

Translating default priors from linear mixed models to repeated-measures ANOVA and
paired t -tests

Frederik Aust, Johnny van Doorn, & Julia M. Haaf

University of Amsterdam

The work reported here is ongoing.

Frederik Aust, Johnny van Doorn, and Julia Haaf, Psychological Methods, Psychology Department, University of Amsterdam.

Frederik Aust was supported by an Advanced ERC grant to Eric-Jan Wagenmakers (743086 UNIFY) and Julia M. Haaf was supported by a Veni grant from the NWO (VI.Veni.201G.019).

All code and intermediate results are available at <https://github.com/crsh/bf-defpfit/>.

The authors made the following contributions. Frederik Aust: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Software, Visualization, Writing - original draft; Johnny van Doorn: Software, Validation, Writing - original draft; Julia M. Haaf: Funding acquisition, Methodology, Supervision, Validation, Writing - original draft.

Correspondence concerning this article should be addressed to Frederik Aust, Psychology Department, University of Amsterdam, Nieuwe Achtergracht 129-B, 1018 WT Amsterdam, The Netherlands. E-mail: f.aust@uva.nl

Translating default priors from linear mixed models to repeated-measures ANOVA and paired t -tests

van Doorn, Aust, Haaf, Stefan, and Wagenmakers (2021) show that Bayes factors that quantify evidence for fixed effects in mixed models with standardized effect size parameterization (Rouder, Morey, Speckman, & Province, 2012) change when responses are aggregated. Their example assumed a simple repeated-measures design with I participants responding K times in each of two conditions. The maximal model for these data has random intercepts α_i and random slopes θ_i ,

$$y_{ijk} \sim \mathcal{N}(\mu + \sigma_\epsilon(\alpha_i + x_j(\nu + \theta_i)), \sigma_\epsilon^2)$$

$$\alpha_i \sim \mathcal{N}(0, g_\alpha)$$

$$\nu \sim \mathcal{N}(0, g_\nu)$$

$$\theta_i \sim \mathcal{N}(0, g_\theta)$$

$$g_\alpha \sim \mathcal{IG}(0.5, 0.5 \, r_\alpha^2)$$

$$g_\nu \sim \mathcal{IG}(0.5, 0.5 \, r_\nu^2)$$

$$g_\theta \sim \mathcal{IG}(0.5, 0.5 \, r_\theta^2)$$

$$(\mu, \sigma_\epsilon^2) \propto 1/\sigma_\epsilon^2.$$

Heck and Bockting (2021) speculate that it an approximate adjustment to the prior may be based known decrease in the error variance as a result of aggregating n trials, $\sigma'_\epsilon = \frac{\sigma_\epsilon}{\sqrt{n}}$. Singmann et al. (2022) argue that an appropriate adjustment of priors for aggregation is difficult.

It is important to distinguish between two types of aggregation: complete and partial aggregation. Complete aggregation refers to the case where participants contribute only one observation to each level of the factor of interest. Partial aggregation, in contrast, reduces the number of observations but leaves at least two observation for each level of the factor. For illustration, consider the case of a 2×2 repeated-measures design: Following aggregation, each participant contributes one observation to each cell of the design. While this constitutes complete aggregation for the interaction term, multiple observation per participant remain for the analyses of the main effects—observations are pooled over levels of the other factor. Both types of aggregation affect residual variance differently.

Partial aggregation

Because in the models considered here priors are placed on standardized effect sizes, a reduction of σ_ϵ increases prior plausibility of relatively large effects. The mixed model becomes

$$y_{ijk} \sim \mathcal{N}(\mu + \sigma_\epsilon(\alpha_i + x_j(\nu + \theta_i)), \sigma_\epsilon^2)$$

$$\alpha_i \sim \mathcal{N}(0, g'_\alpha)$$

$$\nu \sim \mathcal{N}(0, g'_\nu)$$

$$\theta_i \sim \mathcal{N}(0, g'_\theta),$$

which implies the following priors,

$$\nu \sim \mathcal{N}(0, g_\nu/\sqrt{n})$$

$$g'_\alpha \sim \mathcal{IG}(0.5, 0.5 \, r_\alpha^2/\sqrt{n})$$

$$g'_\theta \sim \mathcal{IG}(0.5, 0.5 \, r_\theta^2/\sqrt{n}).$$

This suggests that equivalent Bayes factors can be obtained by adjusting the prior scales accordingly, $r\iota^2 = r^2\sqrt{n}$. To test whether this prior adjustment works as intended across all levels of aggregation, we conducted a simulation.

