Supplemental Materials for

Repeated measures analysis of variance, linear mixed models, and latent growth curve model for group comparison studies contaminated by outliers

Fabio Mason, Eva Cantoni and Paolo Ghisletta

Technicalities of the LMM Estimation Methods

[1]

To introduce the mathematical details of the estimating methods we consider in this paper, it is useful to introduce some additional notation definitions. First, it is useful to consider an alternative generalized parametrisation for the variance component parameters of the LMM of Equations (**??**), namely $\boldsymbol{\zeta} = \boldsymbol{\theta}/\sigma_\varepsilon^2$ such that we can separate the effect of σ_ε^2 and $\boldsymbol{\zeta}$ in $Var(\boldsymbol{y}_i) = (\sigma_\varepsilon^2 Z_i \Sigma(\boldsymbol{\zeta}) Z_i^T + \sigma_\varepsilon^2 I) = \sigma_\varepsilon^2 \Omega(\boldsymbol{\zeta})$. Second, it is convenient to rewrite the model to work with random effects centered around zero and with diagonal constant variance. It amounts at considering the model

$$\boldsymbol{y_i} = X_i \boldsymbol{\gamma} + Z_i U_b(\boldsymbol{\zeta}) \tilde{\boldsymbol{b}}_i + \boldsymbol{\varepsilon}_i, \tag{S1}$$

with $\tilde{\boldsymbol{b}}_i \sim \mathcal{N}(\boldsymbol{0}, \sigma_\varepsilon^2 I)$ and $U_b(\boldsymbol{\zeta})$ such that $\boldsymbol{b}_i = U_b(\boldsymbol{\zeta}) \tilde{\boldsymbol{b}}_i$ and $U_b(\boldsymbol{\zeta}) U_b(\boldsymbol{\zeta})^T = \Omega(\boldsymbol{\zeta})$.

Gaussian ML Estimator

The ML estimator of $\boldsymbol{\gamma}$ under the Gaussian assumption maximizes the log-likelihood. For given $\boldsymbol{\gamma}, \sigma_\varepsilon^2$ and $\boldsymbol{\zeta}$, the best linear unbiased predictor (BLUP) of the random effects $\tilde{\boldsymbol{b}}_i$ are computed as $\tilde{\boldsymbol{b}}_i = \sigma_\varepsilon^2 U_b(\boldsymbol{\zeta})^T Z^T \Omega(\boldsymbol{\gamma})^{-1} (\boldsymbol{y} - X\boldsymbol{\gamma})$. It can be shown that, given σ_ε^2 and $\boldsymbol{\zeta}$, $(\hat{\boldsymbol{\gamma}}, \hat{\tilde{\boldsymbol{b}}})$ are the solution of the Henderson equations (**?**)

$$\begin{bmatrix} X^T X & X^T Z U_b(\boldsymbol{\zeta}) \\ U_b(\boldsymbol{\zeta})^T Z^T X & U_b(\boldsymbol{\zeta})^T Z^T Z U_b(\boldsymbol{\zeta}) + \sigma_\varepsilon^2 I \end{bmatrix} \begin{bmatrix} \boldsymbol{\gamma} \\ \tilde{\boldsymbol{b}} \end{bmatrix} = \begin{bmatrix} X^T \boldsymbol{y} \\ U_b^T Z^T \boldsymbol{y} \end{bmatrix}. \tag{S2}$$

[1] From "Supplemental Material for Parametric and semi-parametric bootstrap-based confidence intervals for robust linear mixed models [Equations and Figures]", by Mason, F., Cantoni, E., & Ghisletta, P. (2021). PsychOpen GOLD. `https://doi.org/10.23668/psycharchives.5302`.

General Classes of Robust Estimators

In this Section we introduce some general classes of robust estimators, upon which the robust estimators for LMM are built. To do so, we consider a generic setting with independent random variables $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$, with $\boldsymbol{u} = (\boldsymbol{u}_1^T, \ldots, \boldsymbol{u}_n^T)^T$, and where the interest is in the estimation of a parameter $\boldsymbol{\beta}$.

An M-estimator for $\boldsymbol{\beta}$ is defined as the solution of a set of estimating equations

$$\sum_{i=1}^{n} \Psi(\boldsymbol{u}_i; \boldsymbol{\beta}) = 0, \tag{S3}$$

under mild regularity conditions on Ψ, see **?**. If $\Psi(t)$ is the derivative of a function $\rho(t)$, solving the above equations corresponds to minimizing $\sum_{i=1}^{n} \rho(\boldsymbol{u}_i; \boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$.

For a well-behaved function ρ_1, an M-scale estimator $s(\boldsymbol{u})$ is defined by the value of s satisfying

$$\frac{1}{n} \sum_{i=1}^{n} \rho_1 \left(\frac{\boldsymbol{u}_i}{s} \right) = d, \tag{S4}$$

where $d = E_{\chi^2_K}(\rho_1(V))$ with $V \sim \chi^2_K$, a chi-square distribution with K degrees of freedom.

Given an M-scale estimator s, that is an estimator satisfying (S4), an S-estimator of $\boldsymbol{\beta}$ is defined as the minimizer of $\det(Var(\boldsymbol{u}))$ or equivalently as

$$\text{argmin}_{\beta}\ s(\boldsymbol{u}) = \text{argmin}_{\beta}\ s(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n). \tag{S5}$$

On the other hand, given an M-scale estimator s as per (S4), a τ estimator is defined as

$$\text{argmin}_{\beta}\ \tau^2(\boldsymbol{u}) = \text{argmin}_{\beta}\ s^2(\boldsymbol{u}) \frac{1}{n} \sum_{i=1}^{n} \rho_2 \left(\frac{u_i}{s(\boldsymbol{u})} \right), \tag{S6}$$

where ρ_2 is a well-behaved function.

