# Méthodes « sparse » en régression et ACP

**Gilbert Saporta**
Chaire de Statistique Appliquée & CEDRIC
CNAM
292 rue Saint Martin, F-75003 Paris

gilbert.saporta@cnam.fr
http://cedric.cnam.fr/~saporta

- Une combinaison linéaire de trop de variables est ininterprétable
- Méthodes « sparse » : rechercher des combinaisons comportant un nombre important de coefficients nuls
  - Compromis à trouver entre capacité prédictive et « sparsité »
- Méthodes récentes: article fondateur 1996
  - Packages R essentiellement

# 1. Le LASSO

http://www-stat.stanford.edu/~tibs/lasso.html

## The Lasso Page

**L1-constrained fitting for statistics and data mining**

The Lasso is a shrinkage and selection method for linear regression.

It minimizes the usual sum of squared errors, with a bound on the sum of the absolute values of the coefficients.

- Critère voisin de la ridge:

$$\|y - Xb\|^2 \quad \text{sous} \sum_{j=1}^{p} |b_j| < c$$

$$\hat{\beta} = \arg\min \|y - Xb\|^2 - \lambda \sum_{j=1}^{p} |b_j|$$

- Pénalité $L_1$ au lieu de $L_2$. Pas de solution analytique
- Si c est petit, certains coefficients seront nuls ⇒ sélection
- Si $c > \sum_{j=1}^{p} |b_{jols}|$ on retrouve la régression multiple usuelle

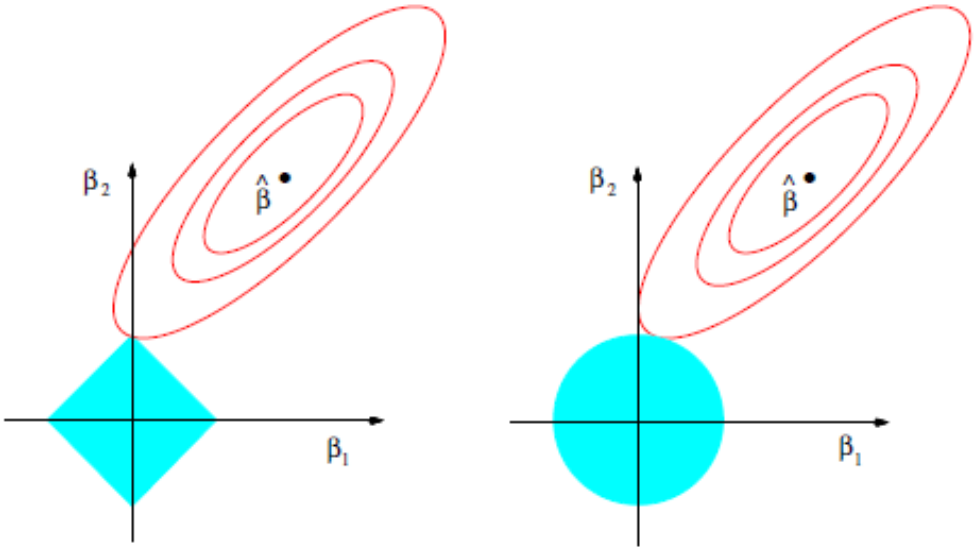



**FIGURE 3.11.** *Estimation picture for the lasso (left) and ridge regression (right). Shown are contours of the error and constraint functions. The solid blue areas are the constraint regions* $|\beta_1| + |\beta_2| \leq t$ *and* $\beta_1^2 + \beta_2^2 \leq t^2$, *respectively, while the red ellipses are the contours of the least squares error function.*

# Lasso et PROC GLMSELECT

```
ods graphics on;
proc glmselect data=bagnole plots=all;
model prix=cyl puis lon lar poids vitesse /
selection=lasso (stop=7 choose=BIC);
run;
```

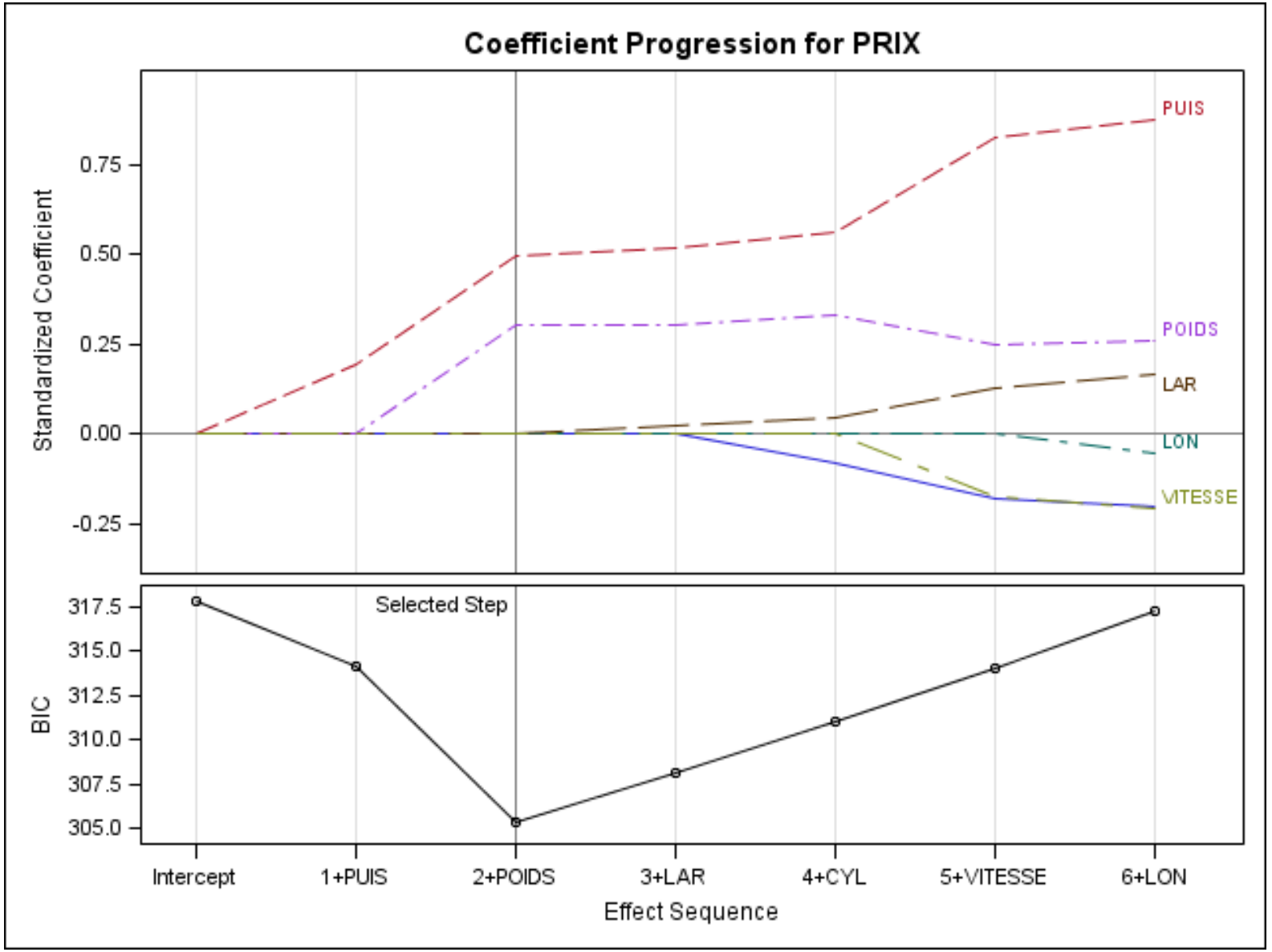
Données pour regression
The GLMSELECT Procedure

