ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

D MATH
Department of Mathematics
Seminar für Statistik

Robust mixed effects models: simulations and limitations

Manuel Koller

Icors 2012, University of Vermont

Outline

1. What kind of models are (not) covered?
2. Generalization to non-diagonal $\boldsymbol{V}_b(\underline{\theta})$.
3. Calculation of the expectations.
4. Properties: bias, sensitivity curves, breakdown, efficiency.
5. R-package *robustlmm*.

1 Recap: the model

$$\underline{Y} = \boldsymbol{X}\underline{\beta} + \boldsymbol{Z}\underline{B} + \underline{\varepsilon}\ ,$$

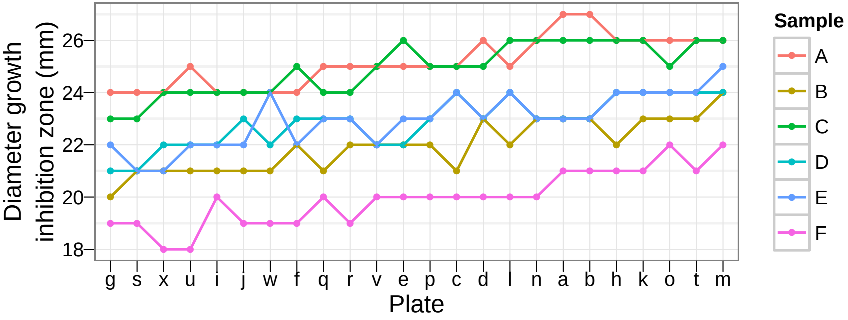
$$\underline{B} \sim \mathcal{N}_q(\underline{0}, \boldsymbol{V}_b(\underline{\theta})), \quad \underline{\varepsilon} \sim \mathcal{N}_n\left(\underline{0}, \sigma^2 \boldsymbol{I}\right), \quad \underline{B} \perp \underline{\varepsilon}\ .$$

$$\Rightarrow \mathbf{Cov}(\underline{Y}) = \boldsymbol{Z}\boldsymbol{V}_b(\underline{\theta})\boldsymbol{Z}^\mathsf{T} + \sigma^2 \boldsymbol{I}$$

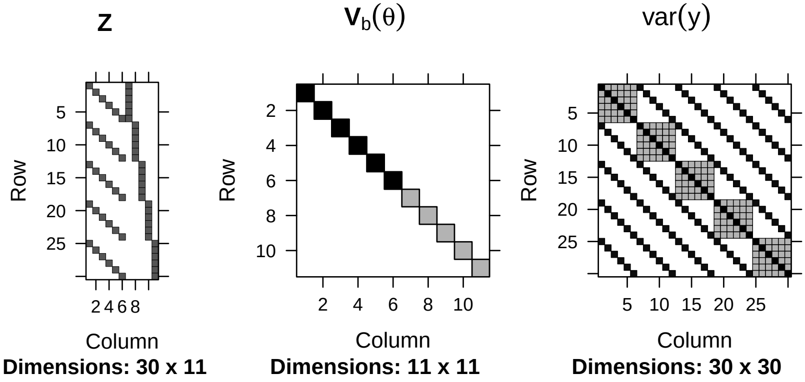
2 What kind of models are (not) covered?

Two-Way Anova (crossed)

Data from an experiment *to assess the variability between samples of penicillin by the* B. subtilis *method*.

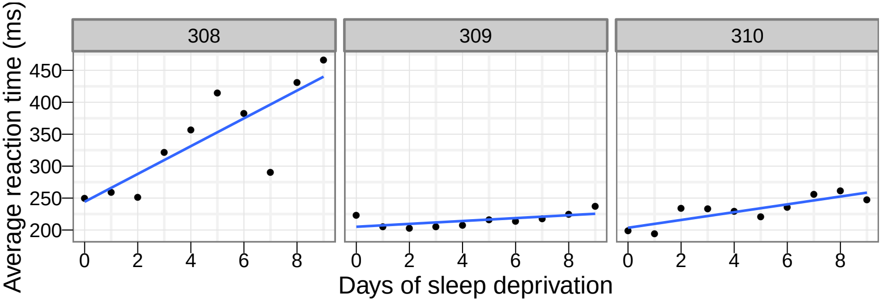


Two-Way Anova (crossed)