The ρ_1 and ρ_2 functions above are usually chosen from a set of available alternatives one can find in the literature. All of these functions depends on one or more tuning constants that need to be set. Guidelines, based on efficiency and or robustness arguments, are usually provided. It is common to denote $\psi_k(u) = \frac{\partial \rho_k(u)}{\partial u}$ $(k = 1, 2)$, the derivative of ρ_k. Amongst the most popular ρ functions, we can cite the Huber and Tukey's function.

We also introduce the definition of the Mahalanobis distance between two vectors $\boldsymbol{u}$ and $\boldsymbol{c}$ (a measure of center for $\boldsymbol{u}$) of dimension g:

$$m(\boldsymbol{u}, \boldsymbol{c}, \Theta) = (\boldsymbol{u} - \boldsymbol{c})^T \Theta^{-1} (\boldsymbol{u} - \boldsymbol{c}). \tag{S7}$$

where Θ (a measure of dispersion for $\boldsymbol{u}$) is an $g \times g$ matrix.

The robustness of estimators can be characterized with respect to bounded influence and high breakdown point. The first concept is quantified by the influence function (**?**), which measures the maximal bias produced by an estimator under contamination. Bounded influence estimators are those for which this bias is limited (bounded). The second concept represents the proportion of contamination that an estimator can tolerate, that is without producing arbitrarily large results. If this proportion is large, the estimator is called a high-breakdown estimator.

For appropriate choices of the Ψ and ρ functions involved in their definitions, the above discussed estimators are robust in the sense that they have a bounded influence function. In addition, the S-estimators and the τ estimators guarantee high-breakdown, but at the price of reduced efficiency.

Robust Estimators for Linear Regression

Consider a linear model

$$y_i = \boldsymbol{x}_i^T \boldsymbol{\gamma} + \varepsilon_i, \tag{S8}$$

for independent observations y_i and where $\boldsymbol{x}_i$ are a set of covariates. The error term ε_i is assumed to be normally distributed around 0 and with constant variance σ^2. The general definitions of the robust estimators in the previous section are applied to $u_i = (y_i - \boldsymbol{x}_i^T \boldsymbol{\gamma})$ to obtain M, S and τ estimators for the linear model.

In the specific context of the linear mixed model, the MM-estimator is a refinement of the above estimators to ensure both high efficiency and high-breakdown (?). It is the sequence of an M-scale estimator followed by an M-estimator of location. Another group of robust estimators for the linear model setting replaces the least squares criterion by a different one, which is less sensitive to outliers. For instance, The L1 estimator minimizes the absolute value of residuals rather than squared residuals. The least median of squares (LMS) and the trimmed least squares (LTS) estimators minimize a scale measure of residuals that is insensitive to large values: the median of the absolute residuals for the former, the trimmed least squares for the latter, see ?. Chapters 4 and 5 of ? are devoted to robust estimators and provide a detailed account of the approaches introduced above. The concept of M, MM, S, and τ estimators can be extended beyond the linear model setting, see next Section. The other estimators discussed above are specific to the linear model and cannot be easily adapted to other settings.

? S-Estimator

? developed the S-estimator for the LMM setting in (S1) by applying its general definition to Mahalanobis distances $m_i = m(\boldsymbol{y}_i, X_i\boldsymbol{\gamma}, \Omega(\boldsymbol{\zeta}))$. More precisely, given an M-scale estimator s satisfying

$$\frac{1}{n}\sum_{i=1}^{n}\rho_1\left(\frac{\sqrt{m(\boldsymbol{y}_i, X_i\boldsymbol{\gamma}, \Omega(\boldsymbol{\zeta}))}}{s}\right) = d, \tag{S9}$$

their proposal solves

$$(\widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\zeta}}) = \text{argmin}_{\gamma,\zeta}\ s\Big(\sqrt{m(\boldsymbol{y}_1, X_1\boldsymbol{\gamma}, \Omega(\boldsymbol{\zeta}))}, \ldots, \sqrt{m(\boldsymbol{y}_n, X_n\boldsymbol{\gamma}, \Omega(\boldsymbol{\zeta}))}\Big), \tag{S10}$$

and then compute

$$\widehat{\sigma}_{\varepsilon}^2 = \frac{s^2\left(\sqrt{m(\boldsymbol{y}_1, X_1\widehat{\boldsymbol{\gamma}}, \Omega(\widehat{\boldsymbol{\zeta}}))}, \ldots, \sqrt{m(\boldsymbol{y}_n, X_n\widehat{\boldsymbol{\gamma}}, \Omega(\widehat{\boldsymbol{\zeta}}))}\right)}{s_0^2}, \tag{S11}$$

with s_0 satisfying $E(\rho_1(V/s_0)) = d$.

The ? ρ function is used by ?. It can be seen as a translated Tukey's function.

For diagnostic purposes, after fitting the model, one can look at the Mahalanobis distances m_i or at weights (between 0 and 1) for each subject i, defined as

$$w(m_i) = \frac{\frac{\partial \rho_1(m_i)}{\partial m_i}}{m_i} = \frac{\psi_1(m_i)}{m_i}. \tag{S12}$$

These weights appear in the estimating equations, and therefore in the final S-estimator. Large Mahalanobis distances will indicate potential outliers and will translate into weights being small, as described by Equation (S12).

? Composite τ-Estimators

The proposal of **?** builds on the philosophy of composite estimators by defining a τ-estimator for pairs of observations of the same subject. It also builds on **?** in that it defines the estimator through Mahalanobis distances.

