

Supervised-Component versus PLS regression

The case of GLMMs with autoregressive random effect

Jocelyn CHAUVET

joint work with **Catherine TROTTIER** and **Xavier BRY**

CASI 2018, Galway,
May 18



A simple Gaussian model

► $y \sim \mathcal{N}_n(\mu = X\beta, \Sigma = \text{Id}_n)$

► $X = \left[\underbrace{x^1 \dots x^{10}}_{\substack{\text{large bundle} \\ \hookrightarrow \text{noise}}} \quad \underbrace{x^{11} \dots x^{15}}_{\substack{\text{small bundle} \\ \hookrightarrow \text{predicts } y}} \quad \underbrace{x^{16} \dots x^{20}}_{\substack{\text{small bundle} \\ \hookrightarrow \text{predicts } y}} \right]$

👉 PLSR vs "Supervised Component Regression"
(a more flexible way to build components)

A simple Gaussian model

- ▶ $y \sim \mathcal{N}_n(\mu = X\beta, \Sigma = \text{Id}_n)$
- ▶ $X = \left[\underbrace{x^1 \dots x^{10}}_{\substack{\text{large bundle} \\ \hookrightarrow \text{noise}}} \underbrace{x^{11} \dots x^{15}}_{\substack{\text{small bundle} \\ \hookrightarrow \text{predicts } y}} \underbrace{x^{16} \dots x^{20}}_{\substack{\text{small bundle} \\ \hookrightarrow \text{predicts } y}} \right]$

👉 PLSR vs "Supervised Component Regression"
(a more flexible way to build components)

- ▶ Supervised Component Regression also available for Bernoulli, binomial and Poisson responses (**R package : SCGLR**)

How to extend it to GLMMs with both individual- and time-specific random effects ?

- 1 Data, motivation and model definition
- 2 A new regularisation framework
- 3 Simulation study
- 4 Conclusions

We consider balanced panel data with:

- ▶ N individuals ...
- ▶ ... observed at the same R time-points

Notations:

- ▶ $\mathbf{y}_{NR \times 1}$: response vector
- ▶ $\mathbf{X}_{NR \times p}$: design matrix of the **many and redundant** explanatory variables

We consider balanced panel data with:

- ▶ N individuals ...
- ▶ ... observed at the same R time-points

Notations:

- ▶ $\mathbf{y}_{NR \times 1}$: response vector
- ▶ $\mathbf{X}_{NR \times p}$: design matrix of the **many and redundant** explanatory variables

Difficulties

- High level of correlation among the explanatory variables
→ **Regularisation** is needed
- Individual-specific and time-specific effects
→ Need to take into account the induced **complex dependence structure**

Example of real data

↪ **Econometrics:** all companies share a common economic climate (latent phenomenon) which tends to persist over time...

Example of real data

↪ **Econometrics**: all companies share a common economic climate (latent phenomenon) which tends to persist over time...

In general: data with

- ▶ a **dependence within individuals** on which data is repeatedly collected
- ▶ a **serially correlated time-specific effect** shared by all the individuals

The method we propose must

- take into account the dependence structure:
 - ↪ Within-individual dependence modelled by a **random effect with independent levels**
 - ↪ Time dependence modelled by a **random effect with AR(1) levels**
 - ↪ **GLMM** in order to deal with non-Gaussian response (e.g. count or binary response)
- handle the high correlations among the explanatory variables
 - ↪ **Ridge-based regularisation**
 - ↪ **Supervised component-based regularisation**

GLMM framework

$$Y_i | \xi \stackrel{\text{iid}}{\sim} F \text{ belonging to the exponential family}$$

$$g(\underbrace{\mathbb{E}(Y | \xi)}_{\mu}) = \eta = X\beta + U_1\xi^1 + U_2\xi^2$$

