₁ Translating default priors from linear mixed models to repeated-measures ANOVA and

₂ paired t-tests

₃ Frederik Aust, Johnny van Doorn, & Julia M. Haaf

₄ University of Amsterdam

₅ The work reported here is ongoing.

₆ Author Note

⁷   Frederik Aust, Johnny van Doorn, and Julia Haaf, Psychological Methods, Psychology

⁸ Department, University of Amsterdam.

¹⁸   Correspondence concerning this article should be addressed to Frederik Aust,

¹⁹ Psychology Department, University of Amsterdam, Nieuwe Achtergracht 129-B, 1018 WT

²⁰ Amsterdam, The Netherlands. E-mail: f.aust@uva.nl

21   Translating default priors from linear mixed models to repeated-measures ANOVA and

22   paired t-tests

23   van Doorn, Aust, Haaf, Stefan, and Wagenmakers (2021) show that Bayes factors that

24   quantify evidence for fixed effects in mixed models with standardized effect size

25   parameterization (Rouder, Morey, Speckman, & Province, 2012) change when responses are

26   aggregated. Their example assumed a simple repeated-measures design with I participants

27   responding K times in each of two conditions. The maximal model for these data has

28   random intercepts α_i and random slopes θ_i,

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

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

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

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

29   Heck and Bockting (2021) speculate that it an approximate adjustment to the prior

30   may be based known decrease in the error variance as a result of aggregating n trials,

31    $\sigma\prime_\epsilon = \frac{\sigma_\epsilon}{\sqrt{n}}$. Singmann et al. (2022) argue that an appropriate adjustment of priors for

32   aggregation is difficult.

₃₃   It is important to distinguish between two types of aggregation: complete and partial

₃₄ aggregation. Complete aggregation refers to the case where participants contribute only one

₃₅ observation to each level of the factor of interest. Partial aggregation, in contrast, reduces

₃₆ the number of observations but leaves at least two observation for each level of the factor.

₃₇ For illustration, consider the case of a 2×2 repeated-measures design: Following

₃₈ aggregation, each participant contributes one observation to each cell of the design. While

₃₉ this constitutes complete aggregation for the interaction term, multiple observation per

₄₀ participant remain for the analyses of the main effects—observations are pooled over levels of

₄₁ the other factor. Both types of aggregation affect residual variance differently.

Partial aggregation

₄₃   Because in the models considered here priors are placed on standardized effect sizes, a

₄₄ reduction of σ_ϵ increases prior plausibility of relatively large effects. The mixed model

₄₅ becomes

$$y\prime_{ijk} \sim \mathcal{N}(\mu + \sigma\prime_\epsilon(\alpha\prime_i + x_j(\nu\prime + \theta\prime_i)), \sigma\prime_\epsilon^2)$$

$$\alpha\prime_i \sim \mathcal{N}(0, g\prime_\alpha)$$
$$\nu\prime \sim \mathcal{N}(0, g\prime_\nu)$$
$$\theta\prime_i \sim \mathcal{N}(0, g\prime_\theta),$$

₄₆   which implies the following priors,

$$\nu\prime \sim \mathcal{N}(0, g_\nu/\sqrt{n})$$
$$g\prime_\alpha \sim \mathcal{IG}(0.5, 0.5 \ r_\alpha^2/\sqrt{n})$$
$$g\prime_\theta \sim \mathcal{IG}(0.5, 0.5 \ r_\theta^2/\sqrt{n}).$$

⁴⁷   This suggests that equivalent Bayes factors can be obtained by adjusting the prior

⁴⁸ scales accordingly, $r'^2 = r^2\sqrt{n}$. To test whether this prior adjustment works as intended

⁴⁹ across all levels of aggregation, we conducted a simulation.