Let us introduce the composite τ estimator in more details. We first define the (j, l)-pairwise squared Mahalanobis distance for subject i as

$$m_i^{jl} = m_i^{jl}(\boldsymbol{\gamma}, \boldsymbol{\zeta}) = m(\boldsymbol{y}_i^{jl}, (X_i\boldsymbol{\gamma})^{jl}, \Omega^{jl}(\boldsymbol{\zeta})), \tag{S13}$$

where $\boldsymbol{y}_i^{jl} = (y_{ij}, y_{il})^T$, $(X_i\boldsymbol{\gamma})^{jl}$ is the bivariate vector containing the j-th and l-th elements of $X_i\boldsymbol{\gamma}$, and

$$\Omega^{jl}(\boldsymbol{\zeta}) = \begin{pmatrix} \omega_{jj}(\boldsymbol{\zeta}) & \omega_{jl}\boldsymbol{\zeta}) \\ \omega_{lj}(\boldsymbol{\zeta}) & \omega_{ll}(\boldsymbol{\zeta}) \end{pmatrix}, \tag{S14}$$

with $\omega_{rs}(\boldsymbol{\zeta})$ being the (r, s) element of the matrix $\Omega(\boldsymbol{\zeta})$.

Following the idea of working with Mahalanobis distances, and given an M-scale estimator $s_{jl}(\boldsymbol{\gamma}, \boldsymbol{\zeta}) = s\left(\sqrt{m_1(\boldsymbol{y}_1^{jl}, (X_i\boldsymbol{\gamma})^{jl}, \Omega^{jl}(\boldsymbol{\zeta}))}, \ldots, \sqrt{m_n(\boldsymbol{y}_n^{jl}, (X_i\boldsymbol{\gamma})^{jl}, \Omega^{jl}(\boldsymbol{\zeta}))}\right)$ for each (j, l) pair and a function ρ_2, the final composite τ criterion is defined by summing over all the possible pairs and subjects

$$T(\boldsymbol{\gamma},\boldsymbol{\zeta}) = \sum_{j=1}^{J-1}\sum_{l=j+1}^{J}\tau_{jl}(\boldsymbol{\gamma},\boldsymbol{\zeta}) = \sum_{J=1}^{J-1}\sum_{l=j+1}^{J} s_{jl}(\boldsymbol{\gamma},\boldsymbol{\zeta})\frac{1}{n}\sum_{i=1}^{n}\rho_2\left(\frac{\sqrt{m_i^{jl}(\boldsymbol{\gamma},\boldsymbol{\zeta})}}{s_{jl}}\right), \tag{S15}$$

where $J = J_i$ is the number of observations per subject (same for all i). The composite τ-estimator of $(\boldsymbol{\gamma},\boldsymbol{\zeta})$ is then defined as the minimizer of $T(\boldsymbol{\gamma},\boldsymbol{\zeta})$ and σ_ε^2 is estimated by solving

$$\frac{2}{J(J-1)n}\sum_{i=1}^{n}\sum_{j=1}^{J-1}\sum_{l=j+1}^{J}\rho_1\left(\frac{(\boldsymbol{y}_i^{jl}-(X_i\hat{\boldsymbol{\gamma}})^{jl})^T\Omega^{jl}(\hat{\boldsymbol{\zeta}})^{-1}(\boldsymbol{y}_i^{jl}-(X_i\hat{\boldsymbol{\gamma}})^{jl})}{\widehat{\sigma}_\varepsilon^2 s_0}\right) = d, \tag{S16}$$

with s_0 satisfying $E(\rho_1(V/s_0)) = d$.

The fitting procedure includes the definition of weights that can be used for diagnostic purposes. They can either be defined based on ρ_1 or ρ_2, for each subject i and each couple of observation (j,l), as follow

$$W_{k,i}^{jl} = W_k\left(\frac{m_i^{jl}(\boldsymbol{\gamma},\boldsymbol{\zeta})}{s_{jl}^2(\boldsymbol{\gamma},\boldsymbol{\zeta})}\right), \tag{S17}$$

where $W_k(x) = \frac{\partial\rho_k(\sqrt{x})}{\partial x}$ and $k = 1,2$. Therefore there are as many weights as there are couple of observations.

? suggest to choose ρ_1 and ρ_2 in the family of functions introduced by **?**.

? DAStau Estimator

The proposal of **?** modifies the ML estimation Equations (S2) for $(\boldsymbol{\gamma},\tilde{\boldsymbol{b}})$ (given σ_ε and $\boldsymbol{\zeta}$) by introducing weight matrices W_e and W_b to downweight both types of outliers :

$$\begin{bmatrix} X^TW_eX & X^TW_eZU_b \\ U_b^TZ^TW_eX & U_b^TZ^TW_eZU_b+\sigma_\varepsilon^2\Lambda_bW_b \end{bmatrix}\begin{bmatrix}\boldsymbol{\gamma}\\ \tilde{\boldsymbol{b}}\end{bmatrix} = \begin{bmatrix} X^TW_e\boldsymbol{y} \\ U_b^TZ^TW_e\boldsymbol{y}\end{bmatrix}, \tag{S18}$$

where $\Lambda_b = \mathrm{diag}(\lambda_e/\lambda_{b,j})_{J=1,\dots,q}$ is a diagonal matrix containing different scaling factors (for more details see e.g. **?**). The diagonal matrix $W_e = \mathrm{diag}(w_e(\boldsymbol{\varepsilon}_{11}/\sigma_\varepsilon),\dots,w_e(\boldsymbol{\varepsilon}_{nJ}/\sigma_\varepsilon))$, where $\varepsilon_i = \varepsilon_i(\boldsymbol{\gamma}, \tilde{\boldsymbol{b}}) = \boldsymbol{y_i} - X_i\boldsymbol{\gamma} - Z_iU_b(\boldsymbol{\zeta})\tilde{\boldsymbol{b}}_i$, contains observation weights defined by