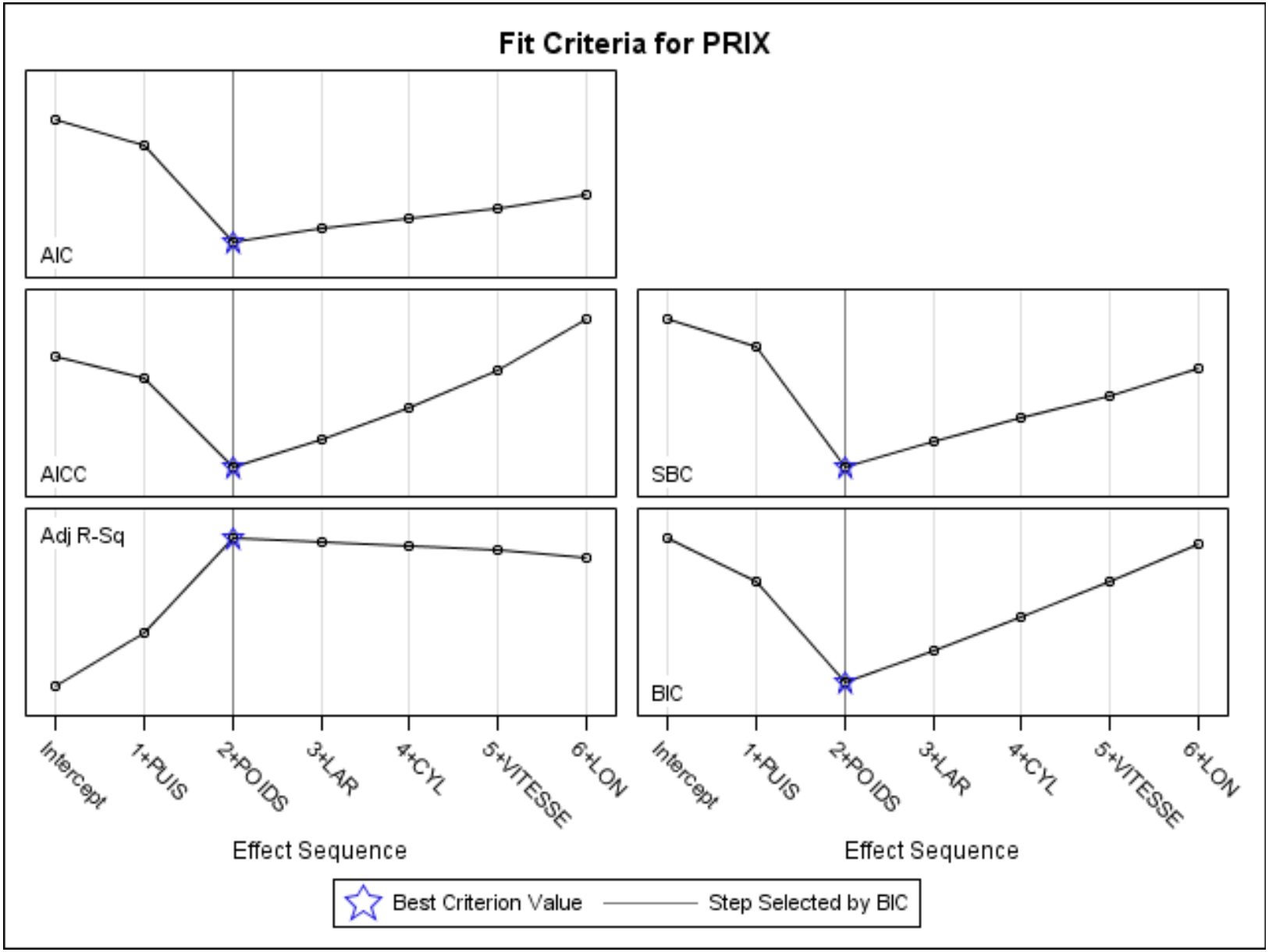
LASSO Selection Summary

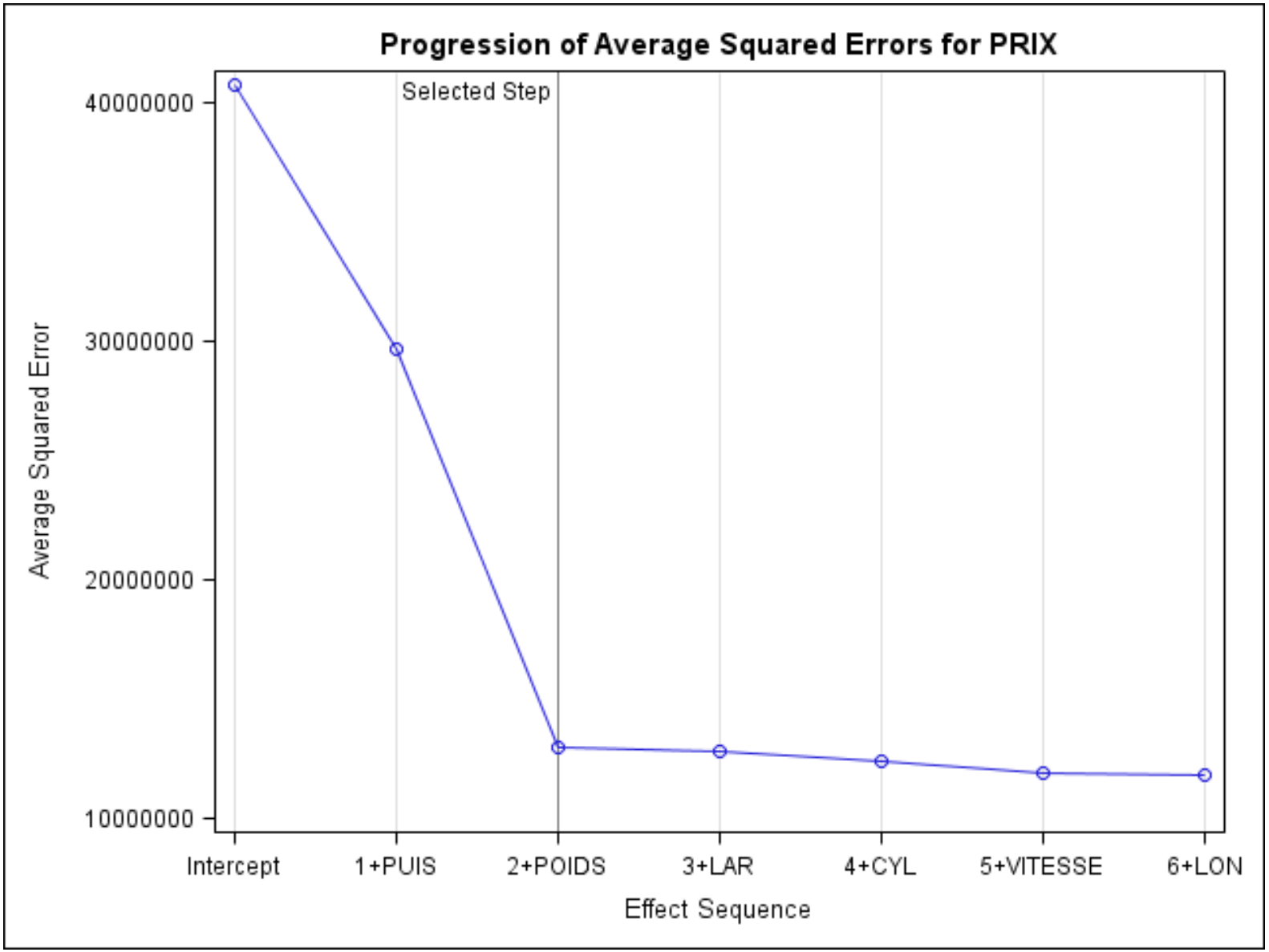
| Step | Effect Entered | Effect Removed | Number Effects In | BIC |
|---|---|---|---|---|
| 0 | Intercept | | 1 | 317.8324 |
| 1 | PUIS | | 2 | 314.0991 |
| 2 | POIDS | | 3 | 305.3416* |
| 3 | LAR | | 4 | 308.1679 |
| 4 | CYL | | 5 | 311.0317 |
| 5 | VITESSE | | 6 | 314.0364 |
| 6 | LON | | 7 | 317.3025 |

* Optimal Value Of Criterion

Selection stopped because all effects are in the final model.


Coefficient Progression for PRIX
Standardized Coefficient
0.75
0.50
0.25
0.00
-0.25
PUIS
POIDS
LAR
LON
VITESSE
BIC
317.5
315.0
312.5
310.0
307.5
305.0
Selected Step
Intercept
1+PUIS
2+POIDS
3+LAR
4+CYL
5+VITESSE
6+LON
Effect Sequence


Fit Criteria for PRIX
AIC
AICC
SBC
Adj R-Sq
BIC
Intercept
1+PUIS
2+POIDS
3+LAR
4+CYL
5+VITESSE
6+LON
Effect Sequence
Effect Sequence
Best Criterion Value
Step Selected by BIC


Progression of Average Squared Errors for PRIX
Selected Step
40000000
30000000
20000000
10000000
Average Squared Error
Intercept
1+PUIS
2+POIDS
3+LAR
4+CYL
5+VITESSE
6+LON
Effect Sequence

The GLMSELECT Procedure
Selected Model

The selected model, based on BIC, is the model at Step 2.

Effects: Intercept PUIS POIDS

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value |
|---|---|---|---|---|
| Model | 2 | 499772877 | 249886439 | 15.99 |
| Error | 15 | 234382913 | 15625528 | |
| Corrected Total | 17 | 734155790 | | |

| | |
|---|---|
| Root MSE | 3952.91380 |
| Dependent Mean | 34159 |
| R-Square | 0.6807 |
| Adj R-Sq | 0.6382 |
| AIC | 320.87771 |
| AICC | 323.95463 |
| BIC | 305.34160 |
| C(p) | 0.07232 |
| SBC | 303.54882 |

Parameter Estimates

| Parameter | DF | Estimate |
|---|---|---|
| Intercept | 1 | 5002.288413 |
| PUIS | 1 | 159.803389 |
| POIDS | 1 | 14.492675 |

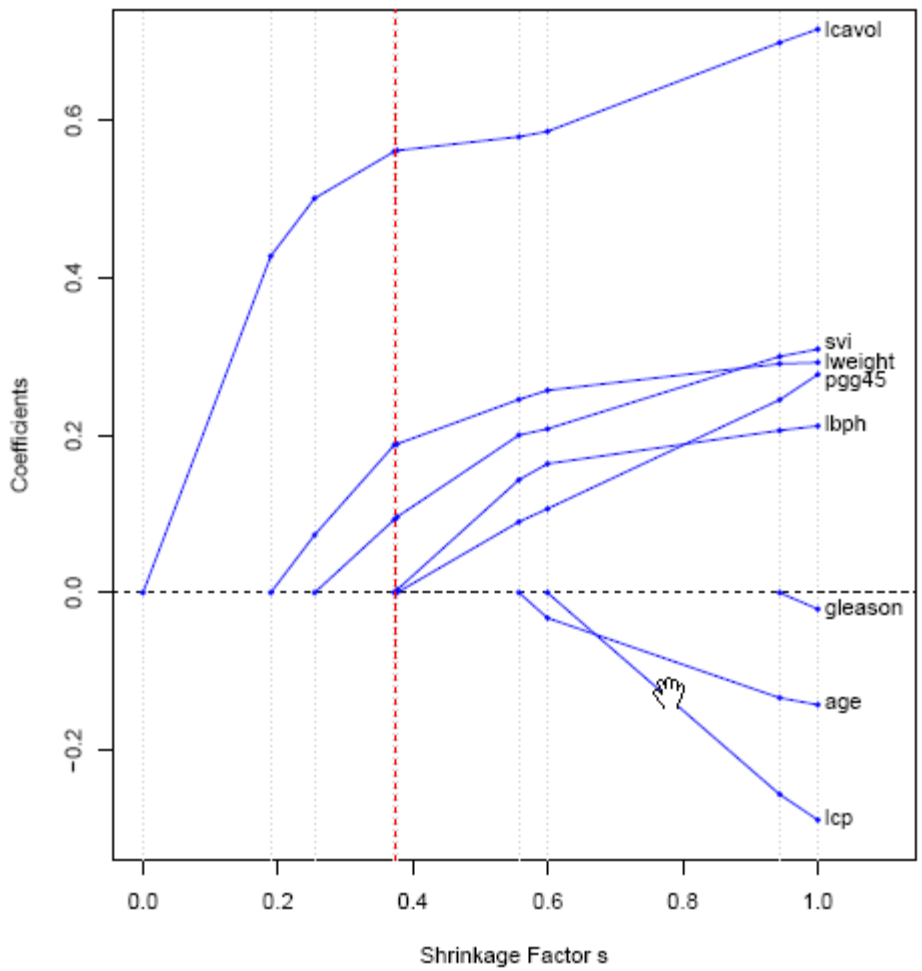
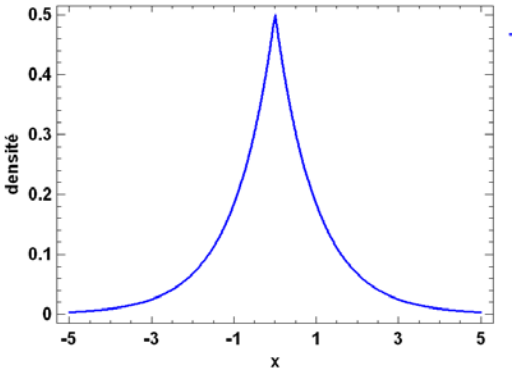



**FIGURE 3.10.** *Profiles of lasso coefficients, as the tuning parameter $t$ is varied. Coefficients are plotted versus $s = t/\sum_1^p |\hat{\beta}_j|$. A vertical line is drawn at $s = 0.36$, the value chosen by cross-validation. Compare Figure 3.8 on page 9; the lasso profiles hit zero, while those for ridge do not. The profiles are piece-wise linear, and so are computed only at the points displayed;*

- Interprétation bayesienne:
  - *a priori* de Laplace ou double exponentielle sur chaque $\beta_j$

$$f(\beta_j) = \frac{1}{2\tau}\exp\left(-\frac{|\beta_j|}{\tau}\right)$$




- Estimateur non linéaire

- Degré de liberté estimé par le nombre de coefficients non nuls (Zou et al. , 2007)
  - Possibilité d'utiliser des critères de type AIC ou BIC pour choisir le paramètre

$$\lambda_{opt} = \arg\min_{\lambda} \left( \frac{\left\| y - X\hat{\beta} \right\|^2}{n\sigma^2} + \frac{2}{n}\hat{df} \right) \text{ pour l'AIC}$$

$$\lambda_{opt} = \arg\min_{\lambda} \left( \frac{\left\| y - X\hat{\beta} \right\|^2}{n\sigma^2} + \frac{\log(n)}{n}\hat{df} \right) \text{ pour BIC}$$

- Avantages et inconvénients
- Pour:
  - Le  Lasso rétrécit les coefficients vers zéro de façon continue.
  - Produit un modèle parcimonieux.
  - Est une méthode de sélection.
- Contre:
  - le nombre de variables sélectionnées est limité par n
    - Inadapté au cas des puces à ADN  n(arrays)<<p(genes)
  - Choisit une seule variable dans un groupe de variables très corrélée

# Une variante « lasso hybrid selection » pour obliger SAS à faire de la validation croisée avec le critère PRESS

```
proc glmselect data=bagnole plots=all;
model prix=cyl puis lon lar poids vitesse /
 selection=lasso  (stop=7 lscoeffs choose=Press);
run;
```

**LSCOEFFS**

requests a hybrid version of the LAR and LASSO methods, where the sequence of models is determined by the LAR or LASSO algorithm but the coefficients of the parameters for the model at any step are determined by using ordinary least squares.

The GLMSELECT Procedure
Selected Model

The selected model, based on PRESS, is the model at Step 2.