- ▶ β fixed effect vector
- ▶ $\xi^1 = (\xi_1^1, \xi_2^1, \dots, \xi_N^1)^\top$ the "individual-specific" random effect vector, $U_1 = \text{Id}_N \otimes \mathbf{1}_R$ the associated design matrix
- ▶ $\xi^2 = (\xi_1^2, \xi_2^2, \dots, \xi_R^2)^\top$ the "time-specific" random effect vector, $U_2 = \mathbf{1}_N \otimes \text{Id}_R$ the associated design matrix

Random effects

$$\mathbf{y} = (y_{\mathbf{1}\mathbf{1}}, y_{\mathbf{1}\mathbf{2}}, \dots, y_{\mathbf{1}\mathbf{R}}, \\ y_{\mathbf{2}\mathbf{1}}, y_{\mathbf{2}\mathbf{2}}, \dots, y_{\mathbf{2}\mathbf{R}}, \dots, \\ y_{\mathbf{N}\mathbf{1}}, y_{\mathbf{N}\mathbf{2}}, \dots, y_{\mathbf{N}\mathbf{R}})^{\top}$$

- ▶ $\boldsymbol{\xi}^1 = (\xi_1^1, \xi_2^1, \dots, \xi_N^1)^{\top} \sim \mathcal{N}_N(\mathbf{0}, \mathbf{D}_1), \mathbf{D}_1 = \sigma_1^2 \mathbf{A}_1$
- ▶ $\boldsymbol{\xi}^2 = (\xi_1^2, \xi_2^2, \dots, \xi_R^2)^{\top} \sim \mathcal{N}_R(\mathbf{0}, \mathbf{D}_2), \mathbf{D}_2 = \sigma_2^2 \mathbf{A}_2(\rho),$

$$\mathbf{A}_2(\rho) = \left(\frac{\rho^{|i-j|}}{1 - \rho^2} \right)_{1 \leq i, j \leq R}$$

- ▶ $\boldsymbol{\xi}^1 \perp \boldsymbol{\xi}^2$

1 Data, motivation and model definition

2 A new regularisation framework

- (Ridge-) penalised EM
- Component-based EM
- The particular case of GLMMs

3 Simulation study

4 Conclusions

Principle of penalised EM algorithm



Green, P.J. (1990) *On use of the EM for penalized likelihood estimation.*
Journal of the Royal Statistical Society. Series B (Methodological), 443-452.

$$\begin{aligned} \text{E} : Q_{\text{pen}}(\theta \mid \theta^{[t]}) &:= \mathbb{E}_{\xi|y} \left[\mathcal{L}_{\text{pen}}(\theta; y, \xi) \mid \theta^{[t]} \right] \\ \text{M} : \theta^{[t+1]} &\leftarrow \arg \max_{\theta} Q_{\text{pen}}(\theta \mid \theta^{[t]}) \end{aligned}$$

Principle of penalised EM algorithm



Green, P.J. (1990) *On use of the EM for penalized likelihood estimation.*
Journal of the Royal Statistical Society. Series B (Methodological), 443-452.

$$\begin{aligned} \text{E} : Q_{\text{pen}}(\theta \mid \theta^{[t]}) &:= \mathbb{E}_{\xi \mid y} [\mathcal{L}_{\text{pen}}(\theta; y, \xi) \mid \theta^{[t]}] \\ \text{M} : \theta^{[t+1]} &\leftarrow \arg \max_{\theta} Q_{\text{pen}}(\theta \mid \theta^{[t]}) \end{aligned}$$

Usual penalised complete log-likelihood

$$\mathcal{L}_{\text{pen}}(\theta; y, \xi) = \mathcal{L}(\theta; y, \xi) - \lambda \text{pen}(\beta)$$

$$\text{pen}(\beta) = \begin{cases} \|\beta\|_1 \\ \|\beta\|_2^2 = \beta^\top \beta \\ \alpha \|\beta\|_2^2 + (1 - \alpha) \|\beta\|_1, & 0 \leq \alpha \leq 1 \end{cases}$$

