ETH
Swiss Federal Institute of Technology Zurich

Seminar for Statistics

Department of Mathematics

Semester Paper Winter 2023

Linus Kühne

Aggregation as a Simple Alternative to Mixed Models

Submission Date: November 15, 2023

Advisor: Dr. Markus Kalisch

Abstract

Real-world data with repeated measures often presents a challenge in statistical analysis due to the presence of "dependence clusters" caused by multiple measurements taken on the same individual. A popular approach to modeling data with repeated measures is to use mixed models, which explicitly account for this dependence structure by incorporating random individual-specific effects into the model.

In this semester project, we compare mixed models with a simpler approach for modeling data with repeated measures. We propose fitting a separate linear regression model to the measurements corresponding to each individual in the data, and then aggregating the estimated slopes across individuals. A t-test allows us to make inference with respect to the population slope via confidence intervals.

Simulation experiments are conducted to compare the performance of the proposed approach of aggregating regressions to mixed models in various situations. The results show that the aggregated regressions yield comparable performance to mixed models in terms of coverage probability and confidence interval width around the true slope, while at the same time being simpler, faster, and more easily interpretable. Additionally, our approach is more robust to some model misspecifications. Mixed models only show better performance in the situations we considered when they can benefit from stricter model assumptions than the aggregated regression approach.

Overall, aggregated regressions provide a simple and effective alternative to mixed models for analyzing repeated measures data.

Keywords: mixed models, repeated measures, linear regression, statistical analysis

Acknowledgement

I extend my gratitude to the supervisor of this semester thesis, Dr. Markus Kalisch. It was a pleasure working on this project with his competent guidance.

The topic of this semester project, the aggregated regression method, and our approach of comparing the different models (using specific simulations as well as statistical tests) were his suggestion.

Contents

Chapter 1

Introduction to Mixed Models

Consider the following scenario. As in a linear regression context, we collect n measurements $y_i \in \mathbb{R}$ for $i = 1, \dots, n$ of a response variable, and corresponding measurements of an explanatory variable $x_i \in \mathbb{R}$ for $i = 1, \dots, n$. A synthetic data set with which we will work is visualized in Fig. 1.1 (left).

A common starting point for modelling such data could be the classical linear model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i,$$

where $\beta_0, \beta_1 \in \mathbb{R}$ are unknown parameters and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ are independent and identically distributed errors with variance $\sigma^2 > 0$. The linear regression line after finding estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ is shown in Fig. 1.1 (right).

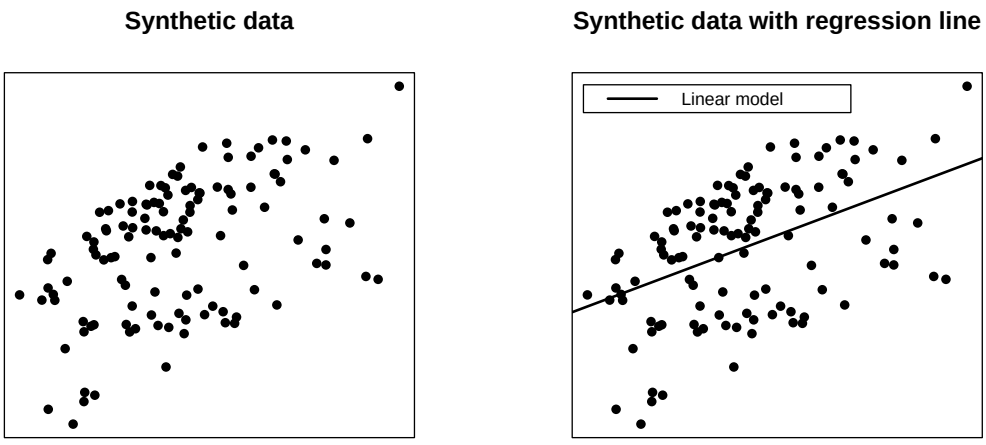


Figure 1.1: Visualizing our data set and the classical linear model fit

However, the classical linear model is not appropriate for this synthetic data set, because this data set mimics data with *repeated measures*. As an example, think of y_i as the reaction time of a person after x_i days of limited sleep. The data set contains such measurements for multiple individuals, but also several measurements per individual. In fact, we can label the data as

$$\left(y_i^{(j)}, x_i^{(j)}\right), \quad i = 1, \dots n_{\text{meas}}, \quad j = 1, \dots, n_{\text{ind}},$$

where n_{meas} is the number of measurements per individual and n_{ind} is the number of different individuals considered.

In this setting, the assumption of i.i.d. errors which we make in the classical linear model is violated (Fahrmeir, Kneib, Lang, and Marx, 2013, Subsection 7.7.1). Viewing the explanatory variable as fixed, the repeated measurements $(y_1^{(j)}, \ldots, y_{n_{\text{meas}}}^{(j)})$ of an individual $j \in \{1, \ldots, n_{\text{ind}}\}$ are not independent. Therefore, there are several "dependence clusters" in the data. Our artificial data set from Fig. 1.1 has been generated with this setting in mind. There, we used $n_{\text{ind}} = 3$ and $n_{\text{meas}} = 40$. Fig. 1.2 (left) shows the same data but plotted with different colors per individual.

To model such repeated measures data, a common choice is a so-called *mixed model*. The idea behind mixed models is to include the *fixed effects* β_0 and β_1 as well as *random effects* in the modelling, which differ for each "cluster" in the data (Fahrmeir et al., 2013, Section 2.2). The next sections outline two versions of mixed models. Section 1.1 deals with *random intercept (RI) models*, and Section 1.2 with *random coefficient models*. Chapter 2 then compares these two models with the approach of fitting a separate regression line for each dependence cluster. In Chapter 3, this simpler approach is compared with mixed models in difficult situations concerning the data.

1.1 Random Intercept Models

Carefully looking at the data in Fig. 1.2 (left), it seems as if the different clusters each follow a linear relationship with similar slopes but different intercepts. In the context of the sleep deprivation example, one (made-up) explanation could be that individuals have naturally different reaction times, but lack of sleep has similar negative effects.

Therefore, we model the data with a *random intercept model*, which is of the form (Fahrmeir et al., 2013, Subsection 7.1.1)

$$y_i^{(j)} = \beta_0 + \beta_1 x_i^{(j)} + \gamma_0^{(j)} + \epsilon_i^{(j)}. \tag{1.1.0.1}$$

Recall that $j = 1, \ldots, n_{\text{ind}}$ is the index for specific individuals, and $i = 1, \ldots, n_{\text{meas}}$ is the index for the i^{th} repeated measurement. The parameters β_0 and β_1 are the fixed population intercept and slope, respectively, which are used across all clusters. We assume i.i.d. errors $\epsilon_i^{(j)} \sim \mathcal{N}(0, \sigma^2)$ for all i and j, as in the classical linear model.

Additionally, we include $\gamma_0^{(j)}$, which is the "individual-specific" random deviation from the fixed intercept β_0 for the individual j. Note that it does not depend on i. We assume that the individuals j represented in the data are chosen randomly from a population. Therefore, $\gamma_0^{(j)}$ is modelled randomly, with distribution

$$\gamma_0^{(j)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \tau_0^2) \quad \text{for } j = 1, \ldots, n_{\text{ind}}.$$

Note that the random intercepts $\gamma_0^{(j)}$ have mean zero, because the population mean is already included in the fixed intercept β_0. We assume that $\epsilon_i^{(j)}$ and $\gamma_0^{(j)}$ are mutually independent for all j.

When the random intercept model (1.1.0.1) is fitted to a data set, estimates for $\beta_0, \beta_1, \sigma^2$,

and τ_0^2 are computed as well as the observed random effects $\hat{\gamma}_0^{(j)}$ corresponding to our data. In R, the function `lmer` from the package **lme4** uses maximum likelihood or restricted maximum likelihood to do this, as described in Bates, Mächler, Bolker, and Walker (2015).

