

Regularising Generalised Linear Mixed Models with an autoregressive random effect

Jocelyn CHAUVET

joint work with Catherine TROTTIER and Xavier BRY

IWSM 2017, Groningen,
July 7



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

Random effects

$$y = (y_{11}, y_{12}, \dots, y_{1R}, \\
 y_{21}, y_{22}, \dots, y_{2R}, \dots \dots \\
 y_{N1}, y_{N2}, \dots, y_{NR})^T$$

- ▶ $\xi^1 = (\xi_1^1, \xi_2^1, \dots, \xi_N^1)^T \sim \mathcal{N}_N(0, \sigma_1^2 \text{Id}_N)$
- ▶ $\xi^2 = (\xi_1^2, \xi_2^2, \dots, \xi_R^2)^T \sim \mathcal{N}_R(0, \Sigma), \Sigma_{ij} = \sigma_2^2 \frac{\rho^{|i-j|}}{1 - \rho^2}$
- ▶ $\xi^1 \perp \xi^2$

- 1 Data, motivation and model definition
- 2 A new regularisation framework
 - Ridge-penalised EM algorithm for LMMs
 - Component-based regularisation of LMMs
 - Extension to GLMMs
- 3 Simulation study
- 4 Conclusions

↪ Focus on LMMs

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.

$$E : Q_{\text{pen}}(\theta | \theta^{[t]}) := \mathbb{E}_{\xi|y} [L_{\text{pen}}(\theta; y, \xi) | \theta^{[t]}]$$

$$M : \theta^{[t+1]} \leftarrow \arg \max_{\theta} Q_{\text{pen}}(\theta | \theta^{[t]})$$

↪ Focus on LMMs

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.

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

Usual penalised complete log-likelihood

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

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

↪ **Focus on LMMs**

Ridge-based regularisation

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

$$\text{GCV} : \lambda^{[t]} \leftarrow \arg \min_{\lambda} \left\{ \text{GCV}(\lambda) = \frac{n^{-1} \left\| y - H_{\lambda}^{[t]} y \right\|^2}{\left[1 - n^{-1} \text{tr} \left(H_{\lambda}^{[t]} \right) \right]^2} \right\}$$

$$\text{E} : Q_{\text{pen}} \left(\theta \mid \theta^{[t]} \right) := \mathbb{E}_{\xi|y} \left[L(\theta; y, \xi) - \lambda^{[t]} \beta^T \beta \mid \theta^{[t]} \right]$$

$$\text{M} : \theta^{[t+1]} \leftarrow \arg \max_{\theta} Q_{\text{pen}} \left(\theta \mid \theta^{[t]} \right)$$



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

$$Y_i | \xi \stackrel{\text{iid}}{\sim} F \text{ belongs to the exponential family}$$
$$g(\mathbb{E}(Y | \xi)) = \eta = \cancel{\mathbf{X}\beta} + U_1\xi^1 + U_2\xi^2$$

replaced with

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

extended to

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

↪ Focus on LMMs

Regularised complete log-likelihood

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

$$L_{\text{reg}}(\theta; y, \xi) = (1 - s) L(\theta; y, \xi) + s \phi(u)$$

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

A few words about $\phi(u)$

- ▶ General formula :

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

- ▶ Particular case :

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

How many bundles do you see ?



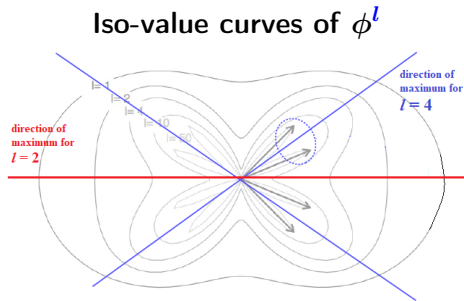
A few words about $\phi(u)$

- ▶ General formula :

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

- ▶ Particular case :

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



Ridge-based penalisation

$$L_{\text{pen}}(\theta; y, \xi) = \\ L(\theta; y, \xi) - \lambda \beta^T \beta$$