Ridge-based regularisation

↪ EM algorithm, $\theta = (\beta, \sigma_1^2, \sigma_2^2, \rho)$ and $\xi = (\xi^1, \xi^2)$

$$\text{E} : \mathcal{Q}_{\text{ridge}}(\theta, \lambda | \theta^{[t]}) := \mathbb{E}_{\xi|y} \left[\mathcal{L}(\theta; y, \xi) - \lambda \beta^\top \beta | \theta^{[t]} \right]$$

$$\text{M} : \begin{cases} \lambda^{[t+1]} \leftarrow \text{GCV}^{[t+1]}(\lambda) \\ \theta^{[t+1]} \leftarrow \arg \max_{\theta} \mathcal{Q}_{\text{ridge}}(\theta, \lambda^{[t+1]} | \theta^{[t]}) \end{cases}$$



Eliot, M., Ferguson, J., Reilly, M.P. and Foulkes, A.S. (2011) *Ridge Regression for Longitudinal Biomarker Data*. The International Journal of Biostatistics, **7**, 1–11.

Component-based regularisation

↪ New linear predictors

$Y_i | \xi \stackrel{\text{iid}}{\sim} F$ belongs to the exponential family

$$g(\mathbb{E}(Y | \xi)) = \eta = \cancel{X\beta} + U_1\xi^1 + U_2\xi^2$$

replaced with

$$\eta = (Xu)\gamma + U_1\xi^1 + U_2\xi^2 \quad \text{for a single component}$$

extended to

$$\eta = \sum_{k=1}^{K^*} (Xu_k)\gamma_k + U_1\xi^1 + U_2\xi^2 \quad \text{for } K^* \text{ components}$$

Complete log-likelihood for supervised component regularisation

With $\theta = (\mathbf{u}, \gamma, \sigma_1^2, \sigma_2^2, \rho)$ and a trade-off parameter $s \in [0, 1]$

$$\mathcal{L}_{\text{sc}}(\theta; \mathbf{y}, \xi) = (1 - s) \mathcal{L}(\theta; \mathbf{y}, \xi) + s \phi(\mathbf{u})$$

- **Log-likelihood** : measures (inter alia) the probability that observations \mathbf{y} have been generated from component $\mathbf{f} = \mathbf{X}\mathbf{u}$
- **Structural relevance criterion** : measures the closeness of component \mathbf{f} to the strongest structures of \mathbf{X}

A few words about the structural relevance criterion

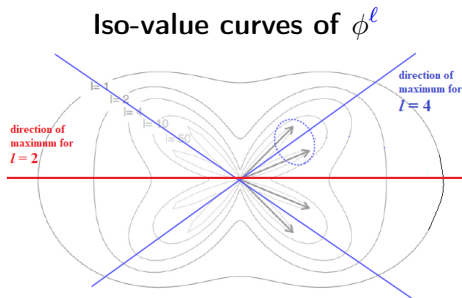
How many bundles do you see ?

$$\phi(\mathbf{u}) = \left(\sum_{j=1}^p \left[\text{cor}^2(\mathbf{X}\mathbf{u}, \mathbf{x}^j) \right]^{\ell} \right)^{\frac{1}{\ell}}$$



A few words about the structural relevance criterion

$$\phi(\mathbf{u}) = \left(\sum_{j=1}^p \left[\text{cor}^2(\mathbf{X}\mathbf{u}, \mathbf{x}^j) \right]^\ell \right)^{\frac{1}{\ell}}$$



Ridge-based penalisation

$$\mathcal{L}(\theta; y, \xi) - \lambda \beta^\top \beta$$

- ▶ **Penalises** the "large" coefficients
- ▶ Sees the high correlations among the explanatory variables as **pure nuisance**
- ▶ η hard to interpret