$$w_e(e) = \begin{cases} \psi_e(e)/e & \text{if} \quad e \neq 0 \\ \psi_e'(0) & \text{if} \quad e = 0, \end{cases} \tag{S19}$$

and the diagonal matrix $W_b = \text{diag}(w_b(m_1), \ldots, w_b(m_J))$ contains weights at the subject level defined by

$$w_b(m) = \begin{cases} \psi_b(\sqrt{m})/\sqrt{m} & \text{if} \quad m \neq 0 \\ \psi_b'(0) & \text{if} \quad m = 0, \end{cases} \tag{S20}$$

where $m_i = \tilde{\boldsymbol{b}}_i^T\tilde{\boldsymbol{b}}_i$ are the Mahalanobis distances for $\tilde{\boldsymbol{b}}_i$.

The inspection of the weights in W_e and W_b give useful information about the observation and the subjects, respectively, that deviate from the bulk of the data.

Two additional sets of estimating equations are used to obtain estimates of σ_ε and $\boldsymbol{\zeta}$. The estimate of the scale parameter σ_ε is obtained following **?** as the solution of

$$\sum_{i=1}^{n}\sum_{j=1}^{J_i} \tau_{e,ij}^2 \; w_e^{(\sigma)}\left(\frac{\hat{\varepsilon}_{ij}}{\tau_{e,ij}\sigma_\varepsilon}\right)\left[\left(\frac{\hat{\varepsilon}_{ij}}{\tau_{e,ij}\sigma_\varepsilon}\right)^2 - \kappa_e^{(\sigma)}\right] = 0, \tag{S21}$$

where $\hat{\boldsymbol{\varepsilon}}_i = \varepsilon_i(\hat{\boldsymbol{\gamma}}, \hat{\tilde{\boldsymbol{b}}})$ and τ_{ij} are such that

$$E\left[w_e^{(\sigma)}\left(\frac{\hat{\varepsilon}_{ij}}{\tau_{ij}\sigma_\varepsilon}\right)\left(\frac{\hat{\varepsilon}_{ij}}{\tau_{ij}\sigma_\varepsilon}\right)^2 - \kappa_e^{(\sigma)} \; w_e^{(\sigma)}\left(\frac{\hat{\varepsilon}_{ij}}{\tau_{ij}\sigma_\varepsilon}\right)\right] = 0, \tag{S22}$$

with

$$\kappa_e^{(\sigma)} = \frac{E\left[w_e^{(\sigma)}(\varepsilon)\;\varepsilon^2\right]}{E\left[w_e^{(\sigma)}(\varepsilon)\right]} \tag{S23}$$

and where $w_e^{(\sigma)}(x) = (\psi^{(\sigma)}(x)/x)^2$ and $w_e^{(\sigma)}(0) = \psi^{(\sigma)\prime}(0)$.

The estimation of the variance-covariance parameters $\boldsymbol{\zeta}$ distinguishes between the case where $U_b(\boldsymbol{\zeta})$ is diagonal versus the case where $U_b(\boldsymbol{\zeta})$ is block-diagonal. In the first case, estimating $\boldsymbol{\zeta}$ is essentially a scale estimation problem on $\tilde{\boldsymbol{b}}$ and the estimating equations are the same as Equations (S21)-(S23) above, with $\hat{\varepsilon}_{ij}$ replaced by $\hat{\hat{b}}_{ij}$. For the other case, the estimating equations have to be adapted to take care of the block structure, impacting the mathematical expressions, but keeping the underlying philosophy of the approach.

The computation of the consistency terms $\kappa_e^{(\sigma)}$ in Equation (S21) and its counterpart for the estimation of $\boldsymbol{\zeta}$ is difficult for complicated models. The authors provide two alternatives: an accurate but slow numerical quadrature (DAStau), or a faster direct approximation, which is less accurate (DASvar). The last option is the only available for complex models with correlated random effects with more than one correlation term.

The smoothed Huber's ψ function is the proposal suggested within this framework.

Technicalities of the Robust LGM

Here, we present the M-estimator proposed by **?**. Let the sample $\boldsymbol{y}_1, ..., \boldsymbol{y}_n$ with dimension J from a multivariate distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ. The first step is to estimate robustly these two quantities. The M-estimator of $\boldsymbol{\mu}$ is defined as:

$$\widehat{\boldsymbol{\mu}} = \frac{\sum_{i=1}^{n} w_{i1}\boldsymbol{y}_i}{\sum_{i=1}^{n} w_{i1}}, \tag{S24}$$

and Σ as:

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n} w_{i2}(\boldsymbol{y}_i - \widehat{\boldsymbol{\mu}})(\boldsymbol{y}_i - \widehat{\boldsymbol{\mu}})' \tag{S25}$$

where $w_{i1} = w_1(d_i)$ and $w_{i2} = w_2(d_i)$ are Huber weights defined by:

$$w_1(d_i) = \begin{cases} 1 & \text{if} \quad d_i \leq 1 \\ \kappa/d_i & \text{if} \quad d_i > \kappa \end{cases} \quad \text{and } w_2(d_i) = {w_1}^2/\tau, \tag{S26}$$

where d_i is the Mahalanobis distance defined as:

$$d_i^2 = (\boldsymbol{y}_i - \widehat{\boldsymbol{\mu}})'\widehat{\boldsymbol{\Sigma}}(\boldsymbol{y}_i - \widehat{\boldsymbol{\mu}}). \tag{S27}$$

Let define $q = (\chi^2_J)^{-1}(1-\varphi)$, the (1-$\varphi$)-quantile of the χ^2_J distribution, then $\kappa = \sqrt{q}$ and

$$\tau = \left(J \cdot \mathbb{P}(\chi^2_{J+2} \leq q) + q \cdot \varphi\right)/J, \tag{S28}$$

where $\mathbb{P}(\chi^2_{J+2} \leq q)$ is the probability that a χ^2_{J+2} distributed random variable is smaller or equal to q.