Simulation

We simulated $K = 100$ trials for $I = 20$ participants ($\mu = 1$; $\sigma_\alpha = 0.50$; $\nu = \{0, 0.2, 0.4\}$; $\sigma_\theta = \{0.1, 0.25, 0.5, 1, 2, 4, 10\}$; $\sigma_\epsilon = 1$). For each combination of simulation parameters, we created six datasets of varying levels of aggregation, $n = \{1, 2, 5, 10, 25, 50\}$, where $n = 1$ indicates no aggregation and $n = 50$ indicates that of the initial 100 trials two means remain per cell. Like van Doorn et al. (2021), we generated the data deterministically such that condition and participant means as well as standard errors were identical across all levels of aggregation. Bayes factors quantify evidence for the maximal model against a model that omits the fixed effect of condition, $\nu = 0$.

Results. Our results confirm that the prior adjustment works reasonably well, Figure 1. However, when an effect was present, the random slope variance σ_θ^2 was substantially smaller than the error variance σ_ϵ^2 , and when only two observations per cell remained, we found Bayes factors to be inflated. This inflation, however, was negligible for small and inconsequential for large Bayes factors. Thus, in most cases the adjustment of priors for partial aggregation is simple and effective.

Complete aggregation

Adjusting priors for complete aggregation is more challenging. When data are fully aggregated data (i.e., $n = K$), the random slopes variance σ_θ^2 becomes confounded with the error variance σ_ϵ^2 . In mixed models the random slope variance is characterized by a probability distribution, which prohibits an exact adjustment of the prior by a simple scaling constant. Here, we explore the adequacy of an approximate adjustment using a point value for the random slope variance.