Simulation

⁵¹   We simulated $K = 100$ trials for $I = 20$ participants ($\mu = 1$; $\sigma_\alpha = 0.50$;

⁵² $\nu = \{0, 0.20, and 0.40\}$; $\sigma_\theta = \{0.10, 0.25, 0.50, 1, 2, 4, and 10\}$; $\sigma_\epsilon = 1$). For each combination

⁵³ of simulation parameters, we created six datasets of varying levels of aggregation,

⁵⁴ $n = \{1, 2, 5, 10, 25, 50\}$, where $n = 1$ indicates no aggregation and $n = 50$ indicates that of

⁵⁵ the initial 100 trials two means remain per cell. Like van Doorn et al. (2021), we generated

⁵⁶ the data deterministically such that condition and participant means as well as standard

⁵⁷ errors were identical across all levels of aggregation. Bayes factors quantify evidence for the

⁵⁸ maximal model against a model that omits the fixed effect of condition, $\nu = 0$.

⁵⁹   **Results.**   Our results confirm that the prior adjustment works reasonably well,

⁶⁰ Figure 1. However, when an effect was present, the random slope variance σ_θ^2 was

⁶¹ substantially smaller than the error variance σ_ϵ^2, and when only two observations per cell

⁶² remained, we found Bayes factors to be inflated. This inflation, however, was negligible for

⁶³ small and inconsequential for large Bayes factors. Thus, in most cases the adjustment of

⁶⁴ priors for partial aggregation is simple and effective.

Complete aggregation

⁶⁶   Adjusting priors for complete aggregation is more challenging. When data are fully

⁶⁷ aggregated data (i.e., $n = K$), the random slopes variance σ_θ^2 becomes confounded with the

⁶⁸ error variance σ_ϵ^2. In mixed models the random slope variance is characterized by a

⁶⁹ probability distribution, which prohibits an exact adjustment of the prior by a simple scaling

⁷⁰ constant. Here, we explore the adequacy of an approximate adjustment using a point value
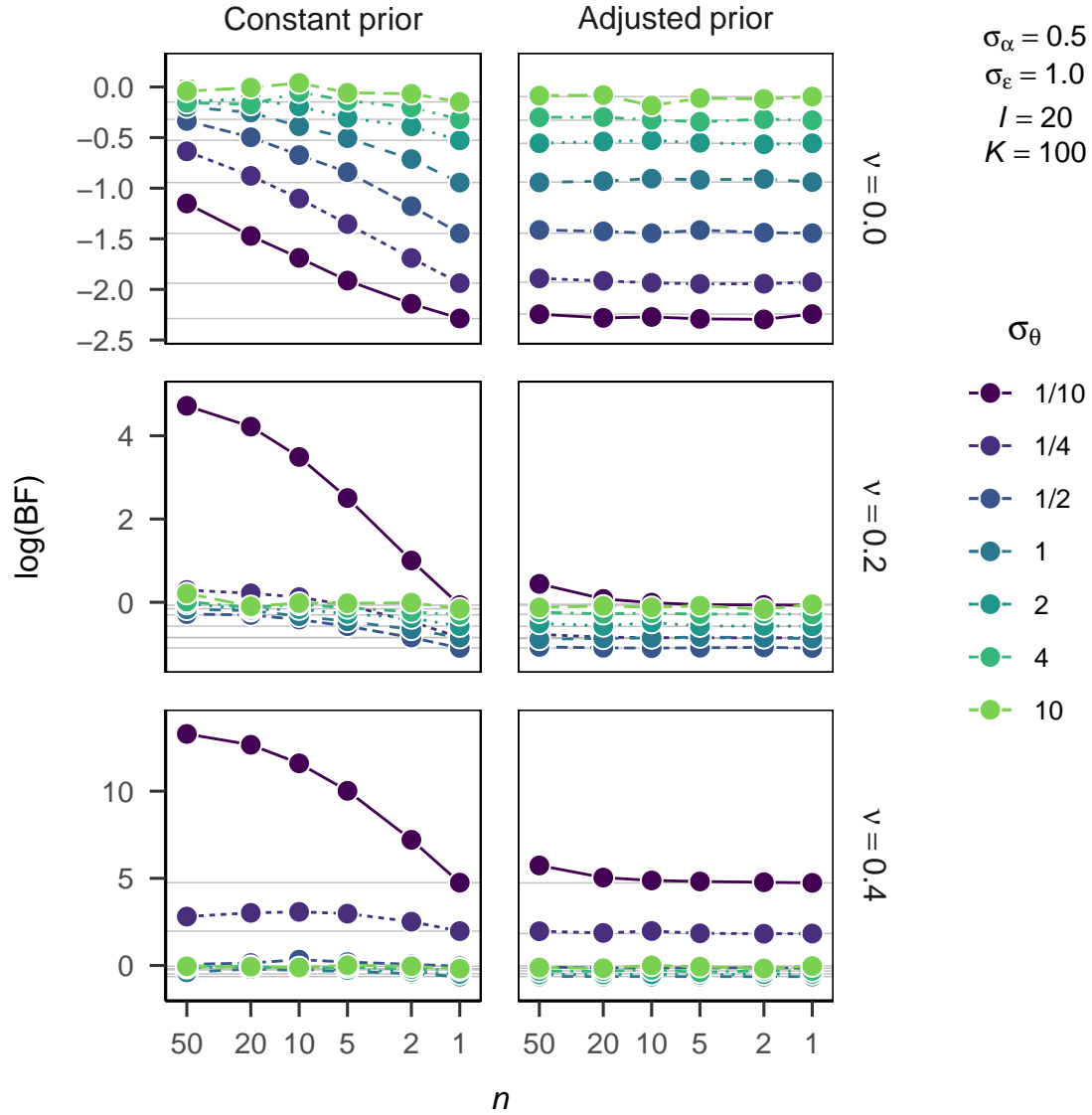
⁷¹ for the random slope variance.