Effects: Intercept PUIS POIDS

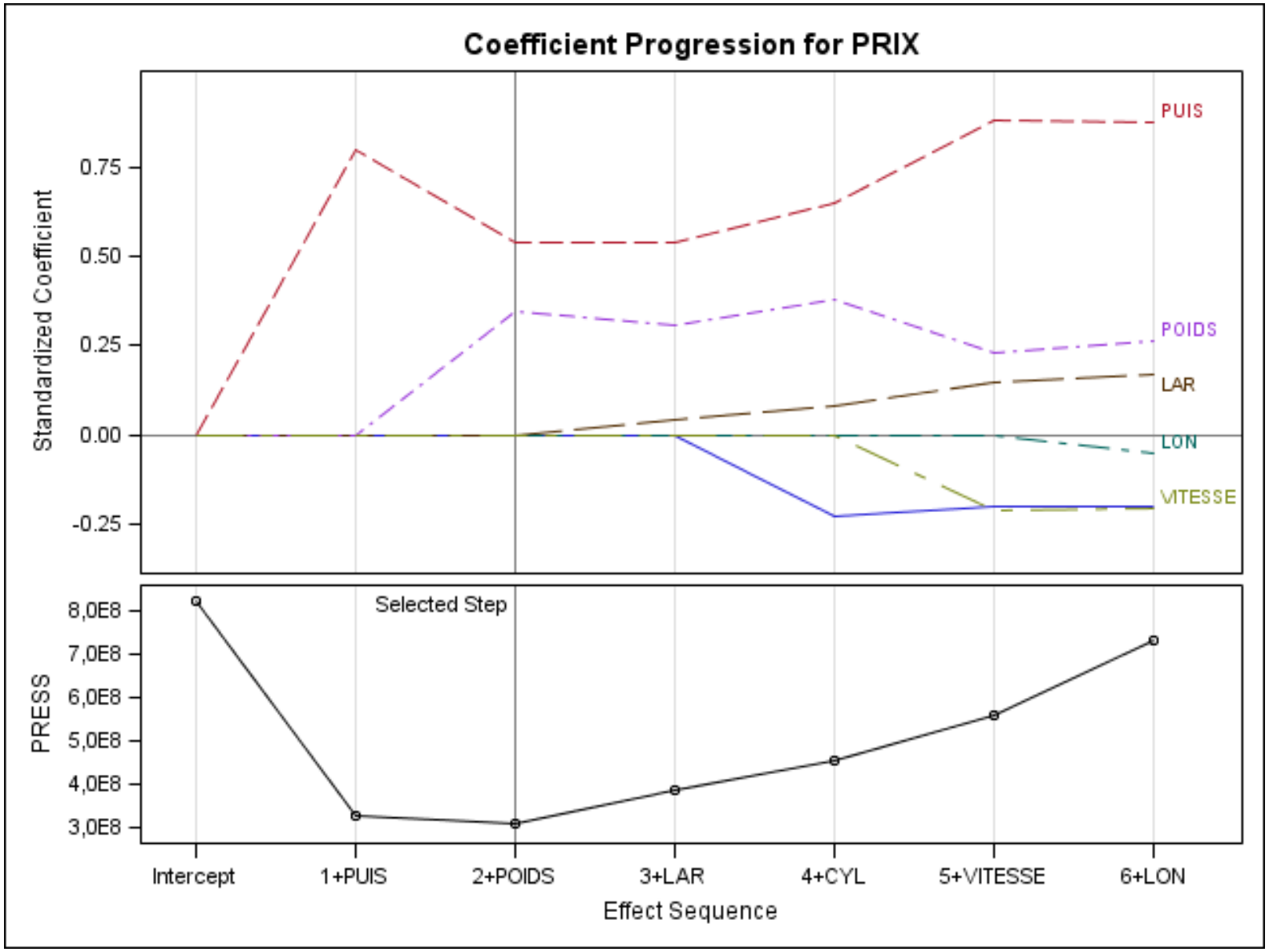
Analysis of Variance

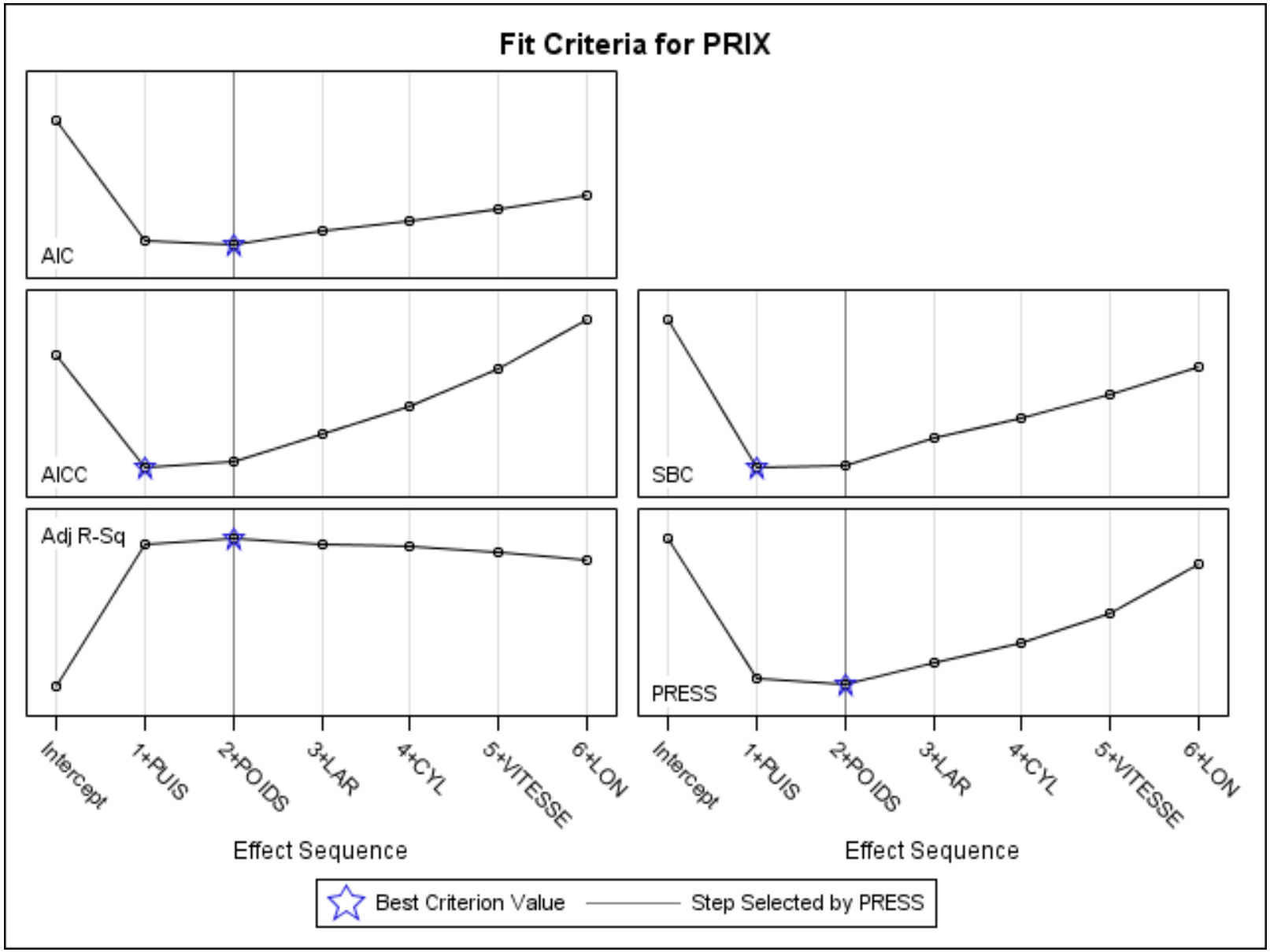
| Source | DF | Sum of Squares | Mean Square | F Value |
|---|---|---|---|---|
| Model | 2 | 504091154 | 252045577 | 16.43 |
| Error | 15 | 230064636 | 15337642 | |
| Corrected Total | 17 | 734155790 | | |

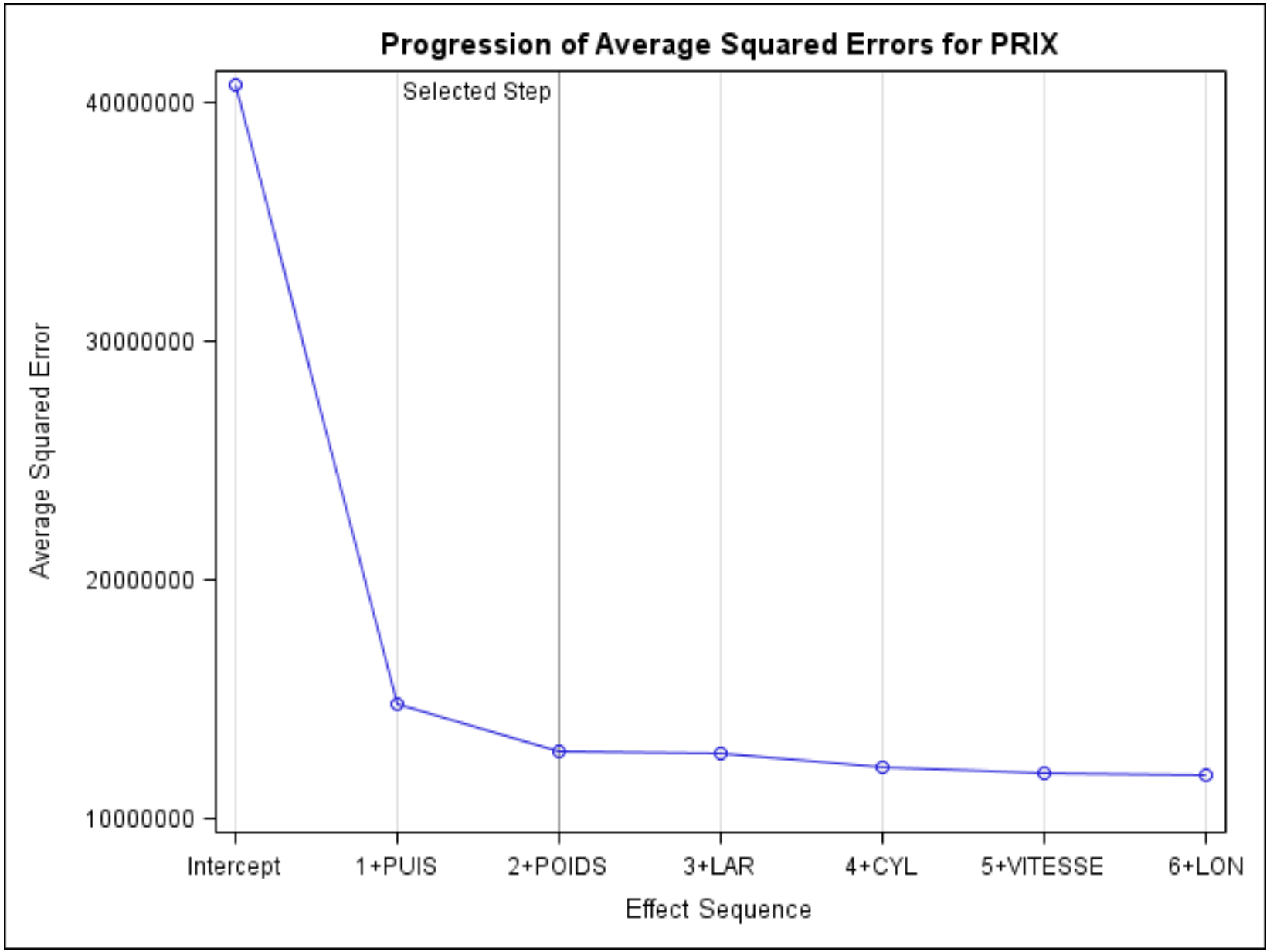
| | |
|---|---|
| Root MSE | 3916.33023 |
| Dependent Mean | 34159 |
| R-Square | 0.6866 |
| Adj R-Sq | 0.6448 |
| AIC | 320.54298 |
| AICC | 323.61991 |
| PRESS | 308496438 |
| SBC | 303.21410 |

Parameter Estimates

| Parameter | DF | Estimate |
|---|---|---|
| Intercept | 1 | 1775.601201 |
| PUIS | 1 | 172.967225 |
| POIDS | 1 | 16.451161 |


Coefficient Progression for PRIX
Standardized Coefficient
0.75
0.50
0.25
0.00
-0.25
PUIS
POIDS
LAR
LON
VITESSE
PRESS
8,0E8
7,0E8
6,0E8
5,0E8
4,0E8
3,0E8
Selected Step
Intercept
1+PUIS
2+POIDS
3+LAR
4+CYL
5+VITESSE
6+LON
Effect Sequence


Fit Criteria for PRIX
AIC
AICC
SBC
Adj R-Sq
PRESS
Intercept
1+PUIS
2+POIDS
3+LAR
4+CYL
5+VITESSE
6+LON
Effect Sequence
Effect Sequence
Best Criterion Value
Step Selected by PRESS


Progression of Average Squared Errors for PRIX
Selected Step
40000000
30000000
20000000
10000000
Average Squared Error
Intercept
1+PUIS
2+POIDS
3+LAR
4+CYL
5+VITESSE
6+LON
Effect Sequence

# 2. Elastic net

- Combine les pénalités de la ridge et du lasso

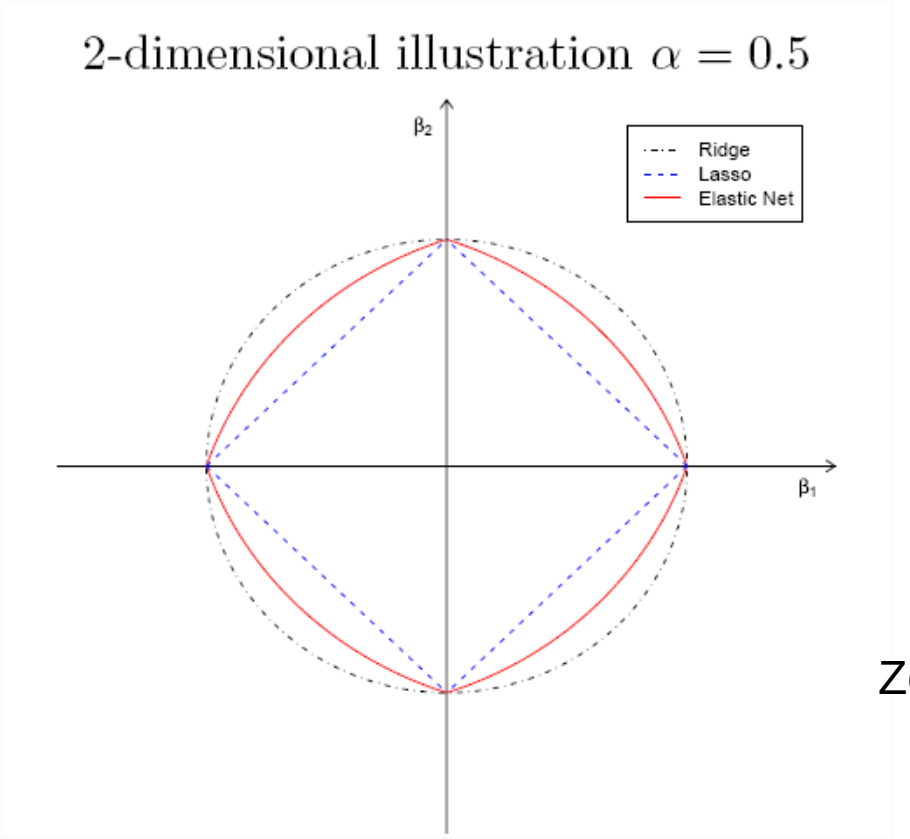
$$\min\left(\|y - Xb\|^2 + \lambda_2 \|b\|^2 + \lambda_1 \|b\|_1\right)$$

- autre formulation:

$$\min \|y - Xb\|^2 + \lambda \sum_{j=1}^{p} \left(\alpha b_j^2 + (1-\alpha)\left|b_j\right|\right)$$