Fig. 1.2 (right) again shows our data set together with the standard regression line (black) as in Fig. 1.1 (right). Additionally, for each individual j, the individual regression line with slope $\hat{\beta}_1$ and intercept $\hat{\beta}_0 + \hat{\gamma}_0^{(j)}$ is plotted in the corresponding colors. Visually, this model provides an excellent fit for this data set. In fact, the synthetic data has been generated from exactly such a random intercept model as in (1.1.0.1).

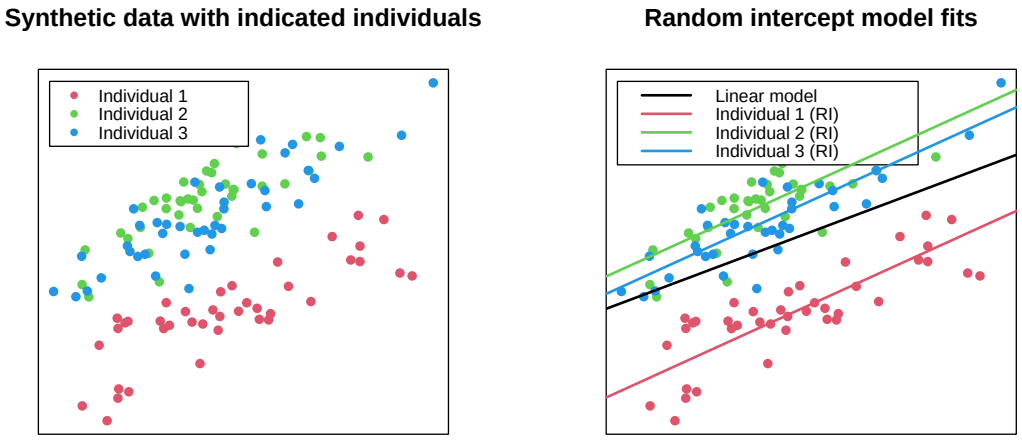


Figure 1.2: Visualization of the different individuals in our data set (left) and the corresponding random intercept (RI) model fits (right)

If we are given a new data point x not in the data set on which the model has been fitted, we might want to predict the corresponding y. If we don't know for which individual j this prediction is made (or if it is for a new person altogether), we use the fitted values $\hat{\beta}_0$ and $\hat{\beta}_1$, and exclude the random effects (i.e., use their mean zero). If, on the other hand, we want to predict for a known individual $j \in \{1, \dots, n_{\text{ind}}\}$, we can perform much better by including the estimated random effect $\hat{\gamma}_0^{(j)}$ in the prediction.

1.1.1 Fitting a Random Intercept Model

We will now discuss how the parameters β_0 and β_1 are estimated and how we find the observed random effects $\hat{\gamma}_0^{(j)}$, which correspond to the data. Our derivation follows Fahrmeir et al. (2013), Section 7.1, where a slightly more general model is considered, but is adapted to the random intercept model. Recall first our model:

$$\begin{aligned} y_i^{(j)} &= \beta_0 + \beta_1 x_i^{(j)} + \gamma_0^{(j)} + \epsilon_i^{(j)} \\ \epsilon_i^{(j)} &\overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2) \\ \gamma_0^{(j)} &\overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \tau_0^2) \\ i &= 1, \dots, n_{\text{meas}}, \quad j = 1, \dots, n_{\text{ind}}. \end{aligned}$$

Thus, we can write everything in vector form:

$$y^{(j)} = (y_1^{(j)}, \dots, y_{n_{\text{meas}}}^{(j)})^T = X^{(j)}\beta + U^{(j)}\gamma_0^{(j)} + \epsilon^{(j)},$$

where

$$X^{(j)} := \begin{pmatrix} 1 & x_1^{(j)} \\ \vdots & \vdots \\ 1 & x_{n_{\text{meas}}}^{(j)} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \quad U^{(j)} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \quad \epsilon^{(j)} = \begin{pmatrix} \epsilon_1^{(j)} \\ \vdots \\ \epsilon_{n_{\text{meas}}}^{(j)} \end{pmatrix}.$$

Note that the measurements $y_l^{(j)}$ and $y_k^{(j)}$ $(k \neq l)$ for a certain individual j satisfy

$$\text{Cov}\Big(y_l^{(j)},\ y_k^{(j)}\Big) = \tau_0^2 \quad \text{and} \quad \text{Var}\Big(y_l^{(j)}\Big) = \text{Var}\Big(y_k^{(j)}\Big) = \tau_0^2 + \sigma^2.$$

For different individuals $j \neq j'$, $\text{Cov}\Big(y_l^{(j)},\ y_k^{(j')}\Big) = 0\ \forall l, k$. Thus, it follows that

$$y^{(j)} \sim \mathcal{N}\Big(X^{(j)}\beta,\ \Sigma^{(j)}\Big), \quad \text{where } \Sigma^{(j)} = \begin{pmatrix} \tau_0^2 + \sigma^2 & \tau_0^2 & \cdots & \tau_0^2 \\ \tau_0^2 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \tau_0^2 \\ \tau_0^2 & \cdots & \tau_0^2 & \tau_0^2 + \sigma^2 \end{pmatrix}.$$

Going one step further from the vector form, we can write

$$y := \Big(y_1^{(1)}, \ldots, y_{n_{\text{meas}}}^{(1)}, y_1^{(2)}, \ldots, y_{n_{\text{meas}}}^{(2)}, \ldots, y_1^{(n_{\text{ind}})}, \ldots, y_{n_{\text{meas}}}^{(n_{\text{ind}})}\Big)$$

as

$$y = X\beta + U\gamma_0 + \epsilon,$$

where

$$X := \begin{pmatrix} X^{(1)} \\ \vdots \\ X^{(n_{\text{ind}})} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \quad U = \text{diag}\Big(U^{(1)}, \ldots, U^{(n_{\text{ind}})}\Big) = \begin{pmatrix} 1 & & & & \\ \vdots & \in \mathbb{R}^{n_{\text{ind}}} & & & \\ 1 & & & & \\ & 1 & & & \\ & \vdots & & & \\ & 1 & & & \\ & & & \ddots & \\ & & & & 1 \\ & & & & \vdots \\ & & & & 1 \end{pmatrix},$$

$$\gamma_0 = \begin{pmatrix} \gamma_0^{(1)} \\ \vdots \\ \gamma_0^{(n_{\text{ind}})} \end{pmatrix}, \quad \epsilon = \begin{pmatrix} \epsilon^{(1)} \\ \vdots \\ \epsilon^{(n_{\text{ind}})} \end{pmatrix} \sim \mathcal{N}(0, \sigma^2 I_n),$$

where $n = n_{\text{ind}} \cdot n_{\text{meas}}$. We therefore end up with the *conditional model*

$$\begin{aligned} y \mid \gamma_0 &\sim \mathcal{N}(X\beta + U\gamma_0,\ \sigma^2 I_n) \\ \gamma_0 &\sim \mathcal{N}(0,\ \tau_0^2 I_{n_{\text{ind}}}) \end{aligned}$$

and the *marginal model*

$$y \sim \mathcal{N}\Big(X\beta,\ \underbrace{\text{diag}\Big(\Sigma^{(1)}, \ldots, \Sigma^{(n_{\text{ind}})}\Big)}_{=:\Sigma}\Big).$$

Assume now that we know the parameters (σ^2, τ_0^2). The log-likelihood of $p(y, \gamma_0) = p(y \mid \gamma_0)p(\gamma_0)$ is

$$-\frac{1}{2\sigma^2}(y - X\beta - U\gamma_0)^T(y - X\beta - U\gamma_0) - \frac{1}{2\tau_0^2}\gamma_0^T\gamma_0.$$

