MAXIMUM LIKELIHOOD ESTIMATION IN NONLINEAR MIXED-EFFECTS MODELS

Marc LAVIELLE & Estelle KUHN
University Paris-Sud

http://www.math.u-psud.fr/~lavielle
Marc.Lavielle@math.u-psud.fr

A first simple model

Observations

Instants of observations

$$y_{ij} = f(\varphi_i, t_{ij}) + \varepsilon_{ij}$$

Individual parameters (unknown)

Random fluctuations

A first simple model

$$N(0, \sigma_{\varepsilon}^{2})$$

$$y_{ij} = f(\varphi_i, t_{ij}) + \varepsilon_{ij}$$

$$\varphi_i = \mu + b_i$$

Fixed effects

Random effects

$$N(0, \Gamma)$$

A first simple model

Fit the model

$$y_{ij} = f(\varphi_i, t_{ij}) + \varepsilon_{ij}$$

C

Estimate the parameters

$$\theta = (\mu, \Gamma, \sigma_\varepsilon^2)$$

The Maximum Likelihood Estimate

$\hat{\theta}^{MLE}$ maximizes the likelihood of the observations $g(y;\theta)$

$\hat{\theta}^{MLE}$ possesses very good statistical properties :

- converges to θ^* when the number n of observations goes to infinity
- is asymptotically normally distributed
- is asymptotically efficient

Without any prior knowledge on θ^* ,
the MLE is usually the "best estimate" of θ^*

The Maximum Likelihood Estimate

$\hat{\theta}^{MLE}$ maximizes the likelihood of the observations $g(y;\theta)$

$$g(y;\theta) = \prod_{i=1}^{n} g(y_i;\theta) = \prod_{i=1}^{n} \int p(\varphi_i, y_i;\theta)\, d\varphi_i$$

$$g(y_i;\theta) = (2\pi\sigma^2)^{-\frac{n_i}{2}} (2\pi|\Gamma|)^{-\frac{1}{2}} \int \exp\left\{ -\frac{1}{2\sigma^2} \|y_i - f(\varphi_i)\|^2 - \frac{1}{2}(\varphi_i - \mu)^t \Gamma^{-1} (\varphi_i - \mu) \right\} d\varphi_i$$

The MLE of θ cannot be computed in a closed-form when f is a nonlinear function of φ

Approximation of the likelihood

- Approximation by the likelihood of a linear mixed-effects model
  first-order Taylor expansion of the model function *f*

  (Sheiner & Beal, 1980 ; Lindstrom & Bates, 1990)

- Laplacian approximation of the likelihood function

  (Pinheiro & Bates, 1995)

- Adaptative Gaussian quadrature approximation

  (Pinheiro & Bates, 1995)

- Simulated Pseudo-Maximum Likelihood (SPML)

  (Concordet & Nuñez, 2002)

The EM algorithm

Expectation-step :

compute $$Q_k(\theta) = E\left\{\log(p(y,\varphi;\theta) \mid y;\theta^{(k)})\right\}$$

Maximization-step :

compute $$\theta^{(k+1)} = Arg\max_{\theta} Q_k(\theta)$$

Under very general conditions,
the EM algorithm converges to
a (local or global) maximum
of the likelihood $g(y;\theta)$

The SAEM algorithm
(Stochastic Approximation version of EM)

Simulation-step :

draw $\varphi^{(k)}$ with the conditional distribution $h(\varphi \mid y;\theta^{(k)})$

(or perform some iterations of MCMC, using the transition kernel $\Pi\left(\varphi^{(k-1)},\bullet;\theta^{(k)}\right)$)

Approximation-step :

compute $Q_k(\theta) = Q_{k-1}(\theta) + \gamma_k\left(\log(p(y,\varphi^{(k)};\theta) - Q_{k-1}(\theta)\right)$

(where $0 < \gamma_k \leq 1$ and $\gamma_k \xrightarrow[k\to\infty]{} 0$)

Maximization-step :