The estimates $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ are computed iteratively, because w_{i1} and w_{i2} depend on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. In a second step the LGM is estimated by ML applied on these robust estimates $\widehat{\boldsymbol{\mu}}$ and $\widehat{\Sigma}$.

The Tolerance Data Application

The parameter values used in the simulation study are inspired by the `Tolerance` dataset, often used for didactic purposes (**?**), and originally discussed in **?**. During five years, once per year (from `time` $= 0$ to `time` $= 4$), a sample of $N = 16$ children (identified by the variable `id`, with initial age of 11 years) scored their engagement in nine deviant behaviors, each rated on a four-point scale (from 1 to 4). The original authors analyzed the mean behavior score across the nine scales as continuous, and called it `tolerance`. For didactic purposes, **?** created a dichotomous variable `group`, based on the median-split of the `exposure` variable, which is the mean of the proportions of the respondents' self-reported close friends involved in the nine deviant behaviors. The low and high exposure children had a `group` value of 0 and 1, respectively. The data set has no missing values and is balanced, with each subject i contributing exactly five scores ($J_i = J = 5$).

To depict the group difference we may resort to grouped boxplots, typically presented in the RMANOVA framework. Figure S1 shows how the two groups differ at each time point. Subjects in the `group` $= 0$ and `group` $= 1$ (8 observations in each group) are presented in light grey/empty circles and solid black/filled circles, respectively. We can clearly see that, across time, the latter group has greater `tolerance` scores than the former. Another advantage of this representation is the focus on possible outliers, such as a `tolerance` value of nearly 2.0 in `group` $= 0$ at `time` $= 0$, or 3.5 in `group` $= 1$ at `time` $= 3$.

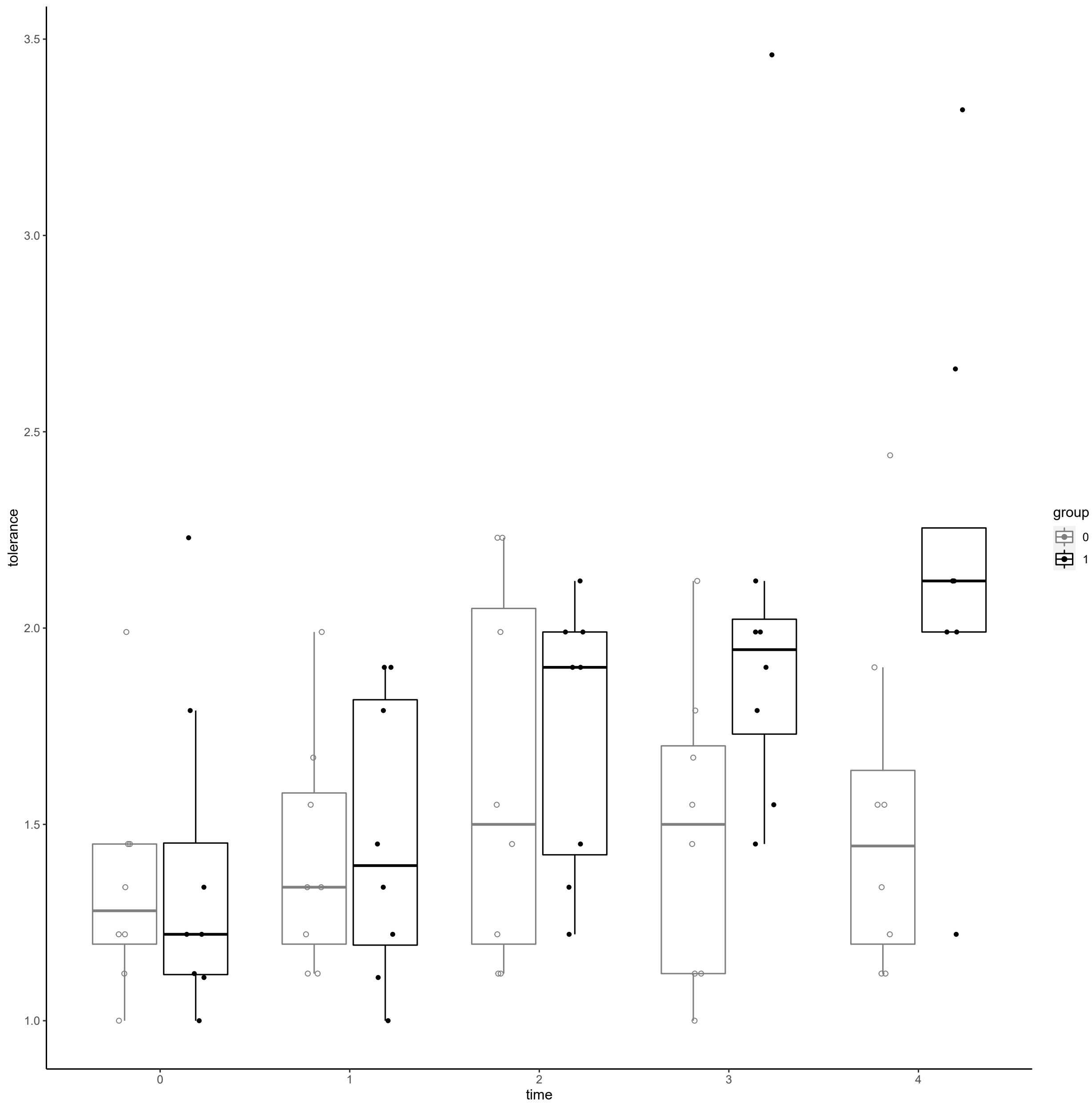

Figure S1
The Tolerance data represented by grouped boxplots. The tolerance to deviant behavior (`tolerance`) as a function of waves of assessment (`time`) and `group`: 0 in grey/empty circles and 1 in black/filled dots.