Hence, maximizing the log-likelihood is equivalent to minimizing

$$\|y - X\beta - U\gamma_0\|_2^2 + \frac{\sigma^2}{\tau_0^2}\|\gamma_0\|_2^2,$$

which can be seen as a penalized regression. This then leads to the estimators (see Fahrmeir et al. (2013), Subsection 7.3.1, for details)

$$\begin{aligned}\hat{\beta} &= (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}y \\ \hat{\gamma}_0 &= \tau_0^2 U^T\Sigma^{-1}(y - X\hat{\beta}).\end{aligned} \tag{1.1.1.1}$$

Note that $\Sigma = \Sigma(\sigma^2, \tau_0^2)$ depends on the unknown parameters. To find them, we use maximum likelihood with the marginal model

$$y \sim \mathcal{N}\Big(X\beta,\ \Sigma(\sigma^2, \tau_0^2)\Big).$$

Up to additive constants, the log-likelihood is

$$l(\beta, \sigma^2, \tau_0^2) = -\frac{1}{2}\Big(\log|\det\Sigma(\sigma^2, \tau_0^2)| + (y - X\beta)^T\Sigma(\sigma^2, \tau_0^2)^{-1}(y - X\beta)\Big).$$

Maximizing this with respect to β while keeping (σ^2, τ_0^2) fixed yields

$$\hat{\beta}(\sigma^2, \tau_0^2) = \hat{\beta} = (X^T\Sigma(\sigma^2, \tau_0^2)^{-1}X)^{-1}X^T\Sigma(\sigma^2, \tau_0^2)^{-1}y.$$

Then, $\hat{\beta}(\sigma^2, \tau_0^2)$ is inserted into l, yielding the so-called *profile log-likelihood*

$$l_p(\sigma^2, \tau_0^2) = -\frac{1}{2}\Big(\log|\det\Sigma(\sigma^2, \tau_0^2)| + (y - X\hat{\beta}(\sigma^2, \tau_0^2))^T\Sigma(\sigma^2, \tau_0^2)^{-1}(y - X\hat{\beta}(\sigma^2, \tau_0^2))\Big).$$

Maximizing $l_p(\sigma^2, \tau_0^2)$ with respect to σ^2 and τ_0^2 results in the ML estimators $\hat{\sigma}^2_{\text{MLE}}$ and $\hat{\tau}^2_{0,\text{MLE}}$. The final estimates $\hat{\beta}$ and $\hat{\gamma}_0$ are then obtained by inserting $\hat{\sigma}^2_{\text{MLE}}$ and $\hat{\tau}^2_{0,\text{MLE}}$ into Equation (1.1.1.1).

1.2 Random Coefficient Models

By going one step further from random intercept models, we arrive at random coefficient models. In the example concerning sleep deprivation and its effect on reaction time in the previous Section 1.1, we have assumed that different individuals have naturally different reaction times, which change to a similar degree under lack of sleep. This is inherent in model (1.1.0.1), where the slope β_1 stays the same for all individuals.

We have thus neglected the possibility that people might have different tolerances and reactions to sleep deprivation. Therefore, the straightforward extension of model (1.1.0.1) is to include another random effect concerning the slope. Hence, we can write down the random coefficient model (Fahrmeir et al., 2013, Subsection 7.1.2) as

$$y_i^{(j)} = \beta_0 + \beta_1 x_i^{(j)} + \gamma_0^{(j)} + \gamma_1^{(j)} x_i^{(j)} + \epsilon_i^{(j)},$$

where, in addition to the parameters from the random intercept model (1.1.0.1), $\gamma_1^{(j)}$ is the random "individual-specific" deviation from the population slope β_1. The random effects are modelled via a Gaussian distribution

$$\begin{pmatrix} \gamma_0^{(j)} \\ \gamma_1^{(j)} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_0^2 & \tau_{0,1} \\ \tau_{0,1} & \tau_1^2 \end{pmatrix} \right)$$

for some values $\tau_{0,1} \in \mathbb{R}$, $\tau_0^2 > 0$, $\tau_1^2 > 0$.

The random coefficient model is often also called random intercept random slope (RIRS) model.

Chapter 2

Default Simulation Experiment

In this simulation experiment, we compare the random intercept and random coefficient models with the following modelling approach: For each "dependence cluster" in the data (i.e., for all individuals j and the corresponding measurements $x_1^{(j)}, \ldots, x_{n_{\text{meas}}}^{(j)}$), we fit a single standard regression line. Thus, we obtain n_{ind} different slope estimates. We then compare the mean slope with the fitted slope $\hat{\beta}_1$ from the random intercept or random coefficient model. To do this, we separately compare the 95% confidence interval for the slope estimate of the mixed models with the 95% confidence interval obtained from a t-test on the n_{ind} different slopes yielded by the second method (where we aggregate the measurements per individual). We check over many simulation runs whether the coverage frequency of the confidence intervals for the true slope parameter is indeed 95%. We furthermore compare the widths of the confidence intervals. See Section A.1 in the appendix for the code.

2.1 Default Simulation: Random Intercept Model versus Aggregated Regressions

In the experiment, we used 3000 simulation runs. In each run, we generated data from the random intercept model

$$
\begin{aligned}
x_i^{(j)} &\overset{\text{i.i.d.}}{\sim} \text{Uniform}(0, 10) \quad \forall i, j \\
\gamma_0^{(j)} &\overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \tau_0^2) \quad \forall j \\
\epsilon_i^{(j)} &\overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2) \quad \forall i, j \\
y_i^{(j)} &= \beta_0 + \beta_1 x_i^{(j)} + \gamma_0^{(j)} + \epsilon_i^{(j)}
\end{aligned}
$$

with "default" parameters in Table 2.1. For fitting the random intercept model, we used the function `lmer` from the package **lme4** in R (Bates et al. (2015)), which also computes confidence intervals for the estimated parameters. For the aggregated linear regression approach, we used the t-test built into R for the n_{ind} different slope estimates.

slope β_1	intercept β_0	individuals n_{ind}	measurements n_{meas}	τ_0	σ
2	1	50	10	2	1/2

Table 2.1: "Default" model parameters for the random intercept model

After 3000 simulation runs, we obtain the results in Table 2.2:

	random intercept model	aggregated regressions
coverage frequency	0.9460	0.9493
average confidence interval width	0.0321	0.0348

Table 2.2: Default simulation results (random intercept model vs. aggregated regressions)

The number of simulation runs was determined in such a way that the half-widths of a 95% confidence interval corresponding to a binomial test for the coverage probability is at most 1%. Thus, the simulation accuracy of the coverage probability is roughly 1%.

From the results above, it is evident that in the simulation, both the random intercept model and the aggregated regressions yield very comparable results. For a visualization of 50 confidence intervals around the slope $\beta_1 = 2$ computed by both methods, see Fig. 2.1.

Visual inspection of Fig. 2.1 and Table 2.2 might suggest that the confidence interval corresponding to the random intercept model is slightly narrower compared with the confidence interval for the aggregated regressions. We can test whether there is a systematic difference with a paired t-test. Since we also run paired t-tests to evaluate differences between the random intercept or coefficient model and the aggregated regressions in the next chapter, we use a *Bonferroni correction* (see Fahrmeir et al. (2013), Section 3.4). Because we conduct a total of twelve paired t-tests, we use a confidence level of $1 - \alpha/12 = 0.996$ with $\alpha = 5\%$ to obtain an overall confidence of 95%. After 3000 simulation runs, we can reject the null-hypothesis (*The mean difference in the interval widths is 0.*) with a p-value of $< 2.2 \cdot 10^{-16}$. We obtain the confidence interval $[-0.002959, -0.002573]$ for the difference (width RI minus width aggregated regressions). However, it must be noted that the difference, although statistically significant, is negligible for most applications.