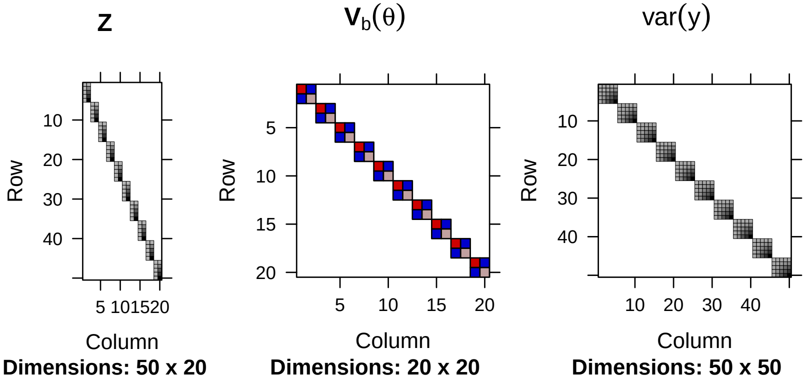
Random Intercept / Slope Model

Data from a study *of the effects of sleep deprivation on reaction time for a number of subjects chosen from a population of long-distance truck drivers.*



(Blue line: least squares fit, subset of data only.)

Random Intercept / Slope Model



Other data / models

Pedigree based models. Relationship matrix does not depend on data and can be incorporated in $\boldsymbol{Z}$, i.e., diagonal $\boldsymbol{V}_b(\underline{\theta})$.

Spatial data. $\boldsymbol{V}_b(\underline{\theta})$ is an $n \times n$ matrix, not in block diagonal form. Use classic ρ_b.

Different error variances. Calculate σ groupwise.

Correlated within (between) group errors. Unclear how to generalize *Design Adaptive Scale*-approach if a special structure is assumed. Simpler variant (next section) seems feasible. Challenging numerics.

(The examples on this slide are not supported by *robustlmm*.)

3 Generalization to block-diagonal $\boldsymbol{V}_b(\underline{\theta})$.

Before:

$$\underline{Y} = \boldsymbol{X}\underline{\beta} + \boldsymbol{Z}\underline{B} + \underline{\varepsilon} \ ,$$

$$\underline{B} \sim \mathcal{N}_q(\underline{0}, \boldsymbol{V}_b(\underline{\theta})), \quad \underline{\varepsilon} \sim \mathcal{N}_n\left(\underline{0}, \sigma^2 \boldsymbol{I}\right), \quad \underline{B} \perp \underline{\varepsilon} \ .$$

Now:

Work with spherical random effects $\underline{b}^*$, $\underline{\theta}$ relative to σ.

$$\underline{Y} = \boldsymbol{X}\underline{\beta} + \boldsymbol{Z}\boldsymbol{U}_b\underline{B}^* + \underline{\varepsilon} \ ,$$

where

$$\underline{B}^* \sim \mathcal{N}_q\left(\underline{0}, \sigma^2 \boldsymbol{I}\right), \quad \sigma^2 \boldsymbol{U}_b \boldsymbol{U}_b^\mathsf{T} = \boldsymbol{V}_b(\underline{\theta}) \ .$$

Let $\underline{e} = \underline{y} - \boldsymbol{X}\underline{\beta} - \boldsymbol{Z}\boldsymbol{U}_b(\underline{\theta})\underline{b}^*$.

