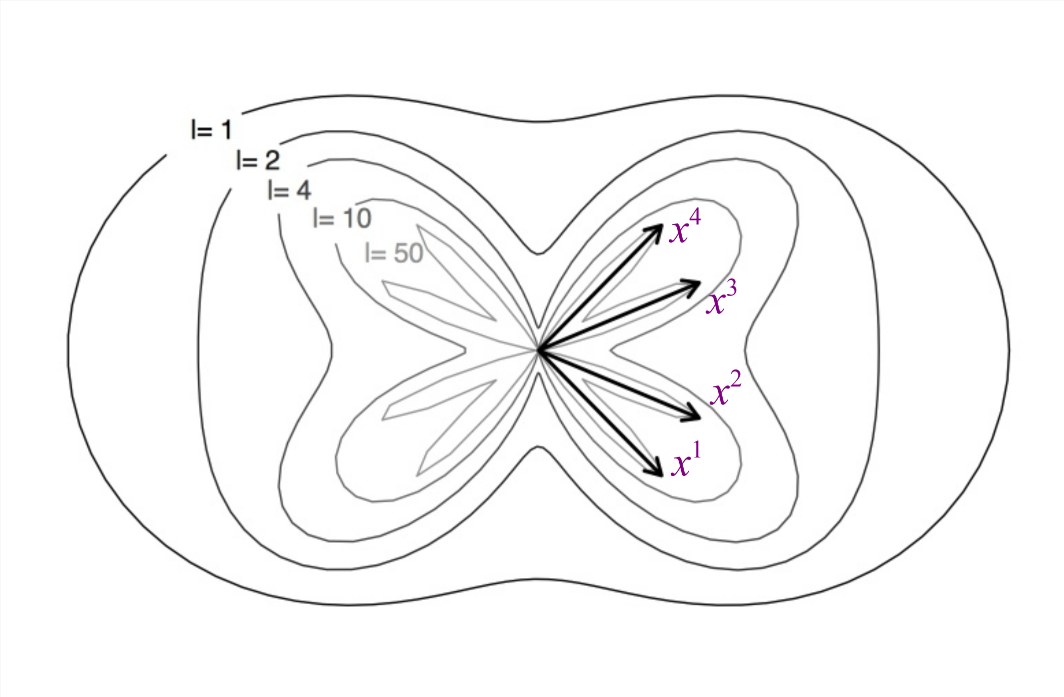
Structural Relevance of components

2. The expression of Structural Relevance

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

Example: 4 variables in a plane...

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



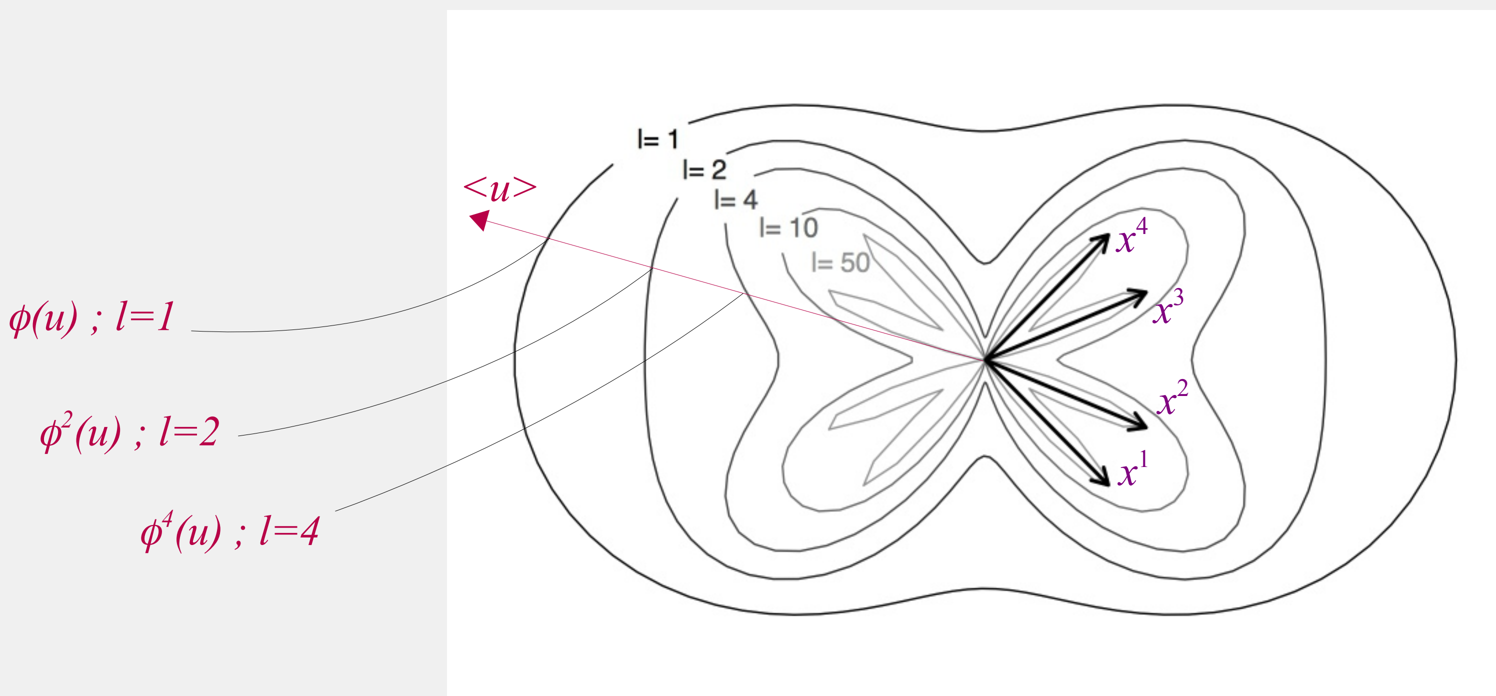
Structural Relevance of components

2. The expression of Structural Relevance

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

Example: 4 variables in a plane...

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



SC-CoxR's mechanism

1. Estimation of a standard Cox-model

1.1. Partial likelihood

Let :

- $R(t)$ denote the set of all individuals at risk at time t;
- δ denote the censoring indicator:

$$\forall i : \delta_i = 1 \text{ if for individual } i\text{, the event occurs at time } y_i$$
$$\delta_i = 0 \text{ if individual } i \text{ is censored at time } y_i$$

Cox (1979) suggested to get $\hat{\beta}$ by maximising on β the following conditional likelihood:
(which is rid of the $h_0(t)$ baseline terms)

$$l_p(\beta) = \prod_{i=1}^{n} \left[\frac{e^{\beta' x_{i,y_i}}}{\sum_{j \in R(y_i)} e^{\beta' x_{j,y_i}}} \right]^{\delta_i}$$

SC-CoxR's mechanism

1. Estimation of a standard Cox-model

1.1. Partial likelihood

Let :

- $R(t)$ denote the set of all individuals at risk at time t;
- δ denote the censoring indicator:

$$\forall i : \delta_i = 1 \text{ if for individual } i, \text{ the event occurs at time } y_i$$
$$\delta_i = 0 \text{ if individual } i \text{ is censored at time } y_i$$

Cox (1979) suggested to get $\hat{\beta}$ by maximising on β the following conditional likelihood:
(which is rid of the $h_0(t)$ baseline terms)

$$l_p(\beta) = \prod_{i=1}^{n} \left[\frac{e^{\beta' x_{i,y_i}}}{\sum_{j \in R(y_i)} e^{\beta' x_{j,y_i}}} \right]^{\delta_i}$$

1.2. Estimation of the baseline hazard

Given $\hat{\beta}$, [Kalbfleisch et al. 1973], [Breslow 1974], among others, proposed an estimation of the Baseline Survival Function, based on it.

SC-CoxR's mechanism

2. Estimation of the single-X component-based Cox Model

2.1. The single-*X* component-based Cox Model

- In the Cox model, X is replaced by $F = XU$, $U = [u_1, ..., u_k]$ where X has been standardised column-wise :

$$h(t; x_{i,t}, z_{i,t}) = h_0(t)\, e^{\alpha' f_{i,t} + \gamma' z_{i,t}}$$

$$= h_0(t)\, e^{\alpha' U' x_{i,t} + \gamma' z_{i,t}}$$

SC-CoxR's mechanism

2. Estimation of the single-X component-based Cox Model

2.1. The single-*X* component-based Cox Model

- In the Cox model, X is replaced by $F = XU$, $U = [u_1, \ldots, u_k]$

  where X has been standardised column-wise :

$$h(t\,;x_{i,t},z_{i,t}) = h_0(t)\,e^{\alpha' f_{i,t} + \gamma' z_{i,t}}$$

$$= h_0(t)\,e^{\alpha' U' x_{i,t} + \gamma' z_{i,t}}$$

both unknown
⇒ non-linear / parameters

SC-CoxR's mechanism

2. Estimation of the single-X component-based Cox Model

2.2. Calculating components

- Component $f^1 = Xu_1$ is sought as the solution of:

$$u_1 = \underset{u}{arg} \underset{\substack{\alpha,\gamma \\ u' M^{-1} u = 1}}{max} \left[\underbrace{\left(l_p(u,\alpha,\gamma) \right)}_{\text{Goodness of fit}}^{1-s} \underbrace{\left(\phi_X(u) \right)}_{\text{SR}}^{s} \right]$$

SC-CoxR's mechanism

2. Estimation of the single-X component-based Cox Model

2.2. Calculating components

- Component $f^1 = Xu_1$ is sought as the solution of:

$$u_1 = \underset{u}{arg} \underset{\substack{\alpha,\gamma \\ u'M^{-1}u=1}}{max} \left[\underbrace{\left(l_p(u,\alpha,\gamma)\right)}_{\text{Goodness of fit}}^{1-s} \underbrace{\left(\phi_X(u)\right)}_{\text{SR}}^{s} \right]$$

$s \in [0;1]$ tunes the importance of the SR with respect to the GOF so that, at the maximum, *relative* variations of GOF and SR compensate:

$$\frac{\nabla l_p(u)}{l_p(u)} = -\frac{s}{1-s} \frac{\nabla \phi(u)}{\phi(u)}$$

SC-CoxR's mechanism

2. Estimation of the single-X component-based Cox Model

2.2. Calculating components

- Component $f^1 = Xu_1$ is sought as the solution of:

$$u_1 = \arg\max_{u} \max_{\substack{\alpha,\gamma \\ u' M^{-1} u = 1}} \left[\underbrace{\left(l_p(u,\alpha,\gamma) \right)}_{\text{Goodness of fit}}{}^{1-s} \underbrace{\left(\phi_X(u) \right)}_{\text{SR}}{}^{s} \right]$$

$s \in [0;1]$ tunes the importance of the SR with respect to the GOF so that, at the maximum, *relative* variations of GOF and SR compensate:

$$\frac{\nabla l_p(u)}{l_p(u)} = -\frac{s}{1-s} \frac{\nabla \phi(u)}{\phi(u)}$$

- A continuum-approach:
  - $s = 0$: the criterion is equal to l_p ; its maximisation leads to the classical Cox Regression
  - $s = 1$: the criterion is equal to $\phi_X(u)$; its maximisation leads to PCA for SR = component-variance and VPI.
  - $0 < s < 1$: the criterion is a trade-off between these extremes, and provides a supervised component-based Cox regression.

SC-CoxR's mechanism

2. Estimation of the single-X component-based Cox Model

2.2. Calculating components

- Calculating the first component:

$$u_1 = \underset{u}{arg} \underset{\substack{\alpha,\gamma \\ u'M^{-1}u=1}}{max} \left[(1-s)\ln\left(\prod_{i=1}^{n} \left[\frac{e^{\alpha u' x_{i,y_i} + \gamma' z_{i,y_i}}}{\sum_{j \in R(y_i)} e^{\alpha u' x_{j,y_i} + \gamma' z_{j,y_i}}} \right]^{\delta_i} \right) + s \ln \phi_X(u) \right]$$

SC-CoxR's mechanism

2. Estimation of the single-X component-based Cox Model

2.2. Calculating components

- Calculating the first component:

$$u_1 = \underset{u}{arg} \underset{\substack{\alpha,\gamma \\ u'M^{-1}u=1}}{max} \left[(1-s) \ln \left(\prod_{i=1}^{n} \left[\frac{e^{\alpha u' x_{i,y_i} + \gamma' z_{i,y_i}}}{\sum_{j \in R(y_i)} e^{\alpha u' x_{j,y_i} + \gamma' z_{j,y_i}}} \right]^{\delta_i} \right) + s \ln \phi_X(u) \right]$$

can be done by alternating, until convergence:

1) With a given u: Cox regression on $f = Xu$ and Z
   → update of α, γ

SC-CoxR's mechanism

2. Estimation of the single-X component-based Cox Model

2.2. Calculating components

- Calculating the first component:

$$u_1 = \underset{u}{arg} \underset{\substack{\alpha,\gamma \\ u'M^{-1}u=1}}{max} \left[(1-s) \ln \left(\prod_{i=1}^{n} \left[\frac{e^{\alpha u' x_{i,y_i} + \gamma' z_{i,y_i}}}{\sum_{j \in R(y_i)} e^{\alpha u' x_{j,y_i} + \gamma' z_{j,y_i}}} \right]^{\delta_i} \right) + s \ln \phi_X(u) \right]$$

can be done by alternating, until convergence:

1) With a given u: Cox regression on $f = Xu$ and Z
→ update of α, γ

2) With given α, γ : solving

$$u_1 = arg \underset{u'M^{-1}u=1}{max} \left[(1-s) \ln l_p(u,\alpha,\gamma) + s \ln \phi_X(u) \right]$$

→ update of u

(this step uses the dedicated PING algorithm, detailed later)

SC-CoxR's mechanism

2. Estimation of the single-X component-based Cox Model

2.2. Calculating components

- Calculating further components:

1) Every new component f^k must be uncorrelated with the former ones: $F^{k\text{-}1} = [f^1, \dots , f^{k\text{-}1}]$

N = number of lines of X = number of individuals-at-risk at time-points (i,t)
W = (N, N) diagonal line-weighting matrix

$$\langle f^k | F^{k-1} \rangle_W = 0 \quad \Rightarrow \quad D_k' u_k = 0 \quad \text{with} \quad D_k = X' W F^{k-1}$$

SC-CoxR's mechanism

2. Estimation of the single-X component-based Cox Model

2.2. Calculating components

- Calculating further components:

1) Every new component f^k must be uncorrelated with the former ones: $F^{k\text{-}1} = [f^1, \ldots , f^{k\text{-}1}]$

N = number of lines of X = number of individuals-at-risk at time-points (i,t)
W = (N, N) diagonal line-weighting matrix

$$\langle f^k | F^{k-1} \rangle_W = 0 \quad \Rightarrow \quad D_k' u_k = 0 \quad \text{with} \quad D_k = X' W F^{k-1}$$

Note on individual-weighting:

- Uniform weighting ⇒ each line of an individual ← weight inversely proportional to the number of the individual's lines.

- Weighting proportional to the individual's duration of follow-up ⇒ The weight of each line = proportional to the line's time span.

SC-CoxR's mechanism

2. Estimation of the single-X component-based Cox Model

2.2. Calculating components

- Calculating further components:

2) Former components $F^{k\text{-}1} = [f^1, \ldots, f^{k\text{-}1}]$ must now be included into the extra covariates in order to remove their effect.

$$Z^k := [Z\,;F^{k-1}]$$

$$u_k = \underset{u}{arg} \; \underset{\substack{\alpha,\gamma \\ u' M^{-1} u = 1 \\ D_k' u = 0}}{max} \left[(1-s)\ln\left(\prod_{i=1}^{n} \left[\frac{e^{\alpha u' x_{i,y_i} + \gamma' z^k_{i,y_i}}}{\sum_{j \in R(y_i)} e^{\alpha u' x_{j,y_i} + \gamma' z^k_{j,y_i}}} \right]^{\delta_i} \right) + s \ln \phi_X(u) \right]$$

performed as for u_1, with additional constraint:
$D_k' u = 0$

SC-CoxR's mechanism

3. The PING algorithm

$$\max_{\substack{u \in \mathbb{R}^p,\, u' M^{-1} u = 1 \\ D' u = 0}} h(u)$$

At the solution: $u = M \, \Pi_{D^\perp} \Gamma(u)$, M^{-1} − normed

with $\Pi_{D^\perp} := I - D(D' M D)^{-1} D' M$

Hence an iteration: $\tilde{u}^{[t+1]} = \dfrac{M \, \Pi_{D^\perp} \Gamma(u^{[t]})}{\| M \, \Pi_{D^\perp} \Gamma(u^{[t]}) \|_{M^{-1}}}$; $u^{[t+1]} = arg \max_{arc(u^{[t]}, \tilde{u}^{[t+1]})} h(u)$ (unidimensional)

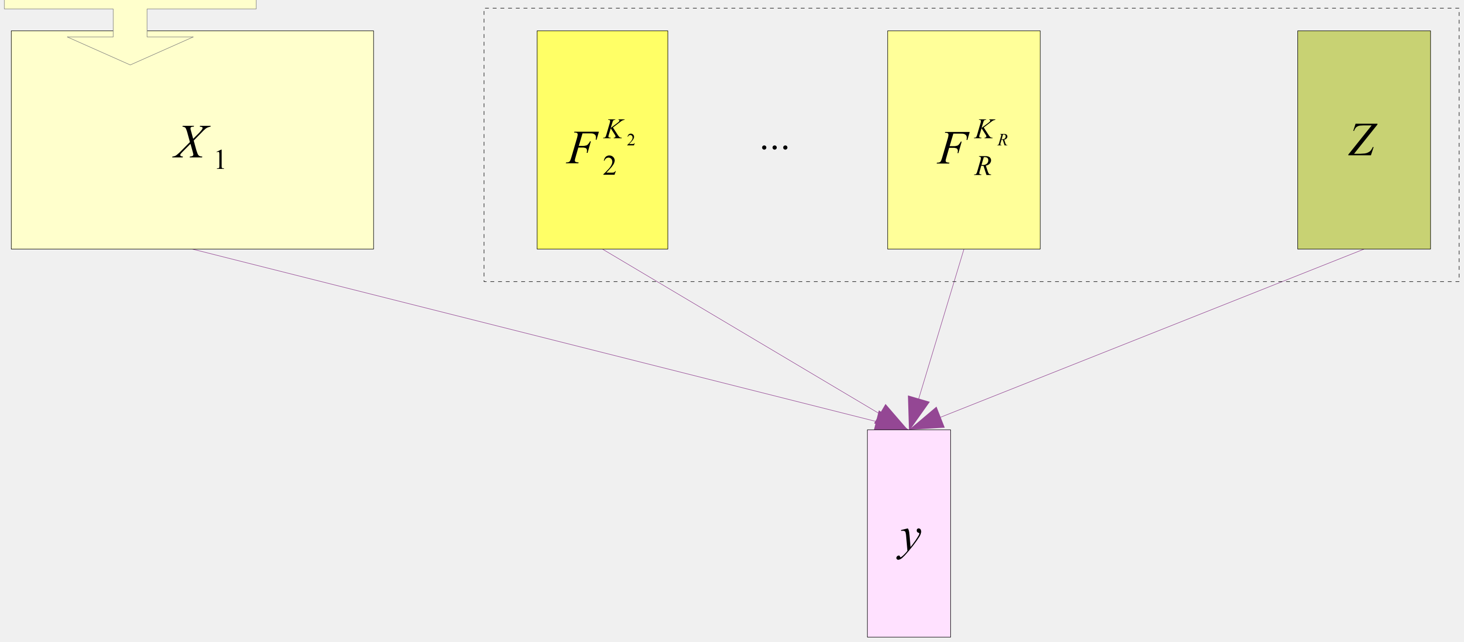
We proved that this iteration follows a direction of ascent.