2.2 Default Simulation: Random Coefficient Model versus Aggregated Regressions

For the same reason as in the simulation of the random intercept model in Section 2.1, we used 3000 simulation runs for the simulation experiment with the random coefficient model. Again, `lmer` from Bates et al. (2015) was used for fitting. Here, we generated synthetic data from the random coefficient model

$$
\begin{aligned}
x_i^{(j)} &\overset{\text{i.i.d.}}{\sim} \text{Uniform}(0, 10) \quad \forall i, j \\
\gamma_0^{(j)} &\overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \tau_0^2) \quad \forall j \\
\gamma_1^{(j)} &\overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \tau_1^2) \quad \forall j \\
\epsilon_i^{(j)} &\overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2) \quad \forall i, j \\
y_i^{(j)} &= \beta_0 + \beta_1 x_i^{(j)} + \gamma_0^{(j)} + \gamma_1^{(j)} x_i^{(j)} + \epsilon_i^{(j)}
\end{aligned}
$$

with "default" parameter values in Table 2.3.

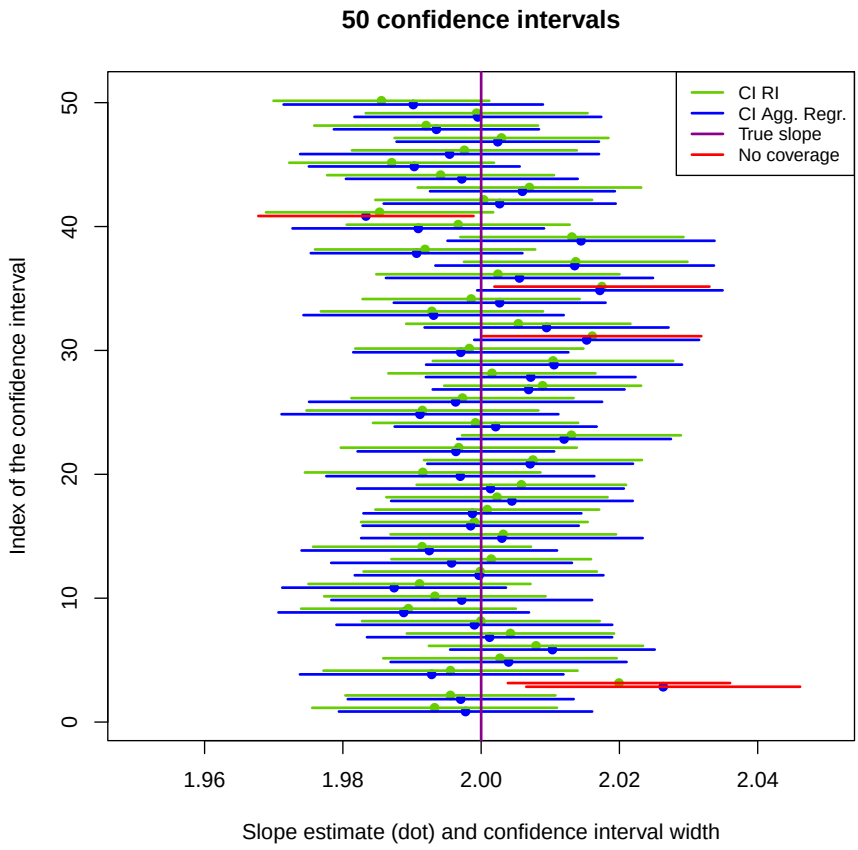


Figure 2.1: Comparison of 50 confidence intervals computed with the random intercept model and aggregated regressions

slope β_1	intercept β_0	individuals n_{ind}	measurements n_{meas}	τ_0	τ_1	σ
2	1	50	10	2	1	1/2

Table 2.3: "Default" model parameters for the random coefficient model

After 3000 simulation runs, we obtain the following results in Table 2.4:

	random coefficient model	aggregated regressions
coverage frequency	0.9510	0.9530
average confidence interval width	0.5568	0.5656

Table 2.4: Default simulation results (random coefficient model vs. aggregated regressions)

Again, both methods perform similarly in the simulation. Note that the average width of both confidence intervals for data from the random coefficient model are considerably wider than for the random intercept model data. This is to be expected, as the data drawn from the random coefficient model includes random effects in the slope and thus contains much more variability. Fig. A.1 in the appendix visualizes 50 confidence intervals in the same fashion as Fig. 2.1 for the random coefficient model.

Again, we conduct a paired t-test to check whether there is a systematic difference in the interval widths associated with both methods. After 3000 simulation runs, we can again reject the null-hypothesis (*The mean difference in the interval widths is 0.*) with a p-value of $< 2.2 \cdot 10^{-16}$. The confidence level is 99.6% to guarantee an overall confidence level of 95% (Bonferroni correction). We obtain the confidence interval $[-0.008916, -0.008812]$ for the difference (width RIRS minus width aggregated regressions). It must be noted that the difference, although statistically significant, is negligible for most applications.

Chapter 3

Model Violations

In this chapter, we conduct experiments testing the models' robustness to "model violations" and extreme situations.

We take a look at the following cases:

i.) **Few measurements per individual:** n_{meas} is small.
In particular, we use $n_{\text{meas}} = 3$.

ii.) **Few individuals in the study:** n_{ind} is small.
In particular, we use $n_{\text{ind}} = 10$.

iii.) **Varying number of measurements per individual:** n_{meas} differs for each $j \in \{1, \dots, n_{\text{ind}}\}$ and we write $n_{\text{meas}}^{(j)}$ for the number of measurements for individual j. In particular, we use randomly drawn samples $n_{\text{meas}}^{(j)} \in \{5, 10, 25, 50, 100, 200\}$.

iv.) **Non-normal errors:** We use i.i.d. centered $\epsilon_i^{(j)} \sim \text{Exp}(1)$ errors.

v.) **Dependent errors:** The errors $(\epsilon_i^{(j)})_{i=1}^{n_{\text{meas}}}$ form an AR(1) process $\forall j$. In particular, $\epsilon_i^{(j)} = 0.7\epsilon_{i-1}^{(j)} + 0.7Z$, where $Z \sim \mathcal{N}(0, 1)$ is independent noise.

3.1 Model Violations: Random Intercept Model versus Aggregated Regressions

We now compare the random intercept model with our method of aggregating several separate regressions. We again use the following default values except when they are changed in a specific model violation i.) to v.):

slope β_1	intercept β_0	individuals n_{ind}	measurements n_{meas}	τ_0	σ
2	1	50	10	2	1/2

Note that these are the same values we used to conduct the simulation experiment in Section 2.1. We compute 3000 simulation runs per model violation experiment. As for the default simulation, a binomial test at 95% confidence reveals that for 3000 simulation runs, we can determine the coverage roughly up to 1% error.

We obtained the following results in Table 3.1. For comparison, we also included the results from Section 2.1 for the default values without violations for comparison. A visual

presentation of the results can be seen in Fig. 3.1. Furthermore, Fig. A.2 in the appendix visualizes the confidence intervals.