Another representation of subject-specific repeated assessments is presented in Figure S2. Here, we see each subject's measurements (empty circles for `group` = 0, filled dots for `group` = 1) and predicted individual linear trajectories across their points (dotted lines). This illustration is typical of the LMM and LGM framework, where a parameter represents the average baseline starting level, and another the implied linear change, or mean slope. We can see that for subjects in `group` = 0 (two top rows) their trajectories tend either to stay stable (flat), or decrease slightly (negative slope), whereas those in `group` = 1 appear to have increasing trajectories. Clearly, the two groups seem to differ with respect to their slopes, but not on their starting levels, as would be expected in a randomized two-group comparison study. To emphasize this, we plotted with solid lines the group-implied trajectories, gray for `group` = 0 and black for `group` = 1. We also plotted each individual's expected linear trajectory with dotted lines (more below).

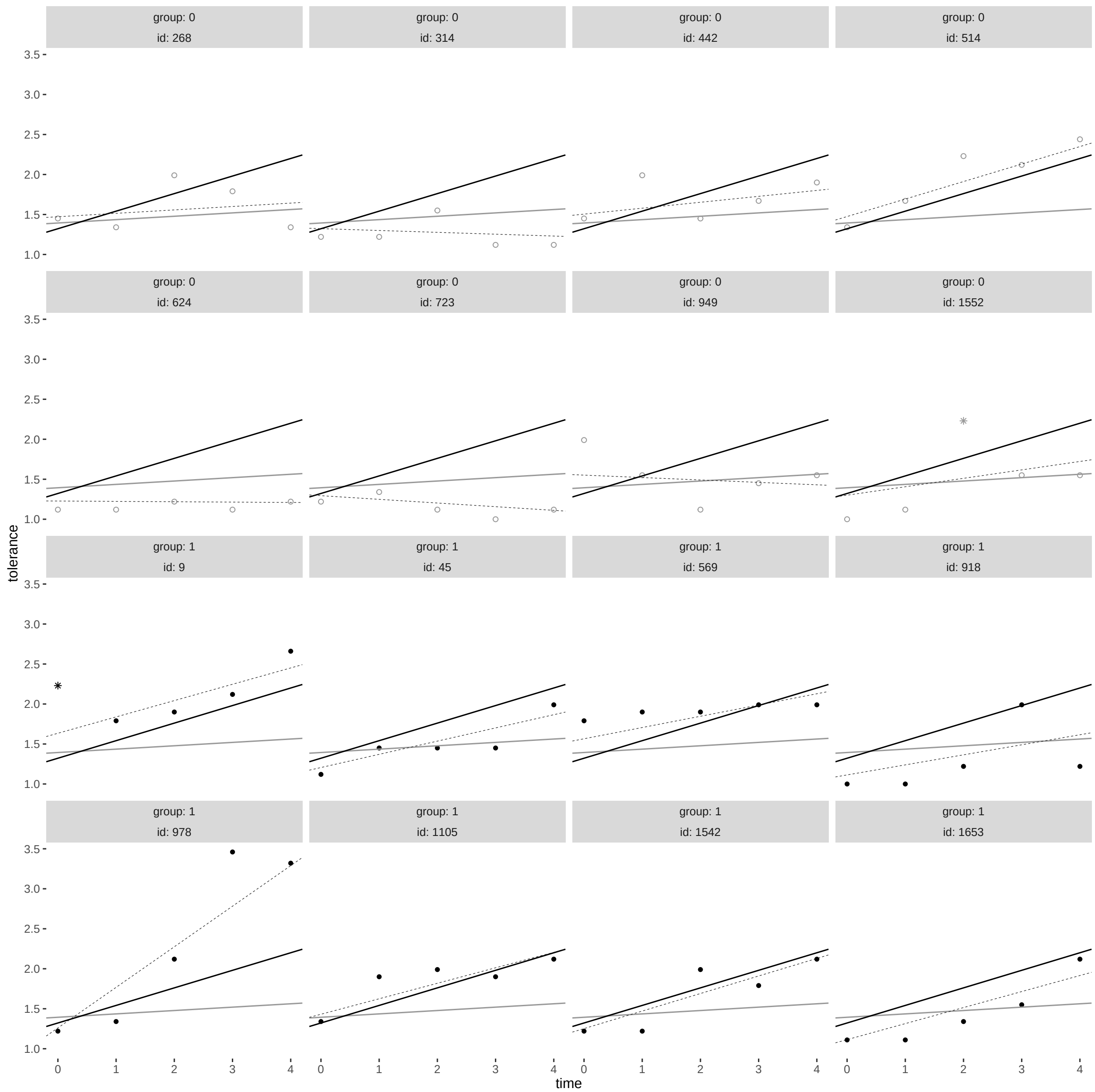

Figure S2

*The individual tolerance to deviant behavior (`tolerance`) as a function of waves of assessment (`time`) and level of exposure to deviant behavior the first year of study (`group = 0` in grey and `group = 1` in black). Subjects are illustrated in separate panels. The regression lines are based on the Maximum Likelihood (ML) estimates of the model in Equation (**??**). The solid lines are the predicted average trajectories for individuals of `group = 0` (in grey) and of `group = 1` (in black). The dotted lines represent the individual predicted trajectories.*

Application of the LMM.