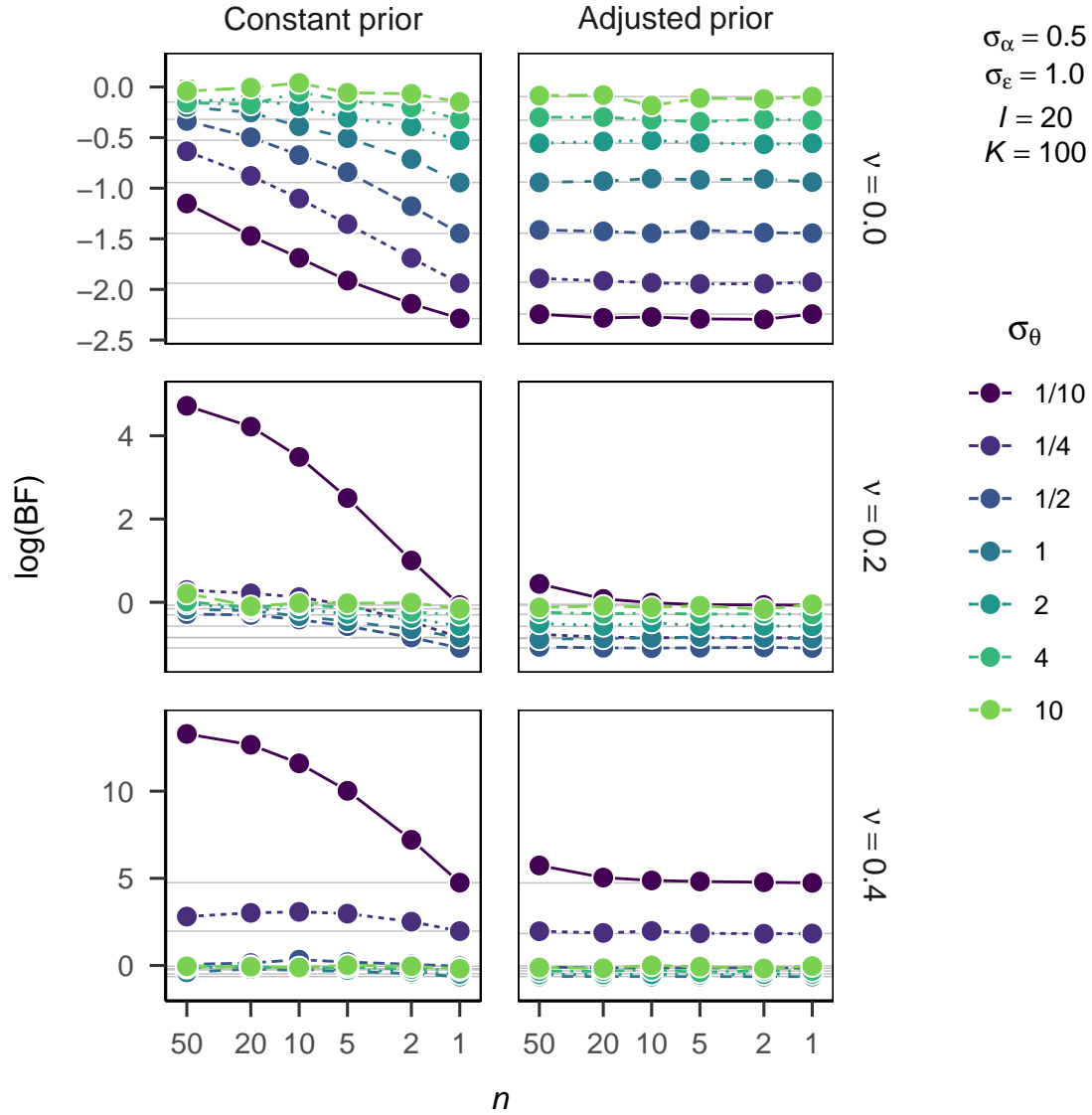


Figure 1. Horizontal lines represent log BF for each level of σ_θ with $n = 1$ (no aggregation) as reference. ν denotes the effect size, σ_α the random slopes variance, σ_θ the random slopes variance, σ_ϵ the error variance, and I the number of participants, and K the number of trials before aggregation.

When aggregating each participant's data to a single observation per cell, the data can analyzed in two ways: By modeling participants' (1) cell means using a one-way repeated-measures ANOVA, or by modeling participants' (2) cell mean differences using a paired t -test.

Repeated-measures ANOVA

Aggregation reduces the maximal model to the following,

$$\begin{aligned}\bar{y}_{ij} &\sim \mathcal{N}(\mu + \sigma_{\epsilon}(\alpha_{\epsilon i} + x_j \nu_{\epsilon}), \sigma_{\epsilon}^2 + \sigma_{\theta}^2/2) \\ \alpha_{\epsilon i} &\sim \mathcal{N}(0, g_{\alpha} \sqrt{\sigma_{\theta}^2}) \\ \nu_{\epsilon} &\sim \mathcal{N}(0, g_{\nu} \sqrt{\sigma_{\theta}^2/2}),\end{aligned}$$

where x_j is an indicator of the condition using orthonormal effect coding, $\pm\sqrt{2}/2$; the random slopes variance σ_{θ}^2 is scaled by the coding used for x_j .

Compared to partial aggregation, the adjustment for the fixed effect requires an additional factor that depends on a weighted ratio of random variance σ_{θ}^2 and error variance σ_{ϵ}^2 ,

$$\begin{aligned}\sqrt{\sigma_{\epsilon}^2/K + \sigma_{\theta}^2/2} &= \sigma_{\epsilon}/\sqrt{K} \sqrt{1 + \frac{K\sigma_{\theta}^2}{2\sigma_{\epsilon}^2}} \\ &= \sigma_{\epsilon}/\sqrt{K} \sqrt{\frac{2\sigma_{\epsilon}^2 + K\sigma_{\theta}^2}{2\sigma_{\epsilon}^2}} \\ &= \sigma_{\epsilon}/\sqrt{K} \sqrt{\frac{2/K + \sigma_{\theta}^2/\sigma_{\epsilon}^2}{2/K}}\end{aligned}$$

For random intercepts, the additional correction factor is obtained by marginalizing over the dummy coded random effect, yielding a weight of 1 for the random slope variance.

85 *t*-Test

86 When analyzing participants' cell mean differences, the random intercept variance
 87 cancels out and requires no prior or adjustment. The prior adjustment for the fixed effect,
 88 however, must account for the different effect coding schemes, i.e. ± 0.5 vs. $\pm\sqrt{2}/2$,

$$\begin{aligned}\bar{y}_{ij\cdot} - \bar{y}_{i\cdot} &\sim \mathcal{N}(\mu + \sigma\iota_{\epsilon}(\alpha_i + 0.5\nu), \sigma\iota_{\epsilon}^2) \\ &\quad - \mathcal{N}(\mu + \sigma\iota_{\epsilon}(\alpha_i - 0.5\nu), \sigma\iota_{\epsilon}^2) \\ &= \mathcal{N}(\sigma\iota_{\epsilon}\nu, 2\sigma\iota_{\epsilon}^2)\end{aligned}$$