	Coverage RI	Coverage Agg.	CI width RI	CI width Agg.
Default	0.9460	0.9493	0.0321	0.0348
i.) n_{meas} small	0.9530	0.9573	0.0681	0.1643
ii.) n_{ind} small	0.9497	0.9507	0.0721	0.0849
iii.) $n_{\text{meas}}^{(j)}$ varying $\forall j$	0.9533	0.9580	0.0121	0.0339
iv.) Exp(1) errors	0.9503	0.9457	0.0640	0.0694
v.) AR(1) errors	0.9500	0.9527	0.0507	0.0552

Table 3.1: Model violation results for the RI model vs. aggregated regressions

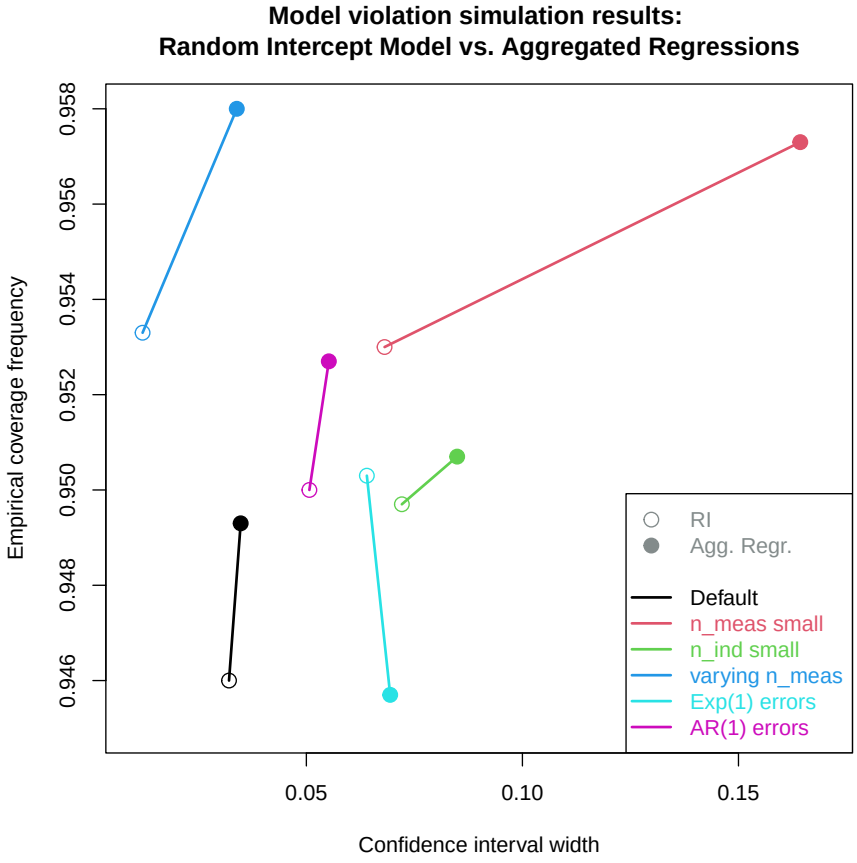


Figure 3.1: Visualization of the simulation results in Table 3.1

From these results, it is visible that both models (random intercept and aggregated regressions) tolerate small n_{ind}, exponential errors, and dependent errors more or less equally well. For both models, these violations result in significantly, but not dramatically, wider confidence intervals.

In violation iii.), the average value of n_{meas} is considerably higher than the default of 10. Thus, both models can fit the slopes in each cluster with less variability on average. However, the confidence interval yielded by the t-test for the aggregated regressions is wider

than for the random intercept model. The different values of $n_{\text{meas}}^{(j)}$ for all $j = 1, \ldots, n_{\text{ind}}$ result in different variances for the slope estimates in the linear models which we aggregate over all j. Thus, this is actually a slight violation of the assumptions behind the t-test, where all random variables over which we average should have the same variance.

The aggregated regression model seems to struggle considerably with small n_{meas} in violation i.). There is also an increase in the width of the confidence interval for the slope in the random intercept model, but it is not that dramatic. See Chapter 4 for a discussion. It would now be interesting to see how small n_{meas} can get until we run into trouble. Fig. 3.2 shows how the confidence interval width changes with respect to a change in n_{meas}. As we can see, the difference between the confidence interval widths of the two models decreases with increasing n_{meas}.

A further point of investigation is the observation that for all model violations as well as the default in Table 3.1, the confidence intervals for the random intercept model are narrower than for the aggregated regressions. Is there a statistically significant difference in width? To answer this, we again conduct a paired t-test on the difference (width RI minus width aggregated regressions). To guarantee a confidence level of 95% over all twelve tests we consider in this project, we conduct each test at confidence level $1-\alpha/12 = 99.6\%$ for $\alpha = 5\%$ (Bonferroni correction). In all model violations, we can reject the null hypothesis that there is no significant difference in the widths between the two methods with a p-value of $< 2.2 \cdot 10^{-16}$. See Table A.1 in the appendix for the exact confidence intervals for the difference in interval width corresponding to each model violation. All of them don't include the value 0. In all cases except model violation i.), however, the difference will be negligible in many applications.

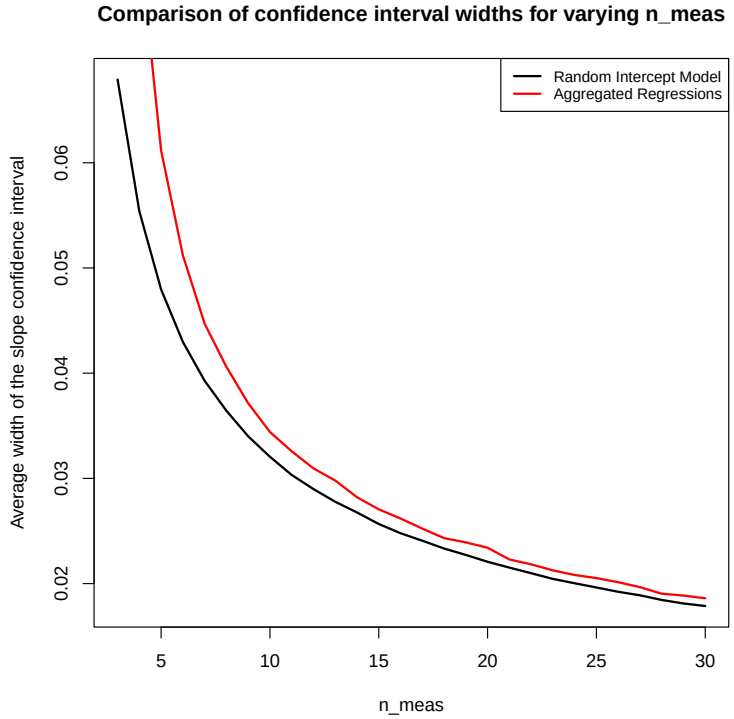


Figure 3.2: Comparison of confidence interval width with changing n_{meas} for the random intercept model and the aggregated regression model

3.2 Model Violations: Random Coefficient Model versus Aggregated Regressions

In this section, we repeat the experiments with the same model violations as for the random intercept model. The following values were used as a default:

slope β_1	intercept β_0	individuals n_{ind}	measurements n_{meas}	τ_0	τ_1	σ
2	1	50	10	2	1	1/2

These are the same values we used to conduct the simulation experiment in Section 2.2. We used 3000 simulation runs per model violation experiment. As explained above, we can thus ensure with high probability that the coverage frequencies are at most 1% off from the truth.

We obtained the values in Table 3.2. For comparison, we also included the results from Section 2.2 for the default values without violations. A visual presentation of the results can be seen in Fig. 3.3. Additionally, A.3 in the appendix plots the confidence intervals.

	Coverage RIRS	Coverage Agg.	CI width RIRS	CI width Agg.
Default	0.9510	0.9530	0.5568	0.5656
i.) n_{meas} small	0.9507	0.9497	0.5683	0.5987
ii.) n_{ind} small	0.9383	0.9570	1.2673	1.3964
iii.) $n_{\text{meas}}^{(j)}$ varying	0.9520	0.9560	0.5585	0.5674
iv.) Exp(1) errors	0.9393	0.9433	0.5612	0.5703
v.) AR(1) errors	0.9497	0.9527	0.5583	0.5672

Table 3.2: Model violation results for the random coefficient (RIRS) model vs. aggregated regressions

Here, both models seem to be very robust to all violations except choosing a small n_{ind}. Interestingly, small values for n_{meas} don't seem to be much more problematic for the aggregated regression method than the random coefficient model. In contrast, the aggregated regressions struggle much more with small n_{meas} than the random intercept model on random intercept data (see Table 3.1).