compute $\theta^{(k+1)} = Arg\max_{\theta} Q_k(\theta)$

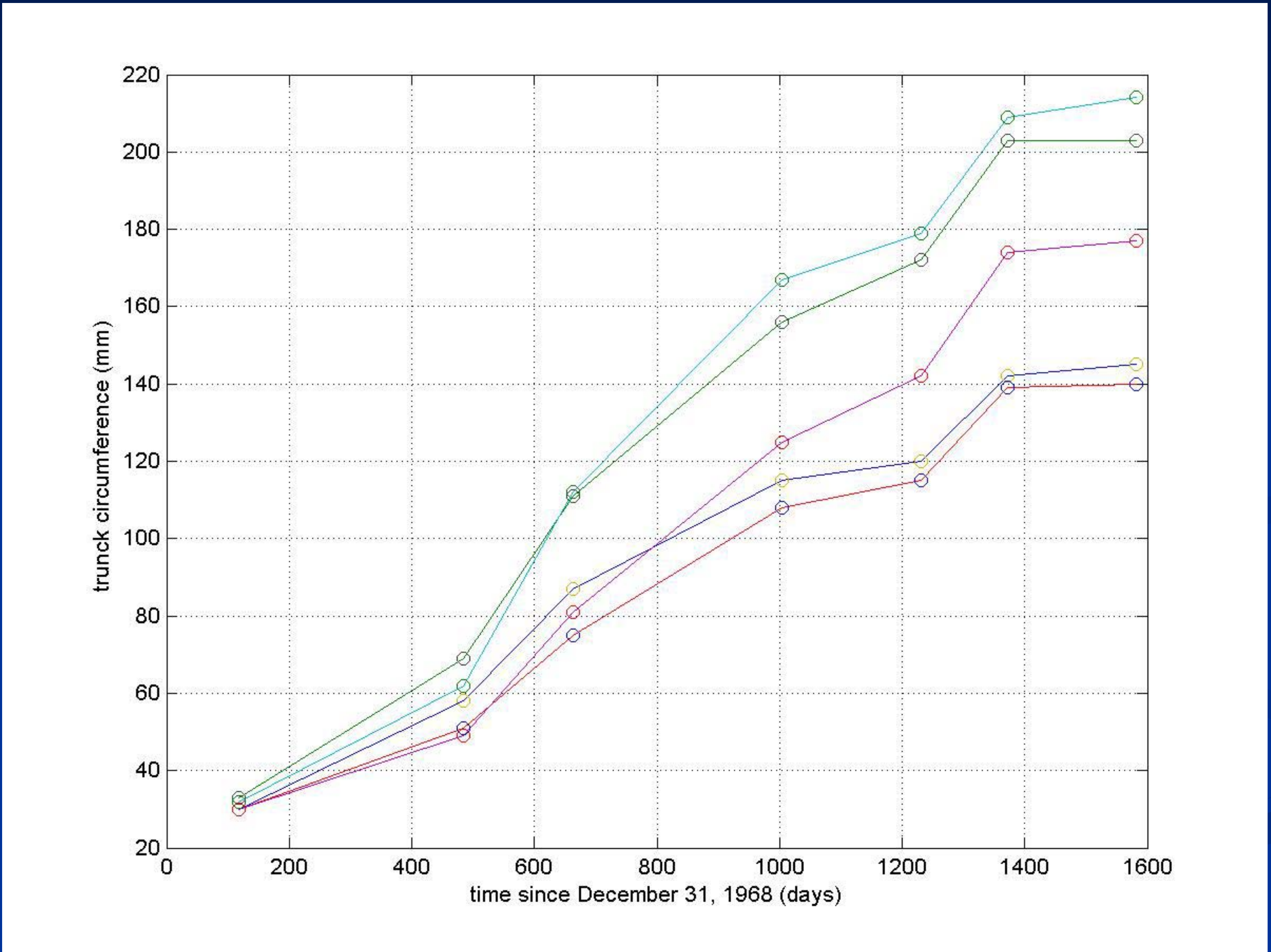
The SAEM algorithm
(Stochastic Approximation version of EM)

Under very general conditions, the SAEM algorithm converges to a (local or global) maximum of the likelihood $g(y;\theta)$

1. Delyon B., Lavielle M., Moulines E. *"Convergence of a stochastic approximation version of the EM algorithm"*, The Annals of Statistics, 1999.

2. Kuhn E., Lavielle M., *"Coupling a stochastic version of the EM algorithm with a MCMC procedure"*, 2003.

Circumference of five orange trees



Circumference of five orange trees

The model:

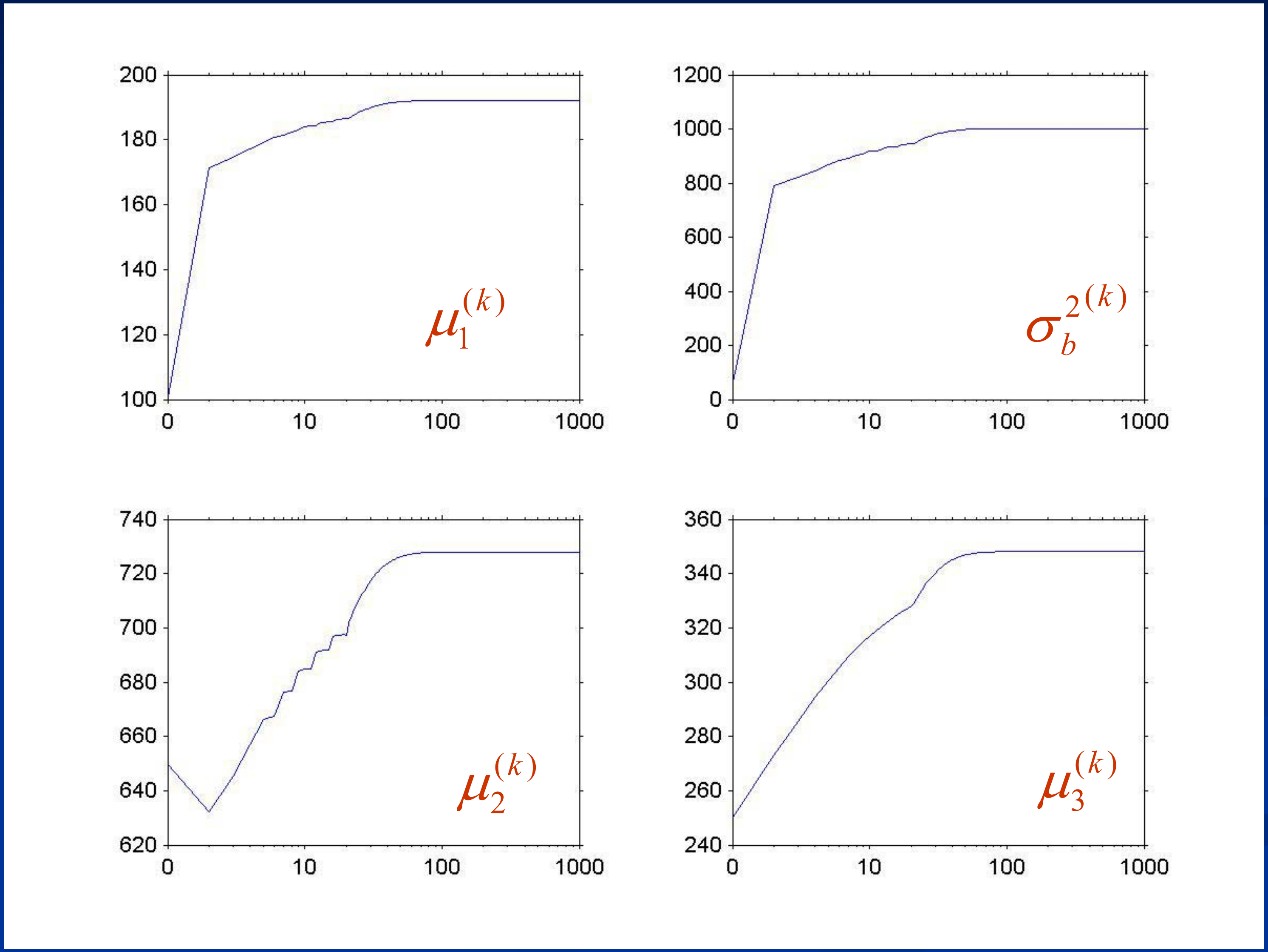
$$y_{ij} = \frac{\mu_1 + b_i}{1 + \exp\left(-\frac{t_j - \mu_2}{\mu_3}\right)} + \varepsilon_{ij}$$

3 fixed effects: (μ_1, μ_2, μ_3)

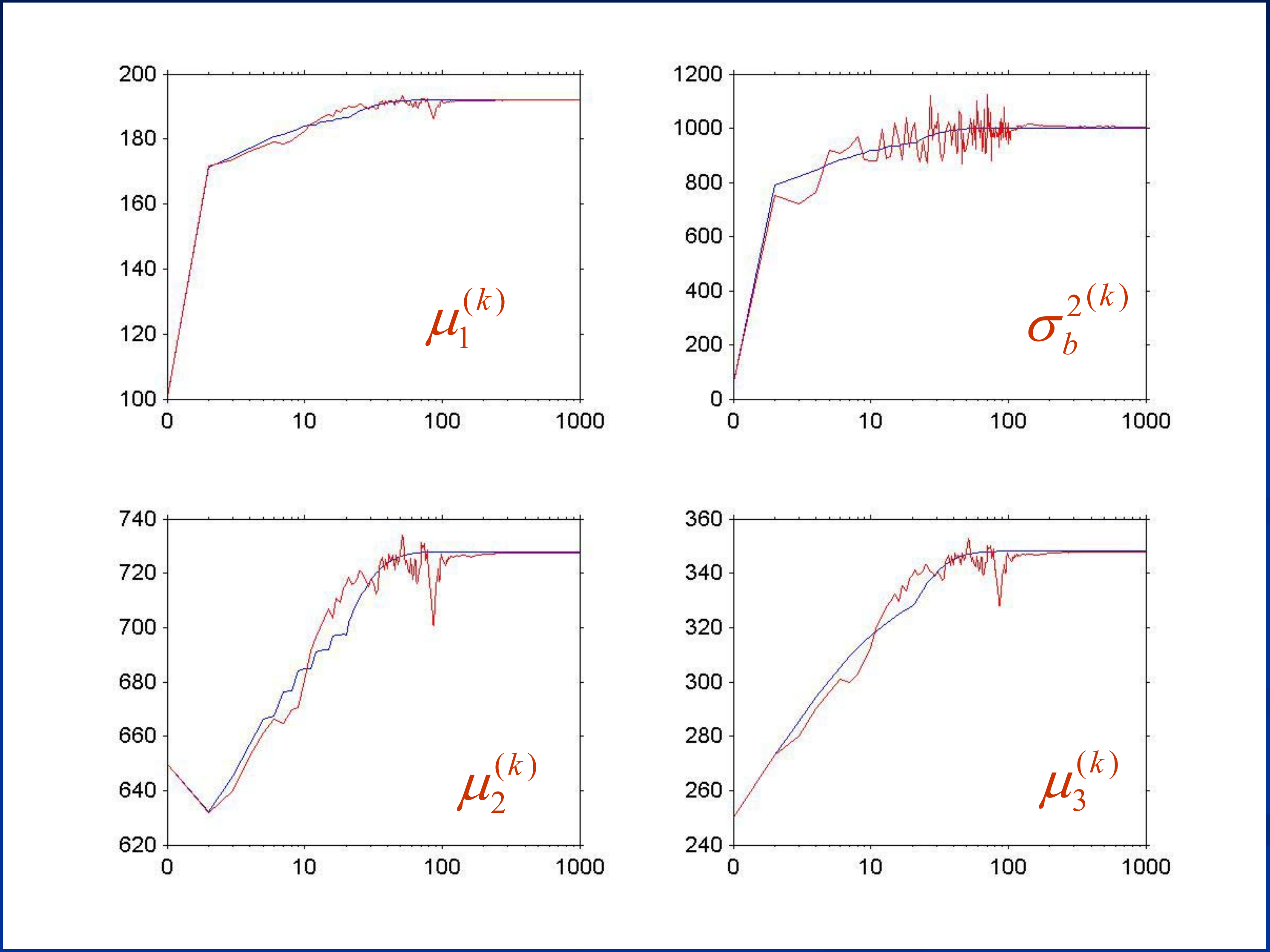
1 random effect: $b_i \sim N\left(0, \sigma_b^2\right)$