Figure 1. Horizontal lines represent $\log \mathrm{BF}$ 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.

⁷² When aggregating each participant's data to a single observation per cell, the data can

⁷³ analyzed in two ways: By modeling participants' (1) cell means using a one-way

⁷⁴ repeated-measures ANOVA, or by modeling participants' (2) cell mean differences using a

⁷⁵ paired t-test.

⁷⁶ **Repeated-measures ANOVA**

⁷⁷ Aggregation reduces the maximal model to the following,

$$\bar{y}_{ij\cdot} \sim \mathcal{N}(\mu + \sigma\prime_\epsilon(\alpha\prime_i + x_j\nu\prime), \sigma\prime_\epsilon^2 + \sigma_\theta^2/2)$$
$$\alpha\prime_i \sim \mathcal{N}(0, g_\alpha\sqrt{\sigma_\theta^2})$$
$$\nu\prime \sim \mathcal{N}(0, g_\nu\sqrt{\sigma_\theta^2/2}),$$

⁷⁸ where x_j is an indicates the condition using orthonormal effect coding, $\pm\sqrt{2}/2$; the

⁷⁹ random slopes variance σ_θ^2 is scaled by the coding used for x_j.

⁸⁰ Compared to partial aggregation, the adjustment for the fixed effect requires an

⁸¹ additional factor that depends on a weighted ratio of random variance σ_θ^2 and error variance

⁸² σ_ϵ^2,

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

⁸³ For random intercepts, the additional correction factor is obtained by marginalizing

⁸⁴ 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. $\pm\sqrt{2}/2$,

$$\bar{y}_{ij\cdot} - \bar{y}_{ij\cdot} \sim \mathcal{N}(\mu + \sigma\prime_\epsilon(\alpha_i + 0.5\nu), \sigma\prime_\epsilon^2)$$
$$- \mathcal{N}(\mu + \sigma\prime_\epsilon(\alpha_i - 0.5\nu), \sigma\prime_\epsilon^2)$$
$$= \mathcal{N}(\sigma\prime_\epsilon\nu, 2\sigma\prime_\epsilon^2)$$