Fig. 3.4 shows the interaction between increasing n_{ind} and confidence interval width for the random coefficient model and the aggregated regression model. As we can see from Table 3.2, this is the "model violation" where both models struggle the most. We see that the confidence interval width decreases in the same fashion for both models. In the log-log plot of the same phenomenon, the widths decrease linearly with increasing n_{ind}. This results from the factor of $\sqrt{n_{\text{ind}}}$ in the t-test statistic, because we conduct a t-test over the n_{ind} slope estimates.

Again, we note that the average random coefficient model confidence interval seems to be narrower than the average aggregated regression confidence interval in Table 3.2. As for the random intercept model, there is a statistically significant difference at level 99.6% (Bonferroni correction). For the exact confidence intervals for the difference in width, we refer to Table A.2 in the appendix.

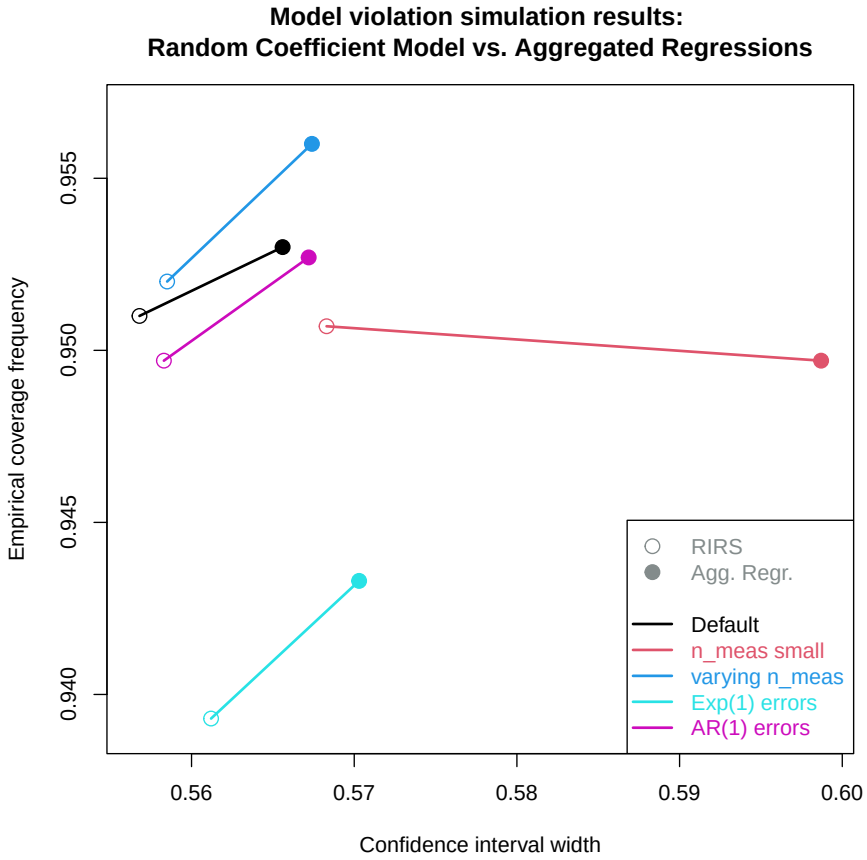


Figure 3.3: Visualization of the simulation results in Table 3.2 (excluding model violation ii.) due to much wider confidence intervals)

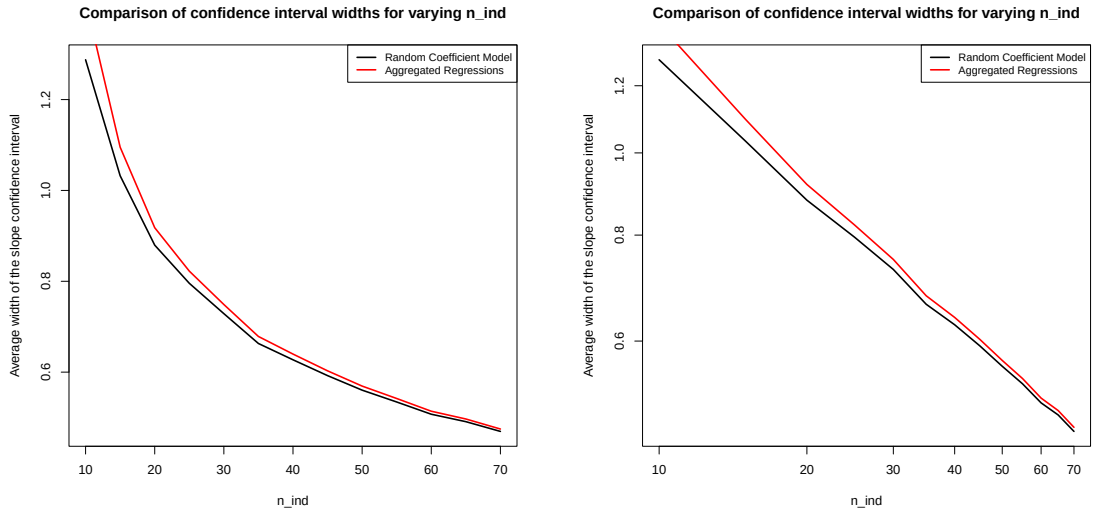


Figure 3.4: Comparison of confidence interval width with changing n_{ind} for the random coefficient model and the aggregated regression model with standard (left) and log-log (right) axes

Chapter 4

Discussion of the Simulations

For the default values, we have shown in Table 2.2 that the aggregated regression approach performs very comparably to the random intercept model, from which the data has been generated. Similarly, Table 2.4 shows that the confidence intervals computed with the aggregated regression t-test are almost as narrow as those obtained from the random coefficient model, from which the data was generated.

From the model violation results in Table 3.1, it follows that both models (random intercept and aggregated regressions) are almost equally adept at handling small numbers of individuals n_{ind}, exponential errors, and dependent errors.

In contrast, a small number of measurements n_{meas} per individual seems to be much harder for the aggregated regressions than the random intercept model. However, one can argue that this is an unfair comparison. If n_{meas} is small, then there is more variance in the slopes fitted by the separate regressions. The random intercept model, on the other hand, has "built-in knowledge" that the true slope is the same across all individuals, thereby yielding estimates with lower variance. This explanation was suggested by the supervisor of this semester project (Kalisch, 2023, private communication).

This is consistent with our observations from Table 3.2. There, the data comes from a random coefficient model with random slopes. The fitted random coefficient model now cannot benefit from stricter assumptions than the aggregated regressions. As a result, the performance of the aggregated regressions when compared with the random coefficient model is almost the same for all considered model violations in Table 3.2, except maybe for small n_{ind}. While we used a different mixed model when going from random intercept to random coefficient data, the aggregated regression approach was not modified in any way.

Now, we consider violation iii.), where we randomly sample $n_{\text{meas}}^{(j)} \in \{5, 10, 25, 50, 100, 200\}$ for all $j = 1, \ldots, n_{\text{ind}}$. There, we observe that for data from the random intercept model, the aggregated regression method yields considerably broader confidence intervals than the random intercept model. Our explanation is twofold.

1) Since $n_{\text{meas}}^{(j)}$ can take on quite small values like 5 or 10, the same explanation as above for violation i.) applies.

2) If $n_{\text{meas}}^{(j)}$ varies greatly for all $j = 1, \ldots, n_{\text{ind}}$ as in violation iii.), then the slopes computed by the linear model have very different variances. If we then apply a t-test, we violate the condition of equal variances in the t-test, yielding a wider interval. This effect is visible for the data coming from the random intercept model. It is, however, not clearly visible for the aggregated regressions on random coefficient data compared with the random coefficient model. It is the author's interpretation that there, the effect is overpowered by the additional randomness in the slopes.