$$\nu \sim \mathcal{N}(0, g_{\nu}\sqrt{\sigma_{\theta}^2/4}).$$

89 Rescaling the prior on ν to the orthonormal scale, $\sqrt{2}\nu$, yields the same adjustment as
 90 for the prior in the repeated-measures ANOVA,

$$\begin{aligned}\sqrt{\sigma_{\epsilon}^2/K + \sigma_{\theta}^2/2} &= \sigma_{\epsilon}/\sqrt{K} \sqrt{1 + \frac{K\sigma_{\theta}^2}{2\sigma_{\epsilon}^2}} \\ &= \sigma_{\epsilon}/\sqrt{K} \sqrt{\frac{2\sigma_{\epsilon}^2 + K\sigma_{\theta}^2}{2\sigma_{\epsilon}^2}} \\ &= \sigma_{\epsilon}/\sqrt{K} \sqrt{\frac{2/K + \sigma_{\theta}^2/\sigma_{\epsilon}^2}{2/K}}.\end{aligned}$$

91 As noted above in the linear mixed model σ_{θ}^2 is characterized by a probability
 92 distribution. To test whether the exact adjustment can be approximated by using a point
 93 value in the equation above, we conducted a second simulation.

94 Simulation

95 We randomly simulated $K = \{5, 25, 100\}$ responses for $I = \{20, 50, 100, 200\}$
 96 participants ($\mu = 1$; $\sigma_{\alpha} = 0.50$; $\nu = \{0, 0.2, 0.5\}$; $\sigma_{\theta} = 0.50$; $\sigma_{\epsilon} = \{0.1, 0.25, 0.5, 1, 2\}$) 10

times each. We analyzed the data using mixed models and conducted corresponding repeated-measures ANOVA and paired t -test as aggregate analyses. Again, Bayes factors quantify evidence for the maximal model against a model that omits the fixed effect of condition, $\nu = 0$.

Results. When the error variance σ_ϵ^2 was small or the sample size I was large, the adjustments worked reasonably well, left panels of Figures 2 and 3. However, compared to the mixed model, in many cases both aggregate analyses produced substantially diverging Bayes factors. The divergence increased as (1) the difference in random slope and error variance increased and (2) the number of participants decreased. For a detailed breakdown of the results by all varied factors see the online supplementary material at <https://github.com/crsh/bf-defprit/>. The divergence of Bayes factors was more pronounced in the repeated-measures ANOVA than in the paired t -test, because priors on both fixed and random intercepts require adjustment.

In an exploratory analysis of the simulation results, we compared several regression models (using AIC and BIC) that expressed the Bayes factors from the mixed model (weighted by the reciprocal of their estimation error) as a function of Bayes factors from the corresponding paired t -test, sample size I , and error variance σ_ϵ^2 . For details refer to the online supplementary material at <https://github.com/crsh/bf-defprit/>. Next, we compared the same set of models for the Bayes factors from repeated-measures ANOVA and found the same winning model. The Bayes factors predicted by the selected model,

$$\begin{aligned} \log \text{BF}_{\text{LMM}} = & b_1 \sigma_\epsilon + b_2 \sqrt{\sigma_\epsilon} + b_3 \sqrt{I} + b_4 \sqrt{I} \sigma_\epsilon + \\ & \log \text{BF}(1 + b_5 \sigma_\epsilon^2 + b_6 \log I + b_7 \sigma_\epsilon^2 \log I), \end{aligned}$$

largely offset divergences between linear mixed model and aggregate analyses, right panels of Figures 2 and 3.