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

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,

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

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 $K = \{5, 25, and 100\}$ responses for $I = \{20, 50, 100, and 200\}$

96 participants ($\mu = 1$; $\sigma_\alpha = 0.50$; $\nu = \{0, 0.20, and 0.50\}$; $\sigma_\theta = 0.50$;

97    $\sigma_\epsilon = \{0.10, 0.25, 0.50, 1, and 2\}$) 10 times each. We analyzed the data using mixed models

98   and conducted corresponding repeated-measures ANOVA and paired t-test as aggregate

99   analyses. Again, Bayes factors quantify evidence for the maximal model against a model

100   that omits the fixed effect of condition, $\nu = 0$.

101   **Results.**   When the error variance σ_ϵ^2 was small or the sample size I was large, the

102   adjustments worked reasonably well, left panels of Figures 2 and 3. However, compared to

103   the mixed model, in many cases both aggregate analyses produced substantially diverging

104   Bayes factors. The divergence increased as (1) the difference in random slope and error

105   variance increased and (2) the number of participants decreased. For a detailed breakdown

106   of the results by all varied factors see the online supplementary material at

107   https://github.com/crsh/bf-defprit/. The divergence of Bayes factors was more pronounced

108   in the repeated-measures ANOVA than in the paired t-test, because priors on both fixed and

109   random intercepts require adjustment.

110   In an exploratory analysis of the simulation results, we compared several regression

111   models (using AIC and BIC) that expressed the Bayes factors from the mixed model

112   (weighted by the reciprocal of their estimation error) as a function of Bayes factors from the

113   corresponding paired t-test, sample size I, and error variance σ_ϵ^2. For details refer to the

114   online supplementary material at https://github.com/crsh/bf-defprit/. Next, we compared

115   the same set of models for the Bayes factors from repeated-measures ANOVA and found the