Component-based regularisation

$$\mathcal{L}(\theta; y, \xi) + \frac{s}{1-s} \phi(u)$$

- ▶ **Gives a bonus** to the most interpretable bundles in X
- ▶ **Takes advantage** of the high correlations among the explanatory variables
- ▶ η **easier to interpret** through decomposition on components

↪ Focus on GLMMs

Starting with the classical Fisher Scoring Algorithm for GLMs, we perform:

LINEARISATION step

- ▶ Linearisation of \mathbf{y}_i at $\boldsymbol{\mu}_i = \mathbb{E}(\mathbf{Y}_i | \boldsymbol{\xi})$:

$$\mathbf{y}_i \simeq \mathbf{z}_i = g(\boldsymbol{\mu}_i) + (\mathbf{y}_i - \boldsymbol{\mu}_i)g'(\boldsymbol{\mu}_i)$$

$$\mathbf{z}_i = \boldsymbol{\eta}_i + \mathbf{e}_i$$

- ▶ Linearised model:

$$\mathcal{M} : \mathbf{z} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}_1\boldsymbol{\xi}^1 + \mathbf{U}_2\boldsymbol{\xi}^2 + \mathbf{e}, \quad \text{with } \mathbb{V}(\mathbf{e}) = \boldsymbol{\Gamma}$$

ESTIMATION step

Penalised/Regularised EM algorithm on \mathcal{M}

Ridge-based penalisation for GLMM-AR(1)

$$\theta = (\beta, \sigma_1^2, \sigma_2^2, \rho)$$

Linearised model

$$\mathcal{M}^{[t]} : z^{[t]} = X\beta + U_1\xi^1 + U_2\xi^2 + e^{[t]}, \quad \text{with } \mathbb{V}(e^{[t]}) = \Gamma^{[t]}$$

Ridge estimation

$$\mathbf{E} : \mathcal{Q}_{\text{ridge}}(\theta, \lambda \mid \theta^{[t]}) := \mathbb{E}_{\xi \mid z^{[t]}} \left[\mathcal{L}(\theta; z^{[t]}, \xi) - \lambda \beta^\top \beta \mid \theta^{[t]} \right]$$

$$\mathbf{M} : \begin{cases} \lambda^{[t+1]} \leftarrow \text{GCV}^{[t+1]}(\lambda) \\ \theta^{[t+1]} \leftarrow \arg \max_{\theta} \mathcal{Q}_{\text{ridge}}(\theta, \lambda^{[t+1]} \mid \theta^{[t]}) \end{cases}$$

Update

Calculate $\xi^{[t+1]}$, $z^{[t+1]}$, $\Gamma^{[t+1]}$ with the updated $\theta^{[t+1]}$

Supervised component-based regularisation for GLMM-AR(1)

$$\theta = (\mathbf{u}, \gamma, \sigma_1^2, \sigma_2^2, \rho)$$

Linearised model

$$\mathcal{M}^{[t]} : \mathbf{z}^{[t]} = (\mathbf{X}\mathbf{u})\gamma + U_1\xi^1 + U_2\xi^2 + e^{[t]}, \quad \text{with } \mathbb{V}(e^{[t]}) = \Gamma^{[t]}$$

SC- estimation

$$\mathbf{E} : \mathcal{Q}_{\text{SC}}(\theta, \theta^{[t]}) := \mathbb{E}_{\xi|\mathbf{z}^{[t]}} \left[(1-s)\mathcal{L}(\theta; \mathbf{z}^{[t]}, \xi) + s\phi(\mathbf{u}) \mid \theta^{[t]} \right]$$

$$\mathbf{M} : \begin{cases} \sigma_1^{2[t+1]}, \sigma_2^{2[t+1]}, \rho^{[t+1]} \text{ computed as previously} \\ \mathbf{u}^{[t+1]} \leftarrow \arg \max_{\mathbf{u}: \|\mathbf{u}\|=1} \mathcal{Q}_{\text{SC}}(\mathbf{u}, \gamma^{[t]} \mid \theta^{[t]}) \\ \gamma^{[t+1]} \leftarrow \arg \max_{\gamma} \mathcal{Q}_{\text{SC}}(\mathbf{u}^{[t+1]}, \gamma \mid \theta^{[t]}) \end{cases}$$