SC-CoxR's mechanism

4. Estimating the Multiple-X model

Iterate over themes until overall convergence:

To calculate components in current theme ... ... consider components of other themes as additional covariates



SC-CoxR's mechanism

5. Assessing the Component Cox model

- Cross-Validation techniques for the Cox Model are provided by [van Houwelingen et al. (2006)]

K-fold subsampling :

Cross-validation quality coefficient of model M : $C_k(M)$

$$C_k(M) = l(\theta_{-k}, M) - l_{-k}(\theta_{-k}, M)$$

k^{ieth} sub-sample

calculated without the k^{ieth} sub-sample

$$C(M) = \frac{1}{K}\sum_{k=1}^{K} C_k(M)$$

SC-CoxR's mechanism

5. Assessing the Component Cox model

- Cross-Validation techniques for the Cox Model are provided by [van Houwelingen et al. (2006)]

  K-fold subsampling :

  Cross-validation quality coefficient of model M : $C_k(M)$

$$C_k(M) = l(\theta_{-k}, M) - l_{-k}(\theta_{-k}, M)$$

k^{ieth} sub-sample

calculated without the k^{ieth} sub-sample

$$C(M) = \frac{1}{K} \sum_{k=1}^{K} C_k(M)$$

- More simply, one can assess the significance of the components by :

  a) calculating the vectors $\{ U_r \}_{r=1,R}$ on a calibration sample C ;

  b) calculating the components' values on a spare test-sample T ;

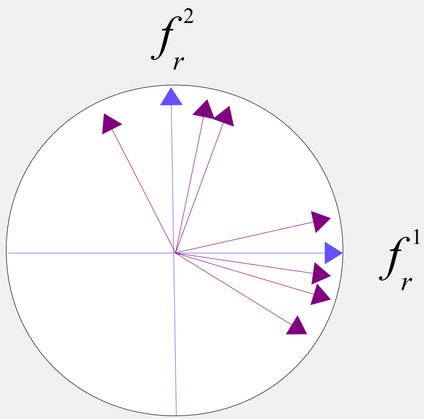
  c) performing a Cox Regression on T, with the associated classical significance-tests.

SC-CoxR's mechanism

6. Outputs

- Correlations of components with variables in each theme → correlation scatterplots

Theme X_r :



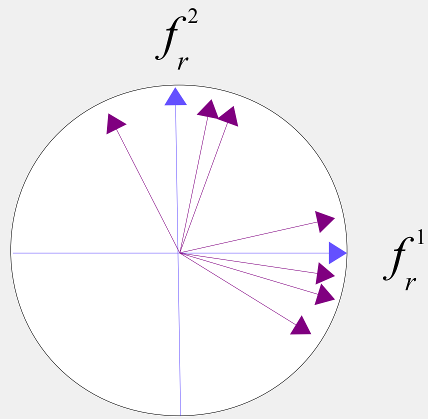
→ component thematic interpretation

SC-CoxR's mechanism

6. Outputs

- Correlations of components with variables in each theme → correlation scatterplots

Theme X_r :



→ component thematic interpretation

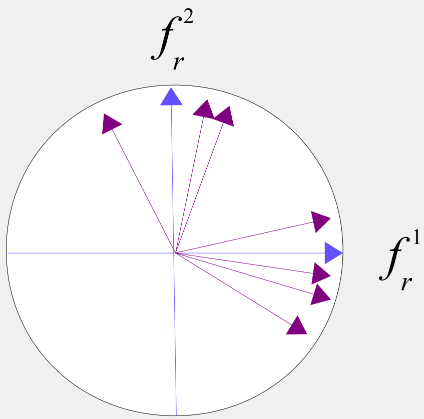
- Cox Regression on components → components' effects ; P-values / confidence interval on test-sample T, or boostrap confidence interval

SC-CoxR's mechanism

6. Outputs

- Correlations of components with variables in each theme → correlation scatterplots

Theme X_r :



→ component thematic interpretation

- Cox Regression on components → components' effects ; P-values / confidence interval on test-sample T, or boostrap confidence interval

- Components' effects + vectors U
  → (regularised) coefficients of original variables in linear predictor
  \+ boostrap confidence interval

Short simulation study

1. Simulation scheme

- Time-span : [0,30] , divided in 30 unit-length elementary intervals.
- Baseline hazard function:

$$h_0(t)=a+b(t-t_m)^2 \quad \text{with} \quad t_m=12\,,\, a=.2\,,\, b=10^{-3}$$

Short simulation study

1. Simulation scheme

- Time-span : [0,30] , divided in 30 unit-length elementary intervals.
- Baseline hazard function:

$$h_0(t) = a + b(t - t_m)^2 \quad \text{with} \quad t_m = 12,\ a = .2,\ b = 10^{-3}$$

- 75 subjects simulated with bundle-structures:

Variables at subject level : $\psi_i^j \sim N(0;1),\ j \in \{1,2,3\},\ i \in \{1,\ldots,75\}$

Variables at subject-time level : $\phi_{it}^j \sim N(0;1),\ j \in \{1,2,3\},\ i \in \{1,\ldots,75\},\ t \in \{1,\ldots,30\}$

Combination : $\forall (i, t, j) \ : \ \xi_{it}^j = \psi_i^j + \phi_{it}^j$

Short simulation study

1. Simulation scheme

- Time-span : [0,30] , divided in 30 unit-length elementary intervals.
- Baseline hazard function:

$$h_0(t)=a+b(t-t_m)^2 \quad \text{with} \quad t_m=12\,,\, a=.2\,,\, b=10^{-3}$$

- 75 subjects simulated with bundle-structures:

Variables at subject level : $\psi_i^j \sim N(0;1),\ j\in\{1{,}2{,}3\},\ i\in\{1,\ldots,75\}$

Variables at subject-time level : $\phi_{it}^j \sim N(0;1),\ j\in\{1{,}2{,}3\},\ i\in\{1,\ldots,75\},\ t\in\{1,\ldots,30\}$

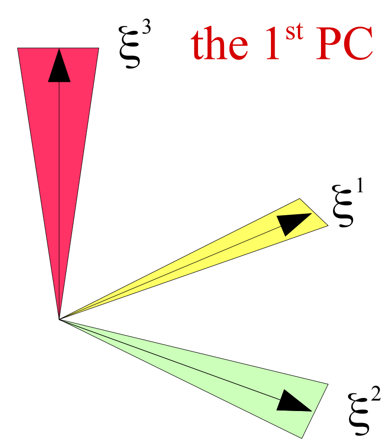
Combination : $\forall (i,t,j)\ :\ \xi_{it}^j=\psi_i^j+\phi_{it}^j$

$\xi^1, \xi^2, \xi^3 \rightarrow$ 3 explanatory variable-bundles:

- B_1 : 4 variables $x^j = \xi^1 + \varepsilon^j$;
- B_2 : 6 variables $x^j = \xi^2 + \varepsilon^j$;
- B_3 : 10 variables $x^j = \xi^3 + \varepsilon^j$;

where $\varepsilon^j = N(0;\sigma^2)$ noise with $\sigma = 0.3$

$+\ B_4$: 20 noise-variables $x^j \sim N(0;1)$



Short simulation study

1. Simulation scheme

- Time-span : [0,30] , divided in 30 unit-length elementary intervals.
- Baseline hazard function:

$$h_0(t)=a+b(t-t_m)^2 \quad \text{with} \quad t_m=12\,,\ a=.2\,,\ b=10^{-3}$$

- 75 subjects simulated with bundle-structures:

Variables at subject level : $\psi_i^j \sim N(0;1),\ j\in\{1,2,3\},\ i\in\{1,\dots,75\}$

Variables at subject-time level : $\phi_{it}^j \sim N(0;1),\ j\in\{1,2,3\},\ i\in\{1,\dots,75\},\ t\in\{1,\dots,30\}$

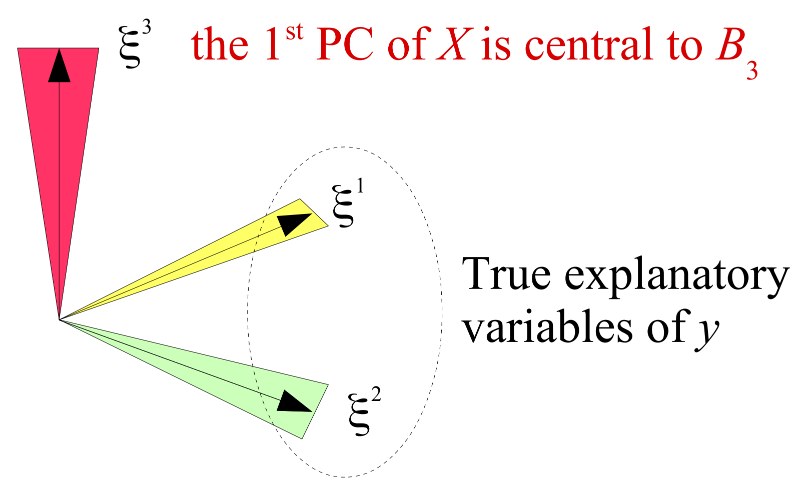
Combination : $\forall(i,t,j)\ :\ \xi_{it}^j=\psi_i^j+\phi_{it}^j$

$\xi^1, \xi^2, \xi^3 \rightarrow$ 3 explanatory variable-bundles:

- B_1 : 4 variables $x^j = \xi^1 + \varepsilon^j$;
- B_2 : 6 variables $x^j = \xi^2 + \varepsilon^j$;
- B_3 : 10 variables $x^j = \xi^3 + \varepsilon^j$;

where $\varepsilon^j = N(0;\sigma^2)$ noise with $\sigma = 0.3$

+ B_4 : 20 noise-variables $x^j \sim N(0;1)$



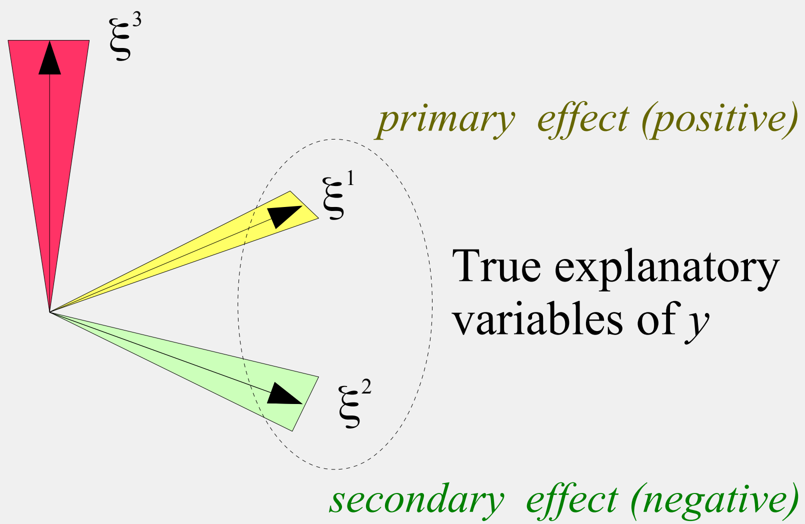
Short simulation study

1. Simulation scheme

• Exponential Survival Time y with hazard function:

$$\forall(i,t,j)\ :\ h_i(t)=h_0(t)\,e^{\eta_{it}} \qquad \text{where} \qquad \eta_{it}=.25+\underbrace{\xi^1_{it}-.5\,\xi^2_{it}}$$

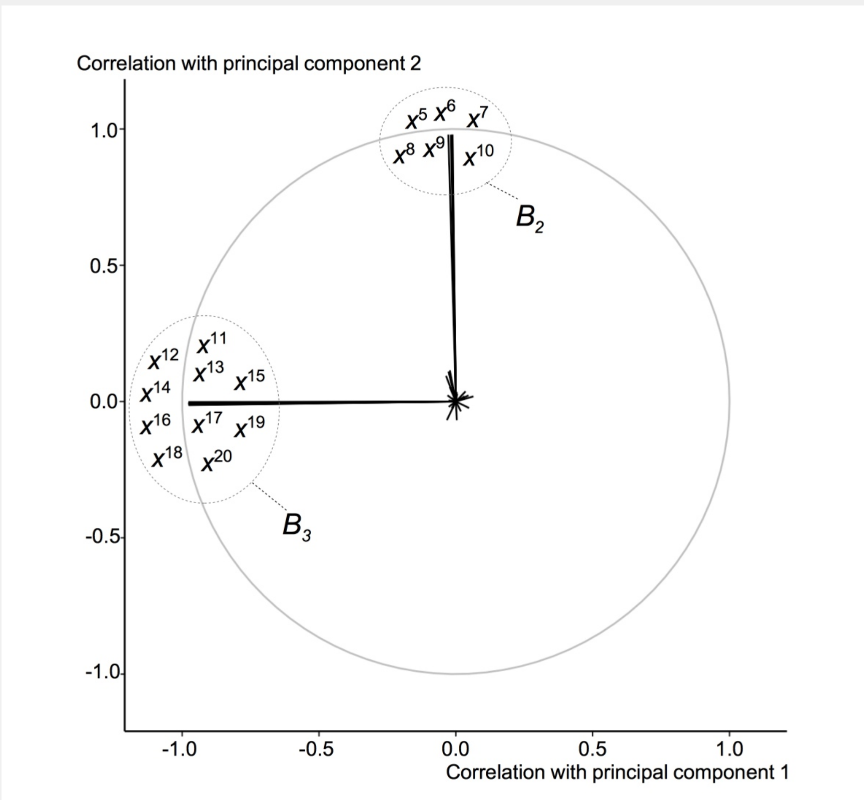
$\Rightarrow$ X_3 (1^{st} PC) is a nuisance variable-bundle.



Short simulation study

2. Results

$s = 1$; $l = 1$; $\tau = 0$ (= PCA)



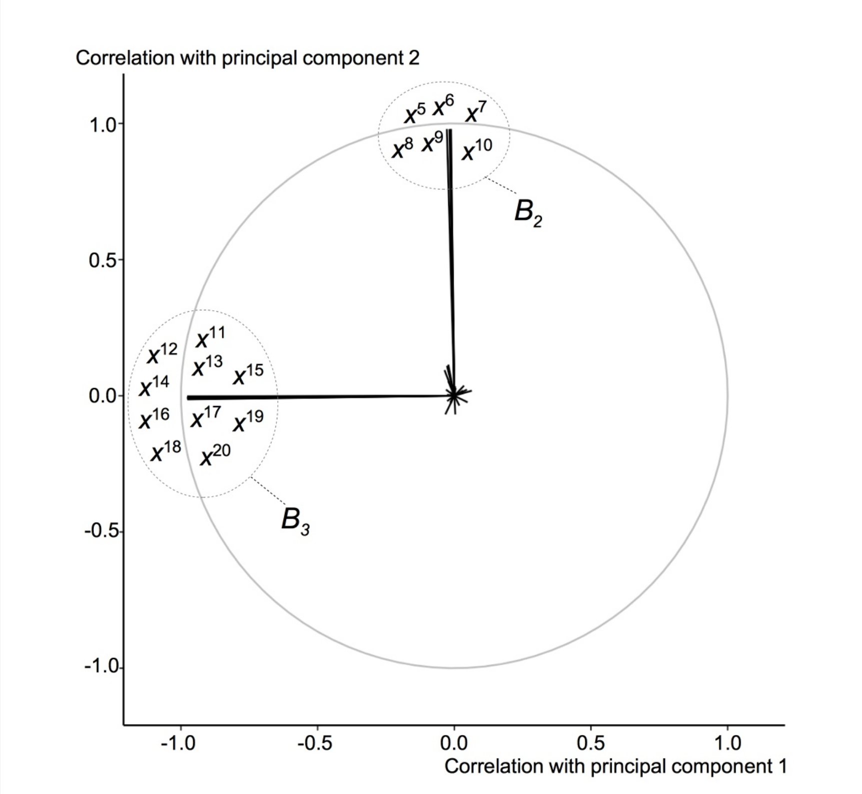
Cox-regression on the components :

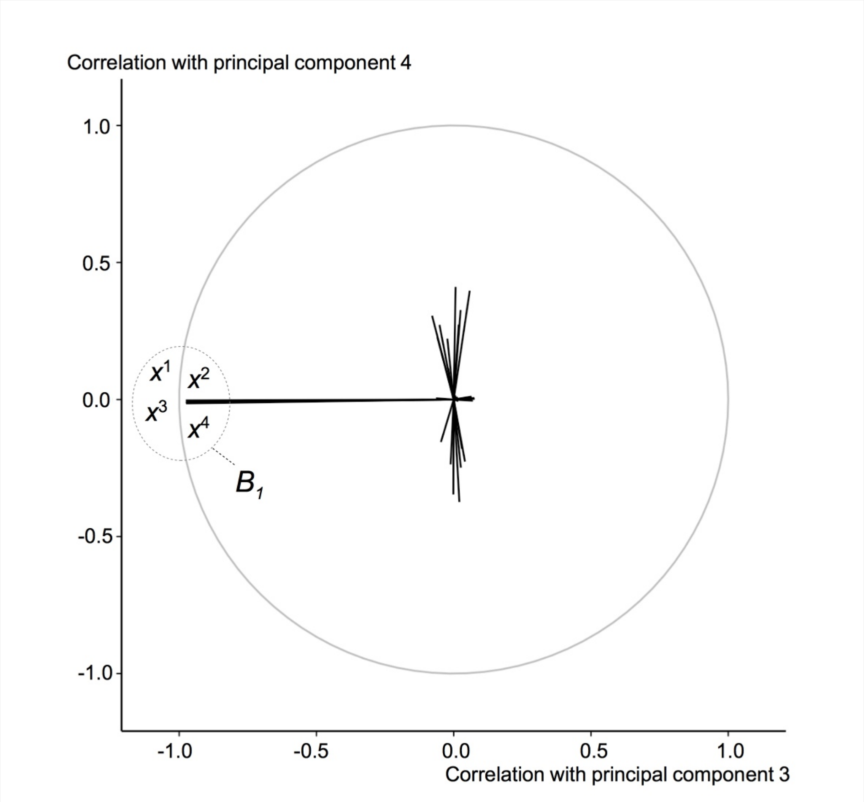
f^1: coefficient = -0.03 ; p=0.830
f^2 : coefficient = -0.42; p=0.004

Short simulation study

2. Results

$s = 1$; $l = 1$; $\tau = 0$ (= PCA)





Cox-regression on the components :