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

Component-based regularisation

$$\frac{s}{1-s} L_{\text{reg}}(\theta; y, \xi) = \\ L(\theta; y, \xi) + \frac{s}{1-s} \phi(\mathbf{u})$$

- ▶ Obj.: **give 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

↪ Extension to GLMMs

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

LINEARISATION step

- ▶ Linearisation of y_i at μ_i :

$$y_i \simeq z_i = g(\mu_i) + (y_i - \mu_i)g'(\mu_i)$$
$$z_i = \eta_i + e_i$$

- ▶ Linearised model:

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

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

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

Ridge estimation

$$\text{GCV} : \lambda^{[t]} \leftarrow \arg \min_{\lambda} \text{GCV}(\lambda)$$

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

$$\mathbf{M} : \theta^{[t+1]} \leftarrow \arg \max_{\theta} \mathcal{Q}_{\text{pen}}(\theta | \theta^{[t]})$$

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)

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

Linearised model

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

SC- estimation

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

$$\mathbf{M} : \blacktriangleright \text{ Given } (\theta \setminus u)^{[t]} = (\gamma^{[t]}, \sigma_1^{2[t]}, \sigma_2^{2[t]}, \rho^{[t]}),$$

$$u^{[t+1]} \longleftarrow \arg \max_{u: \|u\|=1} \mathcal{Q}_{\text{reg}}(u \mid \theta^{[t]})$$

\blacktriangleright Given $u^{[t+1]}$,

$$(\theta \setminus u)^{[t+1]} \longleftarrow \arg \max_{\theta \setminus u} \mathbb{E}_{\xi|z} \left[L(\theta \setminus u; z^{[t]}, \xi) \mid \theta^{[t]} \right]$$

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

- 1 Data, motivation and model definition
- 2 A new regularisation framework
- 3 **Simulation study**
 - Data simulation
 - Parameters' calibration
 - Convergence results
 - Comparison between Ridge- and SC-based regularisations
- 4 Conclusions

Poisson regression with log link

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

▶ $X = [\underbrace{x^1 \dots x^{10}}_{\substack{\text{large bundle} \\ \hookrightarrow \text{noise}}} \quad \underbrace{x^{11} \dots x^{14}}_{\substack{\text{indep. var.} \\ \hookrightarrow \text{noise}}} \quad \underbrace{x^{15} \dots x^{19}}_{\substack{\text{small} \\ \text{bundle}}} \quad \underbrace{x^{20}}_{\substack{\text{very small} \\ \text{bundle}}}]$

\hookrightarrow **predict y**

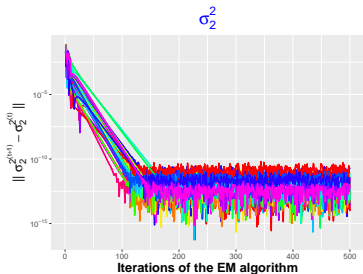
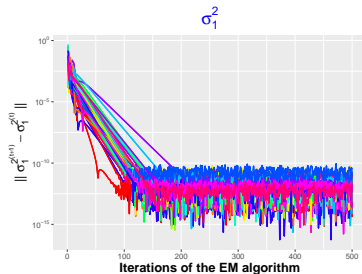
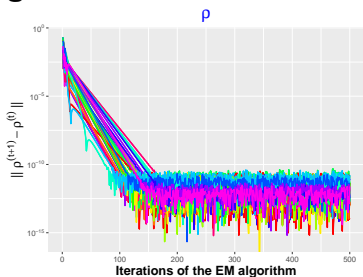
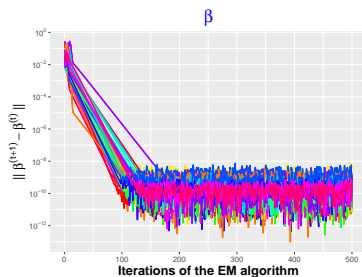
Tuning parameter for Ridge regularisation

- ▶ shrinkage parameter λ
(auto-calibration : GCV at each step of the EM algorithm)