$$\theta = \left(\mu_1, \mu_2, \mu_3, \sigma_b^2, \sigma_\varepsilon^2\right)$$

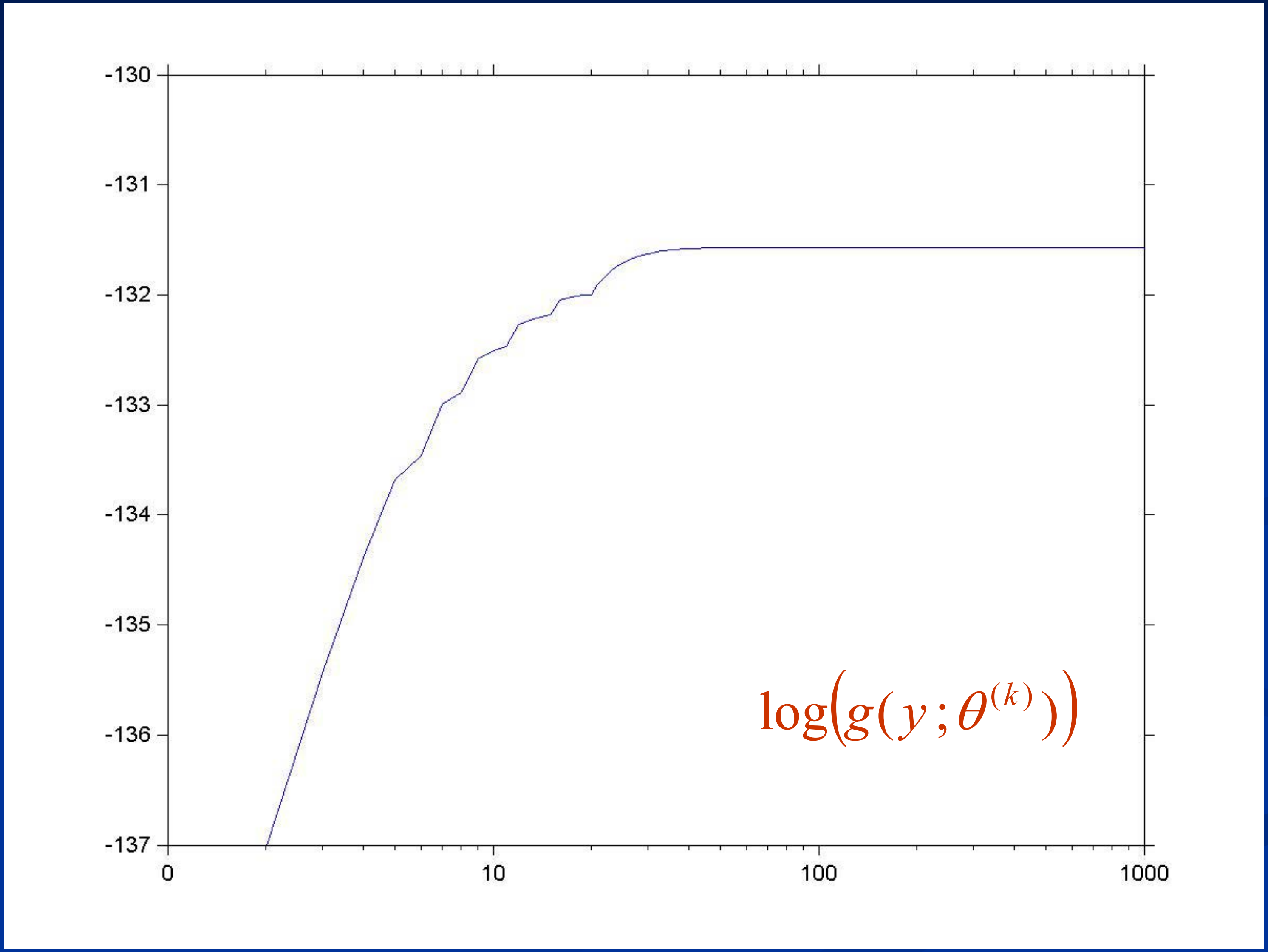
Estimation of the parameters with EM