116   same winning model. The Bayes factors predicted by the selected model,

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

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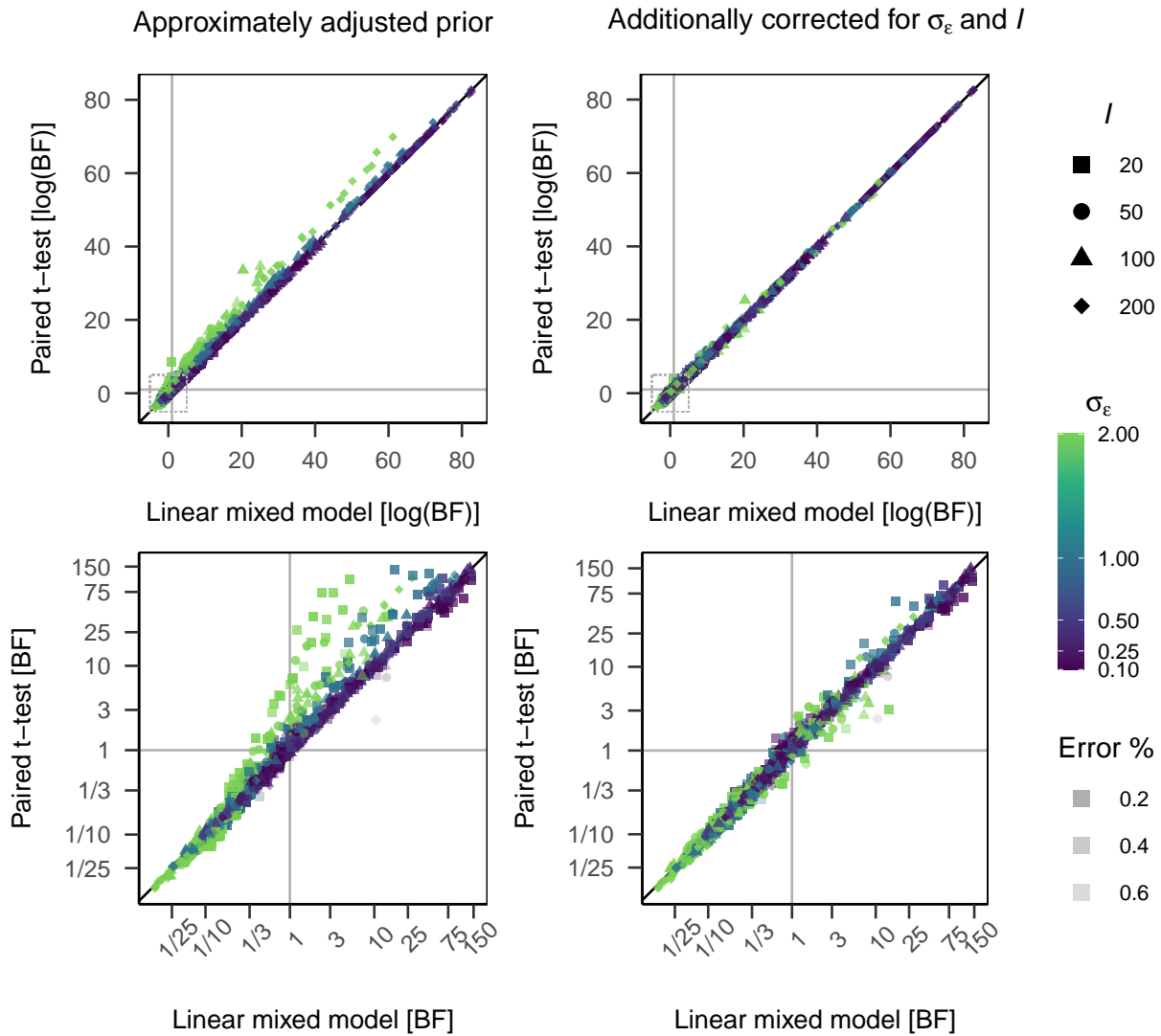
118   panels of Figures 2 and 3.