Assuming $\underline{\theta} = \theta$, the estimating equations then can be written as:

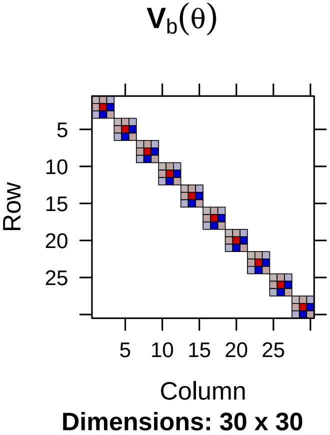
$$\boldsymbol{X}^{\mathsf{T}}\psi_e\left(\frac{\underline{e}}{\sigma}\right) = 0\ ,$$

$$\boldsymbol{U}_b(\underline{\theta})\boldsymbol{Z}^{\mathsf{T}}\psi_e\left(\frac{\underline{e}}{\sigma}\right) - \frac{\lambda_e}{\lambda_b}\psi_b\left(\frac{\underline{b}^*}{\sigma}\right) = 0\ ,$$

$$\sum_{i=1}^{n}\tau_i^{(e)2}\omega_e\left(\frac{e_i}{\tau_i^{(e)}\sigma}\right)\left[\left(\frac{e_i}{\tau_i^{(e)}\sigma}\right)^2 - \kappa_e\right] = 0 \quad \text{and}$$

$$\sum_{j=1}^{q}\tau_j^{(b)2}\omega_b\left(\frac{b_j^*}{\tau_j^{(b)}\sigma}\right)\left[\left(\frac{b_j^*}{\tau_j^{(b)}\sigma}\right)^2 - \kappa_b\right] = 0\ .$$

Introducing blockwise indexing



Indices:

$j = 1, \dots, q$ runs over random effects,

$k = 1, \dots, K$ runs over blocks.

Blocks are of size s, map $k(j)$ maps j to corresponding k.

Split $\underline{b}^*$ into subvectors $\underline{b}_k^*$ of length s corresponding to block k:

$$\underline{b}^* = \begin{pmatrix} \underline{b}_1^* \\ \vdots \\ \underline{b}_K^* \end{pmatrix},$$

Downweight by block, not by observation:

$$\text{replace } \psi_b\left(\frac{\underline{b}^*}{\sigma}\right) \text{ by } \mathbf{W}_b\left(\frac{\underline{b}^*}{\sigma}\right)\left(\frac{\underline{b}^*}{\sigma}\right),$$

where $k(j)$ maps j to block k, and

$$\mathbf{W}_b\left(\frac{\underline{b}^*}{\sigma}\right) = \mathrm{diag}\left(\left(\frac{\psi_b\left(d\left(\underline{b}^*_{k(j)}/\sigma\right)\right)}{d\left(\underline{b}^*_{k(j)}/\sigma\right)}\right)_{j=1,\dots,q}\right), \quad d\left(\underline{b}_k^*\right) = \sqrt{\frac{\underline{b}_k^{*\mathsf{T}}\underline{b}_k^*}{s}}.$$

The estimating equations then become:

$$\boldsymbol{X}^\mathsf{T}\psi_e\left(\frac{\underline{e}}{\sigma}\right) = 0\ ,$$

$$\boldsymbol{U}_b(\underline{\theta})\boldsymbol{Z}^\mathsf{T}\psi_e\left(\frac{\underline{e}}{\sigma}\right) - \frac{\lambda_e}{\lambda_b}\mathbf{W}_b\left(\frac{\underline{b}^*}{\sigma}\right)\left(\frac{\underline{b}^*}{\sigma}\right) = 0\ ,$$

$$\sum_{i=1}^{n}\tau_i^{(e)2}\omega_e\left(\frac{e_i}{\tau_i^{(e)}\sigma}\right)\left[\left(\frac{e_i}{\tau_i^{(e)}\sigma}\right)^2 - \kappa_e\right] = 0\ .$$

However, for θ_l we started with:

$$\underline{b}^{*\mathsf{T}}\boldsymbol{U}_b(\underline{\theta})^{-\mathsf{T}}\left(\frac{\partial\boldsymbol{U}_b(\underline{\theta})}{\partial\theta_l}\right)\underline{b}^* = \ldots \quad \text{(expectation of lhs)}\ .$$

$$\underline{b}^{*\mathsf{T}} \boldsymbol{U}_b(\underline{\theta})^{-\mathsf{T}} \left(\frac{\partial \boldsymbol{U}_b(\underline{\theta})}{\partial \theta_l} \right) \underline{b}^* \; = \; \ldots \quad \text{(expectation of lhs)} \, .$$