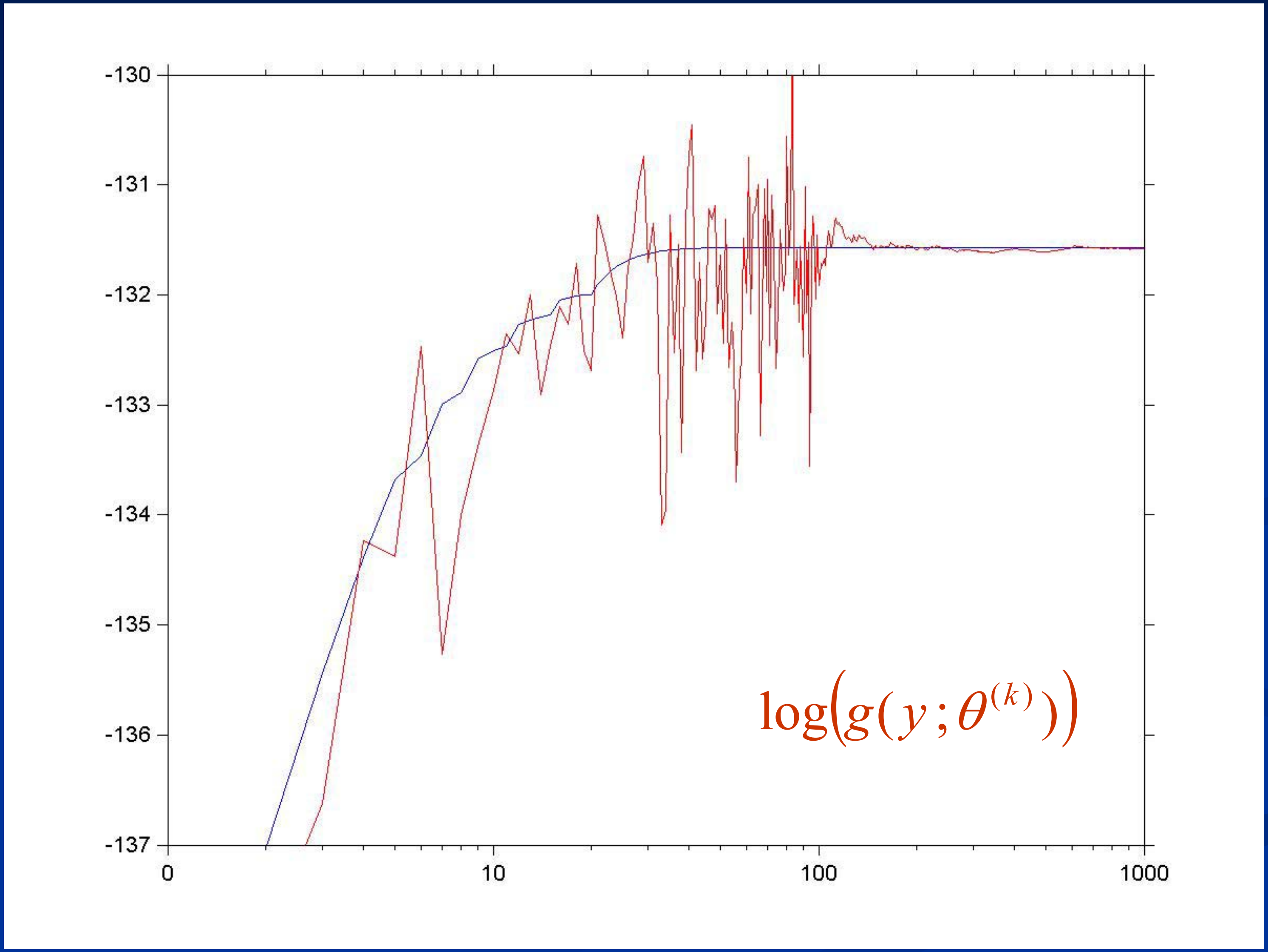
Estimation of the parameters with SAEM



Estimation of the likelihood with EM



Estimation of the likelihood with SAEM



Nonlinear model n°1 (Walker, 1996)