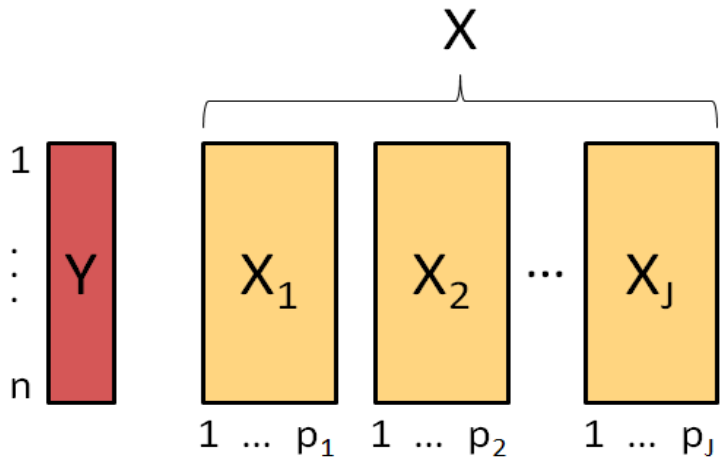
$$avec \quad \alpha = \frac{\lambda_2}{\lambda_1 + \lambda_2}$$

- La partie $L_1$ conduit à un modèle «sparse»
- La partie $L_2$ enlève la limitation sur le nombre de variables retenues et favorise le choix de groupes




Zou et Hastie

# 3. Group lasso



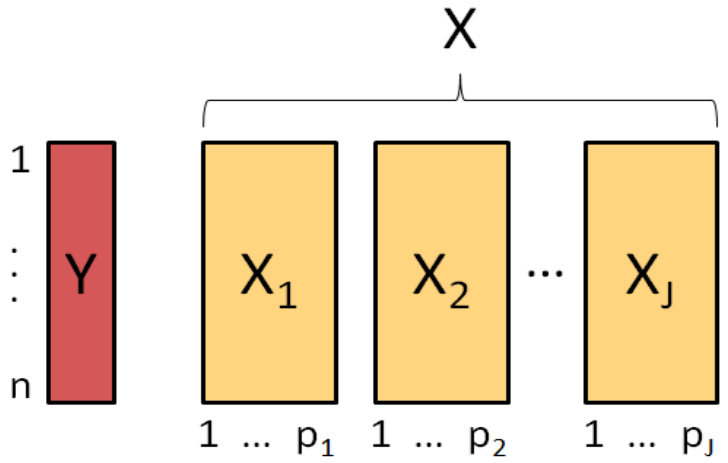

- Permet de sélectionner des groupes de prédicteurs
  - J groupes de $p_j$ variables $X_j$
  - Critère: (Yuan & Lin, 2007)

$$\min_{\boldsymbol{\beta}} \left( \left\| \mathbf{y} - \sum_{j=1}^{J} \mathbf{X}_j \boldsymbol{\beta}_j \right\|^2 + \lambda \sum_{j=1}^{J} \sqrt{p_j} \left\| \boldsymbol{\beta}_j \right\| \right)$$

  - Si $p_j=1$ pour tout j, on retrouve le lasso

# 3.2 Group-lasso

- X matrix divided into J sub-matrices $X_j$ of $p_j$ variables
- **Group Lasso**: extension of Lasso for selecting groups of variables (Yuan & Lin, 2007):




$$\hat{\boldsymbol{\beta}}_{GL} = \arg\min_{\beta} \left\| \mathbf{y} - \sum_{j=1}^{J} \mathbf{X}_j \boldsymbol{\beta}_j \right\|^2 + \lambda \sum_{j=1}^{J} \sqrt{p_j} \left\| \boldsymbol{\beta}_j \right\|$$

If $p_j$=1 for all j, group Lasso = Lasso

- Inconvénient: pas de « sparsité » à l'intérieur des groupes
- Variante de Simon et al. (2012) : sparse group lasso

$$\min_{\boldsymbol{\beta}} \left( \left\| \mathbf{y} - \sum_{j=1}^{J} \mathbf{X}_j \boldsymbol{\beta}_j \right\|^2 + \lambda_1 \sum_{j=1}^{J} \left\| \boldsymbol{\beta}_j \right\| + \lambda_2 \sum_{j=1}^{J} \sum_{i=1}^{p_j} \left| \beta_{ij} \right| \right)$$

- Deux paramètres à régler

# 4. ACP « Sparse »

- Objectif: obtenir des composantes facilement interprétables: beaucoup de coefficients nuls dans les facteurs principaux
- Simple Component Analysis (Vines, 2000)
  - Pas plus de 3 coefficients distincts +, 0, - (Rousson, 2004)

# Exemple

Table 1. Definitions of Variables in Jeffers' Pitprop Data

| *Variable* | *Definition* |
|---|---|
| $x_1$ | Top diameter in inches |
| $x_2$ | Length in inches |
| $x_3$ | Moisture content, % of dry weight |
| $x_4$ | Specific gravity at time of test |
| $x_5$ | Oven-dry specific gravity |
| $x_6$ | Number of annual rings at top |
| $x_7$ | Number of annual rings at bottom |
| $x_8$ | Maximum bow in inches |
| $x_9$ | Distance of point of maximum bow from top in inches |
| $x_{10}$ | Number of knot whorls |
| $x_{11}$ | Length of clear prop from top in inches |
| $x_{12}$ | Average number of knots per whorl |
| $x_{13}$ | Average diameter of the knots in inches |

| PCA | | | | | | Varimax | | | | | | SCA | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **.40** | −.22 | .21 | −.09 | −.08 | .12 | **.53** | .03 | .07 | .00 | .03 | .01 | **.45** | 0 | 0 | 0 | 0 | 0 |
| **.41** | −.19 | .24 | −.10 | −.11 | .16 | **.55** | .03 | .01 | .00 | .03 | −.01 | **.45** | 0 | 0 | 0 | 0 | 0 |
| .12 | **−.54** | −.14 | .08 | **.35** | −.28 | .04 | −.09 | **.71** | −.00 | −.00 | .14 | 0 | 0 | **.71** | 0 | 0 | 0 |
| .17 | **−.46** | **−.35** | .05 | **.36** | −.05 | .01 | .06 | **.69** | .01 | .00 | −.14 | 0 | 0 | **.71** | 0 | 0 | 0 |
| .06 | .17 | **−.48** | .05 | .18 | **.63** | −.11 | .22 | .01 | −.08 | −.05 | **−.79** | 0 | **.5** | 0 | 0 | 0 | **.71** |
| **.28** | .01 | **−.48** | −.06 | **−.32** | .05 | .06 | **.59** | .05 | .20 | −.04 | −.14 | 0 | **.5** | 0 | 0 | 0 | 0 |
| **.40** | .19 | −.25 | −.06 | −.22 | .00 | .15 | **.52** | −.07 | −.03 | .08 | −.05 | 0 | **.5** | 0 | 0 | 0 | 0 |
| **.29** | .19 | .24 | **.29** | .19 | −.06 | .16 | .00 | −.03 | **−.51** | −.05 | .10 | **.45** | 0 | 0 | 0 | 0 | 0 |
| **.36** | −.02 | .21 | .10 | −.11 | .03 | **.37** | .10 | −.05 | −.17 | −.07 | .09 | **.45** | 0 | 0 | 0 | 0 | 0 |
| **.38** | .25 | .12 | −.21 | .16 | −.17 | .18 | .16 | −.04 | **−.28** | **.40** | .13 | **.45** | 0 | 0 | 0 | 0 | 0 |
| .01 | .21 | −.07 | **−.80** | **.34** | −.18 | −.11 | −.03 | .02 | .12 | **.90** | −.04 | 0 | 0 | 0 | 0 | **1** | 0 |
| .12 | **.34** | .09 | .30 | **.60** | .17 | −.12 | −.16 | .04 | **−.70** | .06 | **−.28** | 0 | 0 | 0 | **1** | 0 | 0 |
| .11 | **.31** | **−.33** | .30 | −.08 | **−.63** | **−.40** | **.51** | .08 | −.26 | −.06 | **.45** | 0 | **.5** | 0 | 0 | 0 | **−.71** |

| | | | | | |
|---|---|---|---|---|---|
| 1 | | | | | |
| 0 | 1 | | | | |
| 0 | 0 | 1 | | | |
| 0 | 0 | 0 | 1 | | |
| 0 | 0 | 0 | 0 | 1 | |
| 0 | 0 | 0 | 0 | 0 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| 1 | | | | | |
| .31 | 1 | | | | |
| .27 | .14 | 1 | | | |
| −.32 | −.35 | .17 | 1 | | |
| .14 | .22 | −.13 | −.21 | 1 | |
| .16 | −.17 | −.05 | −.04 | −.04 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| 1 | | | | | |
| .31 | 1 | | | | |
| .17 | .11 | 1 | | | |
| .22 | .18 | −.20 | 1 | | |
| .00 | .08 | −.13 | .03 | 1 | |
| −.11 | −.02 | .09 | −.03 | .07 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| 32.5 | 18.3 | 14.4 | 8.5 | 7.0 | 6.3 |
| 32.5 | 50.7 | 65.2 | 73.7 | 80.7 | 87.0 |

| | | | | | |
|---|---|---|---|---|---|
| 22.5 | 18.6 | 14.5 | 14.2 | 9.6 | 7.6 |
| 22.5 | 39.4 | 52.8 | 63.3 | 72.0 | 78.9 |

| | | | | | |
|---|---|---|---|---|---|
| 25.3 | 17.1 | 14.5 | 7.7 | 7.7 | 6.1 |
| 25.3 | 40.7 | 54.7 | 62.2 | 68.9 | 74.8 |

| | | |
|---|---|---|
| Max cor.= 0 | Opt=100% | |

| | | |
|---|---|---|
| Max cor.= .35 | Opt=90.7% | |

| | | |
|---|---|---|
| Max cor.= .31 | Opt=86.0% | |

Table 6: Summary of six-components solutions for the pitprop data obtained from PCA, Varimax and SCA, where the last three variables have been inverted. From top to bottom are given the matrix of loadings, the correlation matrix between components, the percentage of total variability accounted by each component together with cumulative scores (corrected for correlations), and finally, the maximum correlation between components together with the optimality of the system. Loadings larger or equal to $p^{-1/2} = .277$ are marked in bold face, where $p = 13$ is the number of original variables.

# 4.1 Le SCoTLASS de Joliffe et al (2003)

- Simplified Component Technique -Lasso

$$\max \mathbf{u'Vu} \quad \text{avec} \quad \|\mathbf{u}\|^2 = \mathbf{u'u} = 1 \text{ et } \sum_{j=1}^{p} |u_j| \leq t$$