In Equations (**??**) and (**??**) we introduced the full set of parameters to be estimated with LMM in the chosen design, that is $(\boldsymbol{\gamma}, \sigma^2_\varepsilon, \boldsymbol{\theta}) = (\gamma_0, \gamma_1, \gamma_2, \gamma_3, \sigma^2_\varepsilon, \sigma^2_0, \sigma^2_1, \sigma_{10})$. The eight parameter estimates obtained by each LMM-estimator are presented in Table S1. Note that DAStau did not converge. Its author reported computing difficulties with either complex $\Sigma(\boldsymbol{\theta})$ variance matrices or small elements therein, as in this data set, for which the estimate of σ_1^2 is much smaller than that of σ_0^2 (**?**). Therefore, for DAStau, Table S1 contains estimates for the model with no random slope effects (σ_1^2) nor covariance (σ_{10}). The estimates for the fixed effects $\boldsymbol{\gamma}$ were quite similar across all estimation methods.

Application of the LGMr

We estimated the parameters of the model illustrated in Figure **??** using the **?**'s robust M-estimator. The obtained estimates are presented in Table S1. The estimates for the fixed effects and variance components were virtually equal to the ML estimates. The function `sem` also estimates SE for all parameters, including the variance components. For the fixed effects, the SE are extremely similar to those obtained in the LMM approach (ML, but also S and cTAU). Compared to cTAU, the SE of σ_0^2 was similar, but those of σ_1^2 and σ_{10} were bigger.

Table S1

Parameter and standard error estimates of model from Equations (??) and (??) for the `Tolerance` data.

	ML	S	cTAU	DAStau*	LGMr
γ_0	1.39 (0.11)	1.47 (0.1)	1.33 (0.08)	1.36 (0.12)	1.39 (0.10)
γ_1	-0.07 (0.15)	-0.11 (0.15)	-0.09 (0.15)	-0.38 (0.27)	-0.07 (0.15)
γ_2	0.04 (0.05)	0.04 (0.04)	0.04 (0.04)	0.04 (0.03)	0.04 (0.05)
γ_3	0.18 (0.08)	0.12 (0.06)	0.16 (0.07)	0.16 (0.05)	0.18 (0.07)
σ^2_ε	0.07 (NA)	0.096 (NA)	0.068 (NA)	0.09 (NA)	0.08 (0.02)
σ^2_0	0.05 (NA)	0.07 (0.111)	0.02 (0.041)	0.06 (NA)	0.04 (0.03)
σ^2_1	0.02 (NA)	0.005 (0.005)	0.002 (0.004)	-	0.01 (0.01)
σ_{10}	-0.01 (NA)	0.003 (0.008)	0.009 (0.007)	-	-.004 (0.012)

Note. ML = maximum likelihood; S = Copt and Victoria-Feser's S-Estimator; cTAU = Agostinelli and Yohai's composite τ estimator; DAStau = Koller's DAStau estimator; LGMr = Yuan and Zhong's M-Estimator. Entries represent the parameter estimates, with standard errors in parentheses. "NA" indicates that results are not available. "-" indicates that the parameter could not be estimated in the model including it. *The estimates are obtained from the model without random slope effects.

Additional Simulation Results

In this section we present supplemental results, with a larger effect size for the main parameter of interest ($\gamma_3 = 0.250$, see Figures S3), including the Kenward-Roger (KR) correction associated to ML (see also Figure S4). In sum, we can see that the pattern of results from the simulation study replicates with a bigger effect size for γ_3 and that in this setting, the Satterthwaite and the Kenward-Roger approximations obtain virtually identical results.