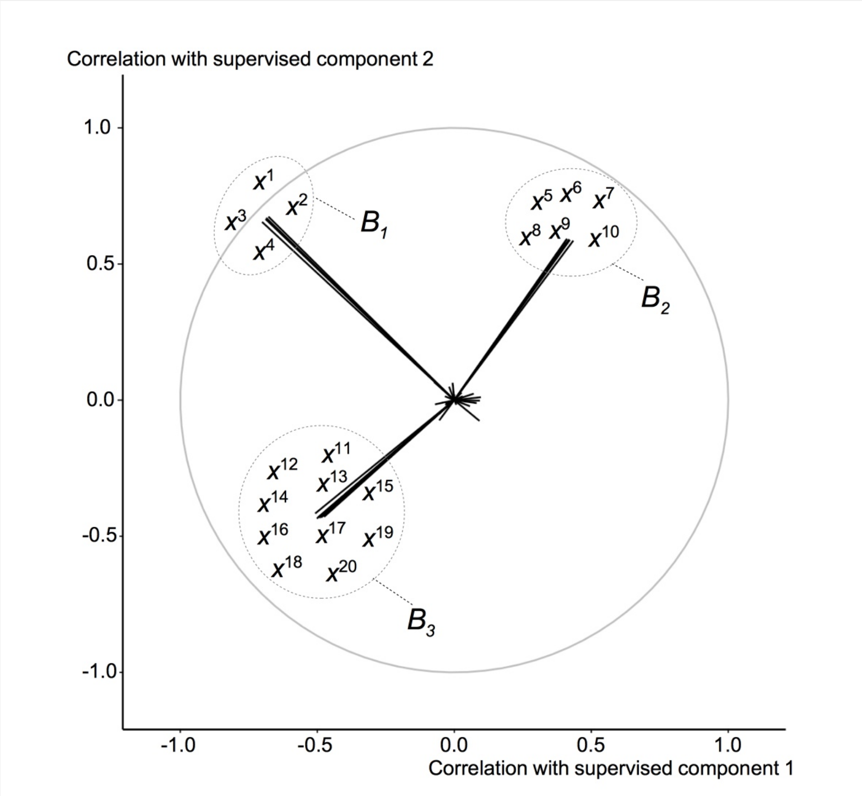
f^1: coefficient = -0.03 ; p=0.830
f^2 : coefficient = -0.42; p=0.004

f^3: coefficient = -1.60 ; $p<10^{-16}$
f^4 : coefficient = -0.09; p=0.49

Short simulation study

2. Results

$s = 0.95$; $l = 1$; $\tau = 0.01$



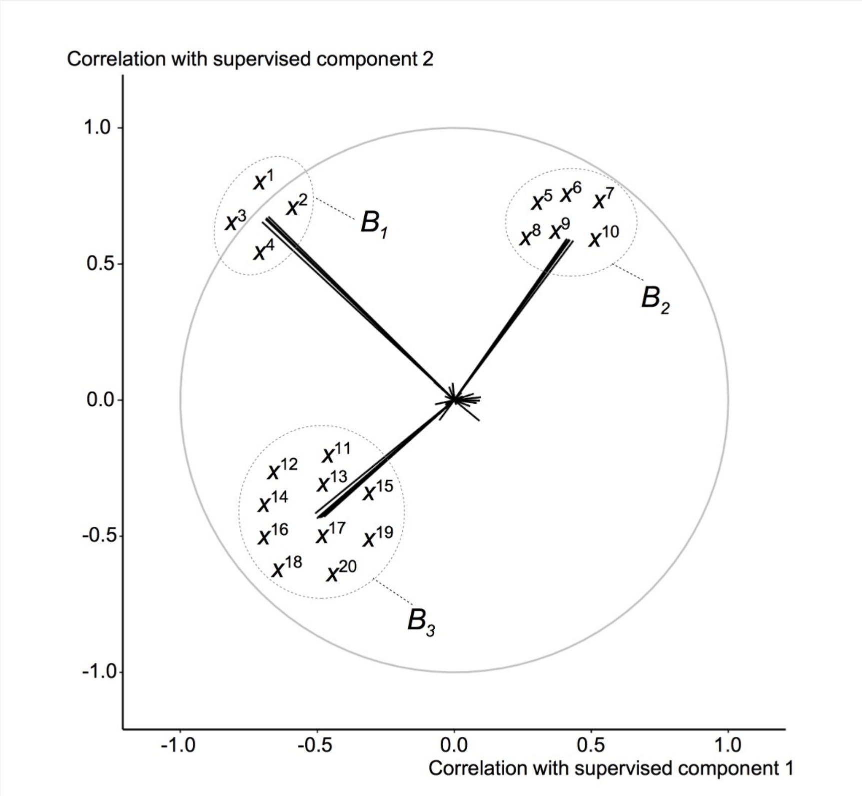
Cox-regression on the components (on test sample):

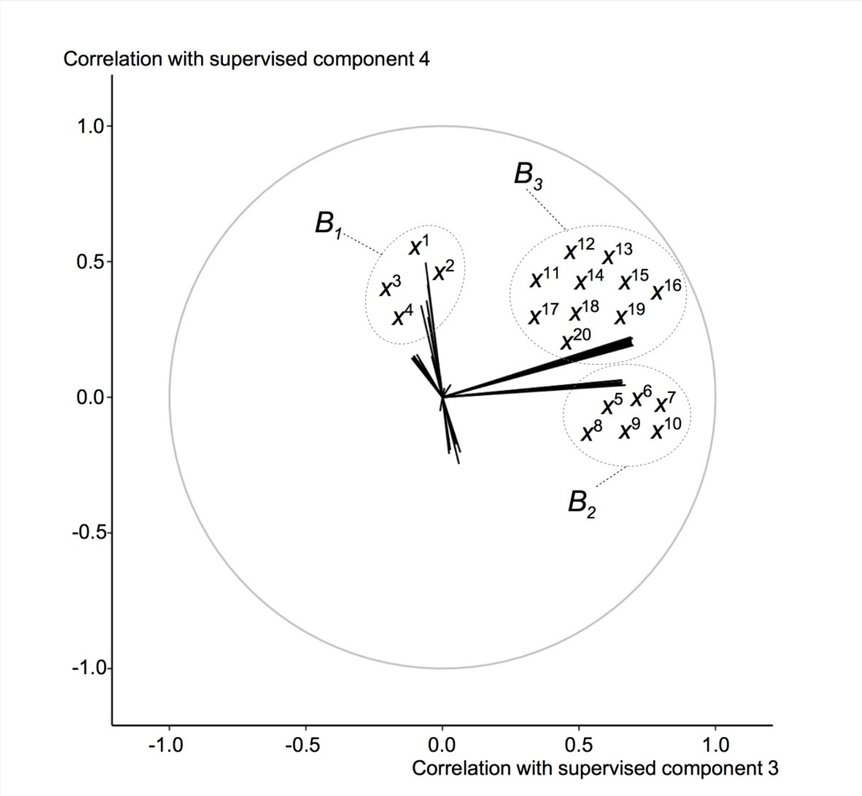
f^1: coefficient = -1.69 ; $p<2.00\ 10^{-16}$
f^2 : coefficient = 0.69; $p=1.49\ 10^{-5}$

Short simulation study

2. Results

$s = 0.95$; $l = 1$; $\tau = 0.01$





Cox-regression on the components (on test sample):

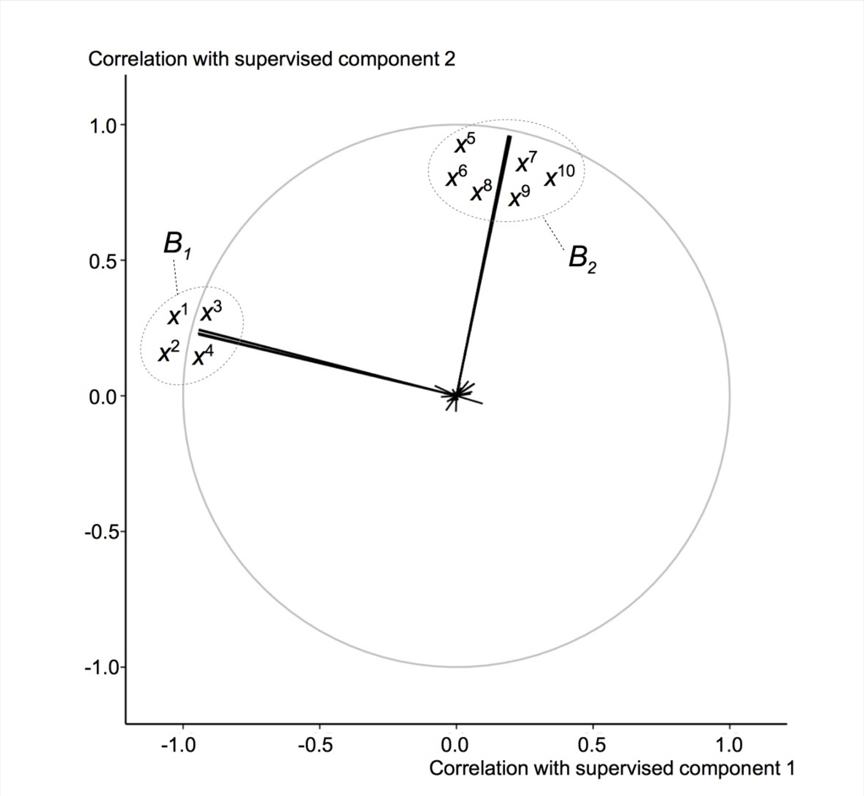
f^1: coefficient = -1.69 ; p<2.00 10^{-16}
f^2 : coefficient = 0.69; p=1.49 10^{-5}

f^3: coefficient = -0.19 ; p=0.19
f^4 : coefficient = -0.09; p=0.56

Short simulation study

2. Results

$s = 0.95$; $l = 4$; $\tau = 0.01$



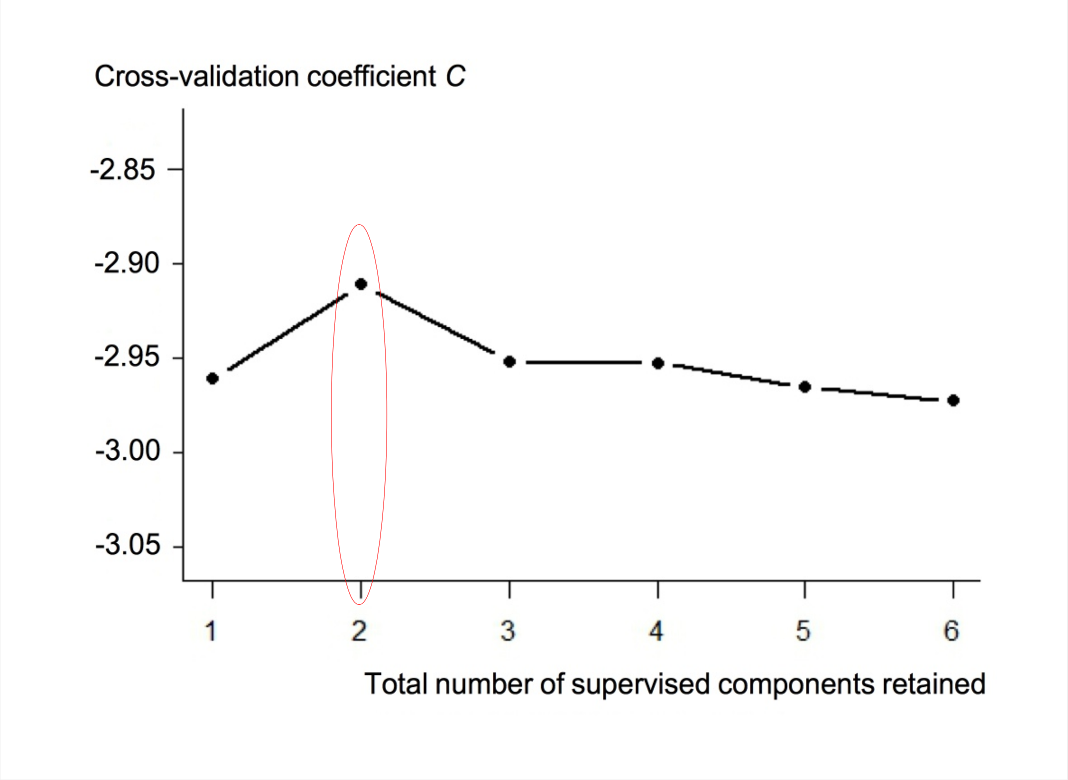
Cox-regression on the components (on test sample):

f^1: coefficient = -1.92 ; $p<2.00\ 10^{-16}$
f^2 : coefficient = -0.27; p=0.068

Short simulation study

2. Results

$s = 0.95$; $l = 4$; $\tau = 0.01$



Right!

Cross-validation performance according
to the number of components retained

Short simulation study

2. Results

The impact of τ (for $s = 0.95$, $l = 4$):

Coefficients with unstable values and signs

		parameter τ tuning regularization									
		$\tau = 0$		$\tau = 0.1$		$\tau = 0.3$		$\tau = 0.5$		$\tau = 0.7$	
		components f^1, f^2									
		f^1	f^2	f^1	f^2	f^1	f^2	f^1	f^2	f^1	f^2
bundle B_1	x^1	-0.19	0.04	-0.22	0.09	-0.26	0.08	-0.30	0.08	-0.35	0.09
	x^2	-0.26	0.00	-0.25	0.05	-0.27	0.06	-0.31	0.07	-0.35	0.08
	x^3	-0.38	0.00	-0.31	0.05	-0.30	0.05	-0.32	0.07	-0.36	0.08
	x^4	-0.13	0.03	-0.22	0.04	-0.26	0.06	-0.30	0.07	-0.35	0.09
bundle B_2	x^5	0.18	0.02	0.11	0.19	0.07	0.20	0.07	0.02	0.07	0.02
	x^6	0.20	-0.02	0.10	0.17	0.07	0.19	0.07	0.02	0.07	0.02
	x^7	0.43	0.05	0.19	0.21	0.11	0.21	0.08	0.02	0.08	0.03
	x^8	-0.12	-0.02	-0.03	0.15	0.01	0.18	0.03	0.02	0.05	0.02
	x^9	-0.31	-0.02	-0.09	0.14	-0.01	0.17	0.02	0.02	0.04	0.02
	x^{10}	-0.16	0.00	-0.03	0.16	0.02	0.18	0.03	0.02	0.05	0.02
bundle B_3	x^{11}	0.20	-0.13	0.13	0.02	0.06	0.02	0.04	0.01	0.03	0.01
	x^{12}	0.24	-0.10	-0.17	-0.02	-0.07	-0.01	0.04	-0.01	-0.02	-0.01
	x^{13}	-0.42	-0.14	-0.04	0.00	-0.02	0.00	-0.01	0.00	0.00	0.00
	x^{14}	-0.10	-0.13	-0.06	0.01	-0.03	0.00	-0.02	0.00	-0.01	0.00
	x^{15}	-0.15	-0.13	-0.05	-0.01	-0.02	0.00	-0.01	0.00	0.00	0.00
	x^{16}	-0.15	-0.14	0.10	0.00	0.04	0.01	0.03	0.00	0.01	0.00
	x^{17}	0.19	-0.13	0.11	0.01	0.05	0.01	0.03	0.00	0.02	0.00
	x^{18}	0.23	-0.11	-0.06	0.02	-0.06	0.01	-0.06	0.01	-0.06	0.01
	x^{19}	-0.06	0.00	-0.03	0.25	-0.03	0.01	-0.03	0.00	-0.02	-0.01
	x^{20}	-0.03	0.00	-0.03	-0.02	-0.03	0.01	-0.03	-0.01	-0.02	-0.01
		correlation of the linear predictor with its estimate									
$\rho(\eta, \hat{\eta})$		0.948		0.965		0.972		0.977		0.982	

Short simulation study

2. Results

The impact of τ (for $s = 0.95$, $l = 4$):

		parameter τ tuning regularization									
		$\tau = 0$		$\tau = 0.1$		$\tau = 0.3$		$\tau = 0.5$		$\tau = 0.7$	
		components f^1, f^2									
		f^1	f^2	f^1	f^2	f^1	f^2	f^1	f^2	f^1	f^2
bundle B_1	x^1	-0.19	0.04	-0.22	0.09	-0.26	0.08	-0.30	0.08	-0.35	0.09
	x^2	-0.26	0.00	-0.25	0.05	-0.27	0.06	-0.31	0.07	-0.35	0.08
	x^3	-0.38	0.00	-0.31	0.05	-0.30	0.05	-0.32	0.07	-0.36	0.08
	x^4	-0.13	0.03	-0.22	0.04	-0.26	0.06	-0.30	0.07	-0.35	0.09
bundle B_2	x^5	0.18	0.02	0.11	0.19	0.07	0.20	0.07	0.02	0.07	0.02
	x^6	0.20	-0.02	0.10	0.17	0.07	0.19	0.07	0.02	0.07	0.02
	x^7	0.43	0.05	0.19	0.21	0.11	0.21	0.08	0.02	0.08	0.03
	x^8	-0.12	-0.02	-0.03	0.15	0.01	0.18	0.03	0.02	0.05	0.02
	x^9	-0.31	-0.02	-0.09	0.14	-0.01	0.17	0.02	0.02	0.04	0.02
	x^{10}	-0.16	0.00	-0.03	0.16	0.02	0.18	0.03	0.02	0.05	0.02
bundle B_3	x^{11}	0.20	-0.13	0.13	0.02	0.06	0.02	0.04	0.01	0.03	0.01
	x^{12}	0.24	-0.10	-0.17	-0.02	-0.07	-0.01	0.04	-0.01	-0.02	-0.01
	x^{13}	-0.42	-0.14	-0.04	0.00	-0.02	0.00	-0.01	0.00	0.00	0.00
	x^{14}	-0.10	-0.13	-0.06	0.01	-0.03	0.00	-0.02	0.00	-0.01	0.00
	x^{15}	-0.15	-0.13	-0.05	-0.01	-0.02	0.00	-0.01	0.00	0.00	0.00
	x^{16}	-0.15	-0.14	0.10	0.00	0.04	0.01	0.03	0.00	0.01	0.00
	x^{17}	0.19	-0.13	0.11	0.01	0.05	0.01	0.03	0.00	0.02	0.00
	x^{18}	0.23	-0.11	-0.06	0.02	-0.06	0.01	-0.06	0.01	-0.06	0.01
	x^{19}	-0.06	0.00	-0.03	0.25	-0.03	0.01	-0.03	0.00	-0.02	-0.01
	x^{20}	-0.03	0.00	-0.03	-0.02	-0.03	0.01	-0.03	-0.01	-0.02	-0.01
		correlation of the linear predictor with its estimate									
$\rho(\eta, \hat{\eta})$		0.948		0.965		0.972		0.977		0.982	

Coefficients with unstable values and signs

Coefficients with stable & even values and signs

Short simulation study

2. Results

The impact of τ (for $s = 0.95$, $l = 4$):

		parameter τ tuning regularization									
		$\tau = 0$		$\tau = 0.1$		$\tau = 0.3$		$\tau = 0.5$		$\tau = 0.7$	
		components f^1, f^2									
		f^1	f^2	f^1	f^2	f^1	f^2	f^1	f^2	f^1	f^2
bundle B_1	x^1	-0.19	0.04	-0.22	0.09	-0.26	0.08	-0.30	0.08	-0.35	0.09
	x^2	-0.26	0.00	-0.25	0.05	-0.27	0.06	-0.31	0.07	-0.35	0.08
	x^3	-0.38	0.00	-0.31	0.05	-0.30	0.05	-0.32	0.07	-0.36	0.08
	x^4	-0.13	0.03	-0.22	0.04	-0.26	0.06	-0.30	0.07	-0.35	0.09
bundle B_2	x^5	0.18	0.02	0.11	0.19	0.07	0.20	0.07	0.02	0.07	0.02
	x^6	0.20	-0.02	0.10	0.17	0.07	0.19	0.07	0.02	0.07	0.02
	x^7	0.43	0.05	0.19	0.21	0.11	0.21	0.08	0.02	0.08	0.03
	x^8	-0.12	-0.02	-0.03	0.15	0.01	0.18	0.03	0.02	0.05	0.02
	x^9	-0.31	-0.02	-0.09	0.14	-0.01	0.17	0.02	0.02	0.04	0.02
	x^{10}	-0.16	0.00	-0.03	0.16	0.02	0.18	0.03	0.02	0.05	0.02
bundle B_3	x^{11}	0.20	-0.13	0.13	0.02	0.06	0.02	0.04	0.01	0.03	0.01
	x^{12}	0.24	-0.10	-0.17	-0.02	-0.07	-0.01	0.04	-0.01	-0.02	-0.01
	x^{13}	-0.42	-0.14	-0.04	0.00	-0.02	0.00	-0.01	0.00	0.00	0.00
	x^{14}	-0.10	-0.13	-0.06	0.01	-0.03	0.00	-0.02	0.00	-0.01	0.00
	x^{15}	-0.15	-0.13	-0.05	-0.01	-0.02	0.00	-0.01	0.00	0.00	0.00
	x^{16}	-0.15	-0.14	0.10	0.00	0.04	0.01	0.03	0.00	0.01	0.00
	x^{17}	0.19	-0.13	0.11	0.01	0.05	0.01	0.03	0.00	0.02	0.00
	x^{18}	0.23	-0.11	-0.06	0.02	-0.06	0.01	-0.06	0.01	-0.06	0.01
	x^{19}	-0.06	0.00	-0.03	0.25	-0.03	0.01	-0.03	0.00	-0.02	-0.01
	x^{20}	-0.03	0.00	-0.03	-0.02	-0.03	0.01	-0.03	-0.01	-0.02	-0.01
		correlation of the linear predictor with its estimate									
$\rho(\eta, \hat{\eta})$		0.948		0.965		0.972		0.977		0.982	