- Modification du critère de l'ACP

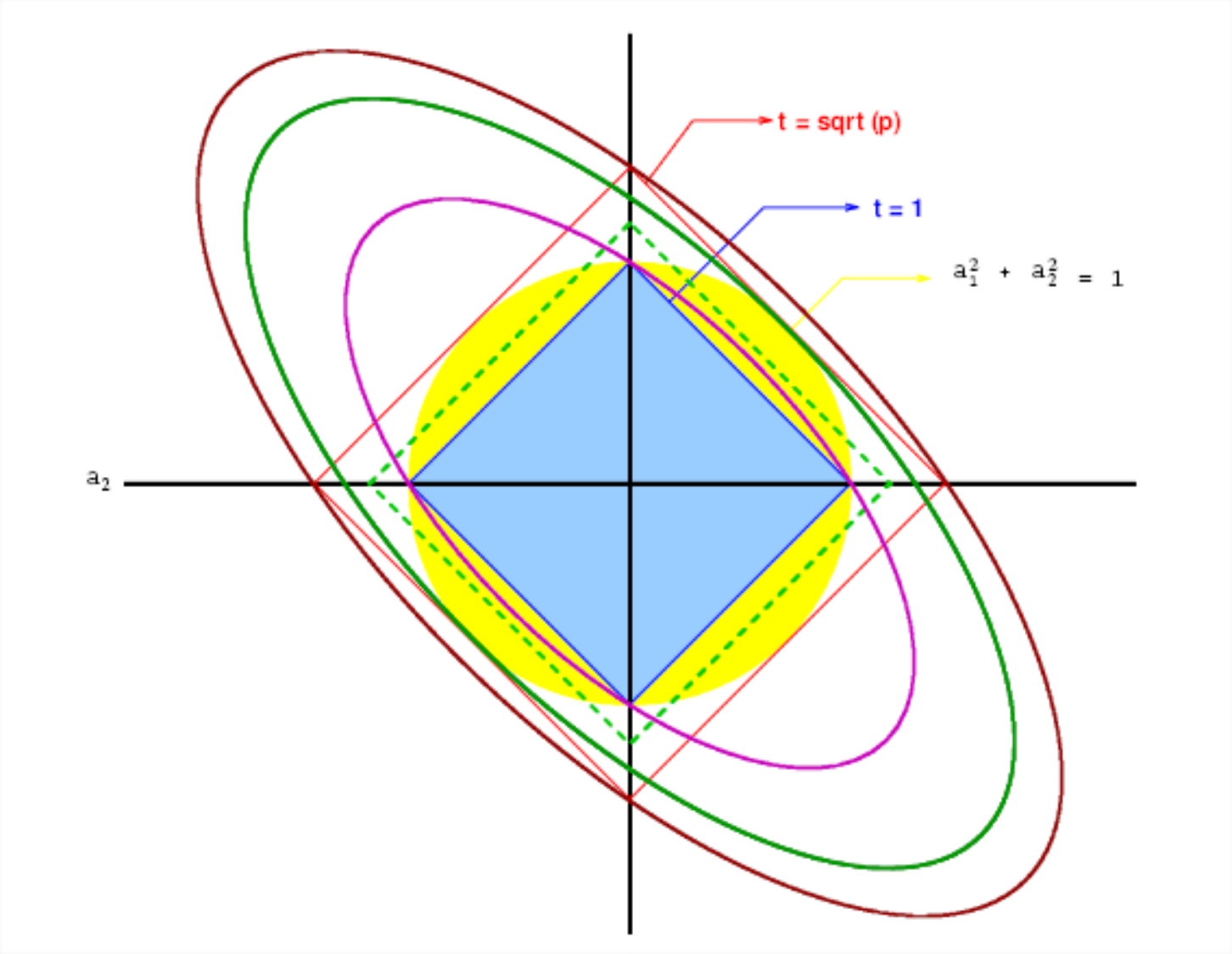
- Propriétés

$t \geq \sqrt{p}$ on retrouve l'ACP

$t < 1$ pas de solution

$t = 1$ un seul coefficient non nul

on cherche $1 < t < \sqrt{p}$ en décroissant depuis $\sqrt{p}$


t = sqrt (p)
t = 1
$a_1^2 + a_2^2 = 1$
$a_2$

| | | Component | | | | | |
|---|---|---|---|---|---|---|---|
| *Technique* | *Variable* | *(1)* | *(2)* | *(3)* | *(4)* | *(5)* | *(6)* |
| SCoTLASS | $x_1$ | 0.558 | 0.085 | −0.093 | −0.107 | 0.056 | 0.017 |
| ($t = 2.25$) | $x_2$ | 0.580 | 0.031 | −0.087 | −0.147 | 0.073 | 0.047 |
| | $x_3$ | 0.000 | 0.647 | −0.129 | 0.215 | −0.064 | −0.101 |
| | $x_4$ | 0.000 | 0.654 | −0.000 | 0.211 | −0.080 | 0.127 |
| | $x_5$ | −0.000 | 0.000 | 0.413 | −0.000 | 0.236 | 0.747 |
| | $x_6$ | 0.001 | 0.208 | 0.529 | −0.022 | −0.108 | 0.033 |
| | $x_7$ | 0.266 | −0.000 | 0.385 | 0.000 | −0.121 | 0.020 |
| | $x_8$ | 0.104 | −0.098 | 0.000 | 0.584 | 0.127 | −0.188 |
| | $x_9$ | 0.372 | −0.000 | −0.000 | 0.019 | 0.142 | −0.060 |
| | $x_{10}$ | 0.364 | −0.154 | 0.000 | 0.212 | −0.296 | 0.000 |
| | $x_{11}$ | −0.000 | 0.099 | −0.000 | 0.000 | 0.879 | −0.156 |
| | $x_{12}$ | −0.000 | 0.241 | −0.001 | −0.699 | −0.044 | −0.186 |
| | $x_{13}$ | −0.000 | 0.026 | −0.608 | −0.026 | −0.016 | 0.561 |

Table 5. Simplicity Factor, Variance, Cumulative Variance and Number of Zero Loadings for Individual Components in PCA, RPCA, and SCoTLASS for Four Values of $t$, Based on the Correlation Matrix for Jeffers' Pitprop Data

| | | Component | | | | | |
|---|---|---|---|---|---|---|---|
| *Technique* | *Measure* | *(1)* | *(2)* | *(3)* | *(4)* | *(5)* | *(6)* |
| PCA | Simplicity factor (varimax) | 0.059 | 0.103 | 0.082 | 0.397 | 0.086 | 0.266 |
| (= SCoTLASS with | Variance (%) | 32.4 | 18.2 | 14.4 | 8.9 | 7.0 | 6.3 |
| $t = \sqrt{13}$) | Cumulative variance (%) | 32.4 | 50.7 | 65.1 | 74.0 | 80.9 | 87.2 |
| RPCA | Simplicity factor (varimax) | 0.362 | 0.428 | 0.199 | 0.595 | 0.131 | 0.343 |
| | Variance (%) | 13.0 | 14.6 | 18.4 | 9.7 | 23.9 | 7.6 |
| | Cumulative variance (%) | 13.0 | 27.6 | 46.0 | 55.7 | 79.6 | 87.2 |
| SCoTLASS | Simplicity factor (varimax) | 0.190 | 0.312 | 0.205 | 0.308 | 0.577 | 0.364 |
| ($t = 2.25$) | Variance (%) | 26.7 | 17.2 | 15.9 | 9.7 | 8.9 | 6.7 |
| | Cumulative variance (%) | 26.7 | 43.9 | 59.8 | 69.4 | 78.4 | 85.0 |
| | Number of zero loadings | 6 | 3 | 5 | 3 | 0 | 1 |
| SCoTLASS | Simplicity factor (varimax) | 0.288 | 0.301 | 0.375 | 0.387 | 0.646 | 0.412 |
| ($t = 2.00$) | Variance (%) | 23.1 | 16.4 | 16.2 | 11.2 | 8.9 | 6.5 |
| | Cumulative variance (%) | 23.1 | 39.5 | 55.8 | 67.0 | 75.9 | 82.3 |
| | Number of zero loadings | 7 | 6 | 2 | 4 | 1 | 2 |
| SCoTLASS | Simplicity factor (varimax) | 0.370 | 0.370 | 0.388 | 0.360 | 0.610 | 0.714 |
| ($t = 1.75$) | Variance (%) | 19.6 | 16.0 | 13.2 | 13.0 | 9.2 | 9.1 |
| | Cumulative variance (%) | 19.6 | 35.6 | 48.7 | 61.8 | 71.0 | 80.1 |
| | Number of zero loadings | 7 | 7 | 7 | 7 | 3 | 0 |
| SCoTLASS | Simplicity factor (varimax) | 0.452 | 0.452 | 0.504 | 0.464 | 0.565 | 0.464 |
| ($t = 1.50$) | Variance (%) | 16.1 | 14.9 | 13.8 | 10.2 | 9.9 | 9.6 |
| | Cumulative variance (%) | 16.1 | 31.0 | 44.9 | 55.1 | 65.0 | 74.5 |
| | Number of zero loadings | 5 | 7 | 2 | 1 | 3 | 5 |

- inconvénients:
  - choix de t peu évident
  - problème non convexe
  - calculs difficiles

# 4.2 la SPCA de Zou et al. (2006)

- Reconstitution du tableau par les composantes et les facteurs principaux **X=CU'**
- Si on régresse la i$^{ème}$ composante principale sur **X,** on trouve $\mathbf{U_i}$ dans les bons cas (pas si p>n par exemple)
- Régression ridge $\hat{\boldsymbol{\beta}}_{ridge} = \arg\min\left(\left\|\mathbf{c}_i - \mathbf{X}\boldsymbol{\beta}\right\|^2 + \lambda\left\|\boldsymbol{\beta}\right\|^2\right)$

$$\mathbf{u}_i = \frac{\hat{\boldsymbol{\beta}}_{ridge}}{\left\|\hat{\boldsymbol{\beta}}_{ridge}\right\|}$$

- Démonstration:

$$\mathbf{X'X} = \mathbf{VDV'} \text{ avec } \mathbf{V'V} = \mathbf{I}$$

$$\hat{\boldsymbol{\beta}}_{i,ridge} = \left(\mathbf{X'X} + k\mathbf{I}\right)^{-1} \mathbf{X'V}\mathbf{v}_i = \frac{d_i}{d_i + k}\mathbf{v}_i$$

- Reconstruction des facteurs (loadings) de l'ACP par une régression ridge
  - Postérieure à l'ACP, ne la remplace pas
- La pénalité ridge n'en est pas vraiment une: sert simplement à reconstruire les composantes

- La S-PCA ajoute une nouvelle pénalité

$$\min\left(\left\|\mathbf{c}_i - \mathbf{X}\boldsymbol{\beta}\right\|^2 + \lambda\left\|\boldsymbol{\beta}\right\|^2 + \lambda_1\left\|\boldsymbol{\beta}\right\|_1\right)$$

$$\left\|\boldsymbol{\beta}\right\|_1 = \sum_{j=1}^{p}\left|\beta_j\right|$$

- Généralisation

$$(\widehat{\mathbf{A}}, \widehat{\mathbf{B}}) \quad = \quad \underset{\mathbf{A},\,\mathbf{B}}{\arg\min} \sum_{i=1}^{n} \|\mathbf{x}_i - \mathbf{A}\mathbf{B}^T\mathbf{x}_i\|^2 + \lambda \sum_{j=1}^{k} \|\beta_j\|^2 + \sum_{j=1}^{k} \lambda_{1,j}\|\beta_j\|_1$$

- algorithme alterné sur A et B . Implémenté en R

Table 3. Pitprops Data: Loadings of the First Six Sparse PCs by SPCA. Empty cells have zero loadings.

| *Variable* | *PC1* | *PC2* | *PC3* | *PC4* | *PC5* | *PC6* |
|---|---|---|---|---|---|---|
| topdiam | −0.477 | | | | | |
| length | −0.476 | | | | | |
| moist | | 0.785 | | | | |
| testsg | | 0.620 | | | | |
| ovensg | 0.177 | | 0.640 | | | |
| ringtop | | | 0.589 | | | |
| ringbut | −0.250 | | 0.492 | | | |
| bowmax | −0.344 | −0.021 | | | | |
| bowdist | −0.416 | | | | | |
| whorls | −0.400 | | | | | |
| clear | | | | −1 | | |
| knots | | 0.013 | | | −1 | |
| diaknot | | | −0.015 | | | 1 |
| Number of nonzero loadings | 7 | 4 | 4 | 1 | 1 | 1 |
| Variance (%) | 28.0 | 14.4 | 15.0 | 7.7 | 7.7 | 7.7 |
| Adjusted variance (%) | 28.0 | 14.0 | 13.3 | 7.4 | 6.8 | 6.2 |
| Cumulative adjusted variance (%) | 28.0 | 42.0 | 55.3 | 62.7 | 69.5 | 75.8 |

$\lambda = 0$ car $n>p$ les $\lambda_1$ de chaque composante sont choisis pour avoir à peu pres la même part de variance expliquée

# 4.3 La sPCA-rSVD de Shen et Huang (2008)

- Part de la SVD de X=UΔV' avec pénalité de type Lasso

$$\mathbf{X}^{(k)} = \sum_{j=1}^{k} d_j \mathbf{u}_j \mathbf{v}'_j$$

$$\min_{\mathbf{u},\mathbf{v}} \left\| \mathbf{X} - \mathbf{u}\mathbf{v}' \right\|^2 + \sum_{j=1}^{p} g_\lambda \left( \left| v_j \right| \right)$$

- L’algorithme ne fait intervenir que la matrice V: application à p>>n
- Critère différent de la SPCA

$$\min_{\tilde{\mathbf{v}}} \{-2\|\mathbf{X}\tilde{\mathbf{v}}\| + \|\tilde{\mathbf{v}}\|^2 + \lambda|\tilde{\mathbf{v}}|_1\},$$

while the same argument yields that SPCA solves

$$\min_{\tilde{\mathbf{v}}}\{-2\|\mathbf{X}^T\mathbf{X}\tilde{\mathbf{v}}\| + \|\mathbf{X}\tilde{\mathbf{v}}\|^2 + \lambda\|\tilde{\mathbf{v}}\|^2 + \lambda_1|\tilde{\mathbf{v}}|_1\}.$$

- Algorithme plus rapide
- Degré de sparsity semble plus fort (choix graphique)