In the diagonal case (assume $\underline{\theta} = \theta$), this reduces to

$$\sum_{j=1}^{q} b_j^{*2} \; = \; \ldots \; ,$$

then apply *Design Adaptive Scale*-approach,

$$\sum_{j=1}^{q} \tau_j^{(b)2} \omega_b \left(\frac{b_j^*}{\tau_j^{(b)} \sigma} \right) \left[\left(\frac{b_j^*}{\tau_j^{(b)} \sigma} \right)^2 - \kappa_b \right] = 0 \, .$$

How can we generalize to the block-diagonal case?

[*Design Adaptive Scale*-approach: work in progress]

We may fall back to,

$$\underline{b}^{*\top}\mathbf{W}_b\left(\frac{\underline{b}^*}{\sigma}\right)^\top \boldsymbol{U}_b(\underline{\theta})^{-\top}\left(\frac{\partial \boldsymbol{U}_b(\underline{\theta})}{\partial\theta_l}\right)\mathbf{W}_b\left(\frac{\underline{b}^*}{\sigma}\right)\underline{b}^* \;=\; \dots\,.$$

4 Calculation of the Expectations

Expand $\underline{\widetilde{\psi}}_e$ and $\underline{\widetilde{\psi}}_b$ linearly around the true $\underline{\beta}$ and $\underline{b}^*$:

$$\underline{\widetilde{\psi}}_e \approx \underline{\psi}_e - \boldsymbol{D}_e \left(\boldsymbol{X} \left(\underline{\widetilde{\beta}} - \underline{\beta} \right) + \boldsymbol{Z} \boldsymbol{U}_b(\underline{\theta}) \left(\underline{\widetilde{b}}^* - \underline{b}^* \right) \right) / \sigma ,$$

$$\underline{\widetilde{\psi}}_b \approx \underline{\psi}_b + \boldsymbol{D}_b \left(\underline{\widetilde{b}}^* - \underline{b}^* \right) / \sigma ,$$

where $\boldsymbol{D}_{.}$ is the expectation of the first derivatives and

$$\underline{\widetilde{\psi}}_e = \psi_e\left(\frac{\underline{\widetilde{e}}}{\sigma} \right), \qquad \underline{\psi}_e \approx \psi_e\left(\frac{\underline{\varepsilon}}{\sigma} \right),$$

$$\underline{\widetilde{\psi}}_b = \mathbf{W}_b\left(\frac{\underline{\widetilde{b}}^*}{\sigma} \right) \frac{\underline{\widetilde{b}}^*}{\sigma} \quad \text{and} \qquad \underline{\psi}_b = \mathbf{W}_b\left(\frac{\underline{b}^*}{\sigma} \right) \frac{\underline{b}^*}{\sigma}.$$

Now insert this into the estimating equations and solve for $\underline{\widetilde{\beta}}$, $\underline{\widetilde{b}}^*$.

We get an approximation of the residuals and the spherical random effects,

$$\underline{\tilde{e}} \approx \underline{\varepsilon} - \sigma \boldsymbol{A}_\theta \underline{\psi}_e - \sigma \boldsymbol{B}_\theta \underline{\psi}_b,$$

$$\underline{\tilde{b}}^* \approx \underline{b}^* - \sigma \boldsymbol{K}_\theta \underline{\psi}_e - \sigma \boldsymbol{L}_\theta \underline{\psi}_b.$$

For **diagonal** $\boldsymbol{V}_b(\underline{\theta})$ this can be simplified for a single component to

$$\begin{aligned}\tilde{b}_j^* &\approx b_j^* - \sigma \underline{L}_{\theta jj} \psi_b\left(\frac{b_j^*}{\sigma}\right) + \sigma h\left(\frac{\underline{\varepsilon}}{\sigma}, \frac{\underline{b}_{[-j]}^*}{\sigma}\right) \\ &\approx b_j^* - \sigma \underline{L}_{\theta jj} \psi_b\left(\frac{b_j^*}{\sigma}\right) + \sigma v\ ,\end{aligned}$$

where h collects terms and $v \sim \mathcal{N}(0, \mathrm{var}(h(\dots)))$.