Coefficients with unstable values and signs

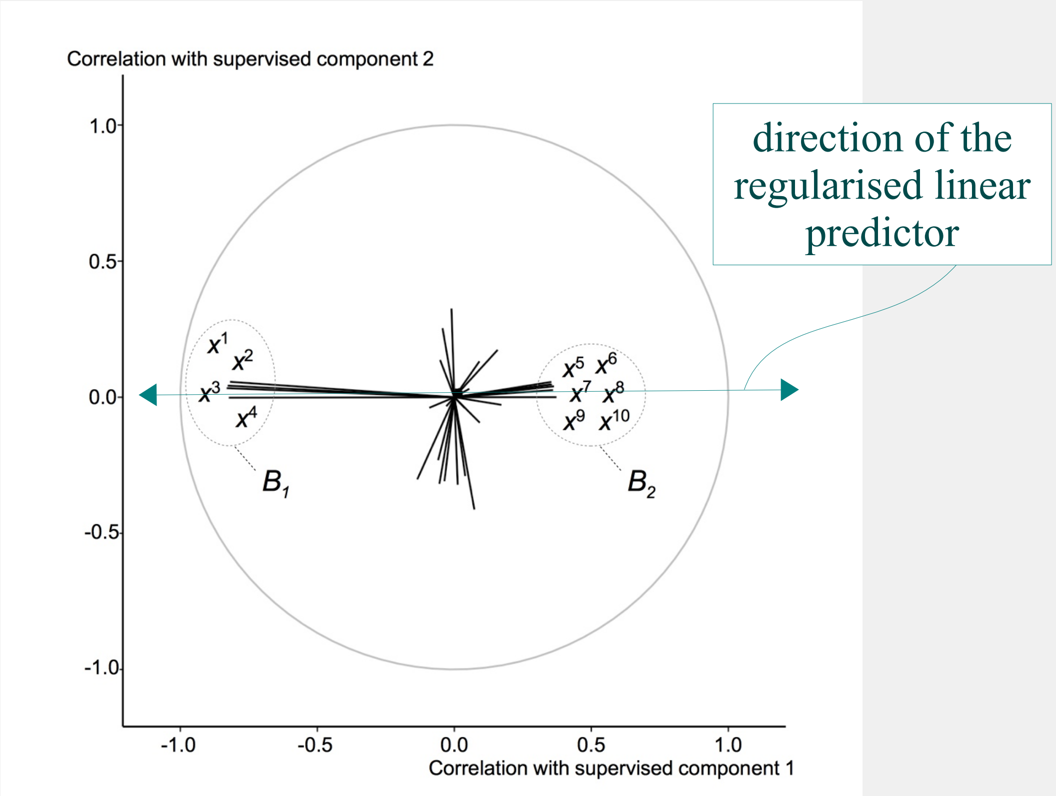
Coefficients with stable & even values and signs

Better fit

Short simulation study

2. Results

$s = 0.00$



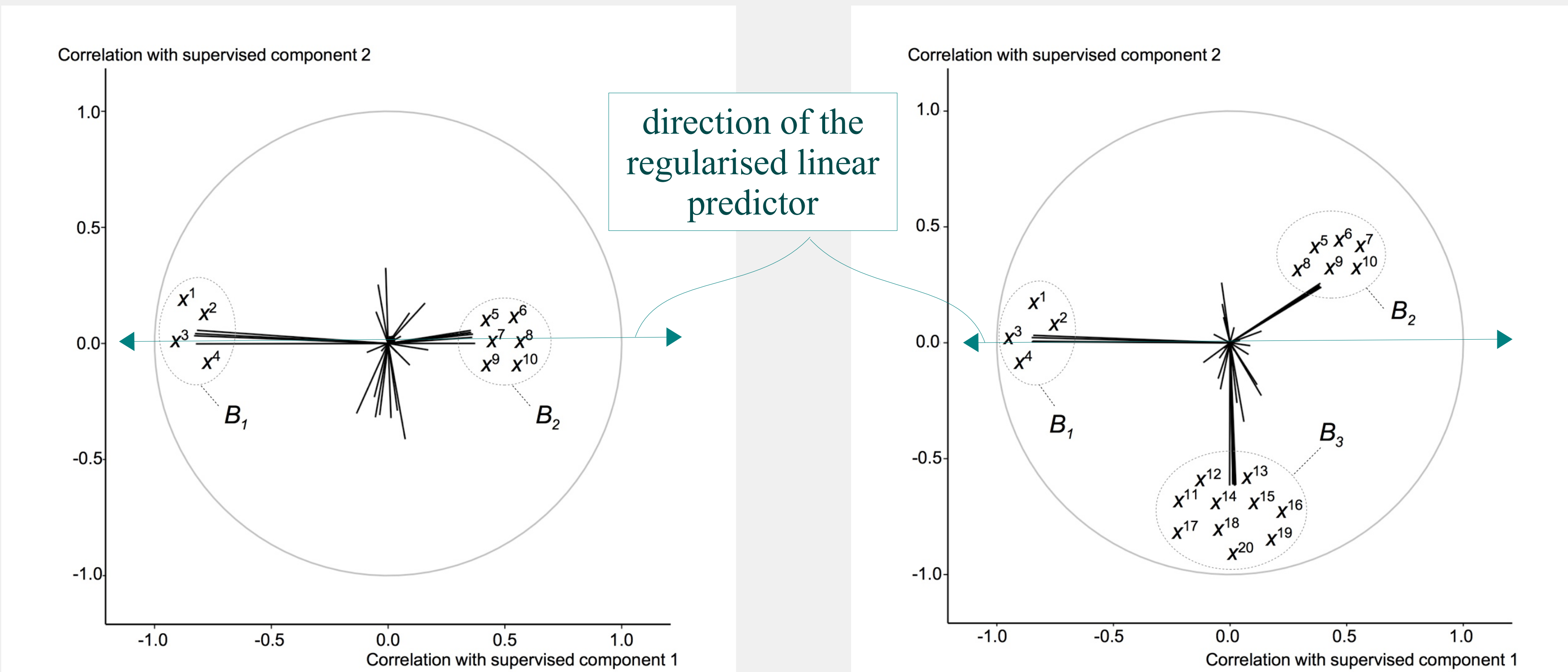
Cox-regression on the components (test sample):

f^1: coefficient = -1.85 ; p<2.00 10^{-16}
f^2 : coefficient = -0.12; p=0.35

Short simulation study

2. Results

$s = 0.00$ $s = 0.1$; $l = 1$; $\tau = 0.01$



Cox-regression on the components (test sample):

f^1: coefficient = -1.85 ; $p<2.00\ 10^{-16}$
f^2 : coefficient = -0.12; p=0.35

f^1: coefficient = -1.83 ; $p<2.00\ 10^{-16}$
f^2 : coefficient = -0.11; p=0.40

An application to life-history analysis

1. The data :

- From the 2001 retrospective survey conducted by Antoine and Fall:
  Crisis, passage to adult age, and family in poor and middle classes in Dakar.

- The subjects: 222 married men born before 1967 and residing in Dakar, Senegal.

- The event under study: the shift from monogamy to polygamy.

  → 55 events (marriages to a second wife).

An application to life-history analysis

1. The data :

- From the 2001 retrospective survey conducted by Antoine and Fall:
  Crisis, passage to adult age, and family in poor and middle classes in Dakar.

- The subjects: 222 married men born before 1967 and residing in Dakar, Senegal.

- The event under study: the shift from monogamy to polygamy.

  → 55 events (marriages to a second wife).

- Covariates: 107 time-varying variables, some of which highly correlated.

  ⇒ direct Cox regression impossible.

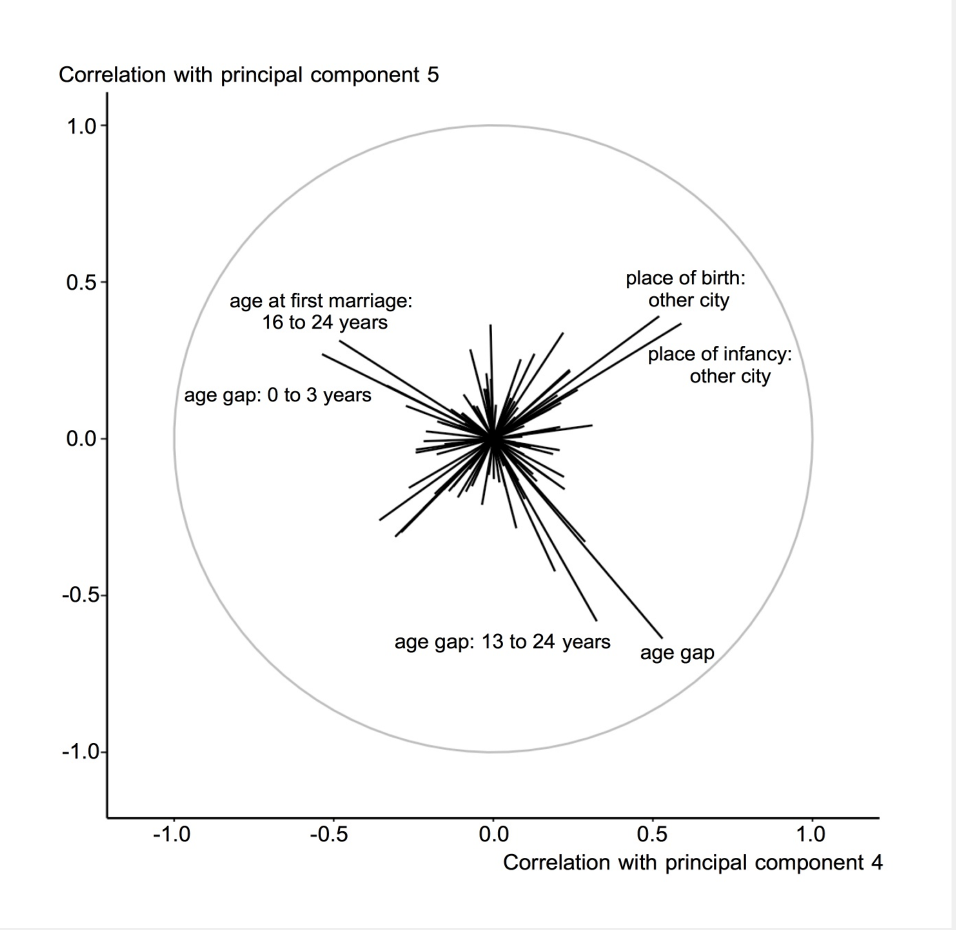
- 0.95-confidence intervals obtained by bootstrap.

An application to life-history analysis

2. Results

$s = 1$, $l = 1$ (PC-CoxR)

Components 4 and 5 have the smallest p-values. Only component 5 has a p-value < 0.05 (0.002).

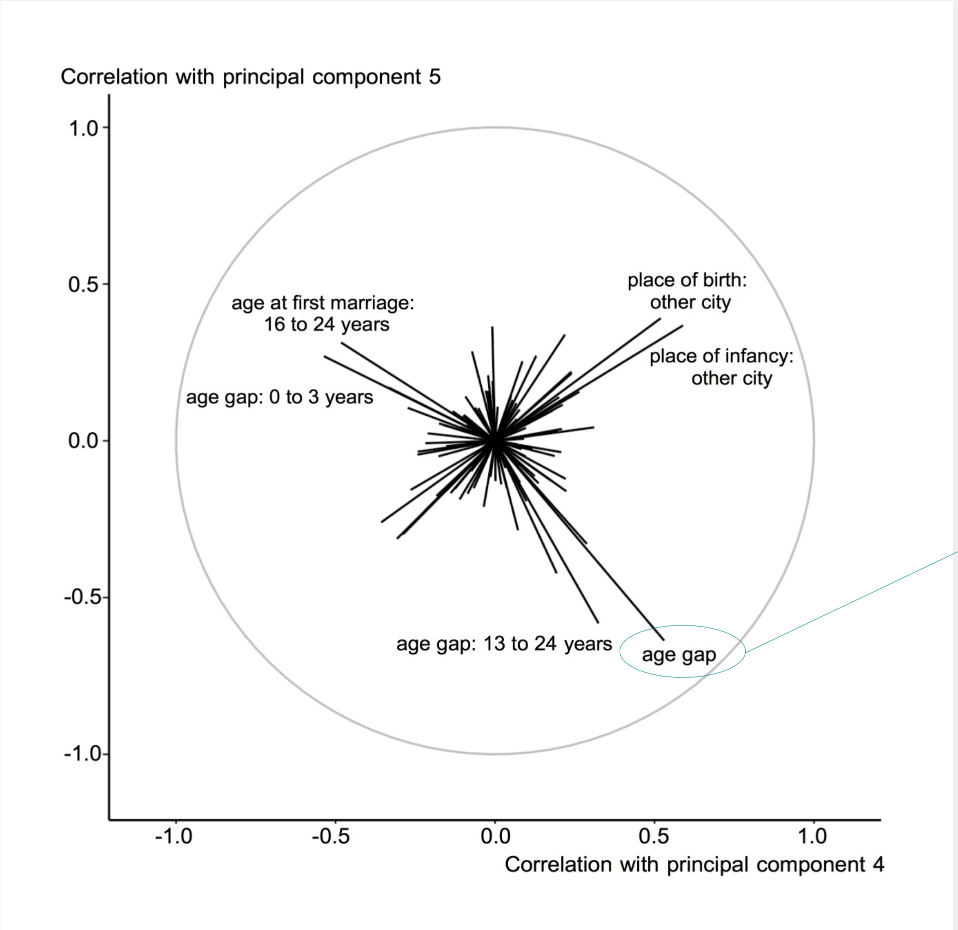


An application to life-history analysis

2. Results

$s = 1$, $l = 1$ (PC-CoxR)

Components 4 and 5 have the smallest p-values. Only component 5 has a p-value < 0.05 (0.002).



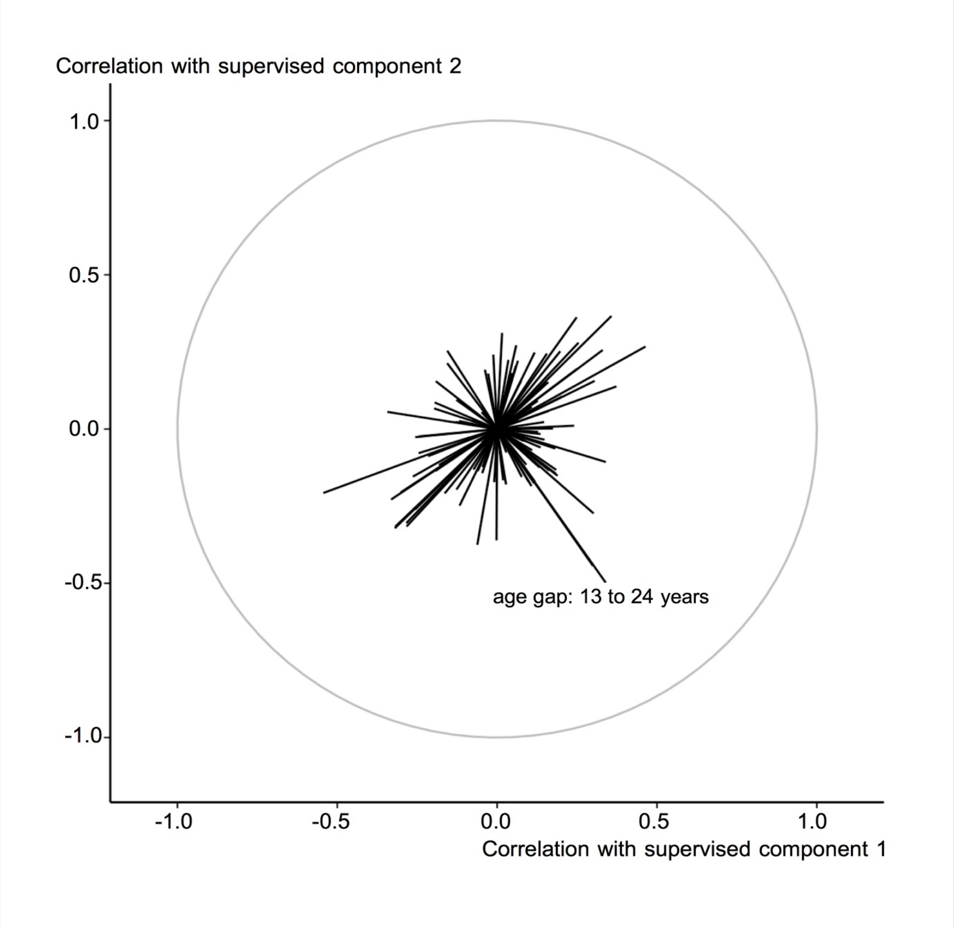
Interpretation is weak.

Only variable with high cosine on the (4,5) plane: age-gap.