Update: Calculate $\xi^{[t+1]}$, $\mathbf{z}^{[t+1]}$, $\Gamma^{[t+1]}$ with the updated $\theta^{[t+1]}$

- 1 Data, motivation and model definition
- 2 A new regularisation framework
- 3 Simulation study
- 4 Conclusions

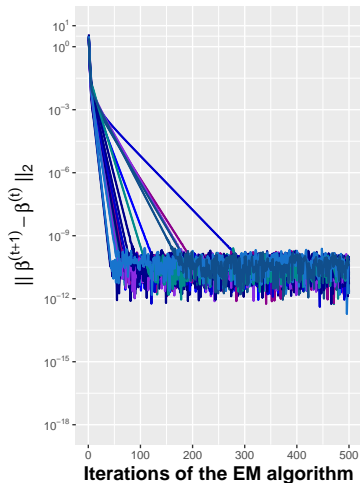
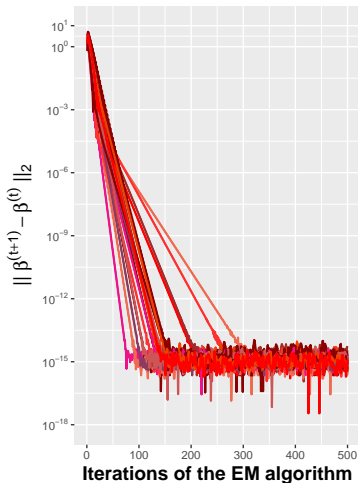
Poisson regression with log link

► $y \sim \mathcal{P}(\lambda = \exp(X\beta + U_1\xi^1 + U_2\xi^2))$

► $X = \left[\underbrace{x^1 \dots x^{10}}_{\substack{\text{large bundle} \\ \hookrightarrow \text{noise}}} \underbrace{x^{11} \dots x^{15}}_{\substack{\text{small bundle} \\ \hookrightarrow \text{predicts } y}} \underbrace{x^{16} \dots x^{20}}_{\substack{\text{small bundle} \\ \hookrightarrow \text{predicts } y}} \right]$

Convergence results

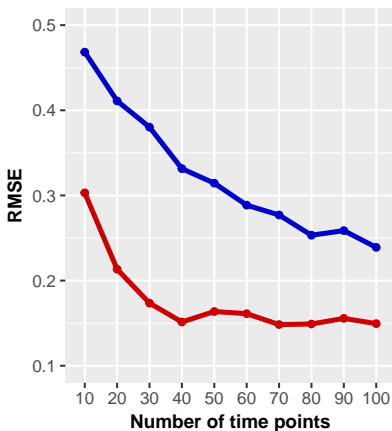
How does convergence go?

Ridge
penaltySupervised Component
regularisation

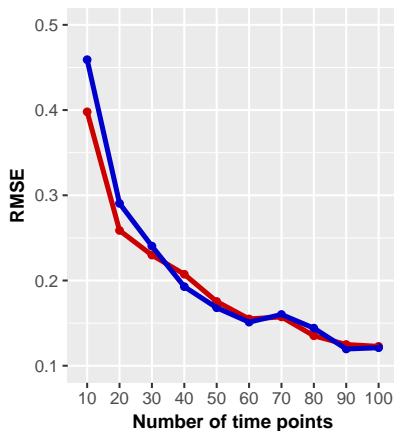
How good are the estimations ?