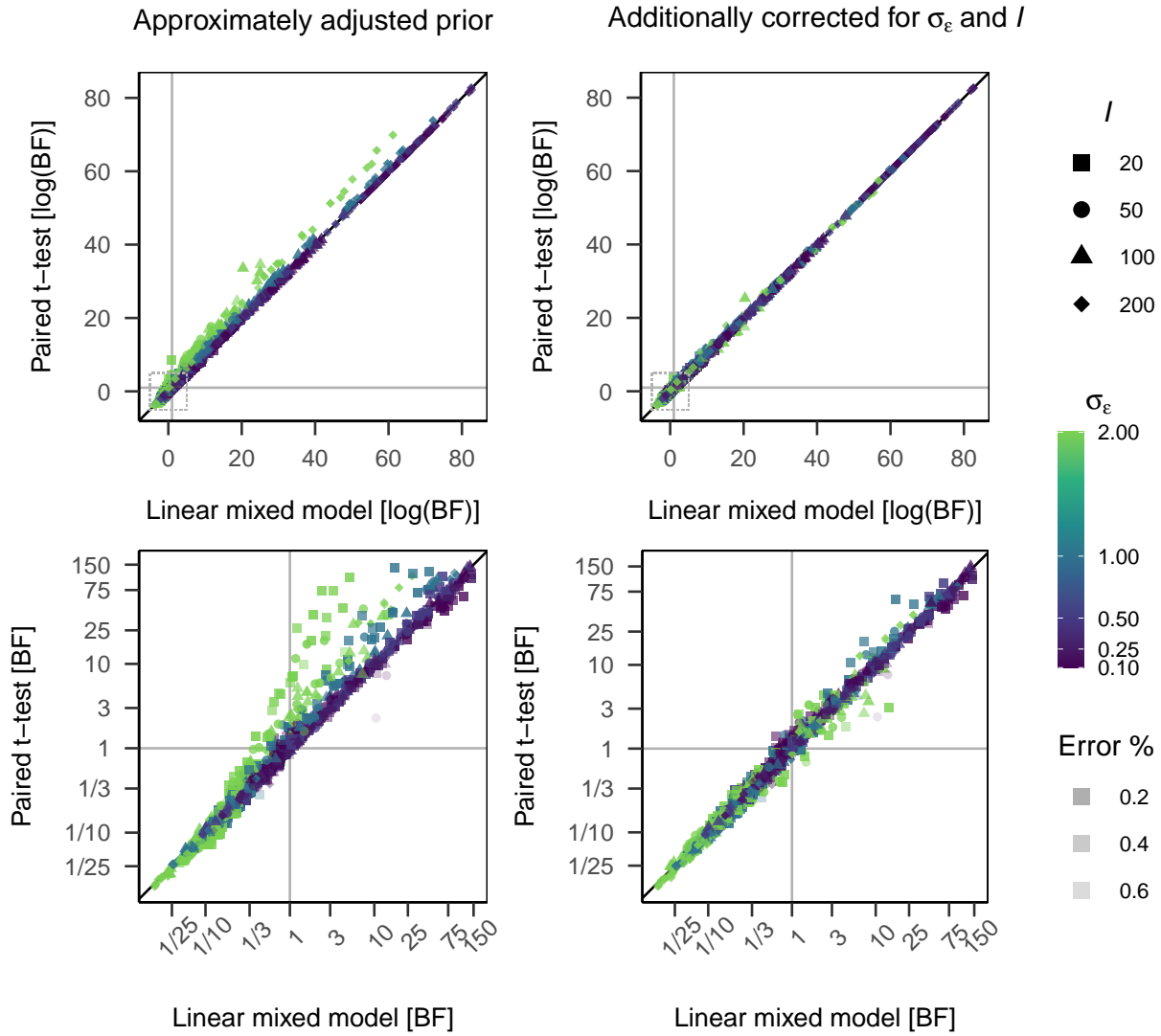


Figure 2. Simulation results showing the effect of the approximate adjustment of the prior scale in paired t -tests on Bayes factors compared to the corresponding mixed model (left) and an additional correction of those Bayes factors by error variance σ_ϵ^2 and number of participants I . The transparency of points represents the reciprocal of the estimation error for mixed model Bayes factors (log scale). The grey dotted square (top) highlights the portion of the plot that is shown in more detail in the bottom panels.