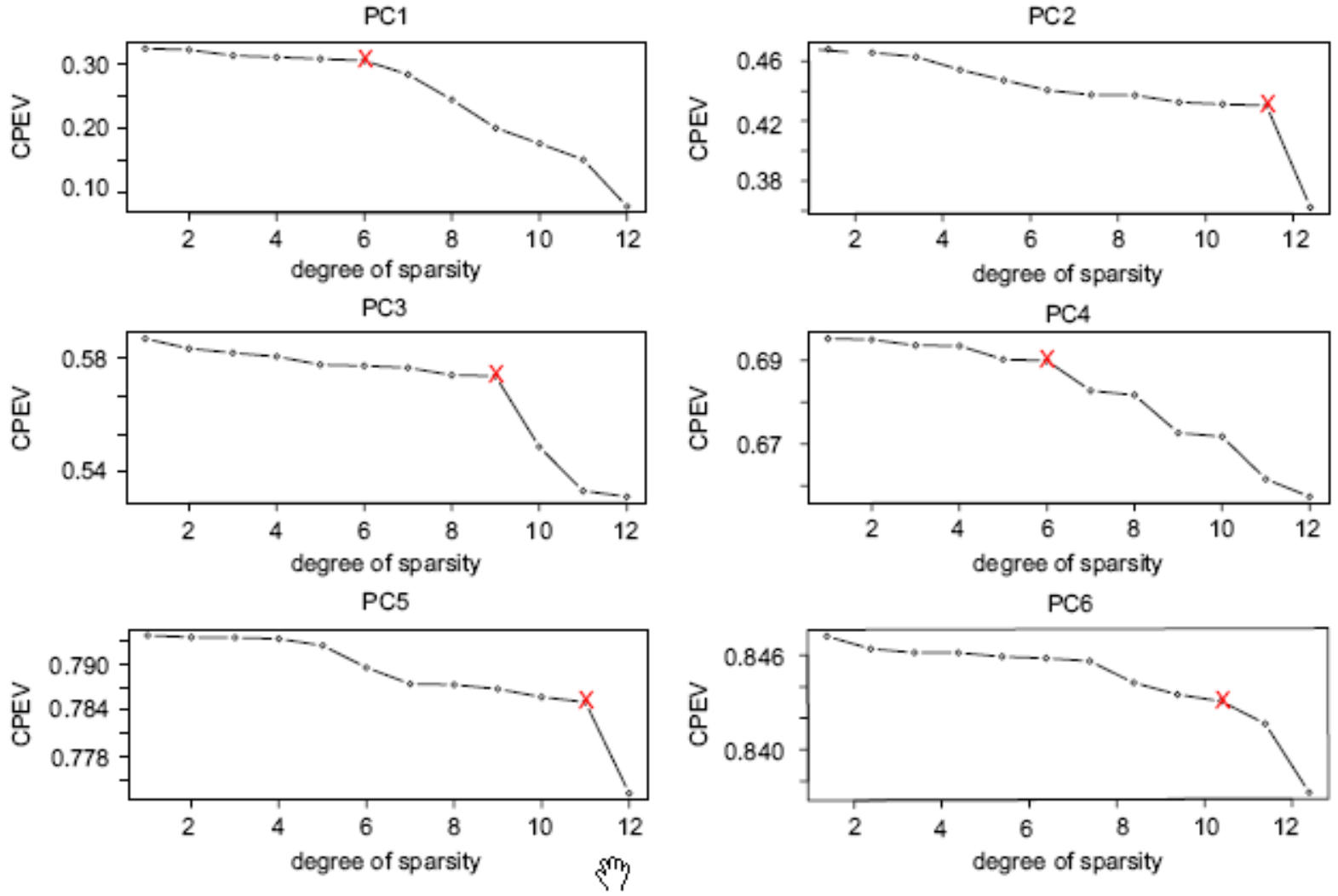



Fig. 2. (Pitprops data) CPEV plot for sPCA-rSVD-soft with selected degrees of sparsity marked.

Table 7
(Pitprops data) Loadings of the first six PCs by PCA and sPCA-rSVD-soft

| Variable | PCA | | | | | | sPCA-rSVD-soft | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 |
| $x_1$ | −0.404 | 0.218 | −0.207 | 0.091 | −0.083 | 0.120 | −0.449 | 0 | 0 | −0.114 | 0 | 0 |
| $x_2$ | −0.406 | 0.186 | −0.235 | 0.103 | −0.113 | 0.163 | −0.460 | 0 | 0 | −0.102 | 0 | 0 |
| $x_3$ | −0.124 | 0.541 | 0.141 | −0.078 | 0.350 | −0.276 | 0 | −0.707 | 0 | 0 | 0 | 0 |
| $x_4$ | −0.173 | 0.456 | 0.352 | −0.055 | 0.356 | −0.054 | 0 | −0.707 | 0 | 0 | 0 | 0 |
| $x_5$ | −0.057 | −0.170 | 0.481 | −0.049 | 0.176 | 0.626 | 0 | 0 | 0.550 | 0 | 0 | −0.744 |
| $x_6$ | −0.284 | −0.014 | 0.475 | 0.063 | −0.316 | 0.052 | −0.199 | 0 | 0.546 | −0.176 | 0 | 0 |
| $x_7$ | −0.400 | −0.190 | 0.253 | 0.065 | −0.215 | 0.003 | −0.399 | 0 | 0.366 | 0 | 0 | 0 |
| $x_8$ | −0.294 | −0.189 | −0.243 | −0.286 | 0.185 | −0.055 | −0.279 | 0 | 0 | 0.422 | 0 | 0 |
| $x_9$ | −0.357 | 0.017 | −0.208 | −0.097 | −0.106 | 0.034 | −0.380 | 0 | 0 | 0 | 0 | 0 |
| $x_{10}$ | −0.379 | −0.248 | −0.119 | 0.205 | 0.156 | −0.173 | −0.407 | 0 | 0 | 0.283 | 0.231 | 0 |
| $x_{11}$ | 0.011 | 0.205 | −0.070 | −0.804 | −0.343 | 0.175 | 0 | 0 | 0 | 0 | −0.973 | 0 |
| $x_{12}$ | 0.115 | 0.343 | 0.092 | 0.301 | −0.600 | −0.170 | 0 | 0 | 0 | −0.785 | 0 | 0.161 |
| $x_{13}$ | 0.113 | 0.309 | −0.326 | 0.303 | 0.080 | 0.626 | 0 | 0 | −0.515 | −0.265 | 0 | −0.648 |
| Sparsity | 0 | 0 | 0 | 0 | 0 | 0 | 6 | 11 | 9 | 6 | 11 | 10 |
| CPEV | 32.5 | 50.7 | 65.2 | 73.7 | 80.7 | 87.0 | 30.6 | 45.0 | 59.0 | 70.0 | 78.5 | 84.5 |

- Perte d’orthogonalité
  - ScotLass: facteurs orthogonaux mais pas les composantes qui sont corrélées
  - S-PCA: ni les facteurs, ni les composantes ne sont orthogonales
  - Nécessité d’ajuster les % de variance expliquée

# 5. Sparse PLS

Combine sélection de variables et modélisation

## 5.1 L'approche de Le Cao et al. (2008)

- PLS-SVD. Une variante de la PLS2 pour Y et X.
  - max cov(Xu,Yv) ⇒ vecteurs propres de X'Y
  - SVD de X'Y mais composantes non orthogonales

- Pénalisation de u et v pour sélectionner dans les deux groupes
- Critère

$$\min_{\mathbf{u},\mathbf{v}} \left\| \mathbf{X}'\mathbf{Y} - \mathbf{u}\mathbf{v}' \right\| + P_{\lambda_1}(\mathbf{u}) + P_{\lambda_2}(\mathbf{v})$$

- $P_\lambda$ est la fonction de « soft thresholding » ou « seuillage doux »

$$P_\lambda(x) = \left( |x| - \lambda \right)_+ sign(x)$$

- Résolution itérative

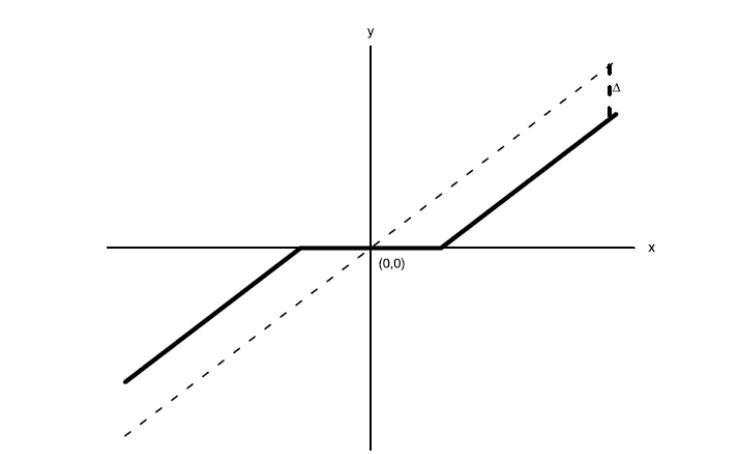



*Figure 1. An illustration of soft-thresholding rule* $y = (|x| - \Delta)_+ Sign(x)$ *with* $\Delta = 1$.

1. $X_0 = X \qquad Y_0 = Y$

2. For h in 1..H:

   (a) Set $\tilde{M}_{h-1} = X_{h-1}^T Y_{h-1}$

   (b) Decompose $\tilde{M}_{h-1}$ and extract the first pair of singular vectors $u_{old} = u_h$ and $v_{old} = v_h$

   (c) Until convergence of $u_{new}$ and $v_{new}$:

      i. $u_{new} = g_{\lambda_2}(\tilde{M}_{h-1} v_{old})$, norm $u_{new}$

      ii. $v_{new} = g_{\lambda_1}(\tilde{M}_{h-1}^T u_{old})$, norm $v_{new}$

      iii. $u_{old} = u_{new}$, $v_{old} = v_{new}$

   (d) $\xi_h = X_{h-1} u_{new} / u'_{new} u_{new}$
   $\omega_h = Y_{h-1} v_{new} / v'_{new} v_{new}$

   (e) $c_h = X_{h-1}^T \xi_h / \xi'_h \xi_h$
   $d_h = Y_{h-1}^T \xi_h / \xi'_h \omega_h$
   $e_h = Y_{h-1}^T \omega_h / \omega'_h \omega_h$

   (f) $X_h = X_{h-1} - \xi_h c'_h$

   (g) $Y_h = Y_{h-1} - \xi_h d'_h$

# 5.2 L'approche de Chun et Keles

- SIMPLS pour Y multiple à q composantes

$$\max_{\mathbf{w}} \left( \mathbf{w'X'YY'Xw} \right) \text{ avec } \left\| \mathbf{w} \right\|^2 = 1$$

- Idée: rechercher un facteur $\mathbf{c}$ sparse mais proche de la solution initiale $\boldsymbol{\alpha}$

$$\min_{\alpha, c} \left( -k\boldsymbol{\alpha}'\mathbf{M}\boldsymbol{\alpha} + (1-k)(\mathbf{c}-\boldsymbol{\alpha})'\mathbf{M}(\mathbf{c}-\boldsymbol{\alpha}) + \lambda_1 \left\| \mathbf{c} \right\|_1 + \lambda_2 \left\| \mathbf{c} \right\|^2 \right)$$

avec $\boldsymbol{\alpha}'\boldsymbol{\alpha} = \mathbf{c}'\mathbf{c} = 1$ et $\mathbf{M} = \mathbf{X'YY'X}$

- Cas particuliers M=X'X
  - $\boldsymbol{\alpha}=\mathbf{c}$ ScotLASS
  - k=0.5 SPCA
- Solution: alterner recherche de $\boldsymbol{\alpha}$ à $\mathbf{c}$ fixé et vice-versa
  - À c fixé

$$\min_{\boldsymbol{\alpha}}\left(-k\boldsymbol{\alpha}'\mathbf{M}\boldsymbol{\alpha}+(1-k)(\mathbf{c}-\boldsymbol{\alpha})'\mathbf{M}(\mathbf{c}-\boldsymbol{\alpha})\right)=$$

$$\min_{\boldsymbol{\alpha}}\left((\mathbf{Y}'\mathbf{X}\boldsymbol{\alpha}-k_1\,\mathbf{Y}'\mathbf{X}\mathbf{c})'(\mathbf{Y}'\mathbf{X}\boldsymbol{\alpha}-k_1\mathbf{Y}'\mathbf{X}\mathbf{c})\right) \text{ avec } k_1=\frac{1-k}{1-2k}$$

$$\boldsymbol{\alpha}=k_1\left(\mathbf{M}+\lambda*\mathbf{I}\right)^{-1}\mathbf{M}\mathbf{c} \text{ où } \lambda* \text{ est solution de } \mathbf{c}'(\mathbf{M}+\lambda\mathbf{I})^{-2}\mathbf{M}\mathbf{c}=k_1^2$$

- A $\alpha$ fixé

$$\min_{\mathbf{c}} \left(\mathbf{Y'Xc} - \mathbf{Y'X\alpha}\right)' \left(\mathbf{Y'Xc} - \mathbf{Y'X\alpha}\right) + \lambda_1 \left\|\mathbf{c}\right\|_1 + \lambda_2 \left\|\mathbf{c}\right\|$$