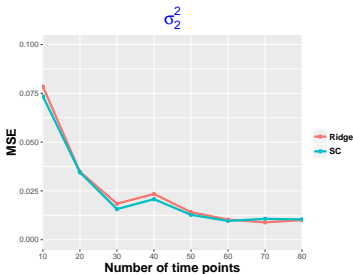
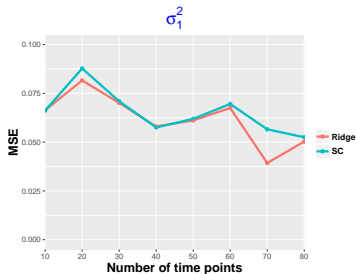
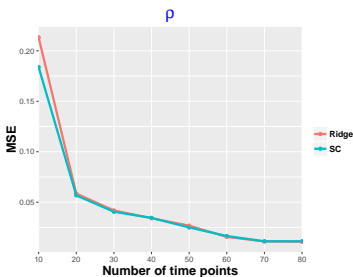
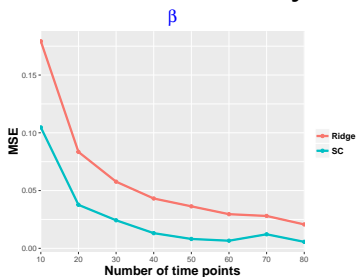
Tuning parameters for SC regularisation

- ▶ trade-off parameter s
 - ▶ bundle locality parameter l
 - ▶ number of components K^*
- } standard cross validation

How does convergence look like?

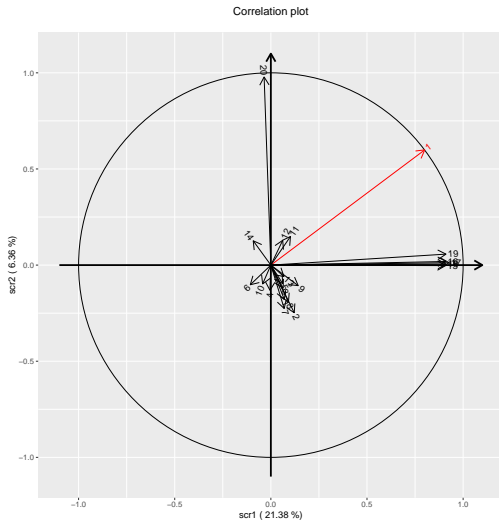


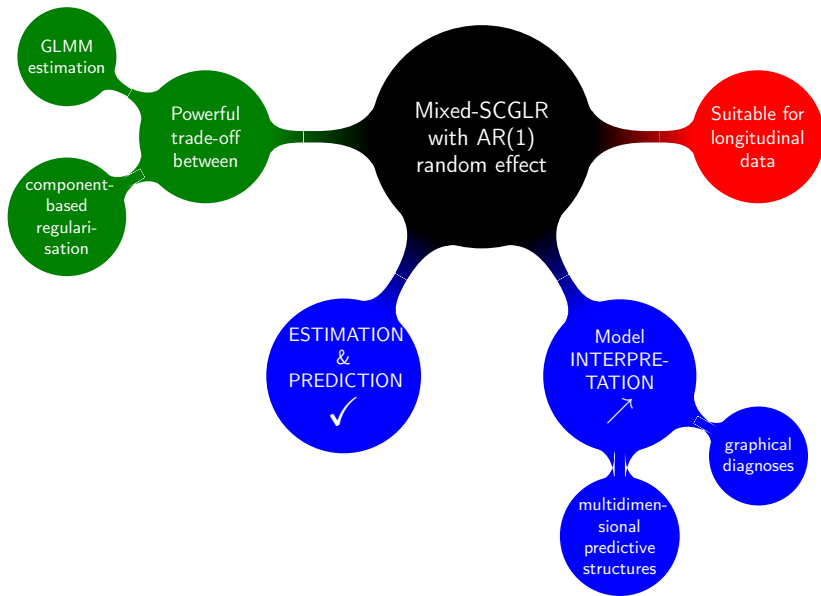
Quality of the estimates








Power for model interpretation

Component plane (1,2)





Bibliography

-  **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.