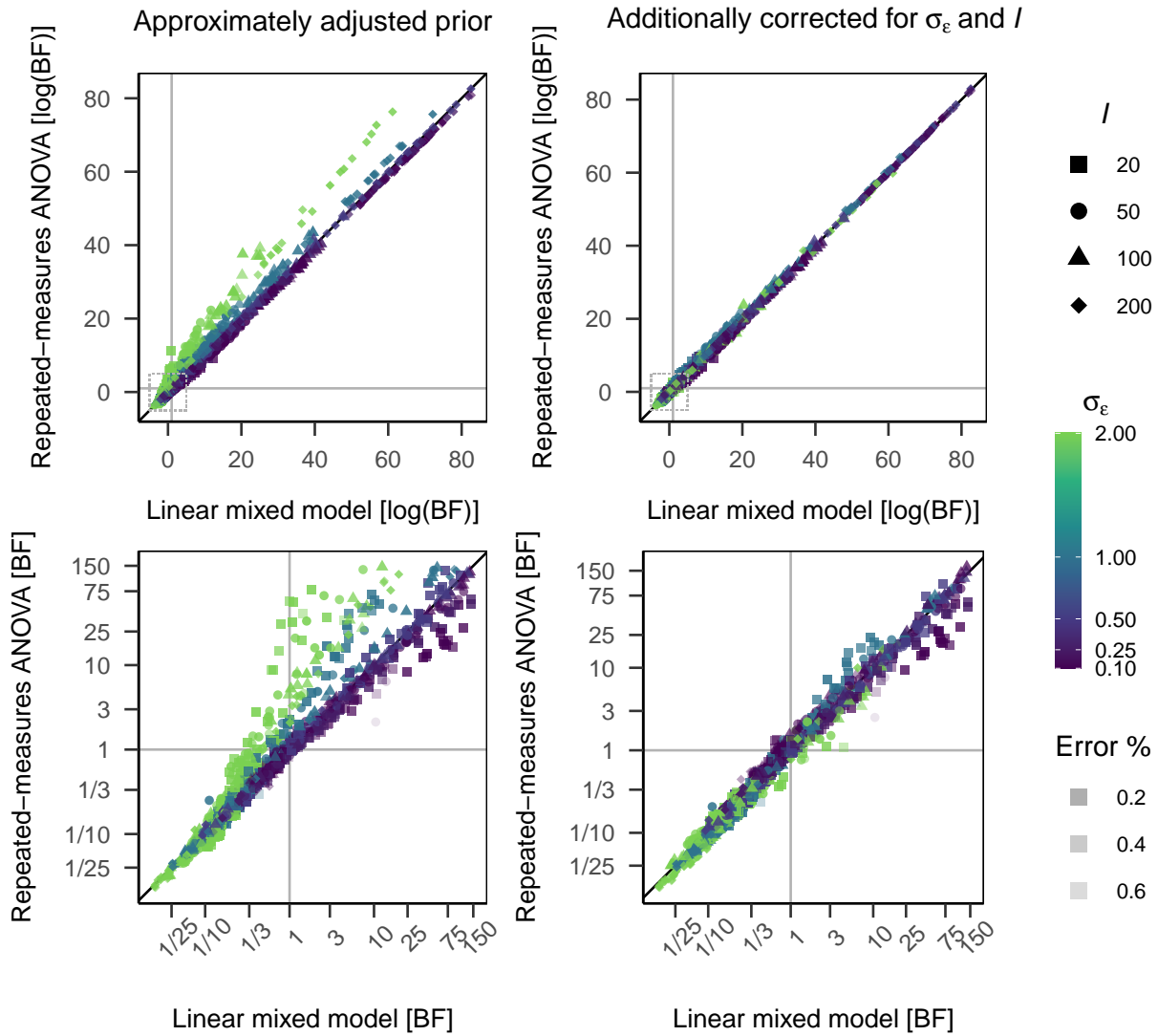


Figure 3. Simulation results showing the effect of the approximate adjustment of the prior scale in repeated-measures ANOVAS on Bayes factors compared to the corresponding mixed model (left) and an additional correction of those Bayes factors by error variance σ_ϵ^2 and number of participants I . The transparency of points represents the reciprocal of the estimation error for mixed model Bayes factors (log scale). The grey dotted square (top) highlights the portion of the plot that is shown in more detail in the bottom panels.

Discussion

We explored approximate translations of default priors between mixed models, repeated-measures ANOVA, and paired t -tests to obtain equivalent Bayes factors. Although these translations worked reasonably well in the case of partial aggregation, they were less successful for complete aggregation. We are currently working on refined adjustments to eliminate the observed discrepancies between mixed models and the corresponding aggregate analyses.

References

- Heck, D. W., & Bockting, F. (2021). Benefits of Bayesian Model Averaging for Mixed-Effects Modeling. *Computational Brain & Behavior*.
<https://doi.org/10.1007/s42113-021-00118-x>
- Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, 56(5), 356–374.
<https://doi.org/10.1016/j.jmp.2012.08.001>
- Singmann, H., Kellen, D., Cox, G. E., Chandramouli, S. H., Davis-Stober, C. P., Dunn, J. C., ... Shiffrin, R. M. (2022). Statistics in the Service of Science: Don't Let the Tail Wag the Dog. *Computational Brain & Behavior*. <https://doi.org/10.1007/s42113-022-00129-2>
- van Doorn, J., Aust, F., Haaf, J. M., Stefan, A. M., & Wagenmakers, E.-J. (2021). Bayes factors for mixed models. *Computational Brain & Behavior*.
<https://doi.org/10.1007/s42113-021-00113-2>