After all the simulations, the author would therefore recommend to use aggregated regressions in general, because the model is simpler, and the separate regressions for each individual are very easily interpretable. As an added benefit, fitting several linear regressions is computationally cheaper than fitting a random coefficient model.

The random intercept model is only clearly preferable in a situation where it is for some reason evident from the physical mechanism generating the data that all dependence clusters or individuals have the same slope. However, ...

A word of caution: Arguably, it is rarely the case that we can make that assumption beyond any doubt. In a final simulation experiment, we now demonstrate how bad the random intercept model can go wrong even if the assumption of a *constant true slope* across all individuals is only slightly wrong. Hence, we fit the random intercept model to data generated by the random coefficient model (implying that the slopes are different across individuals). We used the default values as for the random intercept model:

slope β_1	intercept β_0	individuals n_{ind}	measurements n_{meas}	τ_0	σ
2	1	50	10	2	1/2

For the individual-specific slopes, we added a normally distributed random slope to β_1 with mean 0 and standard deviation

$$\tau_1 \in \{0.00,\ 0.02,\ 0.04,\ 0.06,\ \ldots,\ 0.98,\ 1.00\}.$$

Note that $\tau_1 = 1$ was the default value for testing the random coefficient model. In 500 simulation runs for each value of τ_1, we have drawn data from the random coefficient model, fitted the random intercept model and the aggregated regression model to the data, and computed the widths of the confidence intervals for the slope estimates. Then, we evaluated the coverage frequency over all simulation runs.

The result is plotted in Fig. 4.1. Even though the confidence intervals for the random intercept model stay narrower than for the aggregated regressions, the coverage frequency is very far from 0.95, even for very small τ_1. For $\tau_1 \geq 0.2$, the coverage frequency of the confidence intervals remains below 55%, which is unacceptable. Visually, a standard deviation of $\tau_1 = 0.2$ for the random slopes is almost not noticeable in the plotted data. In Fig. A.4 in the appendix, data from the random coefficient model with our default parameter values and $\tau_1 = 0.2$ is plotted with $n_{\text{ind}} = 4$ and $n_{\text{meas}} = 40$.

Hence, it is not enough if the condition of constant slopes across individuals is only approximately valid to be able to successfully apply the random intercept model in applications. We therefore conclude with the following recommendation:

Concluding recommendation:

In the majority of cases, the aggregated regression approach should be preferred over the random intercept and random coefficient model due to simplicity and interpretability, while at the same time yielding comparable performance.

Secondly, if there is doubt about whether the true slopes are equal for all individuals, the aggregated regression approach should be used, even in case of very minor differences, and not the random intercept model. The random coefficient model is more complex and only offers a very modest performance increase. For the situations considered in our simulation experiment, the random intercept model is only clearly superior to the aggregated regressions if we are dealing with extreme situations concerning n_{meas} and we are absolutely certain that the true slopes are equal for all individuals.

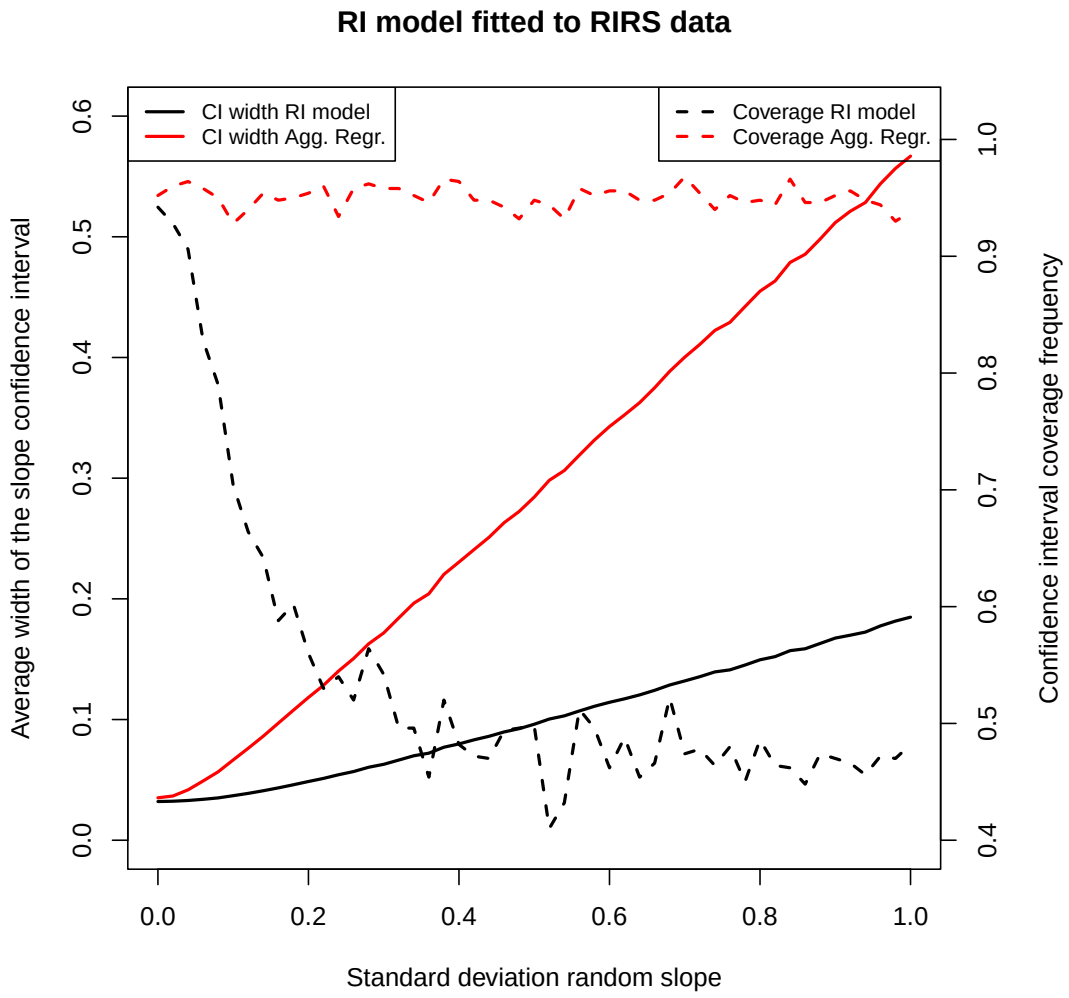


Figure 4.1: Applying the random intercept model to random coefficient model data (an unfortunate idea)

Bibliography

Bates, D., M. Mächler, B. M. Bolker, and S. C. Walker (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software 67*(1), 1–48.

Fahrmeir, L., T. Kneib, S. Lang, and B. Marx (2013). *Regression*. Heidelberg: Springer-Verlag.

Kalisch, M. (2023). Private Communication.