The model:

$$y_{ij} = \varphi_{i1} - \frac{\varphi_{i2}\, t_j}{\varphi_{i3} + t_j} + \varepsilon_{ij}$$

$$\varphi_{i1} = \mu_1 + b_{i1}$$

$$\varphi_{i2} = \mu_2 + b_{i2}$$

$$\varphi_{i3} = \mu_3 + b_{i3}$$

$$\theta = \left(\mu_1, \mu_2, \mu_3, \sigma_{b_1}^2, \sigma_{b_2}^2, \sigma_{b_3}^2, \sigma_{\varepsilon}^2\right)$$

Comparison of different estimates

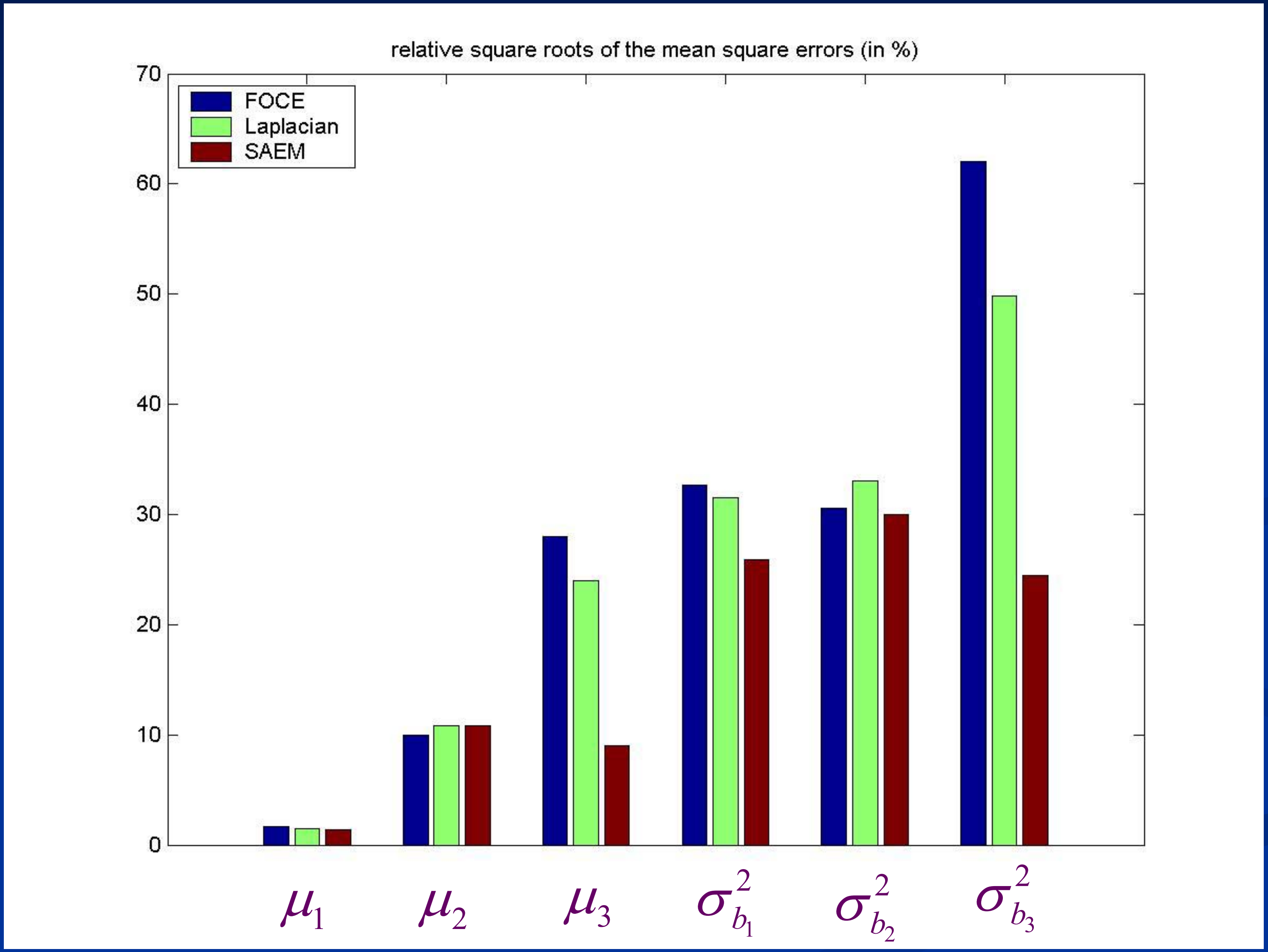
The root mean squared relative error (RMSRE)

$$\sqrt{E\left(\frac{\hat{\theta}-\theta^*}{\theta^*}\right)^2}$$

is estimated by Monte-Carlo:

$$\sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(\frac{\hat{\theta}_m-\theta^*}{\theta^*}\right)^2}$$

Nonlinear model n°1 (Walker, 1996)