Then just insert

$$\widetilde{b}_j^* \approx b_j^* - \sigma \underline{L}_{\theta jj} \psi_b\left(\frac{b_j^*}{\sigma}\right) + \sigma v \ ,$$

into

$$\mathbf{E}\left[\omega_b\left(\frac{\widetilde{b}_j^*}{\tau_j^{(b)}\sigma}\right)\left[\left(\frac{\widetilde{b}_j^*}{\tau_j^{(b)}\sigma}\right)^2 - \kappa_b\right]\right] = 0 \ ,$$

and solve for $\tau_j^{(b)}$.

Need to evaluate 2-dim integral.

In **non-diagonal case**:

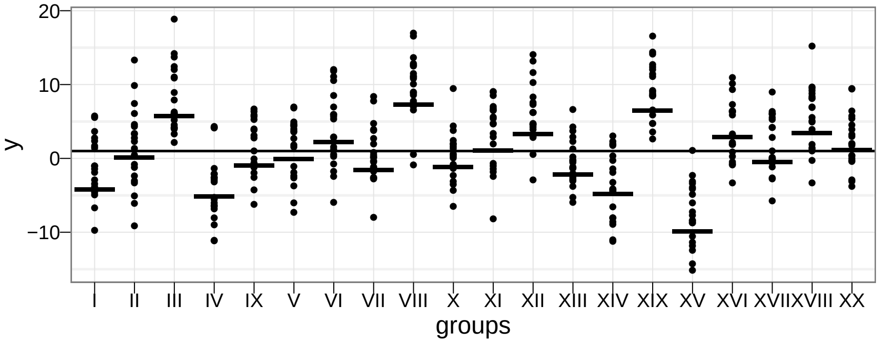
Weights depend on $d(\underline{b}^*_k/\sigma)$, and spherical random effects belonging to the same block are dependent. Simplification leads only to:

$$\tilde{\underline{b}}^*_k \approx \underline{b}^*_k - \sigma \boldsymbol{L}_{\theta kk} \underline{\psi}_{b,k} - \sigma \underline{\boldsymbol{v}}_k .$$

Need to evaluate 2s-dim integral to compute expectations!!

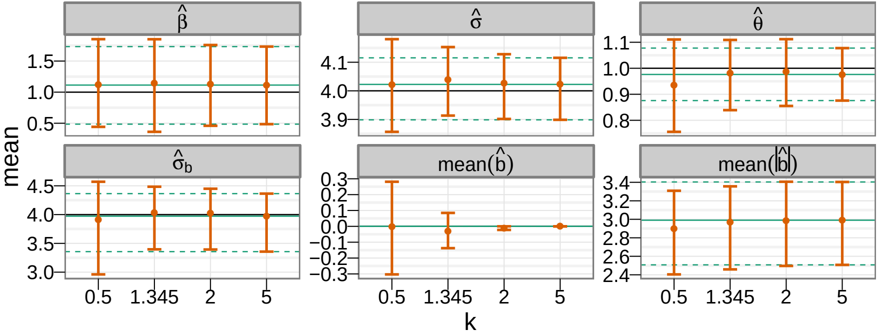
5 Properties of the robust estimator

Study case of balanced one-way ANOVA.