An application to life-history analysis

2. Results

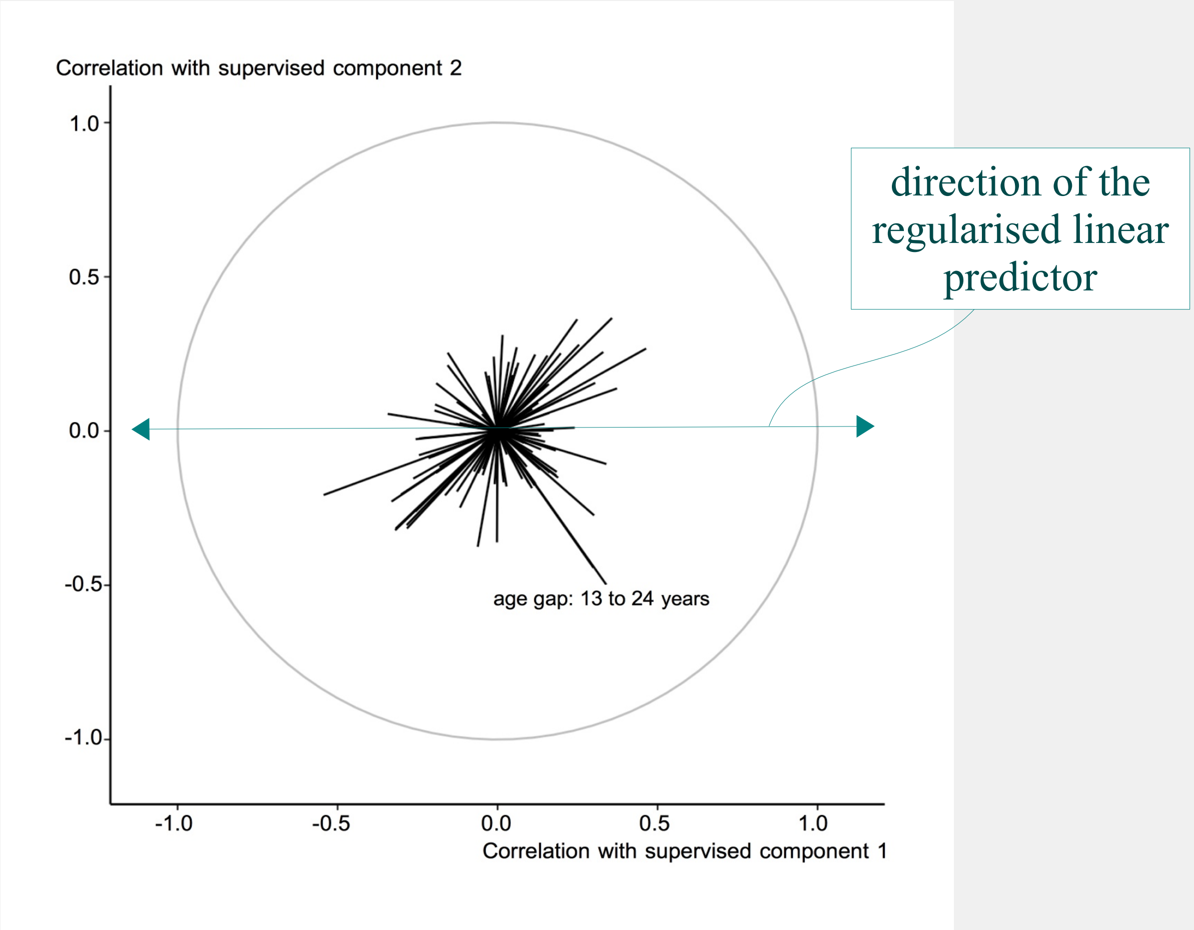
$s = 10^{-3}$; $l = 1$; $\tau = 1$



An application to life-history analysis

2. Results

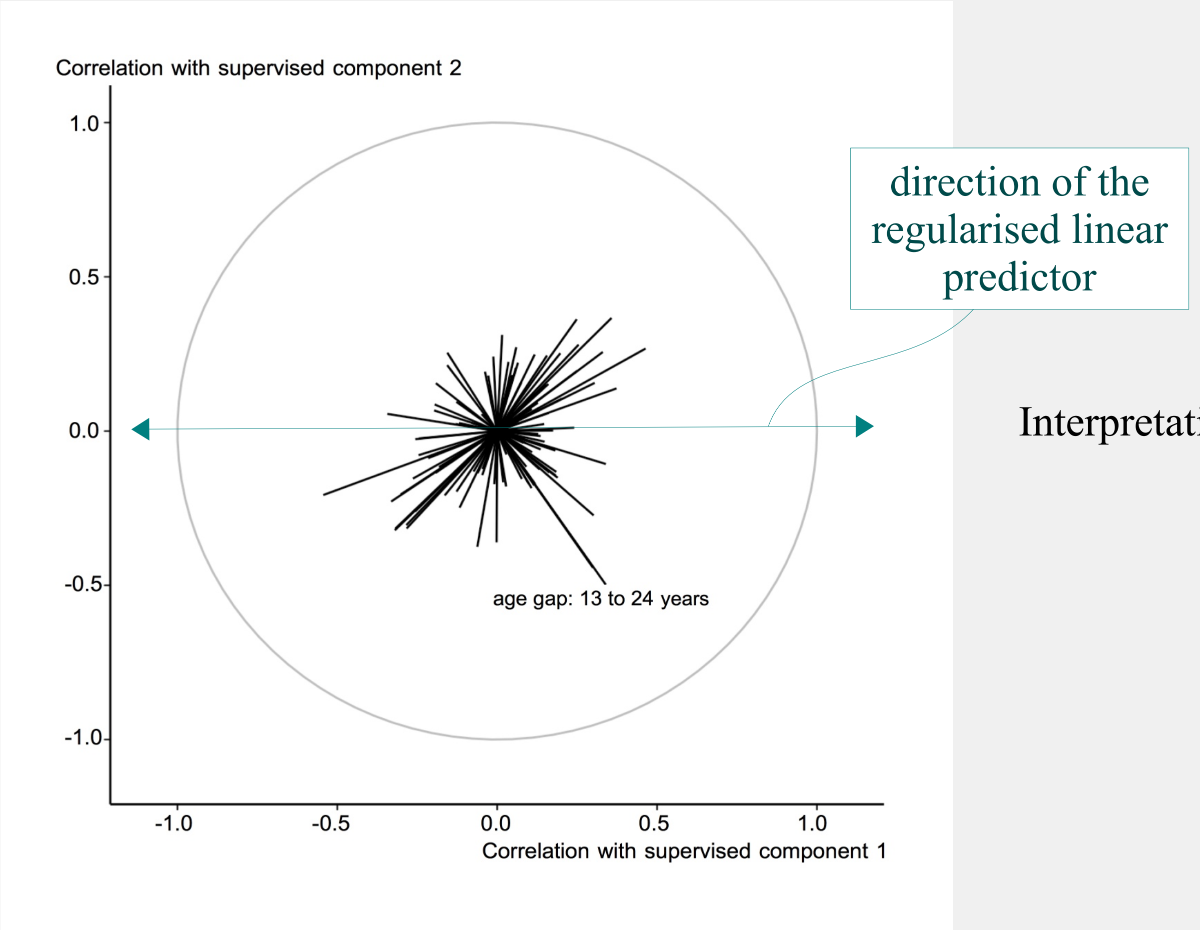
$s = 10^{-3}$; $l = 1$; $\tau = 1$



An application to life-history analysis

2. Results

$s = 10^{-3}$; $l = 1$; $\tau = 1$

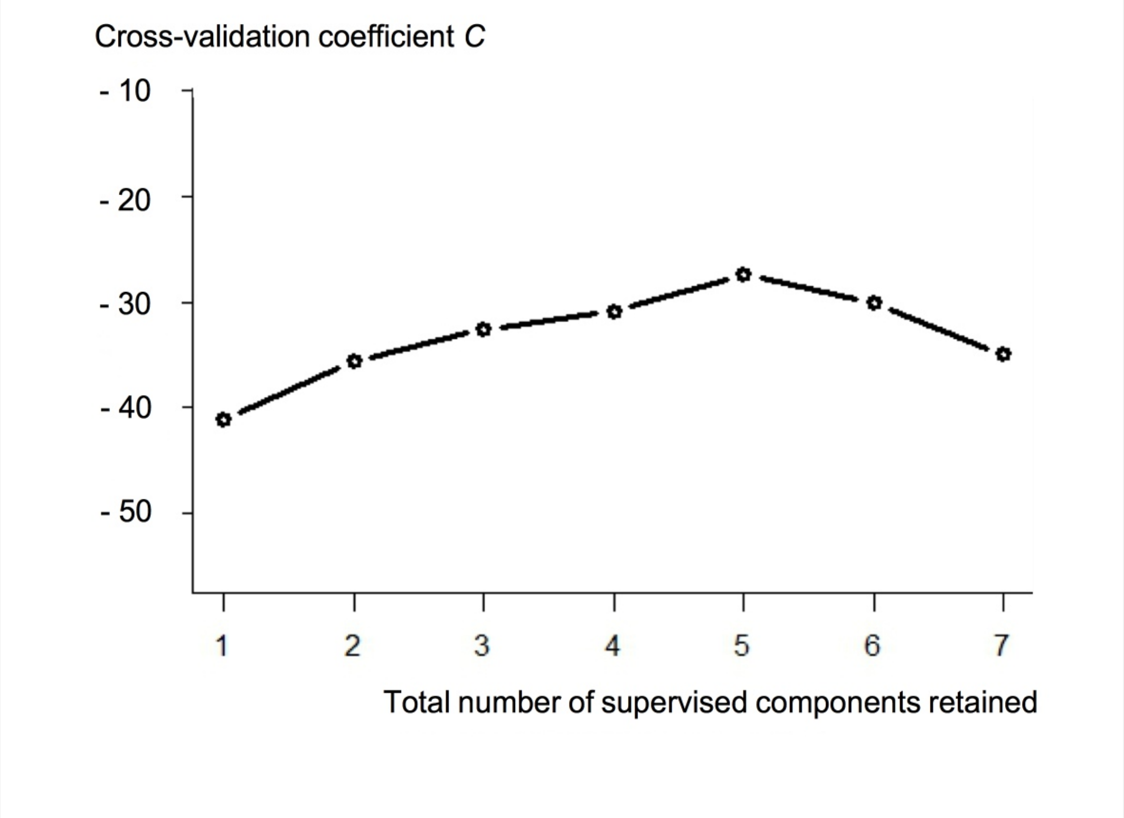


Interpretation is null.

An application to life-history analysis

2. Results

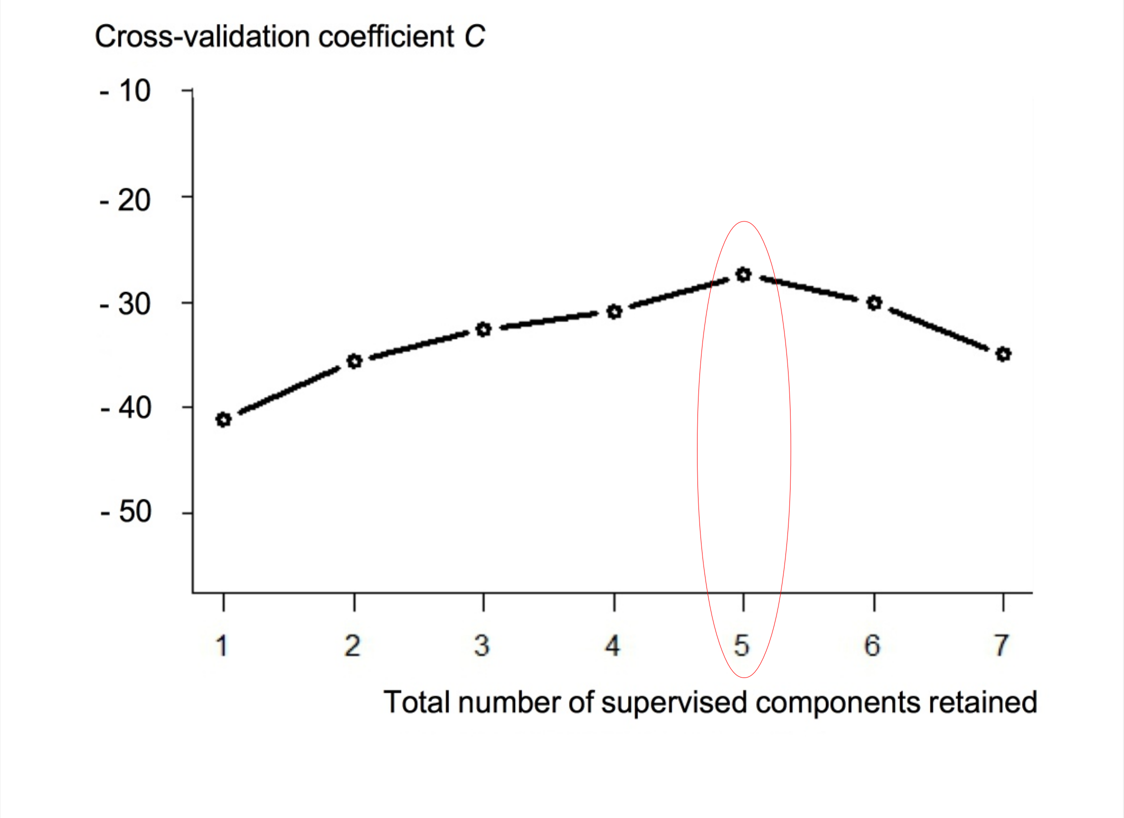
Best values : $s = 0.9$; $l = 8$; $\tau = 1$



An application to life-history analysis

2. Results

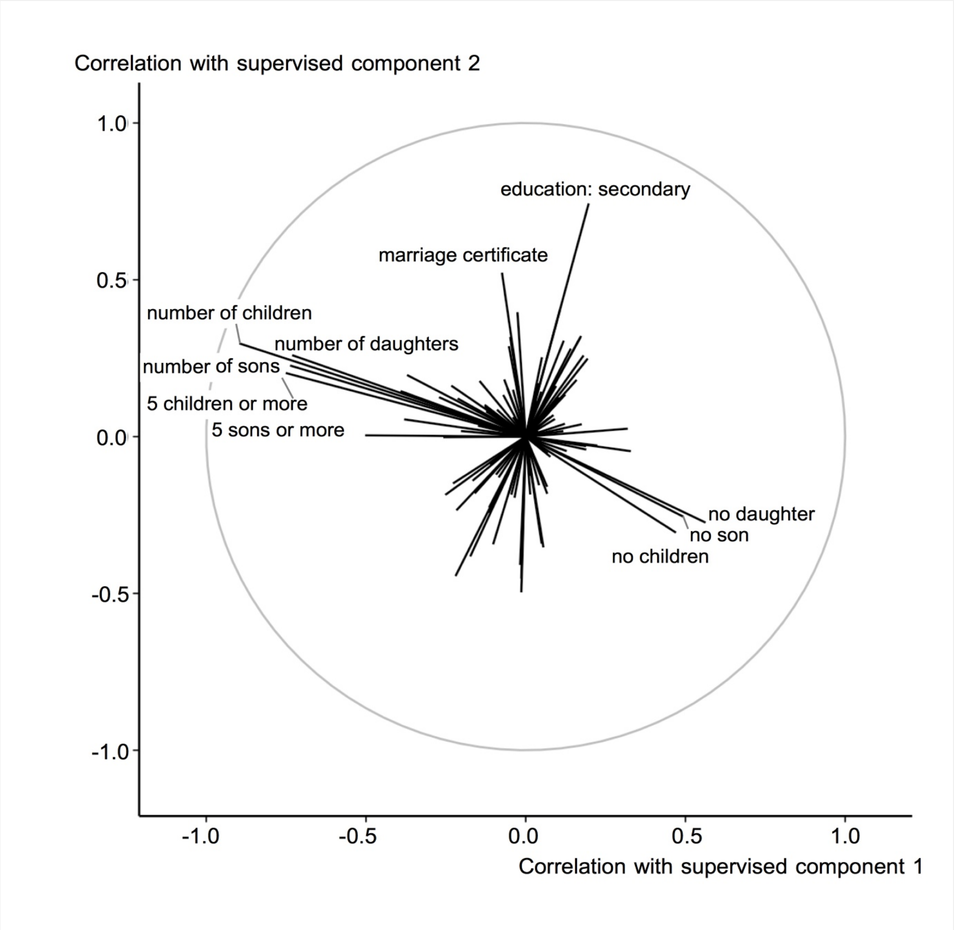
Best values : $s = 0.9$; $l = 8$; $\tau = 1$