Accuracy of the estimates

Fixed effects parameter β



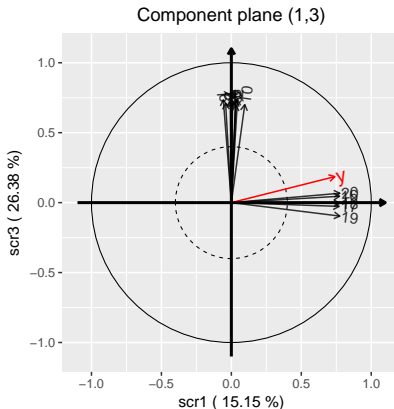
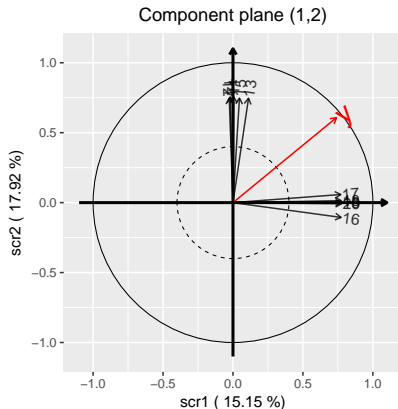
Autocorrelation ρ

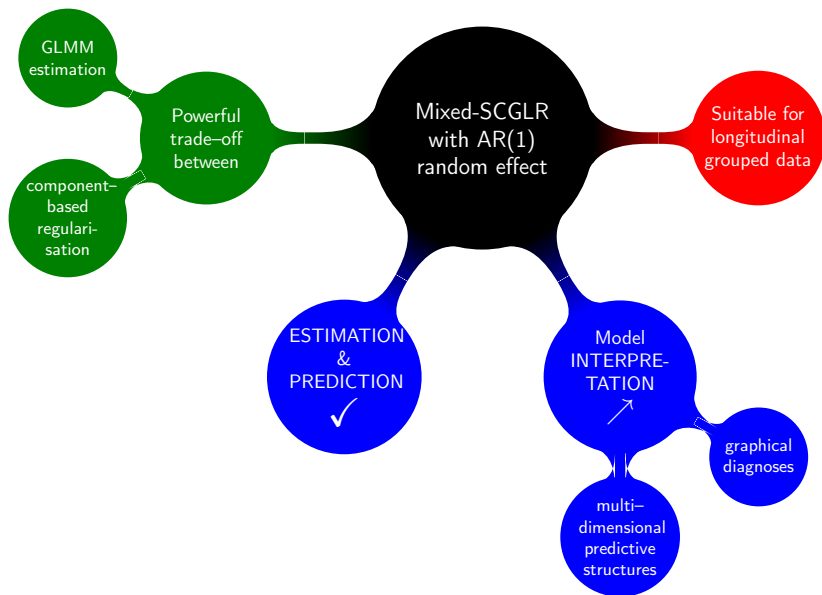


—●— SC —●— ridge

Does the use of SC—regularisation facilitate the model interpretation ?

Power for model interpretation







Bry, X., Trottier, C., Verron, T. and Mortier, F. (2013) *Supervised component generalized linear regression using a PLS-extension of the Fisher scoring algorithm*. Journal of Multivariate Analysis, **119**, 47–60.

+ **Package R : SCGLR**



Chauvet, J., Bry, X., Trottier, C. and Mortier, F. (2016) *Extension to mixed models of the Supervised Component-based Generalised Linear Regression*. In COMPSTAT: Proceedings in Computational Statistics.



Eliot, M., Ferguson, J., Reilly, M.P. and Foulkes, A.S. (2011) *Ridge Regression for Longitudinal Biomarker Data*. The International Journal of Biostatistics, **7**, 1–11.



Green, P.J. (1990) *On use of the EM for penalized likelihood estimation*. Journal of the Royal Statistical Society. Series B (Methodological), 443-452.



Marx, B. D. (1996) *Iteratively reweighted partial least squares estimation for generalized linear regression*. Technometrics, **38**, 4, 374–381.