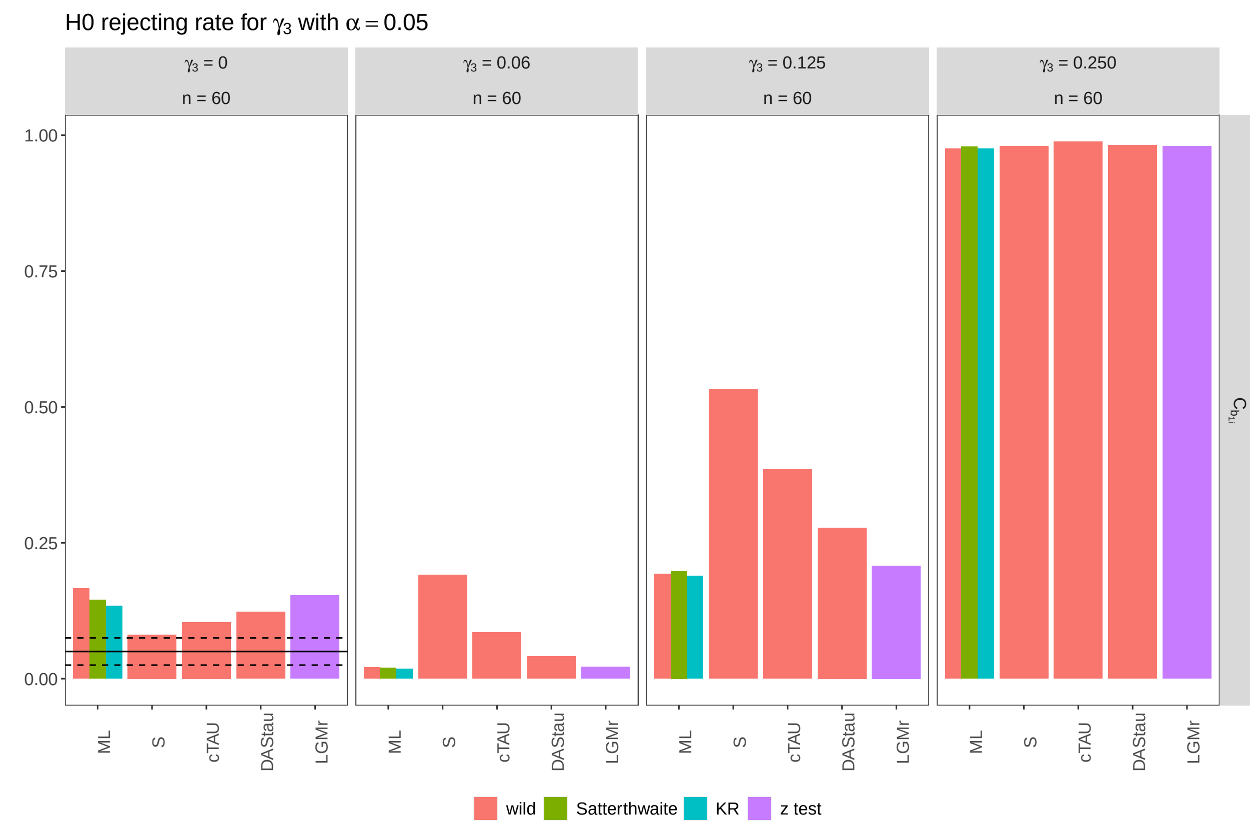

Figure S3

Rejection rates associated to γ_3 in $C_{b_{1i}}$ with $\alpha = 0.05$ as a function of tests/CIs, estimators, γ_3 and N (in columns). Note: ML = maximum likelihood; S = Copt and Victoria-Feser's S-Estimator; cTAU = Agostinelli and Yohai's composite τ; DAStau = Koller's DAStau; LGMr=Yuan and Zhong's M-estimator. Estimators are displayed on the abscissa and tests/CIs are represented by colors.

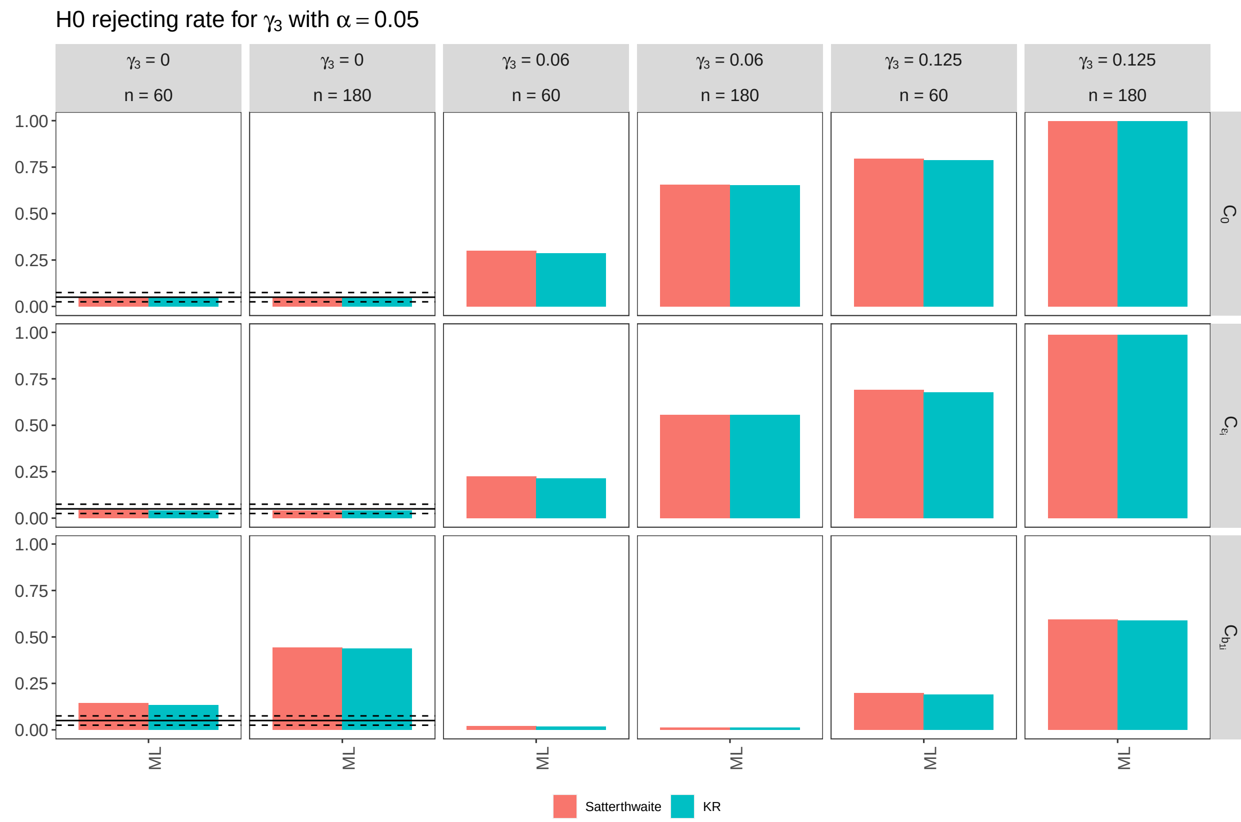

Figure S4

Rejection rates associated to γ_3 with $\alpha = 0.05$ as a function of tests (the t test with Satterthwaite correction in red and with the Kenward and Roger correction in blue), C (in rows), γ_3 and N (in columns). Note: ML = maximum likelihood.