An application to life-history analysis

2. Results

Best values : $s = 0.9$; $l = 8$; $\tau = 1$



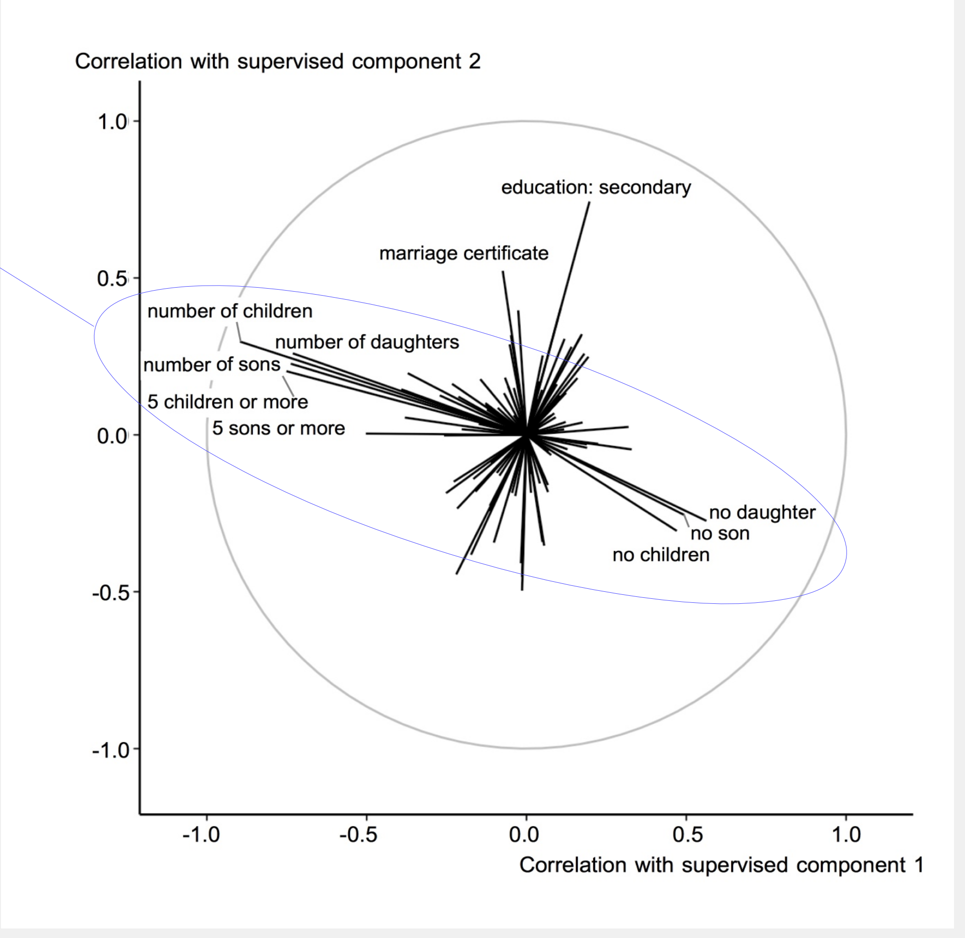
An application to life-history analysis

2. Results

Best values : $s = 0.9$; $l = 8$; $\tau = 1$

Offspring

Offspring size



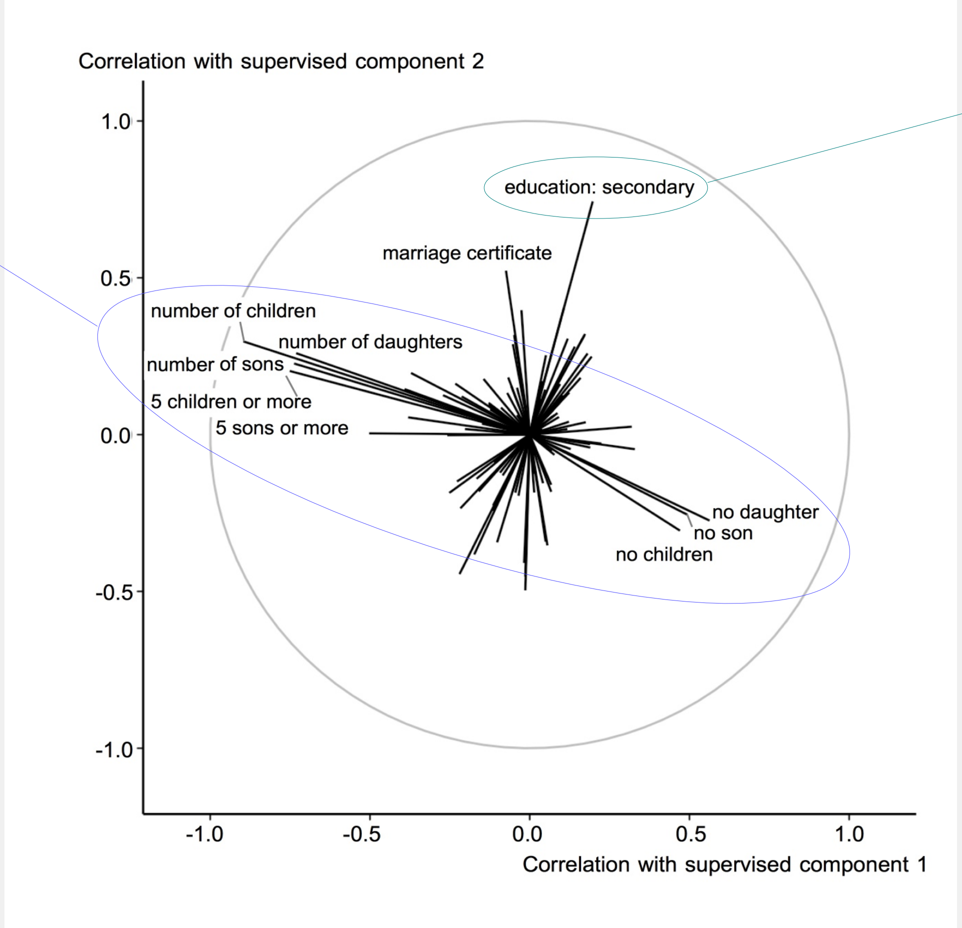
An application to life-history analysis

2. Results

Best values : $s = 0.9$; $l = 8$; $\tau = 1$

Offspring

Offspring size
& high education

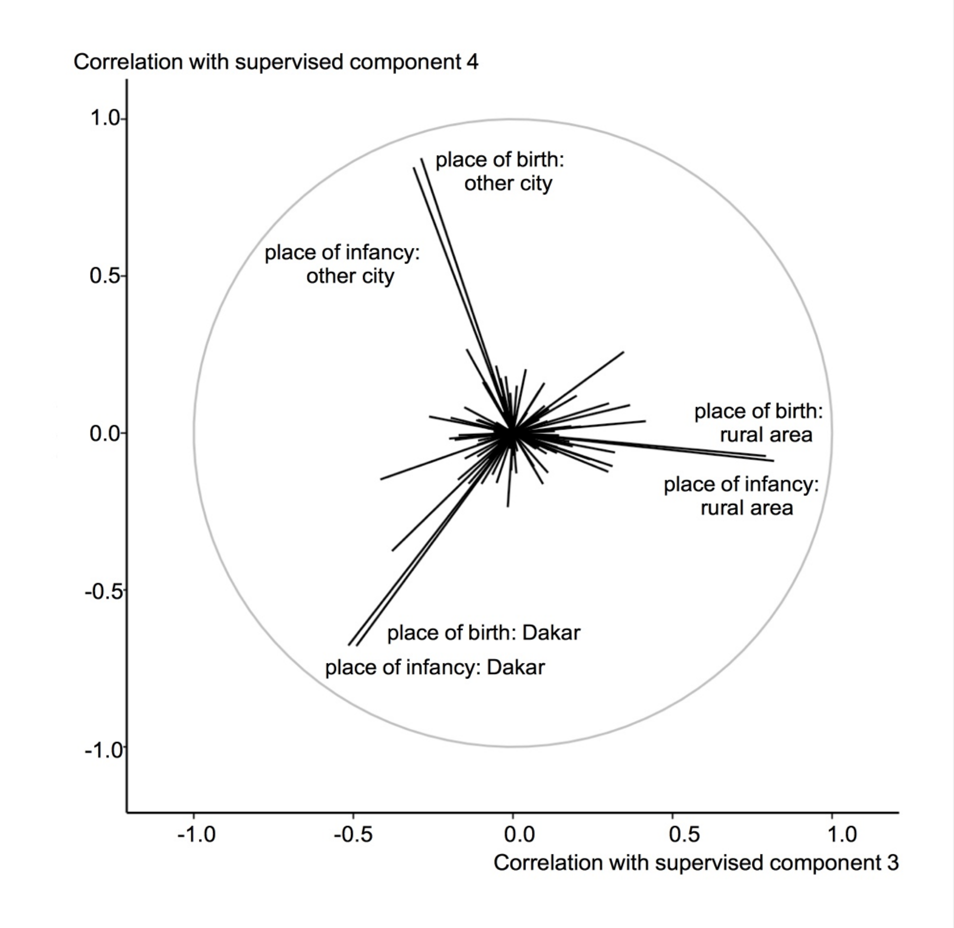


High education

An application to life-history analysis

2. Results

Best values : $s = 0.9$; $l = 8$; $\tau = 1$

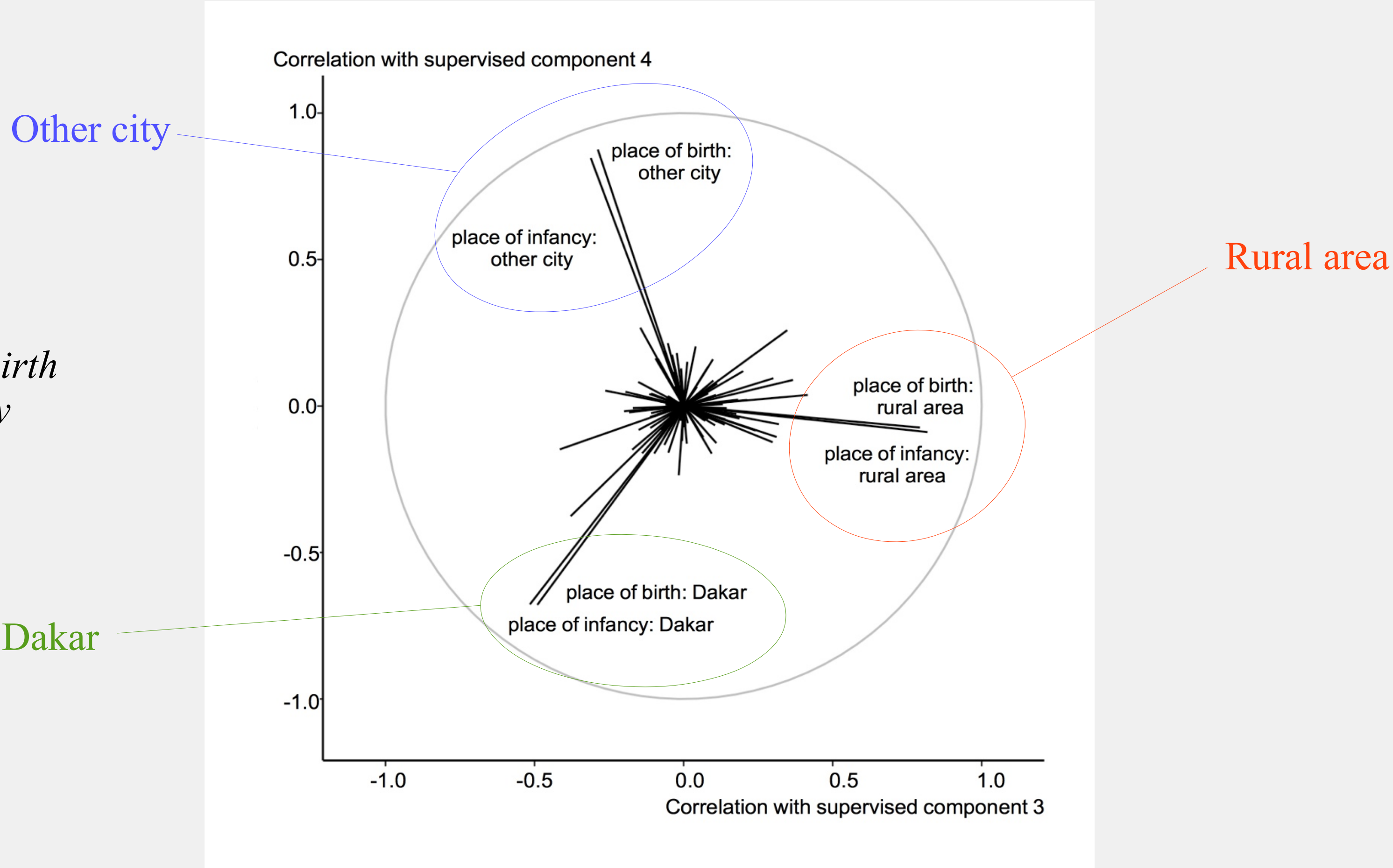


An application to life-history analysis

2. Results

Best values : $s = 0.9$; $l = 8$; $\tau = 1$

Places of birth and infancy



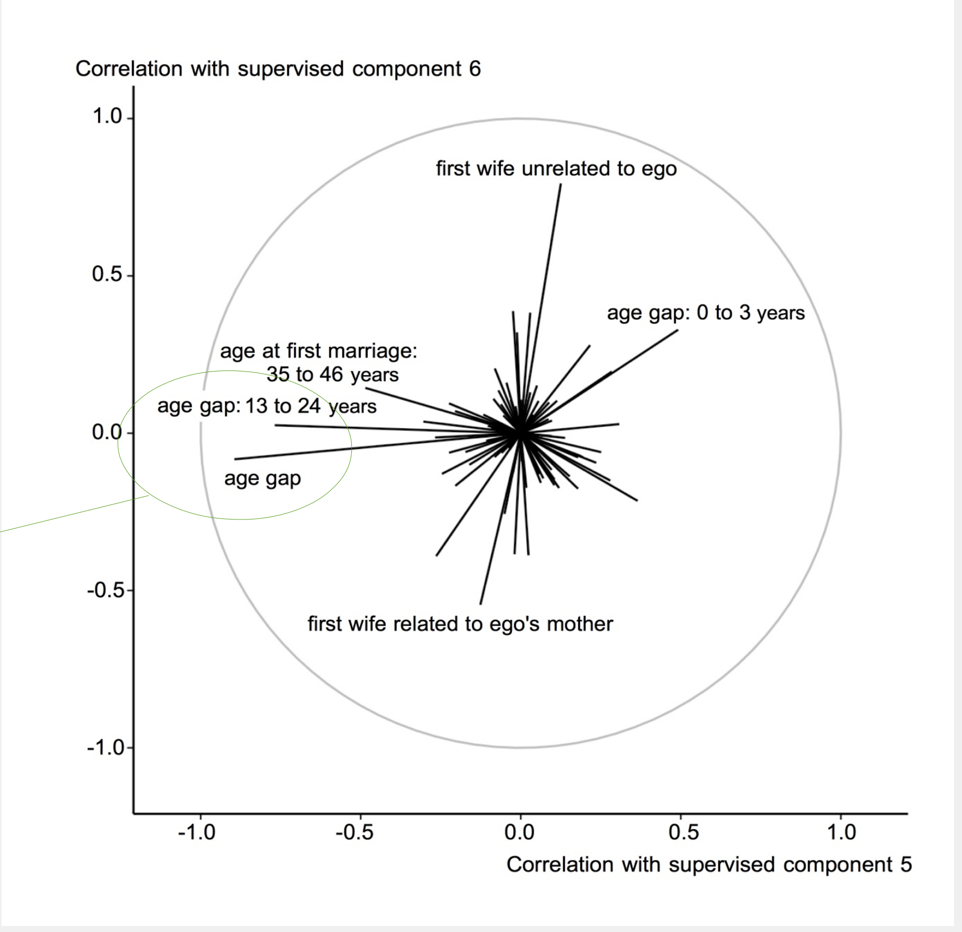
An application to life-history analysis

2. Results

Best values : $s = 0.9$; $l = 8$; $\tau = 1$

Age gap

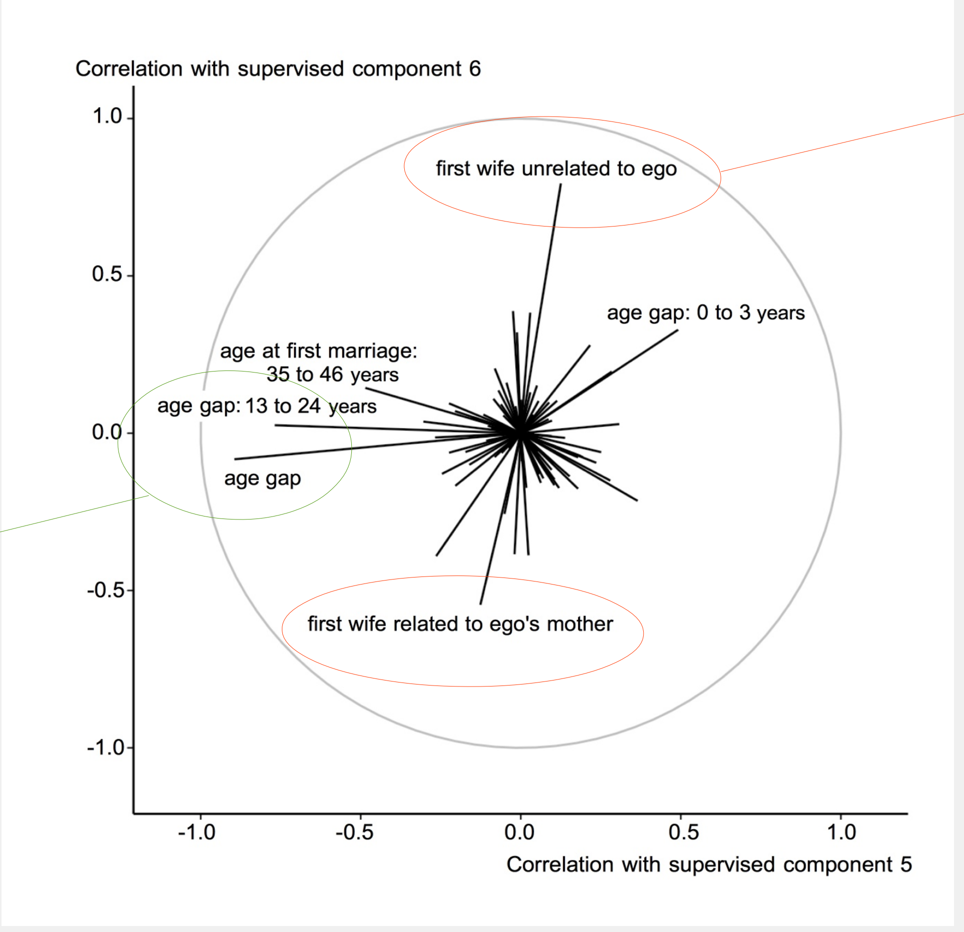
Age gap
(husband-wife 1)



An application to life-history analysis

2. Results

Best values : $s = 0.9$; $l = 8$; $\tau = 1$



Kinship between ego and wife 1

Age gap and spouse-kinship

Age gap (husband-wife 1)

An application to life-history analysis

2. Results

Variable-coefficients
(with 0.95 IC) :

Variable	nation: Senegal	nation: Bissau-Guinea	nation: Guinea	nation: Mali
β	-0.009 [-0.022;0.004]	0.062 [-0.222;0.347]	0.022 [-0.030;0.075]	-0.044 [-0.202;0.113]
$\beta^{(5)}$	0.006 [-0.003;0.016]	0.087 [-0.126;0.300]	-0.014 [-0.035;0.007]	-0.089 [-0.247;0.068]
Variable	nation: Benin	father deceased	mother deceased	parents divorced
β	-0.050 [-0.113;0.013]	-0.020 [-0.352;0.312]	0.128 [-0.388;0.644]	-0.056 [-0.489;0.377]
$\beta^{(5)}$	-0.023 [-0.086;0.040]	-0.033 [-0.647;0.580]	0.150 [-0.490;0.790]	-0.072 [-0.232;0.089]
Variable	marriage-rank	consent	age gap	education: none
β	0.000 [-0.030;0.030]	-0.112 [-1.148;0.923]	-0.208* [-0.237;-0.179]	0.037 [-0.582;0.655]
$\beta^{(5)}$	0.000 [-0.035;0.035]	-0.075 [-1.265;1.116]	-0.414* [-0.450;-0.378]	0.063 [-0.022;0.149]
Variable	education: coranic	education: primary	education: secondary	father education: none
β	0.054 [-0.434;0.542]	0.033 [-0.583;0.649]	-0.099 [-0.342;0.144]	-0.089 [-0.398;0.220]
$\beta^{(5)}$	0.056 [-0.049;0.161]	0.061 [-0.323;0.445]	-0.143* [-0.273;-0.013]	-0.103 [-0.685;0.478]
Variable	father education: coranic	father education: primary	father education: secondary	father education: non-available
β	0.200 [-0.157;0.557]	-0.060 [-0.589;0.468]	-0.047 [-0.635;0.541]	-0.115 [-0.551;0.320]
$\beta^{(5)}$	0.154 [-0.338;0.645]	-0.024 [-0.477;0.429]	-0.025 [-0.125;0.076]	-0.077 [-0.247;0.093]
Variable	mother education: none	mother education: coranic	mother education: primary	mother education: secondary
β	-0.127 [-0.402;0.147]	0.069 [-0.590;0.728]	0.051 [-0.985;1.086]	0.061 [-0.974;1.097]
$\beta^{(5)}$	-0.109 [-0.424;0.205]	-0.014 [-0.574;0.547]	0.101 [-0.343;0.544]	0.094 [-0.349;0.538]
Variable	mother education: non-available	ethnic group: Wolof	ethnic group: Pular	ethnic group: Serer
β	0.065 [-0.710;0.839]	0.078 [-0.303;0.459]	-0.043 [-0.594;0.507]	0.014 [-0.693;0.721]
$\beta^{(5)}$	0.113 [-0.738;0.964]	0.093 [-0.116;0.303]	-0.084 [-0.324;0.156]	0.029 [-0.822;0.880]
Variable	ethnic group: Diola	ethnic group: other	religion: tidjan	religion: murid
β	-0.071 [-0.548;0.406]	-0.017 [-0.368;0.333]	-0.070 [-0.331;0.192]	0.067 [-0.204; 0.339]
$\beta^{(5)}$	-0.053 [-0.545;0.439]	-0.022 [-0.425;0.381]	-0.091 [-0.506;0.324]	0.043 [-0.414;0.500]
Variable	religion: other muslim	religion: christian	age at first marriage: 16 to 24	age at first marriage: 25 to 29
β	0.121 [-0.450;0.693]	-0.133* [-0.205;-0.061]	0.176 [-0.289;0.642]	0.102 [-0.298;0.502]
$\beta^{(5)}$	0.141 [-0.465;0.746]	-0.085* [-0.156;-0.015]	0.221 [-0.156;-0.015]	0.134 [-0.147;0.415]
Variable	age at first marriage: 30 to 34	age at first marriage: 35 to 46	choice of first marriage: ego	choice of first marriage: mutual
β	-0.201* [-0.395;-0.007]	-0.154* [-0.288;-0.021]	-0.020 [-0.371;0.330]	-0.048 [-0.304;0.208]
$\beta^{(5)}$	-0.176 [-0.567;0.215]	-0.300* [-0.563;-0.037]	-0.004 [-0.149;0.142]	-0.037 [-0.274;0.201]
Variable	choice of first marriage: parents	first wife related to ego's father	first wife related to ego's mother	first wife unrelated to ego
β	0.087 [-0.394;0.568]	0.080 [-0.260;0.420]	0.155* [0.071;0.239]	-0.201* [-0.343;-0.058]
$\beta^{(5)}$	0.052 [-0.336;0.439]	0.136 [-0.497;0.769]	0.196 [-0.152;0.543]	-0.283* [-0.312;-0.254]
Variable	age of first wife at marriage: non-available	age of first wife at marriage: 13 to 16	age of first wife at marriage: 17 to 19	age of first wife at marriage: 20 to 24
β	-0.086 [-0.581;0.409]	0.140 [-0.128;0.409]	0.010 [-0.265;0.285]	-0.067 [-0.536;0.402]
$\beta^{(5)}$	-0.107 [-0.226;0.012]	0.089 [-0.377;0.555]	-0.030 [-0.437;0.376]	-0.046 [-0.529;0.436]
Variable	age of first wife at marriage: 25 to 37	place of birth: Dakar	place of birth: rural area	place of birth: other city
β	-0.053 [-0.522;0.415]	-0.087 [-0.263;0.088]	0.139* [0.022;0.256]	-0.053* [-0.103;-0.003]
$\beta^{(5)}$	0.040 [-0.425;0.505]	-0.011 [-0.328;0.307]	0.062* [0.011;0.114]	-0.056 [-0.418;0.306]
Variable	place of infancy: Dakar	place of infancy: rural area	place of infancy: other city	first wife never married
β	-0.160* [-0.292;-0.029]	0.132* [0.009;0.254]	0.043 [-0.284;0.370]	0.027 [-0.410;0.463]
$\beta^{(5)}$	-0.123* [-0.238;-0.008]	0.059* [0.009;0.109]	0.078 [-0.232;0.388]	0.021 [-0.425;0.466]

An application to life-history analysis

2. Results

Variable-coefficients
(with 0.95 IC) :

- The younger ego's wife is relative to him, the lower the risk.