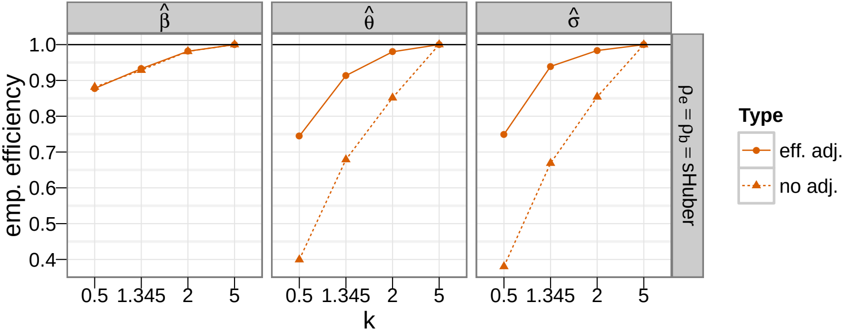
5.1 Bias

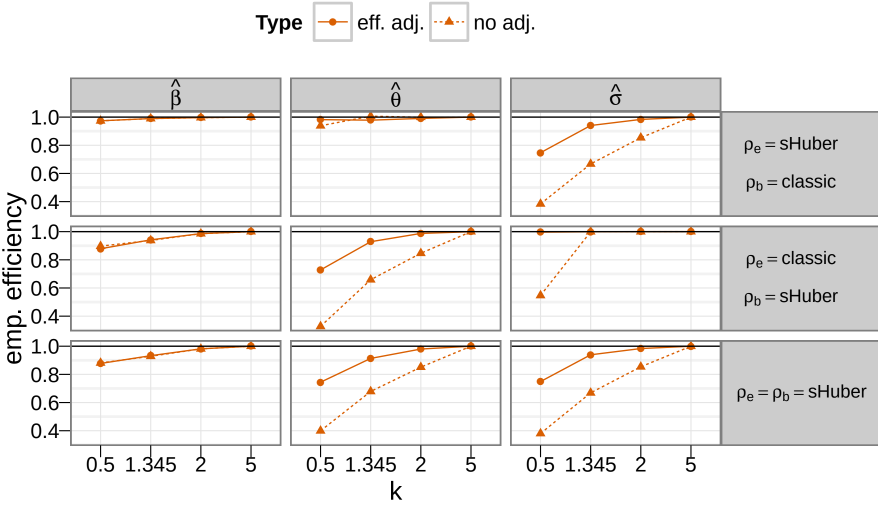
For 100 replicates of balanced one-way ANOVA (20×20), we show the mean and the quartiles of the estimates for varying tuning constants k.



5.2 Efficiency (empirical)

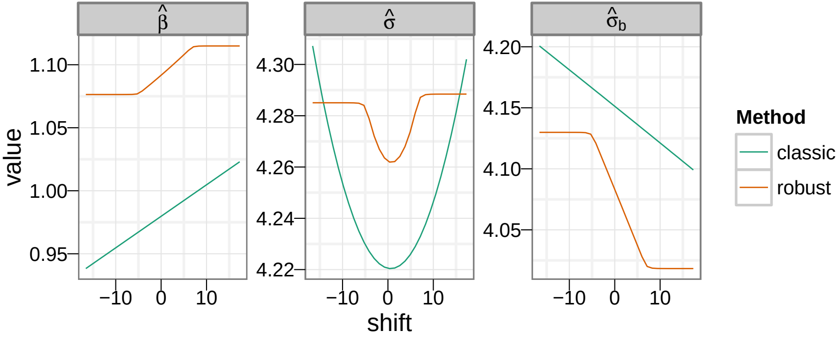
Comparing robust and classical estimates for the same replicates used in the consistency simulation.




Type
eff. adj.
no adj.
β̂
θ̂
σ̂
ρe = sHuber
ρb = classic
ρe = classic
ρb = sHuber
ρe = ρb = sHuber
emp. efficiency
1.0
0.8
0.6
0.4
0.5
1.345
2
5
k

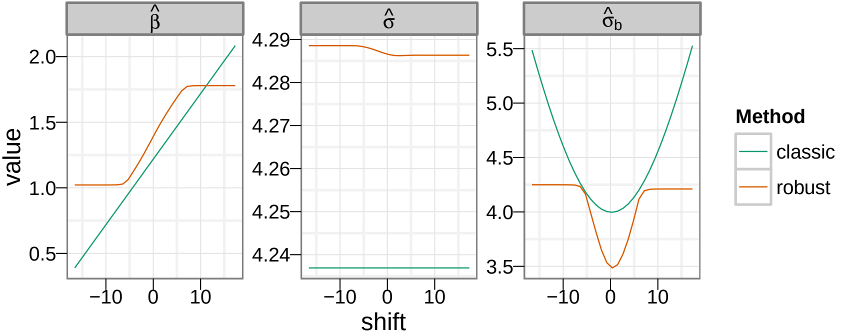
5.3 Sensitivity Curves: Shift a single observation

Take a balanced one-way ANOVA dataset (20×20),
set the true error of observation 1 to *shift*.