Nonlinear model n°2 (Concordet, 2002)

The model:

$$y_{ij} = \varphi_{i1}\left(1 - \exp\left(-\varphi_{i2}\, t_j\right)\right) + \varepsilon_{ij}$$

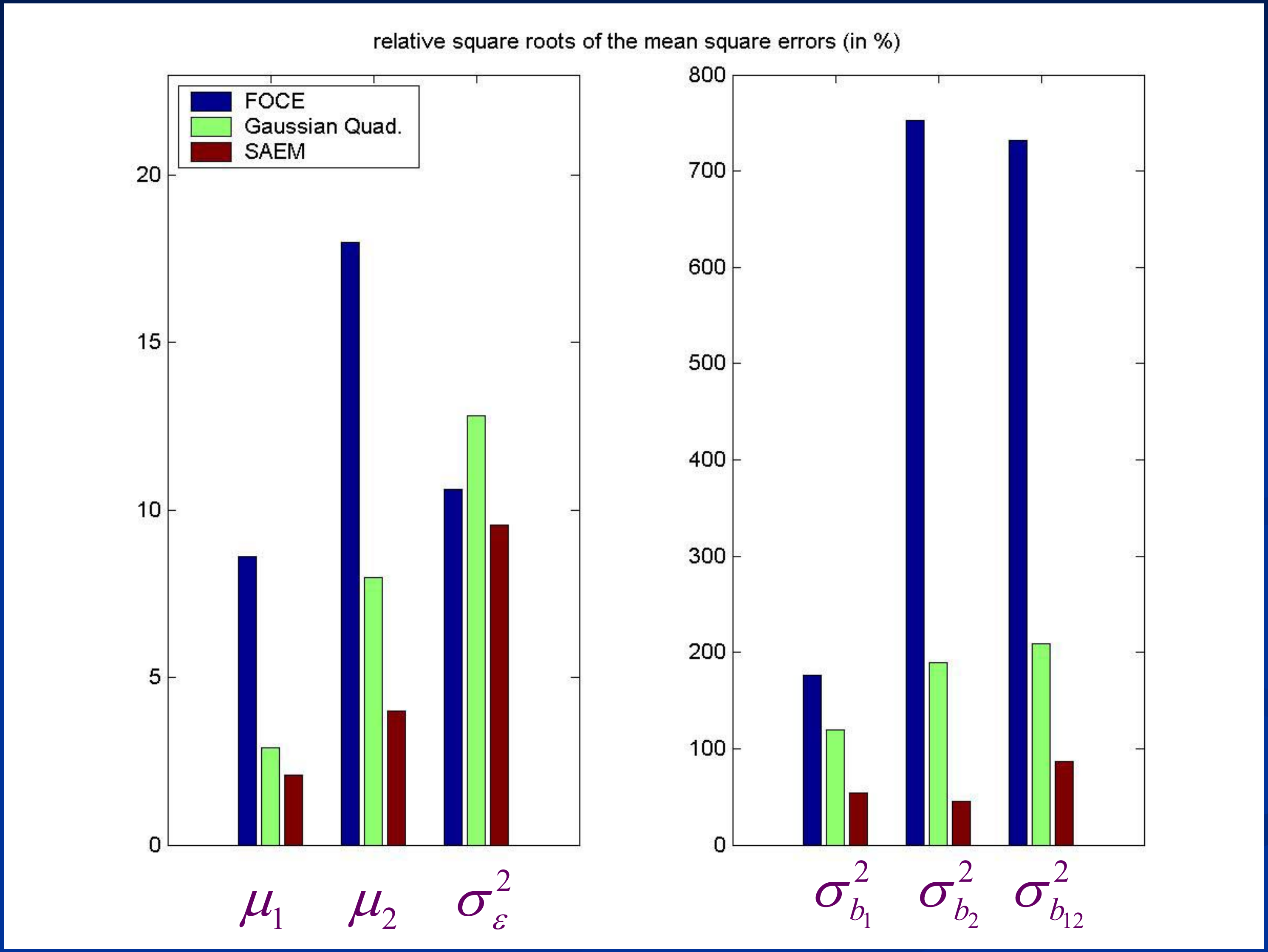
$$\varphi_{i1} = \mu_1 + b_{i1}$$

$$\varphi_{i2} = \mu_2 + b_{i2}$$

$$\Gamma = \begin{pmatrix} \sigma^2_{b_1} & \sigma^2_{b_{12}} \\ \sigma^2_{b_{12}} & \sigma^2_{b_2} \end{pmatrix}$$

$$\theta = \left(\mu_1, \mu_2, \sigma^2_{b_1}, \sigma^2_{b_2}, \sigma^2_{b_{12}}, \sigma^2_{\varepsilon}\right)$$

Nonlinear model n°2 (Concordet, 2002)



Nonlinear model n°3 (Mentré & Gomeni)