Variable	nation: Senegal	nation: Bissau-Guinea	nation: Guinea	nation: Mali
β	-0.009 [-0.022;0.004]	0.062 [-0.222;0.347]	0.022 [-0.030;0.075]	-0.044 [-0.202;0.113]
$\beta^{(5)}$	0.006 [-0.003;0.016]	0.087 [-0.126;0.300]	-0.014 [-0.035;0.007]	-0.089 [-0.247;0.068]
Variable	nation: Benin	father deceased	mother deceased	parents divorced
β	-0.050 [-0.113;0.013]	-0.020 [-0.352;0.312]	0.128 [-0.388;0.644]	-0.056 [-0.489;0.377]
$\beta^{(5)}$	-0.023 [-0.086;0.040]	-0.033 [-0.647;0.580]	0.150 [-0.490;0.790]	-0.072 [-0.232;0.089]
Variable	marriage-rank	consent	age gap	education: none
β	0.000 [-0.030;0.030]	-0.112 [-1.148;0.923]	-0.208* [-0.237;-0.179]	0.037 [-0.582;0.655]
$\beta^{(5)}$	0.000 [-0.035;0.035]	-0.075 [-1.265;1.116]	-0.414* [-0.450;-0.378]	0.063 [-0.022;0.149]
Variable	education: coranic	education: primary	education: secondary	father education: none
β	0.054 [-0.434;0.542]	0.033 [-0.583;0.649]	-0.099 [-0.342;0.144]	-0.089 [-0.398;0.220]
$\beta^{(5)}$	0.056 [-0.049;0.161]	0.061 [-0.323;0.445]	-0.143* [-0.273;-0.013]	-0.103 [-0.685;0.478]
Variable	father education: coranic	father education: primary	father education: secondary	father education: non-available
β	0.200 [-0.157;0.557]	-0.060 [-0.589;0.468]	-0.047 [-0.635;0.541]	-0.115 [-0.551;0.320]
$\beta^{(5)}$	0.154 [-0.338;0.645]	-0.024 [-0.477;0.429]	-0.025 [-0.125;0.076]	-0.077 [-0.247;0.093]
Variable	mother education: none	mother education: coranic	mother education: primary	mother education: secondary
β	-0.127 [-0.402;0.147]	0.069 [-0.590;0.728]	0.051 [-0.985;1.086]	0.061 [-0.974;1.097]
$\beta^{(5)}$	-0.109 [-0.424;0.205]	-0.014 [-0.574;0.547]	0.101 [-0.343;0.544]	0.094 [-0.349;0.538]
Variable	mother education: non-available	ethnic group: Wolof	ethnic group: Pular	ethnic group: Serer
β	0.065 [-0.710;0.839]	0.078 [-0.303;0.459]	-0.043 [-0.594;0.507]	0.014 [-0.693;0.721]
$\beta^{(5)}$	0.113 [-0.738;0.964]	0.093 [-0.116;0.303]	-0.084 [-0.324;0.156]	0.029 [-0.822;0.880]
Variable	ethnic group: Diola	ethnic group: other	religion: tidjan	religion: murid
β	-0.071 [-0.548;0.406]	-0.017 [-0.368;0.333]	-0.070 [-0.331;0.192]	0.067 [-0.204; 0.339]
$\beta^{(5)}$	-0.053 [-0.545;0.439]	-0.022 [-0.425;0.381]	-0.091 [-0.506;0.324]	0.043 [-0.414;0.500]
Variable	religion: other muslim	religion: christian	age at first marriage: 16 to 24	age at first marriage: 25 to 29
β	0.121 [-0.450;0.693]	-0.133* [-0.205;-0.061]	0.176 [-0.289;0.642]	0.102 [-0.298;0.502]
$\beta^{(5)}$	0.141 [-0.465;0.746]	-0.085* [-0.156;-0.015]	0.221 [-0.156;-0.015]	0.134 [-0.147;0.415]
Variable	age at first marriage: 30 to 34	age at first marriage: 35 to 46	choice of first marriage: ego	choice of first marriage: mutual
β	-0.201* [-0.395;-0.007]	-0.154* [-0.288;-0.021]	-0.020 [-0.371;0.330]	-0.048 [-0.304;0.208]
$\beta^{(5)}$	-0.176 [-0.567;0.215]	-0.300* [-0.563;-0.037]	-0.004 [-0.149;0.142]	-0.037 [-0.274;0.201]
Variable	choice of first marriage: parents	first wife related to ego's father	first wife related to ego's mother	first wife unrelated to ego
β	0.087 [-0.394;0.568]	0.080 [-0.260;0.420]	0.155* [0.071;0.239]	-0.201* [-0.343;-0.058]
$\beta^{(5)}$	0.052 [-0.336;0.439]	0.136 [-0.497;0.769]	0.196 [-0.152;0.543]	-0.283* [-0.312;-0.254]
Variable	age of first wife at marriage: non-available	age of first wife at marriage: 13 to 16	age of first wife at marriage: 17 to 19	age of first wife at marriage: 20 to 24
β	-0.086 [-0.581;0.409]	0.140 [-0.128;0.409]	0.010 [-0.265;0.285]	-0.067 [-0.536;0.402]
$\beta^{(5)}$	-0.107 [-0.226;0.012]	0.089 [-0.377;0.555]	-0.030 [-0.437;0.376]	-0.046 [-0.529;0.436]
Variable	age of first wife at marriage: 25 to 37	place of birth: Dakar	place of birth: rural area	place of birth: other city
β	-0.053 [-0.522;0.415]	-0.087 [-0.263;0.088]	0.139* [0.022;0.256]	-0.053* [-0.103;-0.003]
$\beta^{(5)}$	0.040 [-0.425;0.505]	-0.011 [-0.328;0.307]	0.062* [0.011;0.114]	-0.056 [-0.418;0.306]
Variable	place of infancy: Dakar	place of infancy: rural area	place of infancy: other city	first wife never married
β	-0.160* [-0.292;-0.029]	0.132* [0.009;0.254]	0.043 [-0.284;0.370]	0.027 [-0.410;0.463]
$\beta^{(5)}$	-0.123* [-0.238;-0.008]	0.059* [0.009;0.109]	0.078 [-0.232;0.388]	0.021 [-0.425;0.466]

An application to life-history analysis

2. Results

Variable-coefficients (with 0.95 IC) :

- The younger ego's wife is relative to him, the lower the risk.
- The older ego is at first marriage, the lower the risk.

Variable	nation: Senegal	nation: Bissau-Guinea	nation: Guinea	nation: Mali
β	-0.009 [-0.022;0.004]	0.062 [-0.222;0.347]	0.022 [-0.030;0.075]	-0.044 [-0.202;0.113]
$\beta^{(5)}$	0.006 [-0.003;0.016]	0.087 [-0.126;0.300]	-0.014 [-0.035;0.007]	-0.089 [-0.247;0.068]
Variable	nation: Benin	father deceased	mother deceased	parents divorced
β	-0.050 [-0.113;0.013]	-0.020 [-0.352;0.312]	0.128 [-0.388;0.644]	-0.056 [-0.489;0.377]
$\beta^{(5)}$	-0.023 [-0.086;0.040]	-0.033 [-0.647;0.580]	0.150 [-0.490;0.790]	-0.072 [-0.232;0.089]
Variable	marriage-rank	consent	age gap	education: none
β	0.000 [-0.030;0.030]	-0.112 [-1.148;0.923]	-0.208* [-0.237;-0.179]	0.037 [-0.582;0.655]
$\beta^{(5)}$	0.000 [-0.035;0.035]	-0.075 [-1.265;1.116]	-0.414* [-0.450;-0.378]	0.063 [-0.022;0.149]
Variable	education: coranic	education: primary	education: secondary	father education: none
β	0.054 [-0.434;0.542]	0.033 [-0.583;0.649]	-0.099 [-0.342;0.144]	-0.089 [-0.398;0.220]
$\beta^{(5)}$	0.056 [-0.049;0.161]	0.061 [-0.323;0.445]	-0.143* [-0.273;-0.013]	-0.103 [-0.685;0.478]
Variable	father education: coranic	father education: primary	father education: secondary	father education: non-available
β	0.200 [-0.157;0.557]	-0.060 [-0.589;0.468]	-0.047 [-0.635;0.541]	-0.115 [-0.551;0.320]
$\beta^{(5)}$	0.154 [-0.338;0.645]	-0.024 [-0.477;0.429]	-0.025 [-0.125;0.076]	-0.077 [-0.247;0.093]
Variable	mother education: none	mother education: coranic	mother education: primary	mother education: secondary
β	-0.127 [-0.402;0.147]	0.069 [-0.590;0.728]	0.051 [-0.985;1.086]	0.061 [-0.974;1.097]
$\beta^{(5)}$	-0.109 [-0.424;0.205]	-0.014 [-0.574;0.547]	0.101 [-0.343;0.544]	0.094 [-0.349;0.538]
Variable	mother education: non-available	ethnic group: Wolof	ethnic group: Pular	ethnic group: Serer
β	0.065 [-0.710;0.839]	0.078 [-0.303;0.459]	-0.043 [-0.594;0.507]	0.014 [-0.693;0.721]
$\beta^{(5)}$	0.113 [-0.738;0.964]	0.093 [-0.116;0.303]	-0.084 [-0.324;0.156]	0.029 [-0.822;0.880]
Variable	ethnic group: Diola	ethnic group: other	religion: tidjan	religion: murid
β	-0.071 [-0.548;0.406]	-0.017 [-0.368;0.333]	-0.070 [-0.331;0.192]	0.067 [-0.204; 0.339]
$\beta^{(5)}$	-0.053 [-0.545;0.439]	-0.022 [-0.425;0.381]	-0.091 [-0.506;0.324]	0.043 [-0.414;0.500]
Variable	religion: other muslim	religion: christian	age at first marriage: 16 to 24	age at first marriage: 25 to 29
β	0.121 [-0.450;0.693]	-0.133* [-0.205;-0.061]	0.176 [-0.289;0.642]	0.102 [-0.298;0.502]
$\beta^{(5)}$	0.141 [-0.465;0.746]	-0.085* [-0.156;-0.015]	0.221 [-0.156;-0.015]	0.134 [-0.147;0.415]
Variable	age at first marriage: 30 to 34	age at first marriage: 35 to 46	choice of first marriage: ego	choice of first marriage: mutual
β	-0.201* [-0.395;-0.007]	-0.154* [-0.288;-0.021]	-0.020 [-0.371;0.330]	-0.048 [-0.304;0.208]
$\beta^{(5)}$	-0.176 [-0.567;0.215]	-0.300* [-0.563;-0.037]	-0.004 [-0.149;0.142]	-0.037 [-0.274;0.201]
Variable	choice of first marriage: parents	first wife related to ego's father	first wife related to ego's mother	first wife unrelated to ego
β	0.087 [-0.394;0.568]	0.080 [-0.260;0.420]	0.155* [0.071;0.239]	-0.201* [-0.343;-0.058]
$\beta^{(5)}$	0.052 [-0.336;0.439]	0.136 [-0.497;0.769]	0.196 [-0.152;0.543]	-0.283* [-0.312;-0.254]
Variable	age of first wife at marriage: non-available	age of first wife at marriage: 13 to 16	age of first wife at marriage: 17 to 19	age of first wife at marriage: 20 to 24
β	-0.086 [-0.581;0.409]	0.140 [-0.128;0.409]	0.010 [-0.265;0.285]	-0.067 [-0.536;0.402]
$\beta^{(5)}$	-0.107 [-0.226;0.012]	0.089 [-0.377;0.555]	-0.030 [-0.437;0.376]	-0.046 [-0.529;0.436]
Variable	age of first wife at marriage: 25 to 37	place of birth: Dakar	place of birth: rural area	place of birth: other city
β	-0.053 [-0.522;0.415]	-0.087 [-0.263;0.088]	0.139* [0.022;0.256]	-0.053* [-0.103;-0.003]
$\beta^{(5)}$	0.040 [-0.425;0.505]	-0.011 [-0.328;0.307]	0.062* [0.011;0.114]	-0.056 [-0.418;0.306]
Variable	place of infancy: Dakar	place of infancy: rural area	place of infancy: other city	first wife never married
β	-0.160* [-0.292;-0.029]	0.132* [0.009;0.254]	0.043 [-0.284;0.370]	0.027 [-0.410;0.463]
$\beta^{(5)}$	-0.123* [-0.238;-0.008]	0.059* [0.009;0.109]	0.078 [-0.232;0.388]	0.021 [-0.425;0.466]

An application to life-history analysis

2. Results

Variable-coefficients
(with 0.95 IC) :

- The younger ego's wife is relative to him, the lower the risk.
- The older ego is at first marriage, the lower the risk.

- A wife unrelated to ego lowers the risk.
- A wife related to ego's mother increases the risk.

Variable	nation: Senegal	nation: Bissau-Guinea	nation: Guinea	nation: Mali
β	-0.009 [-0.022;0.004]	0.062 [-0.222;0.347]	0.022 [-0.030;0.075]	-0.044 [-0.202;0.113]
$\beta^{(5)}$	0.006 [-0.003;0.016]	0.087 [-0.126;0.300]	-0.014 [-0.035;0.007]	-0.089 [-0.247;0.068]
Variable	nation: Benin	father deceased	mother deceased	parents divorced
β	-0.050 [-0.113;0.013]	-0.020 [-0.352;0.312]	0.128 [-0.388;0.644]	-0.056 [-0.489;0.377]
$\beta^{(5)}$	-0.023 [-0.086;0.040]	-0.033 [-0.647;0.580]	0.150 [-0.490;0.790]	-0.072 [-0.232;0.089]
Variable	marriage-rank	consent	age gap	education: none
β	0.000 [-0.030;0.030]	-0.112 [-1.148;0.923]	-0.208* [-0.237;-0.179]	0.037 [-0.582;0.655]
$\beta^{(5)}$	0.000 [-0.035;0.035]	-0.075 [-1.265;1.116]	-0.414* [-0.450;-0.378]	0.063 [-0.022;0.149]
Variable	education: coranic	education: primary	education: secondary	father education: none
β	0.054 [-0.434;0.542]	0.033 [-0.583;0.649]	-0.099 [-0.342;0.144]	-0.089 [-0.398;0.220]
$\beta^{(5)}$	0.056 [-0.049;0.161]	0.061 [-0.323;0.445]	-0.143* [-0.273;-0.013]	-0.103 [-0.685;0.478]
Variable	father education: coranic	father education: primary	father education: secondary	father education: non-available
β	0.200 [-0.157;0.557]	-0.060 [-0.589;0.468]	-0.047 [-0.635;0.541]	-0.115 [-0.551;0.320]
$\beta^{(5)}$	0.154 [-0.338;0.645]	-0.024 [-0.477;0.429]	-0.025 [-0.125;0.076]	-0.077 [-0.247;0.093]
Variable	mother education: none	mother education: coranic	mother education: primary	mother education: secondary
β	-0.127 [-0.402;0.147]	0.069 [-0.590;0.728]	0.051 [-0.985;1.086]	0.061 [-0.974;1.097]
$\beta^{(5)}$	-0.109 [-0.424;0.205]	-0.014 [-0.574;0.547]	0.101 [-0.343;0.544]	0.094 [-0.349;0.538]
Variable	mother education: non-available	ethnic group: Wolof	ethnic group: Pular	ethnic group: Serer
β	0.065 [-0.710;0.839]	0.078 [-0.303;0.459]	-0.043 [-0.594;0.507]	0.014 [-0.693;0.721]
$\beta^{(5)}$	0.113 [-0.738;0.964]	0.093 [-0.116;0.303]	-0.084 [-0.324;0.156]	0.029 [-0.822;0.880]
Variable	ethnic group: Diola	ethnic group: other	religion: tidjan	religion: murid
β	-0.071 [-0.548;0.406]	-0.017 [-0.368;0.333]	-0.070 [-0.331;0.192]	0.067 [-0.204; 0.339]
$\beta^{(5)}$	-0.053 [-0.545;0.439]	-0.022 [-0.425;0.381]	-0.091 [-0.506;0.324]	0.043 [-0.414;0.500]
Variable	religion: other muslim	religion: christian	age at first marriage: 16 to 24	age at first marriage: 25 to 29
β	0.121 [-0.450;0.693]	-0.133* [-0.205;-0.061]	0.176 [-0.289;0.642]	0.102 [-0.298;0.502]
$\beta^{(5)}$	0.141 [-0.465;0.746]	-0.085* [-0.156;-0.015]	0.221 [-0.156;-0.015]	0.134 [-0.147;0.415]
Variable	age at first marriage: 30 to 34	age at first marriage: 35 to 46	choice of first marriage: ego	choice of first marriage: mutual
β	-0.201* [-0.395;-0.007]	-0.154* [-0.288;-0.021]	-0.020 [-0.371;0.330]	-0.048 [-0.304;0.208]
$\beta^{(5)}$	-0.176 [-0.567;0.215]	-0.300* [-0.563;-0.037]	-0.004 [-0.149;0.142]	-0.037 [-0.274;0.201]
Variable	choice of first marriage: parents	first wife related to ego's father	first wife related to ego's mother	first wife unrelated to ego
β	0.087 [-0.394;0.568]	0.080 [-0.260;0.420]	0.155* [0.071;0.239]	-0.201* [-0.343;-0.058]
$\beta^{(5)}$	0.052 [-0.336;0.439]	0.136 [-0.497;0.769]	0.196 [-0.152;0.543]	-0.283* [-0.312;-0.254]
Variable	age of first wife at marriage: non-available	age of first wife at marriage: 13 to 16	age of first wife at marriage: 17 to 19	age of first wife at marriage: 20 to 24
β	-0.086 [-0.581;0.409]	0.140 [-0.128;0.409]	0.010 [-0.265;0.285]	-0.067 [-0.536;0.402]
$\beta^{(5)}$	-0.107 [-0.226;0.012]	0.089 [-0.377;0.555]	-0.030 [-0.437;0.376]	-0.046 [-0.529;0.436]
Variable	age of first wife at marriage: 25 to 37	place of birth: Dakar	place of birth: rural area	place of birth: other city
β	-0.053 [-0.522;0.415]	-0.087 [-0.263;0.088]	0.139* [0.022;0.256]	-0.053* [-0.103;-0.003]
$\beta^{(5)}$	0.040 [-0.425;0.505]	-0.011 [-0.328;0.307]	0.062* [0.011;0.114]	-0.056 [-0.418;0.306]
Variable	place of infancy: Dakar	place of infancy: rural area	place of infancy: other city	first wife never married
β	-0.160* [-0.292;-0.029]	0.132* [0.009;0.254]	0.043 [-0.284;0.370]	0.027 [-0.410;0.463]
$\beta^{(5)}$	-0.123* [-0.238;-0.008]	0.059* [0.009;0.109]	0.078 [-0.232;0.388]	0.021 [-0.425;0.466]

An application to life-history analysis

2. Results

Variable-coefficients
(with 0.95 IC) :

- The younger ego's wife is relative to him, the lower the risk.
- The older ego is at first marriage, the lower the risk.