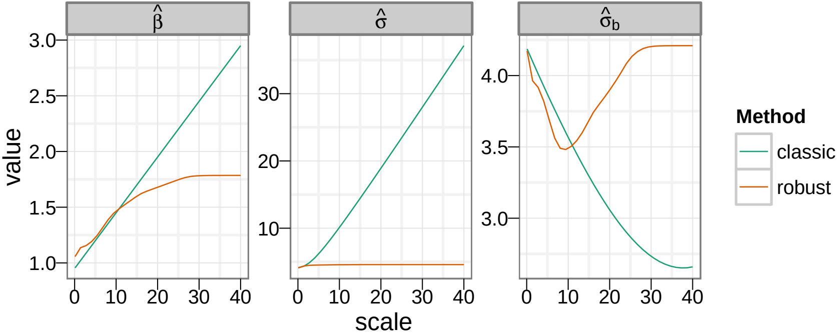
5.4 Sensitivity Curves: Shift a group

Take a balanced one-way ANOVA dataset (20×20),
set the true random effect of group 1 to *shift*.



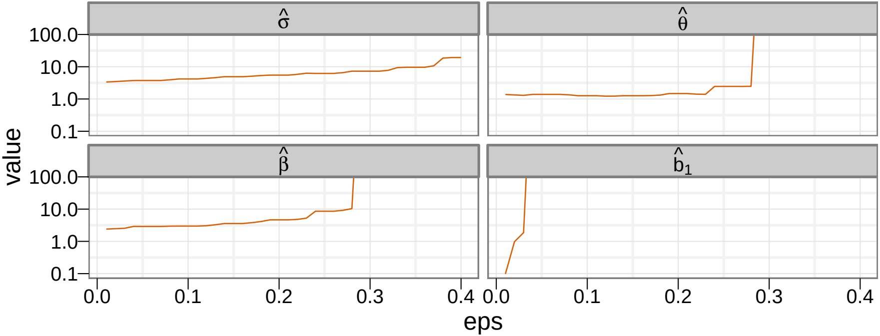
5.5 Sensitivity Curves: Scale a group

Take a balanced one-way ANOVA dataset (20×20),
scale the true errors of the observations in group 1 by *scale*.



5.6 Breakdown

Take a balanced one-way ANOVA dataset (20×5),
contaminate observation after observation, group after group.



6 R-package robustlmm

. . . is now available on github:

https://github.com/kollerma/robustlmm

Thank you for listening!

References

D. M. Bates (2012). *lme4: Mixed-effects modeling with R*, *(http://lme4.r-forge.r-project.org/lMMwR/)*.

M. Koller and W. A. Stahel (2011). Sharpening Wald-type inference in robust regression for small samples. *Computational Statistics & Data Analysis* 55(8), 2504–2515.

J. C. Pinheiro and D. M. Bates (2000). *Mixed-Effects Models in S and S-PLUS*, Springer.

$$\begin{bmatrix} \theta_1 & 0 & 0 \\ \theta_2 & \theta_4 & 0 \\ \theta_3 & \theta_5 & \theta_6 \end{bmatrix}^{-T} = \begin{bmatrix} \theta_1{}^{-1} & -\frac{\theta_2}{\theta_1\theta_4} & \frac{\theta_2\theta_5-\theta_3\theta_4}{\theta_1\theta_4\theta_6} \\ 0 & \theta_4{}^{-1} & -\frac{\theta_5}{\theta_4\theta_6} \\ 0 & 0 & \theta_6{}^{-1} \end{bmatrix}$$

7 Smoothed Huber function