The model:

$$y_{ij} = D \frac{\varphi_{i1}}{\varphi_{i3} - \varphi_{i1}\varphi_{i2}} \left(e^{-\varphi_{i1} t_j} - e^{-\frac{\varphi_{i3}}{\varphi_{i2}} t_j} \right) \left(1 + \varepsilon_{ij}\right)$$

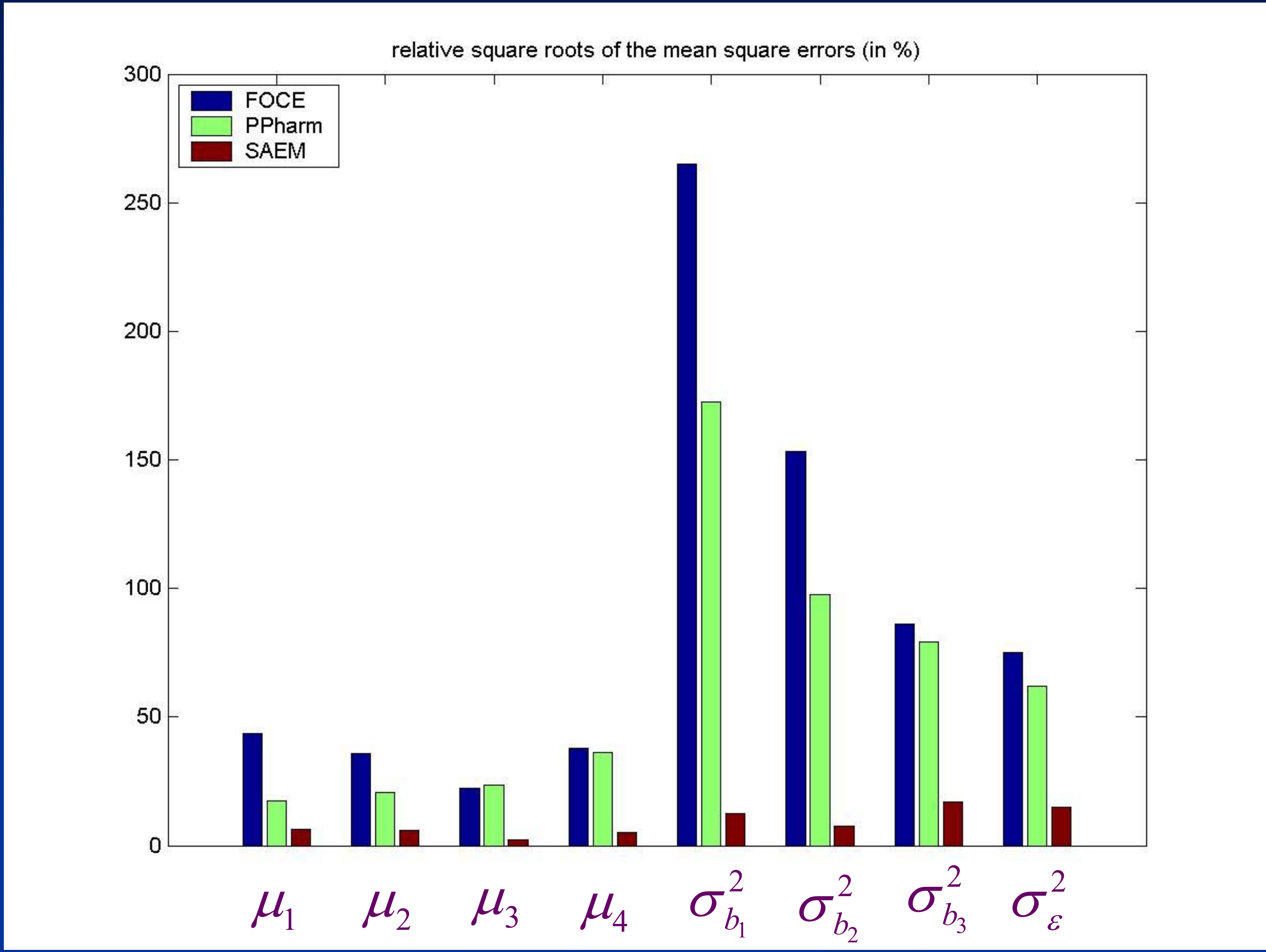
$$\varphi_{i1} = \mu_1 + b_{i1}$$

$$\varphi_{i2} = \mu_2 + b_{i2}$$

$$\varphi_{i3} = \mu_3 - \mu_4 \times age + b_{i3}$$

$$\theta = \left(\mu_1, \mu_2, \mu_3, \mu_4, \sigma_{b_1}^2, \sigma_{b_2}^2, \sigma_{b_3}^2, \sigma_{\varepsilon}^2 \right)$$

Nonlinear model n°3 (Mentré & Gomeni)



Some possible extensions

- Estimation of the likelihood
  (model selection, hypothesis testing, …)

- Estimation of the Fisher information matrix
  (confidence intervals)

- Heteroscedastic models

- Non Gaussian random effects
  (parametric or nonparametric distributions)

Molto grazie …