- A wife unrelated to ego lowers the risk.
- A wife related to ego's mother increases the risk.

- Infancy in Dakar lowers the risk.
- Birth and infancy in a rural area increases the risk.

Variable	nation: Senegal	nation: Bissau-Guinea	nation: Guinea	nation: Mali
β	-0.009 [-0.022;0.004]	0.062 [-0.222;0.347]	0.022 [-0.030;0.075]	-0.044 [-0.202;0.113]
$\beta^{(5)}$	0.006 [-0.003;0.016]	0.087 [-0.126;0.300]	-0.014 [-0.035;0.007]	-0.089 [-0.247;0.068]
Variable	nation: Benin	father deceased	mother deceased	parents divorced
β	-0.050 [-0.113;0.013]	-0.020 [-0.352;0.312]	0.128 [-0.388;0.644]	-0.056 [-0.489;0.377]
$\beta^{(5)}$	-0.023 [-0.086;0.040]	-0.033 [-0.647;0.580]	0.150 [-0.490;0.790]	-0.072 [-0.232;0.089]
Variable	marriage-rank	consent	age gap	education: none
β	0.000 [-0.030;0.030]	-0.112 [-1.148;0.923]	-0.208* [-0.237;-0.179]	0.037 [-0.582;0.655]
$\beta^{(5)}$	0.000 [-0.035;0.035]	-0.075 [-1.265;1.116]	-0.414* [-0.450;-0.378]	0.063 [-0.022;0.149]
Variable	education: coranic	education: primary	education: secondary	father education: none
β	0.054 [-0.434;0.542]	0.033 [-0.583;0.649]	-0.099 [-0.342;0.144]	-0.089 [-0.398;0.220]
$\beta^{(5)}$	0.056 [-0.049;0.161]	0.061 [-0.323;0.445]	-0.143* [-0.273;-0.013]	-0.103 [-0.685;0.478]
Variable	father education: coranic	father education: primary	father education: secondary	father education: non-available
β	0.200 [-0.157;0.557]	-0.060 [-0.589;0.468]	-0.047 [-0.635;0.541]	-0.115 [-0.551;0.320]
$\beta^{(5)}$	0.154 [-0.338;0.645]	-0.024 [-0.477;0.429]	-0.025 [-0.125;0.076]	-0.077 [-0.247;0.093]
Variable	mother education: none	mother education: coranic	mother education: primary	mother education: secondary
β	-0.127 [-0.402;0.147]	0.069 [-0.590;0.728]	0.051 [-0.985;1.086]	0.061 [-0.974;1.097]
$\beta^{(5)}$	-0.109 [-0.424;0.205]	-0.014 [-0.574;0.547]	0.101 [-0.343;0.544]	0.094 [-0.349;0.538]
Variable	mother education: non-available	ethnic group: Wolof	ethnic group: Pular	ethnic group: Serer
β	0.065 [-0.710;0.839]	0.078 [-0.303;0.459]	-0.043 [-0.594;0.507]	0.014 [-0.693;0.721]
$\beta^{(5)}$	0.113 [-0.738;0.964]	0.093 [-0.116;0.303]	-0.084 [-0.324;0.156]	0.029 [-0.822;0.880]
Variable	ethnic group: Diola	ethnic group: other	religion: tidjan	religion: murid
β	-0.071 [-0.548;0.406]	-0.017 [-0.368;0.333]	-0.070 [-0.331;0.192]	0.067 [-0.204; 0.339]
$\beta^{(5)}$	-0.053 [-0.545;0.439]	-0.022 [-0.425;0.381]	-0.091 [-0.506;0.324]	0.043 [-0.414;0.500]
Variable	religion: other muslim	religion: christian	age at first marriage: 16 to 24	age at first marriage: 25 to 29
β	0.121 [-0.450;0.693]	-0.133* [-0.205;-0.061]	0.176 [-0.289;0.642]	0.102 [-0.298;0.502]
$\beta^{(5)}$	0.141 [-0.465;0.746]	-0.085* [-0.156;-0.015]	0.221 [-0.156;-0.015]	0.134 [-0.147;0.415]
Variable	age at first marriage: 30 to 34	age at first marriage: 35 to 46	choice of first marriage: ego	choice of first marriage: mutual
β	-0.201* [-0.395;-0.007]	-0.154* [-0.288;-0.021]	-0.020 [-0.371;0.330]	-0.048 [-0.304;0.208]
$\beta^{(5)}$	-0.176 [-0.567;0.215]	-0.300* [-0.563;-0.037]	-0.004 [-0.149;0.142]	-0.037 [-0.274;0.201]
Variable	choice of first marriage: parents	first wife related to ego's father	first wife related to ego's mother	first wife unrelated to ego
β	0.087 [-0.394;0.568]	0.080 [-0.260;0.420]	0.155* [0.071;0.239]	-0.201* [-0.343;-0.058]
$\beta^{(5)}$	0.052 [-0.336;0.439]	0.136 [-0.497;0.769]	0.196 [-0.152;0.543]	-0.283* [-0.312;-0.254]
Variable	age of first wife at marriage: non-available	age of first wife at marriage: 13 to 16	age of first wife at marriage: 17 to 19	age of first wife at marriage: 20 to 24
β	-0.086 [-0.581;0.409]	0.140 [-0.128;0.409]	0.010 [-0.265;0.285]	-0.067 [-0.536;0.402]
$\beta^{(5)}$	-0.107 [-0.226;0.012]	0.089 [-0.377;0.555]	-0.030 [-0.437;0.376]	-0.046 [-0.529;0.436]
Variable	age of first wife at marriage: 25 to 37	place of birth: Dakar	place of birth: rural area	place of birth: other city
β	-0.053 [-0.522;0.415]	-0.087 [-0.263;0.088]	0.139* [0.022;0.256]	-0.053* [-0.103;-0.003]
$\beta^{(5)}$	0.040 [-0.425;0.505]	-0.011 [-0.328;0.307]	0.062* [0.011;0.114]	-0.056 [-0.418;0.306]
Variable	place of infancy: Dakar	place of infancy: rural area	place of infancy: other city	first wife never married
β	-0.160* [-0.292;-0.029]	0.132* [0.009;0.254]	0.043 [-0.284;0.370]	0.027 [-0.410;0.463]
$\beta^{(5)}$	-0.123* [-0.238;-0.008]	0.059* [0.009;0.109]	0.078 [-0.232;0.388]	0.021 [-0.425;0.466]

An application to life-history analysis

2. Results

Variable-coefficients
(with 0.95 IC) :

Variable	first wife once married	occupation of first wife: house-wife	occupation of first wife: student	occupation of first wife: employee
β	-0.027 [-0.463;0.410]	0.024 [-0.283;0.332]	-0.092 [-0.385;0.202]	-0.065 [-0.441;0.311]
$\beta^{(5)}$	-0.021 [-0.466;0.425]	0.012 [-0.283;0.308]	-0.093 [-0.803;0.617]	-0.050 [-0.487;0.387]
Variable	occupation of first wife: artisan	occupation of first wife: trade	occupation of first wife: agriculture	occupation of first wife: non-available
β	0.071 [-0.985;1.128]	0.058 [-0.862;0.978]	0.250 [-0.807;1.306]	-0.063 [-0.983;0.858]
$\beta^{(5)}$	0.066 [-0.709;0.841]	0.081 [-0.435;0.598]	0.188 [-0.457;0.834]	-0.053 [-0.623;0.517]
Variable	occupation: informal	occupation: employee	occupation: apprentice	occupation: independent
β	-0.004 [-0.309;0.300]	0.133 [-0.142;0.408]	-0.088 * [-0.162;-0.015]	-0.051 [-0.527;0.424]
$\beta^{(5)}$	-0.010 [-0.295;0.275]	0.159 [-0.123;0.440]	-0.071 [-0.542;0.400]	-0.105 [-0.357;0.148]
Variable	occupation: student	occupation: retired	occupation: unemployed	occupation: other inactive
β	-0.039 [-0.371;0.293]	-0.091 [-0.583;0.400]	0.003 [-0.594;0.600]	-0.071 [-1.004;0.863]
$\beta^{(5)}$	-0.062 [-0.284;0.159]	-0.046 [-0.248;0.156]	0.022 [-0.163;0.207]	-0.042 [-0.264;0.180]
Variable	occupation: other with no income	residence: owner	residence: lodger	residence: family
β	-0.097 [-0.818;0.625]	0.021 [-0.333;0.376]	-0.0862 [-0.389;0.216]	0.014 [-0.390;0.418]
$\beta^{(5)}$	-0.078 [-0.325;0.169]	0.028 [-0.095;0.151]	-0.076 [-0.340;0.188]	0.060 [-0.207;0.327]
Variable	residence: husband's parents	residence: other parents	residence: other	number of sons
β	0.040 [-0.363;0.444]	0.114 [-0.290;0.517]	-0.089 [-0.493;0.315]	-0.055 [-0.170;0.060]
$\beta^{(5)}$	0.062 [-0.160;0.284]	0.076 [-0.361;0.513]	-0.133 [-0.400;0.134]	-0.040 [-0.095;0.014]
Variable	number of daughters	no son	1 son	2 sons
β	-0.040 [-0.114;0.034]	0.010 [-0.212;0.419]	-0.054 [-0.352;0.244]	-0.059 [-0.582;0.465]
$\beta^{(5)}$	-0.039 [-0.127;0.050]	0.060 [-0.185;0.306]	-0.062 [-0.258;0.134]	-0.025 [-0.470;0.419]
Variable	3 sons	4 sons	5 sons or more	no daughter
β	-0.023 [-0.850;0.805]	-0.039 [-0.490;0.411]	0.051 [-0.399;0.501]	0.015 [-0.267;0.297]
$\beta^{(5)}$	0.031 [-0.393;0.454]	-0.022 [-0.144;0.101]	0.014 [-0.109;0.137]	-0.003 [-0.130;0.124]
Variable	1 daughter	2 daughters	3 daughters	4 daughters
β	-0.121 [-0.493;0.252]	0.164 [-0.228;0.557]	0.051 [-0.690;0.793]	-0.084 [-0.806;0.638]
$\beta^{(5)}$	-0.076 [-0.245;0.092]	0.141 [-0.003;0.285]	0.037 [-0.603;0.676]	-0.084 [-0.458;0.289]
Variable	5 daughters or more	number of children	no child	1 child
β	-0.085 [-0.807;0.637]	-0.058* [-0.110;-0.007]	0.049 [-0.112;0.210]	0.012 [-0.388;0.411]
$\beta^{(5)}$	-0.072 [-0.569;0.426]	-0.048* [-0.090;-0.006]	-0.009 [-0.279;0.262]	0.014 [-0.491;0.520]
Variable	2 children	3 children	4 children	5 children or more
β	-0.044 [-0.599;0.512]	0.098 [-0.524;0.720]	-0.144 [-1.049;0.761]	0.003 [-0.423;0.430]
$\beta^{(5)}$	-0.023 [-0.501;0.455]	0.129 [-0.286;0.544]	-0.135 [-0.799;0.529]	0.007 [-0.427;0.441]
Variable	no child out of marriage	child out of marriage	age gap: 0 to 3	age gap: 4 to 7
β	-0.017 [-0.692;0.657]	0.017 [-0.657;0.692]	0.121* [0.015;0.227]	-0.053 [-0.363;0.257]
$\beta^{(5)}$	-0.035 [-0.644;0.575]	0.035 [-0.575;0.644]	0.196* [0.018;0.374]	0.025 [-0.354;0.404]
Variable	age gap: 8 to 12	age gap: 13 to 24	marriage certificate	
β	0.147 [-0.359;0.654]	-0.221* [-0.410;-0.032]	-0.138 [-0.571;0.294]	
$\beta^{(5)}$	0.137 [-0.367;0.642]	-0.381* [-0.739;-0.023]	-0.155 [-0.769;0.458]	

An application to life-history analysis

2. Results

Variable-coefficients
(with 0.95 IC) :

• A high number of children lowers the risk.

Variable	first wife once married	occupation of first wife: house-wife	occupation of first wife: student	occupation of first wife: employee
β	-0.027 [-0.463;0.410]	0.024 [-0.283;0.332]	-0.092 [-0.385;0.202]	-0.065 [-0.441;0.311]
$\beta^{(5)}$	-0.021 [-0.466;0.425]	0.012 [-0.283;0.308]	-0.093 [-0.803;0.617]	-0.050 [-0.487;0.387]
Variable	occupation of first wife: artisan	occupation of first wife: trade	occupation of first wife: agriculture	occupation of first wife: non-available
β	0.071 [-0.985;1.128]	0.058 [-0.862;0.978]	0.250 [-0.807;1.306]	-0.063 [-0.983;0.858]
$\beta^{(5)}$	0.066 [-0.709;0.841]	0.081 [-0.435;0.598]	0.188 [-0.457;0.834]	-0.053 [-0.623;0.517]
Variable	occupation: informal	occupation: employee	occupation: apprentice	occupation: independent
β	-0.004 [-0.309;0.300]	0.133 [-0.142;0.408]	-0.088 * [-0.162;-0.015]	-0.051 [-0.527;0.424]
$\beta^{(5)}$	-0.010 [-0.295;0.275]	0.159 [-0.123;0.440]	-0.071 [-0.542;0.400]	-0.105 [-0.357;0.148]
Variable	occupation: student	occupation: retired	occupation: unemployed	occupation: other inactive
β	-0.039 [-0.371;0.293]	-0.091 [-0.583;0.400]	0.003 [-0.594;0.600]	-0.071 [-1.004;0.863]
$\beta^{(5)}$	-0.062 [-0.284;0.159]	-0.046 [-0.248;0.156]	0.022 [-0.163;0.207]	-0.042 [-0.264;0.180]
Variable	occupation: other with no income	residence: owner	residence: lodger	residence: family
β	-0.097 [-0.818;0.625]	0.021 [-0.333;0.376]	-0.0862 [-0.389;0.216]	0.014 [-0.390;0.418]
$\beta^{(5)}$	-0.078 [-0.325;0.169]	0.028 [-0.095;0.151]	-0.076 [-0.340;0.188]	0.060 [-0.207;0.327]
Variable	residence: husband's parents	residence: other parents	residence: other	number of sons
β	0.040 [-0.363;0.444]	0.114 [-0.290;0.517]	-0.089 [-0.493;0.315]	-0.055 [-0.170;0.060]
$\beta^{(5)}$	0.062 [-0.160;0.284]	0.076 [-0.361;0.513]	-0.133 [-0.400;0.134]	-0.040 [-0.095;0.014]
Variable	number of daughters	no son	1 son	2 sons
β	-0.040 [-0.114;0.034]	0.010 [-0.212;0.419]	-0.054 [-0.352;0.244]	-0.059 [-0.582;0.465]
$\beta^{(5)}$	-0.039 [-0.127;0.050]	0.060 [-0.185;0.306]	-0.062 [-0.258;0.134]	-0.025 [-0.470;0.419]
Variable	3 sons	4 sons	5 sons or more	no daughter
β	-0.023 [-0.850;0.805]	-0.039 [-0.490;0.411]	0.051 [-0.399;0.501]	0.015 [-0.267;0.297]
$\beta^{(5)}$	0.031 [-0.393;0.454]	-0.022 [-0.144;0.101]	0.014 [-0.109;0.137]	-0.003 [-0.130;0.124]
Variable	1 daughter	2 daughters	3 daughters	4 daughters
β	-0.121 [-0.493;0.252]	0.164 [-0.228;0.557]	0.051 [-0.690;0.793]	-0.084 [-0.806;0.638]
$\beta^{(5)}$	-0.076 [-0.245;0.092]	0.141 [-0.003;0.285]	0.037 [-0.603;0.676]	-0.084 [-0.458;0.289]
Variable	5 daughters or more	number of children	no child	1 child
β	-0.085 [-0.807;0.637]	-0.058* [-0.110;-0.007]	0.049 [-0.112;0.210]	0.012 [-0.388;0.411]
$\beta^{(5)}$	-0.072 [-0.569;0.426]	-0.048* [-0.090;-0.006]	-0.009 [-0.279;0.262]	0.014 [-0.491;0.520]
Variable	2 children	3 children	4 children	5 children or more
β	-0.044 [-0.599;0.512]	0.098 [-0.524;0.720]	-0.144 [-1.049;0.761]	0.003 [-0.423;0.430]
$\beta^{(5)}$	-0.023 [-0.501;0.455]	0.129 [-0.286;0.544]	-0.135 [-0.799;0.529]	0.007 [-0.427;0.441]
Variable	no child out of marriage	child out of marriage	age gap: 0 to 3	age gap: 4 to 7
β	-0.017 [-0.692;0.657]	0.017 [-0.657;0.692]	0.121* [0.015;0.227]	-0.053 [-0.363;0.257]
$\beta^{(5)}$	-0.035 [-0.644;0.575]	0.035 [-0.575;0.644]	0.196* [0.018;0.374]	0.025 [-0.354;0.404]
Variable	age gap: 8 to 12	age gap: 13 to 24	marriage certificate	
β	0.147 [-0.359;0.654]	-0.221* [-0.410;-0.032]	-0.138 [-0.571;0.294]	
$\beta^{(5)}$	0.137 [-0.367;0.642]	-0.381* [-0.739;-0.023]	-0.155 [-0.769;0.458]	

THE END

Thank you, all

A few references:

• Chauvet J., Trottier C., Bry X. (2019): *Component-based regularisation of multivariate generalised linear mixed models*, Journal of Computational and Graphical Statistics (in press).

• Bry X., Simac T., El Ghachi S., Antoine P. (2018) : *Bridging data exploration and modeling in event-history analysis: the supervised-component Cox regression*, Mathematical Population Studies (in press).

• Bry X., Trottier C., Mortier F., Cornu G. (2018): *Component-based regularisation of a multivariate GLM with a thematic partitioning of the explanatory variables*, Statistical Modeling, SAGE (in press).

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

• Bastien P., Esposito Vinzi V., and Tenenhaus M. (2005). *PLS generalised linear regression*. Computational Statistics & Data Analysis, 48: 17-46

• Bastien P. (2007): *Deviance residuals based PLS regression for censored data in high dimensional setting*, Chemometrics and Intelligent Laboratory Systems, pp. 78-86, 2007

• Bry X., Antoine P. (2004): *Exploring explanatory models ; an event history application* – Population-E 59(6).

• D.R. Cox, (1975): *Partial Likelihood*, Biometrika, vol. 62, p. 269-276

• Breslow, N. E. and Crowley, J. (1974): *A large-sample study of the life table and product limit estimates under random censorship*. Annals of Statistics 2, 437-454.

• Kalbfleisch, J. D. and Prentice, R. L. (1973): *Marginal likelihoods based on Cox's regression and life model*. Biometrika 60, 267-278.

• van Houwelingen HC, Bruinsma T, Hart AAM, van't Veer LJ, Wessels LFA (2006): *Cross-Validated Cox Regression on Microarray Gene Expression Data*. Statistics in Medicine, 25, 3201–3216