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.
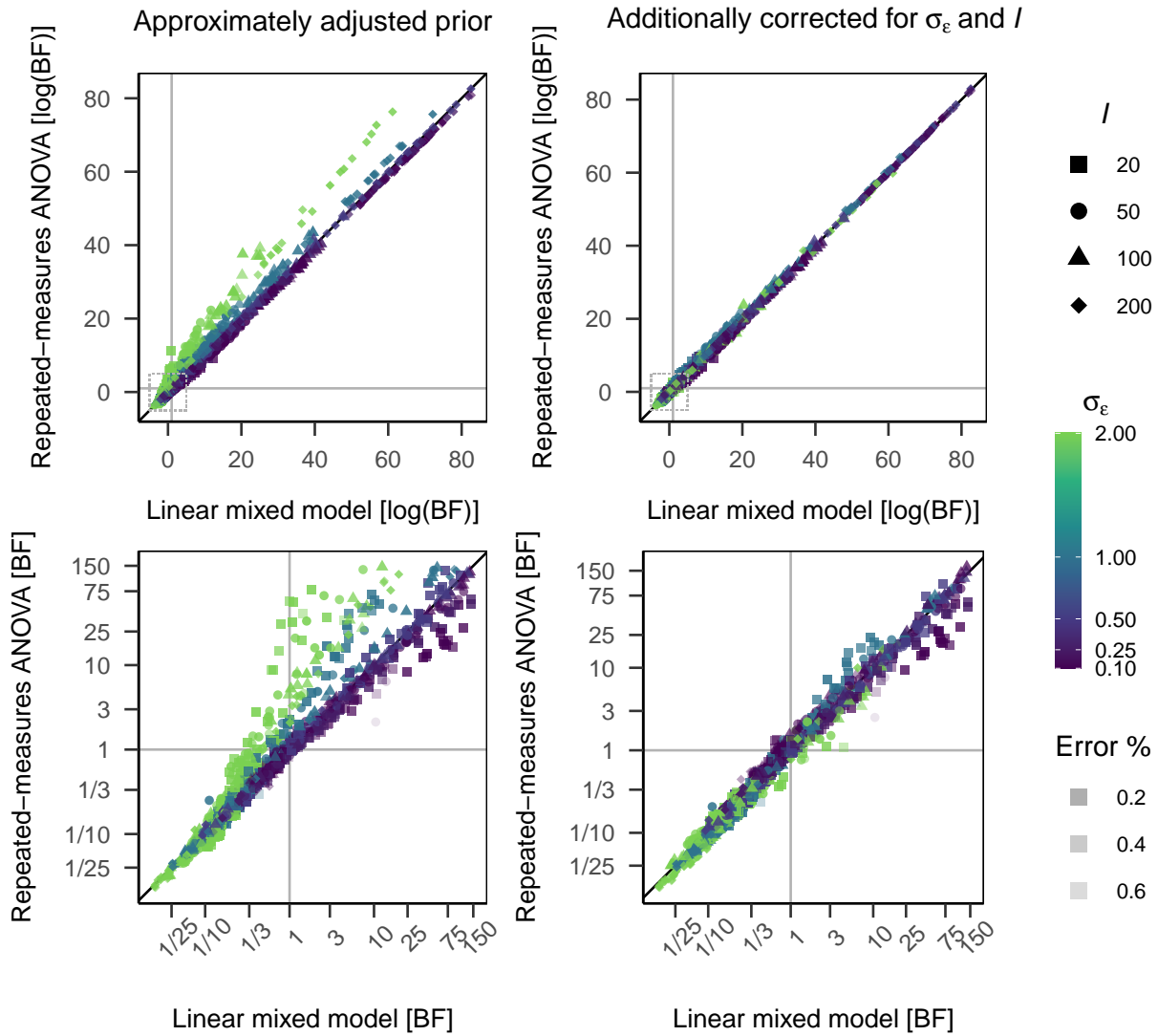
Figure 3. Simulation results showing the effect of the approximate adjustment of the prior scale in repeated-measures ANOVAS on Bayes factors compared to the corresponding mixed model (left) and an additional correction of those Bayes factors by error variance σ_ϵ^2 and number of participants I. The 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

¹¹⁹

¹²⁰   We explored approximate translations of default priors between mixed models,

¹²¹ repeated-measures ANOVA, and paired t-tests to obtain equivalent Bayes factors. Although

¹²² these translations worked reasonably well in the case of partial aggregation, they were less

¹²³ successful for complete aggregation. We are currently working on refined adjustments to

¹²⁴ eliminate the observed discrepancies between mixed models and the corresponding aggregate

¹²⁵ analyses.

References

126

127 Heck, D. W., & Bockting, F. (2021). Benefits of Bayesian Model Averaging for Mixed-Effects

128 Modeling. *Computational Brain & Behavior.*

129 https://doi.org/10.1007/s42113-021-00118-x

130 Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default bayes

131 factors for ANOVA designs. *Journal of Mathematical Psychology, 56*(5), 356–374.

132 https://doi.org/10.1016/j.jmp.2012.08.001

133 Singmann, H., Kellen, D., Cox, G. E., Chandramouli, S. H., Davis-Stober, C. P., Dunn, J. C.,

134 . . . Shiffrin, R. M. (2022). Statistics in the Service of Science: Don't Let the Tail Wag

135 the Dog. *Computational Brain & Behavior.* https://doi.org/10.1007/s42113-022-00129-2

136 van Doorn, J., Aust, F., Haaf, J. M., Stefan, A. M., & Wagenmakers, E.-J. (2021). Bayes

137 factors for mixed models. *Computational Brain & Behavior.*

138 https://doi.org/10.1007/s42113-021-00113-2