Elastic net avec $\mathbf{y}$ remplacé par $\mathbf{Y'X\alpha}$

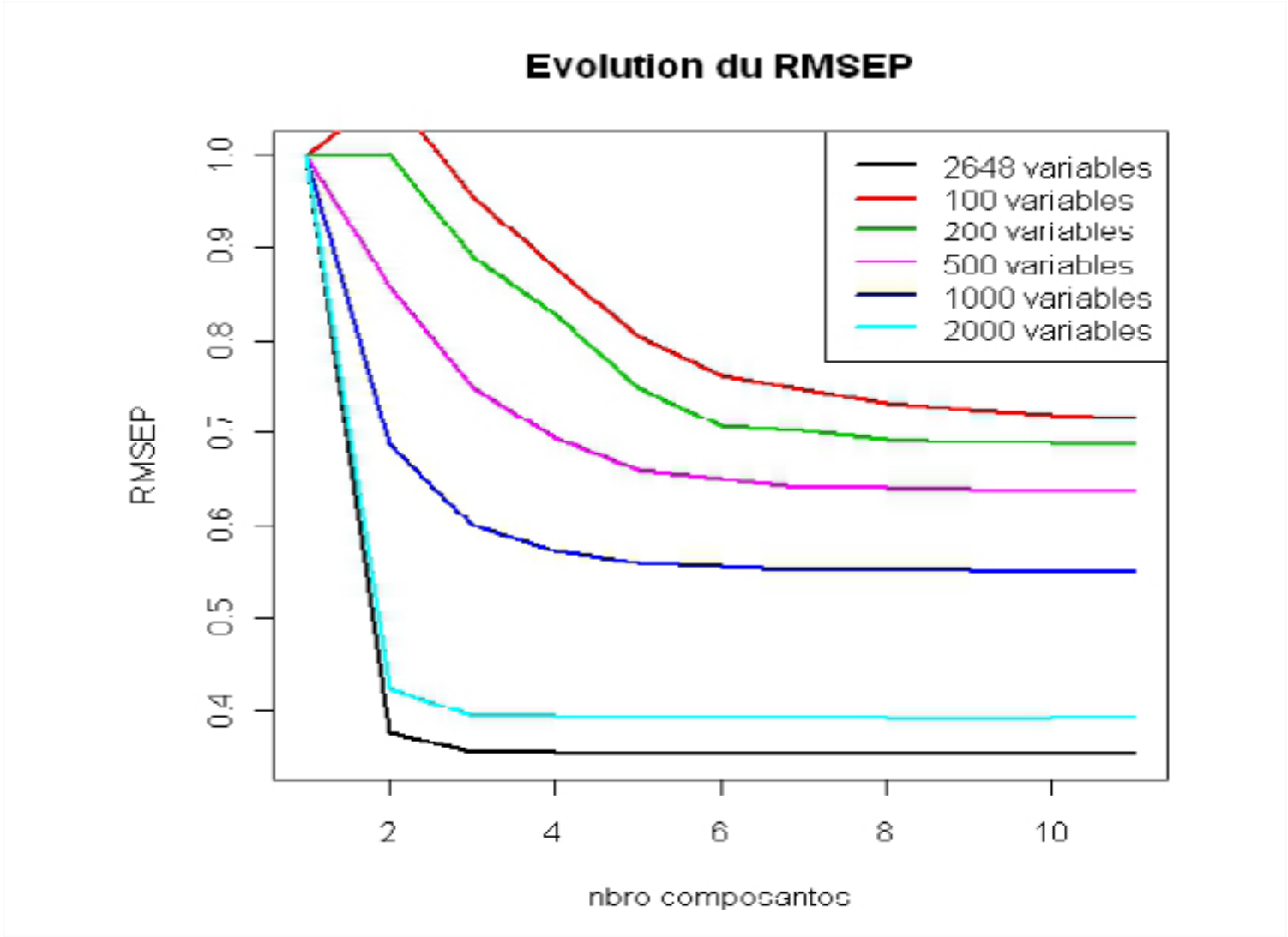
- On initialise avec $\boldsymbol{\alpha}$ solution de la PLS

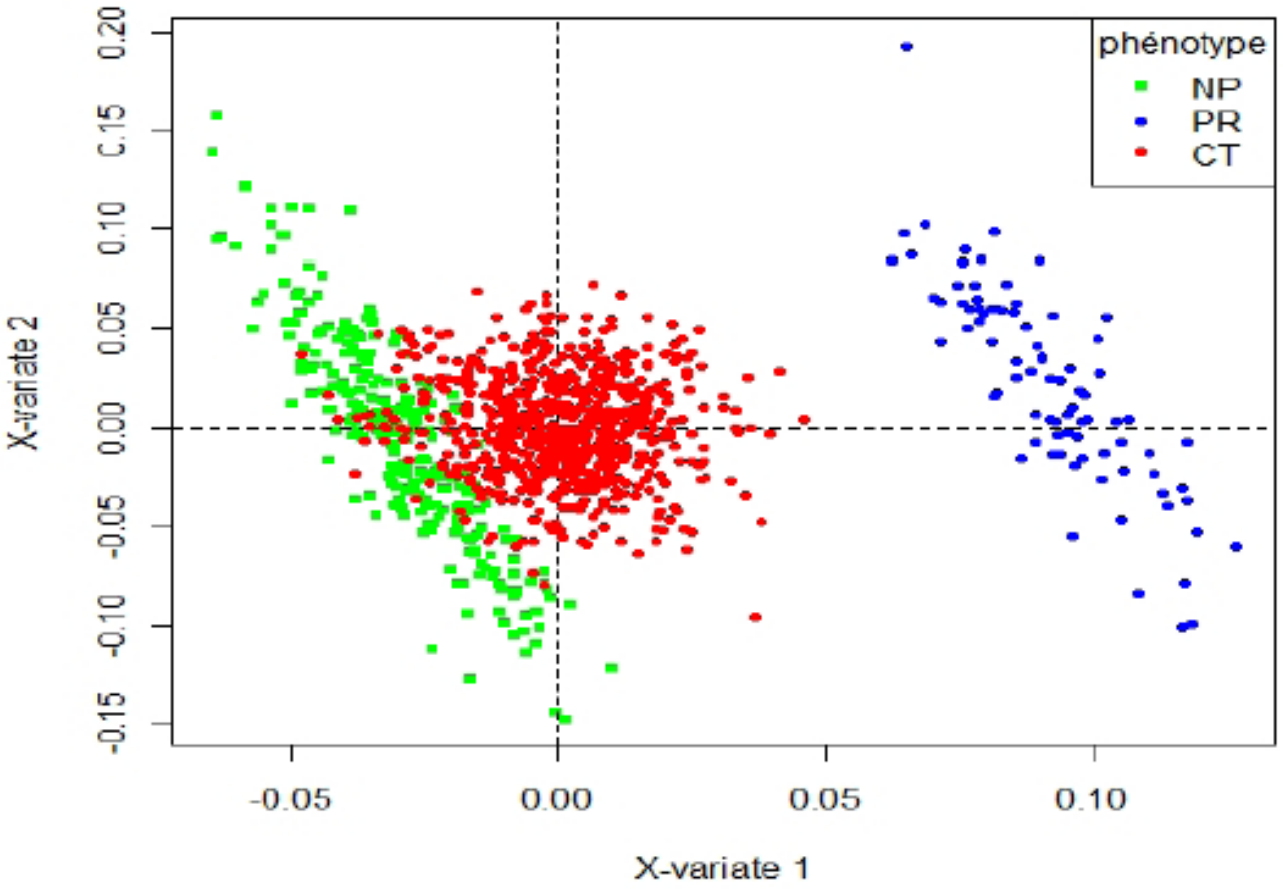
- Pour $\mathbf{y}$ univarié

$$\mathbf{c} = sign(\boldsymbol{\alpha}) \left( \boldsymbol{\alpha} - \frac{\lambda_1}{2} \right)_+$$ seuillage doux du premier facteur PLS $$\boldsymbol{\alpha} = \frac{\mathbf{X'y}}{\left\|\mathbf{X'y}\right\|}$$

# Application: cohorte GRIV du SIDA; Mémoire master C.Wolley 2010

- 305643 variables. 266 individus NP, 84 PR, 697 CT
- Discrimination entre PR et NP
- Prétraitements: p ramené à 2648
- Validation croisée par blocs de 10; 3 composantes S-PLS retenues

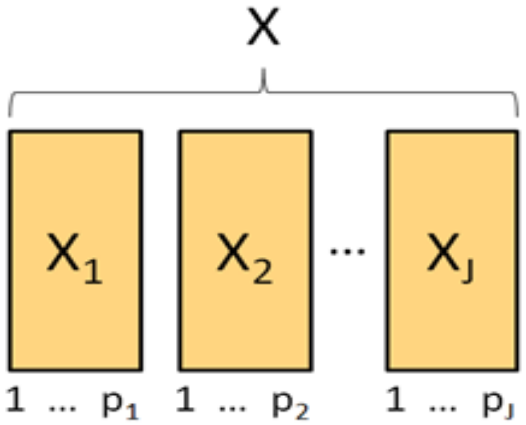

Evolution du RMSEP
RMSEP
0.4
0.5
0.6
0.7
0.8
0.9
1.0
2
4
6
8
10
nbro compocantos
2648 variables
100 variables
200 variables
500 variables
1000 variables
2000 variables


phénotype
NP
PR
CT
X-variate 2
0.20
0.15
0.10
0.05
0.00
-0.05
-0.10
-0.15
-0.05
0.00
0.05
0.10
X-variate 1

Thèse Anne Bernard, 2013

# 6. Group Sparse PCA

Data matrix X divided into J groups $X_j$ of $p_j$ variables, but no Y




**Group Sparse PCA**: compromise between SPCA and group Lasso

**Goal**: select groups of continuous variables (zero coefficients to entire blocks of variables)

**Principle:** replace the penalty function in the SPCA algorithm

$$\hat{\boldsymbol{\beta}} = \arg\min_{\beta} \left\| \mathbf{Z} - \mathbf{X}\boldsymbol{\beta} \right\|^2 + \lambda \left\| \boldsymbol{\beta} \right\|^2 + \lambda_1 \left\| \boldsymbol{\beta} \right\|_1$$

by that defined in the group Lasso

$$\hat{\boldsymbol{\beta}}_{GL} = \arg\min_{\beta} \left\| \mathbf{Z} - \sum_{j=1}^{J} \mathbf{X}_j \boldsymbol{\beta}_j \right\|^2 + \lambda \sum_{j=1}^{J} \sqrt{p_j} \left| \boldsymbol{\beta}_j \right|$$

# 7. Sparse MCA

Original table

| $X_J$ |
|---|
| 1 |
| $p_J$ |
| . |
| . |
| . |
| 3 |

**In MCA:**

Selection of ***1 column*** in the original table (categorical variable $X_j$ )

=

Selection of ***a block of $p_j$ indicator variables*** in the complete disjunctive table

Complete disjunctive table

| $X_{J1}$ | … | $X_{Jpj}$ |
|---|---|---|
| 1 | | 0 |
| 0 | | 1 |
| . | | . |
| . | | . |
| . | | . |
| 0 | | 0 |

**Challenge of Sparse MCA :** select categorical variables, not categories

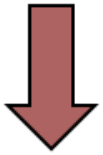
**Principle:** a straightforward extension of Group Sparse PCA for groups of indicator variables, with the chi-square metric .

| Properties | MCA | Sparse MCA |
|---|---|---|
| Uncorrelated Components | TRUE | FALSE |
| Orthogonal loadings | TRUE | FALSE |
| Barycentric property | TRUE | partly TRUE |
| % of inertia | $\lambda_j / tot \times 100$ | $\left\| \tilde{\mathbf{z}}_{\mathbf{j.1,\ldots,j\text{-}1}} \right\|^2$ |
| Total inertia | $\frac{1}{p}\sum_{j=1}^{p} p_j - 1$ | $\sum_{j=1}^{k} \left\| \tilde{\mathbf{z}}_{\mathbf{j.1,\ldots,j\text{-}1}} \right\|^2$ |

$\tilde{\mathbf{z}}_{\mathbf{j.1,\ldots,j\text{-}1}}$ are the residuals after adjusting $\tilde{\mathbf{z}}_{\mathbf{j}}$ for $\tilde{\mathbf{z}}_{\mathbf{1,\ldots,j\text{-}1}}$ (regression projection)

# Toy example: Dogs

| $X_1$ Size | ... | $X_6$ Aggressiveness |
|---|---|---|
| large (L) | | agressive (A) |
| medium (M) | | agressive (A) |
| . | . . . | . |
| . | | . |
| . | | . |
| small (S) | | nonagressive (N) |



| $K_1$ Size | | | ... | $K_6$ Aggressiveness | |
|---|---|---|---|---|---|
| S. | M. | L. | | A | N |
| 0 | 0 | 1 | | 1 | 0 |
| 0 | 1 | 0 | | 1 | 0 |
| . | . | . | . . . | . | . |
| . | . | . | | . | . |
| . | . | . | | . | . |
| 1 | 0 | 0 | | 0 | 1 |

**Data:**

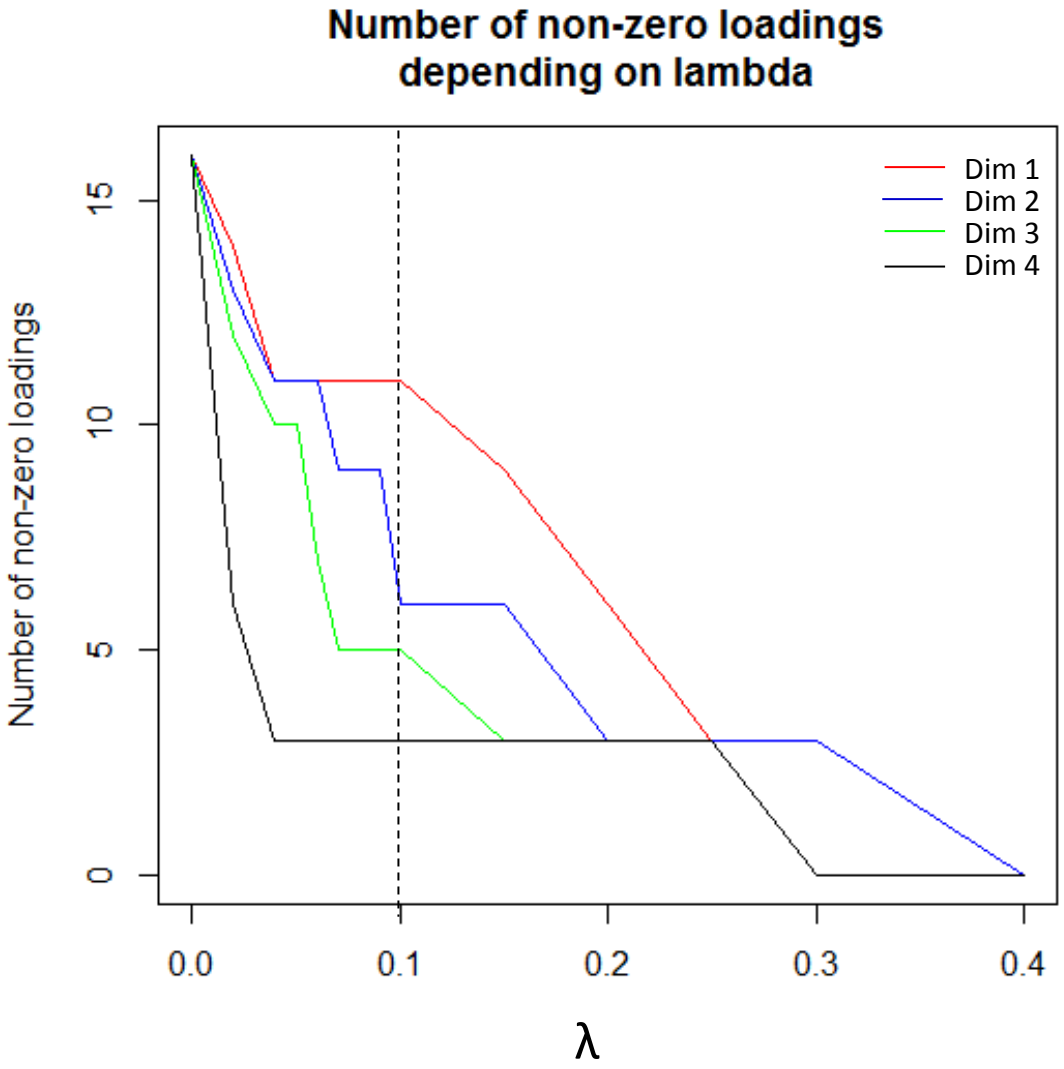
n=27 breeds of dogs
p=6 variables
q=16 (total number of columns)

X : 27 x 6 matrix of categorical variables

K : 27 x 16 complete disjunctive table → $K=(K_1, \ldots, K_6)$

**1 block**
**= 1 $K_j$ matrix**

# Toy example: Dogs




For λ=0.10:

- 11 non-zero loadings on the 1st axis
- 6 non-zero loadings on the 2nd axis
- 5 non-zero loadings on the 3rd axis
- 3 non-zero loadings on the 4th axis

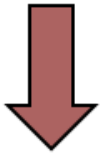
# Toy example: Comparison of the loadings

| SNPs | MCA | | | | Sparse MCA | | | |
|---|---|---|---|---|---|---|---|---|
| | Dim 1 | Dim 2 | Dim 3 | Dim 4 | Dim 1 | Dim 2 | Dim 3 | Dim 4 |
| large | **-0.270** | **0.017** | -0.072 | 0.060 | **-0.399** | **-0.517** | **0.000** | **0.000** |
| medium | **0.222** | **-0.444** | 0.384 | -0.065 | **0.808** | **0.008** | **0.000** | **0.000** |
| small | **0.453** | **0.402** | -0.205 | -0.085 | **-0.331** | **0.610** | **0.000** | **0.000** |
| lightweight | 0.437 | **0.332** | **-0.098** | -0.091 | **0.000** | **0.471** | **0.278** | **0.000** |
| heavy | -0.061 | **-0.265** | **-0.118** | 0.154 | **0.000** | **-0.369** | **0.426** | **0.000** |
| veryheavy | -0.428 | **0.332** | **0.493** | -0.334 | **0.000** | **-0.059** | **-0.860** | **0.000** |
| slow | 0.070 | 0.297 | 0.285 | -0.144 | -0.002 | **0.000** | **0.000** | **0.000** |
| fast | 0.177 | -0.269 | 0.065 | -0.019 | 0.013 | **0.000** | **0.000** | **0.000** |
| veryfast | -0.286 | -0.068 | -0.429 | 0.201 | -0.011 | **0.000** | **0.000** | **0.000** |
| unintelligent | -0.052 | 0.328 | -0.087 | **0.417** | -0.184 | **0.000** | **0.000** | **-0.248** |
| avg intelligent | 0.087 | -0.140 | 0.255 | **0.096** | 0.197 | **0.000** | **0.000** | **-0.488** |
| veryintelligent | -0.118 | -0.134 | -0.437 | **-0.764** | -0.035 | **0.000** | **0.000** | **0.836** |
| unloving | -0.264 | 0.123 | -0.028 | 0.076 | -0.040 | **0.000** | -0.007 | **0.000** |
| veryaffectionate | 0.245 | -0.115 | 0.026 | -0.070 | 0.040 | **0.000** | 0.007 | **0.000** |
| aggressive | -0.113 | 0.079 | 0.053 | -0.034 | **0.000** | **0.000** | **0.000** | **0.000** |
| non-agressive | 0.105 | -0.074 | -0.049 | 0.032 | **0.000** | **0.000** | **0.000** | **0.000** |
| #non-zero loadings | 16 | 16 | 16 | 16 | 11 | 6 | 5 | 3 |
| % inertia | 28.19 | 22.79 | 13.45 | 9.55 | 21.37 | 20.81 | 12.04 | 5.88 |

# Application on genetic data

## Single Nucleotide Polymorphisms

| SNP 1=$X_1$ | ... | SNP 100=$X_{100}$ |
|---|---|---|
| AA | | AB |
| AB | | BB |
| . | . . . | . |
| . | | . |
| . | | . |
| AA | | AA |
| BB | | AA |



| SNP 1=$K_1$ | | | ... | SNP 100=$K_{100}$ | | |
|---|---|---|---|---|---|---|
| AA | AB | BB | | AA | AB | BB |
| 1 | 0 | 0 | | 0 | 1 | 0 |
| 0 | 1 | 0 | | 0 | 0 | 1 |
| . | . | . | . . . | . | . | . |
| . | . | . | | . | . | . |
| . | . | . | | . | . | . |
| 1 | 0 | 0 | | 1 | 0 | 0 |
| 0 | 0 | 1 | | 1 | 0 | 0 |

**Data:**

n=502 individuals

p=100 SNPs (among more than 800 000 of the original data base, 15000 genes)

q=281 (total number of columns)

X : 502 x 100 matrix of qualitative variables

K : 502 x 281 complete disjunctive table → $K=(K_1, \ldots, K_{100})$

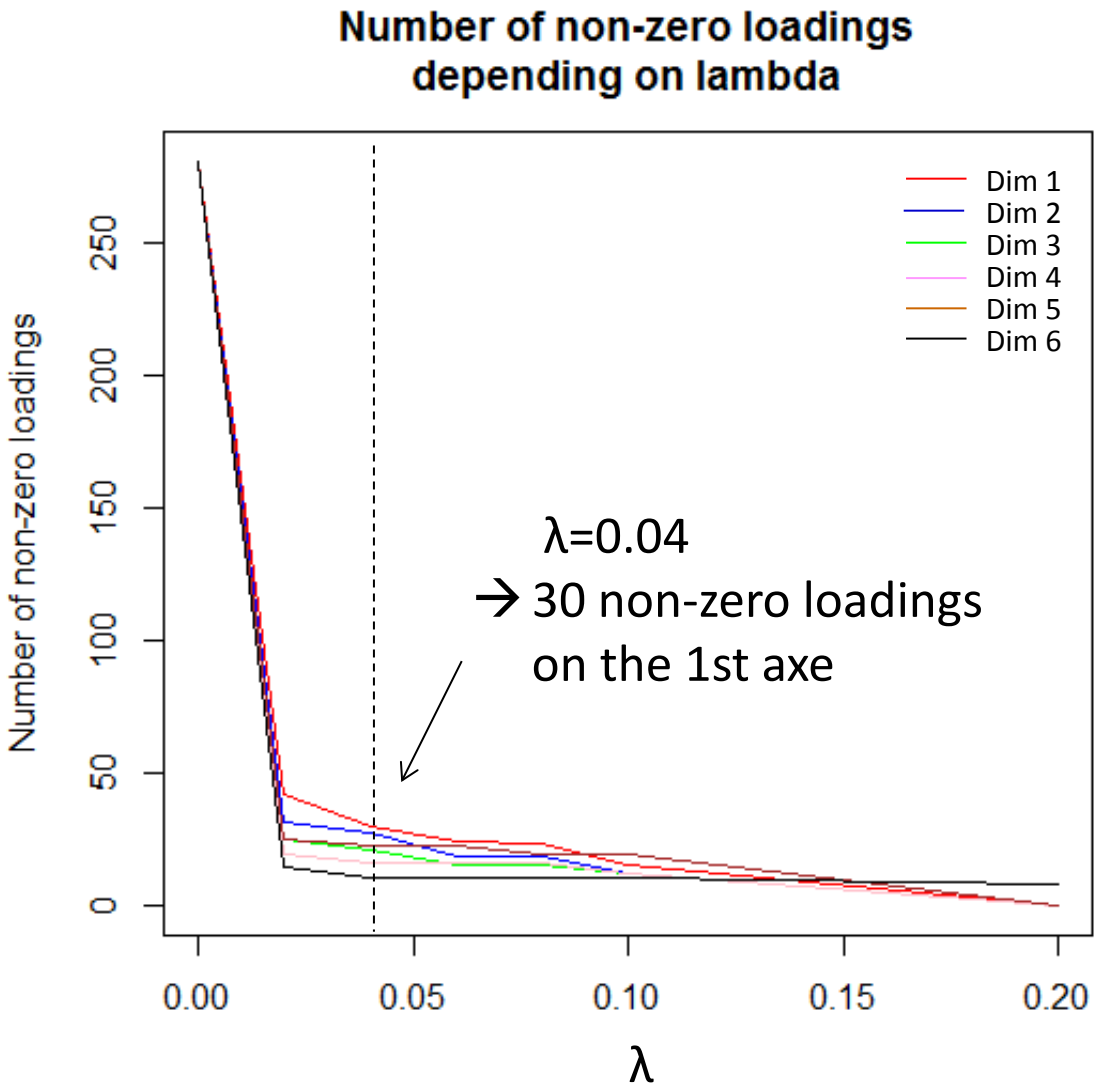
**1 block**

**=**

**1 SNP = 1 $K_j$ matrix**

# Application on genetic data

## Single Nucleotide Polymorphisms



# Application on genetic data

## Comparison of the loadings

| SNPs | MCA | | Sparse MCA | |
|---|---|---|---|---|
| | Dim 1 | Dim 2 | Dim 1 | Dim 2 |
| rs4253711.AA | **-0.323** | -0.043 | **-0.309** | **0.000** |
| rs4253711.AG | **0.009** | 0.016 | **0.057** | **0.000** |
| rs4253711.GG | **0.024** | -0.006 | **0.086** | **0.000** |
| rs4253724.AA | **-0.264** | -0.025 | **-0.424** | **0.000** |
| rs4253724.AT | **0.018** | 0.014 | **0.115** | **0.000** |
| rs4253724.TT | **0.027** | -0.008 | **0.116** | **0.000** |
| rs26722.AG | 0.054 | **-0.421** | **0.000** | **-0.574** |
| rs26722.GG | -0.003 | **0.024** | **0.000** | **0.574** |
| rs35406.AA | -0.002 | **0.024** | **0.000** | **0.241** |
| rs35406.AG | 0.038 | **-0.388** | **0.000** | **-0.241** |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| #non-zero loadings | 281 | 281 | 30 | 24 |
| % inertia | 6.86 | 6.73 | 5.03 | 4.95 |

# Références


- Bernard, A. (2013), *Développement de méthodes statistiques nécessaires à l'analyse de données génomiques*, Thèse de doctorat, CNAM
- Chun, H. and Keles, S. (2010), "Sparse partial least squares for simultaneous dimension reduction and variable selection*", Journal of the Royal Statistical Society - Series B*, Vol. 72, pp. 3–25.
- Friedman, J. H., Hastie, T. and Tibshirani, R. (2010) A note on the group lasso and a sparse group lasso.
- I.T. Jolliffe, N.T. Trendafilov, M. Uddin, A modified principal component technique based on the LASSO, *J. Comput. Graph. Statist*. 12 (2003) 531–547.
- Lê Cao K.-A., Rossouw, D., Robert-Granié C., Besse, P. (2008) A Sparse PLS for Variable Selection when Integrating Omics data *Statistical Applications in Genetics and Molecular Biology:* Vol. **7** : Iss. 1, Article 35.
- Rousson, V. and Gasser, T. (2004), Simple component analysis. *Journal of the Royal Statistical Society*: Series C (Applied Statistics), vol 53,539-555

- Shen, H. and Huang, J. Z. (2008). Sparse principal component analysis via regularized low rank matrix approximation. *Journal of Multivariate Analysis*, 99:1015-1034
- R. Tibshirani, (1996), Regression shrinkage and selection via the lasso, *J. Roy. Statist. Soc. Ser. B* 58 267–288.
- Vines, S.K., (2000) Simple principal components, *Journal of the Royal Statistical Society*: *Series C (Applied Statistics)* Volume 49, 441-451
- Yuan, M. & Lin, Y. (2007), Model selection and estimation in regression with grouped variables, *J. Roy. Statist. Soc. Ser. B ,* 49-67.
- H. Zou, T. Hastie, (2005), Regularization and variable selection via the elastic net, *J. Roy. Statist. Soc. Ser. B 67* 301–320.
- H. Zou, T. Hastie, R. Tibshirani, (2006), Sparse principal component analysis, *J. Comput. Graph. Statist.* 15 265–286.
- H. Zou, T. Hastie, R. Tibshirani, (2007), On the “degrees of freedom” of the lasso, *The Annals of Statistics*, 35, 5, 2173–2192.