Appendix A

Complementary Material

A.1 Code

All the simulations (and the figures) have been implemented in R. The complete code is available here:

https://github.com/LinusKuehne/Aggregation-as-a-Simple-Alternative-to-Mixed-Models

A.2 Model Violations: Confidence Intervals for the Differences in Interval Width

	CI width RI	CI width Agg.	99.6% CI on difference
Default	0.0321	0.0348	$[-0.0030, -0.0026]$
i.) n_{meas} small	0.0681	0.1643	$[-0.1035, -0.0889]$
ii.) n_{ind} small	0.0721	0.0849	$[-0.0139, -0.0117]$
iii.) $n_{\text{meas}}^{(j)}$ varying $\forall j$	0.0121	0.0339	$[-0.0223, -0.0213]$
iv.) Exponential errors	0.0640	0.0694	$[-0.0059, -0.0050]$
v.) Dependent errors	0.0507	0.0552	$[-0.0048, -0.0041]$

Table A.1: Confidence intervals for the difference in confidence interval widths for the RI model vs. aggregated regressions

	CI width RIRS	CI width Agg.	99.6% CI on difference
Default	0.5568	0.5656	$[-0.0089, -0.0088]$
i.) n_{meas} small	0.5683	0.5987	$[-0.0347, -0.0261]$
ii.) n_{ind} small	1.2673	1.3964	$[-0.1306, -0.1275]$
iii.) $n_{\text{meas}}^{(j)}$ varying $\forall j$	0.5585	0.5674	$[-0.0091, -0.0089]$
iv.) Exponential errors	0.5612	0.5703	$[-0.0092, -0.0090]$
v.) Dependent errors	0.5583	0.5672	$[-0.0090, -0.0089]$

Table A.2: Confidence intervals for the difference in confidence interval widths for the RIRS model vs. aggregated regressions

A.3 Complementary Figures

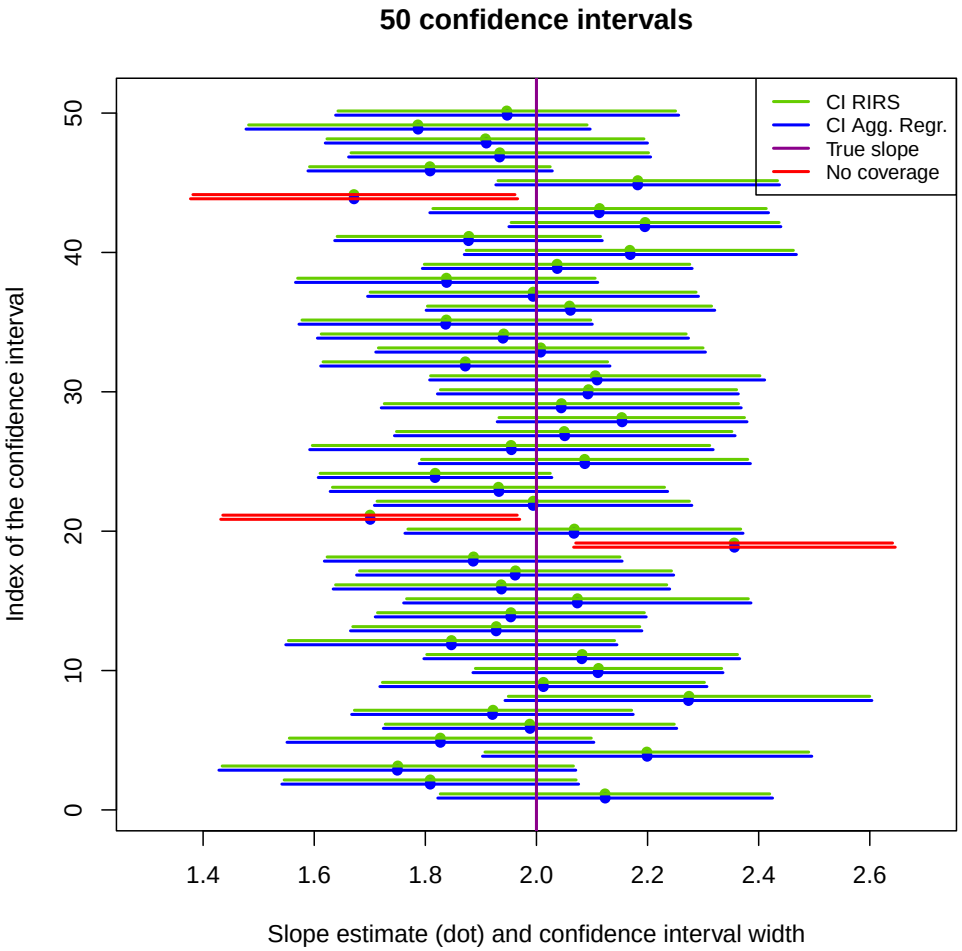


Figure A.1: Comparison of 50 confidence intervals computed with the random coefficient model and aggregated regressions using the default parameters in Table 2.3

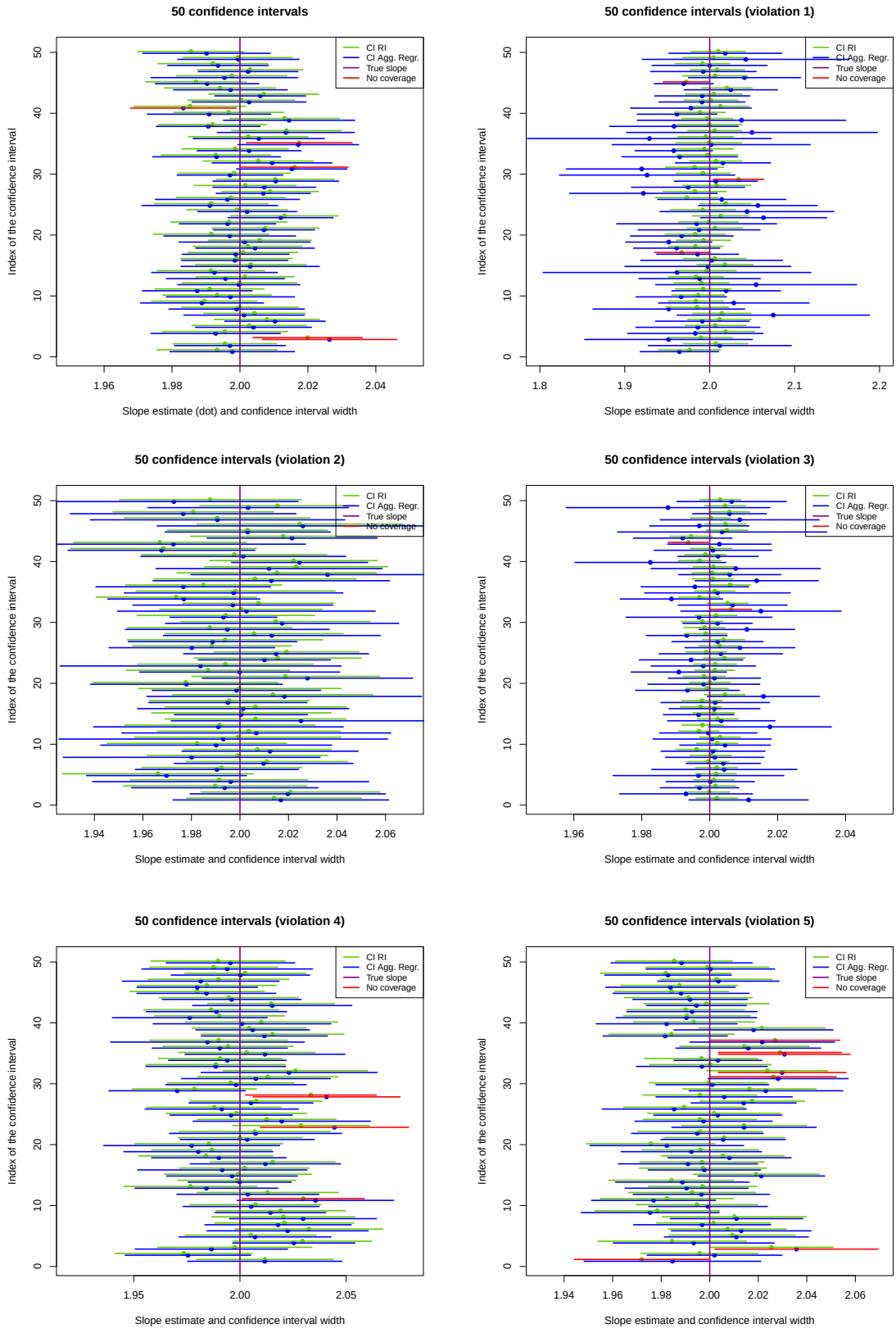


Figure A.2: 50 sample confidence intervals for all violations and the default (top left) comparing the random intercept model (RI) with the aggregated regressions

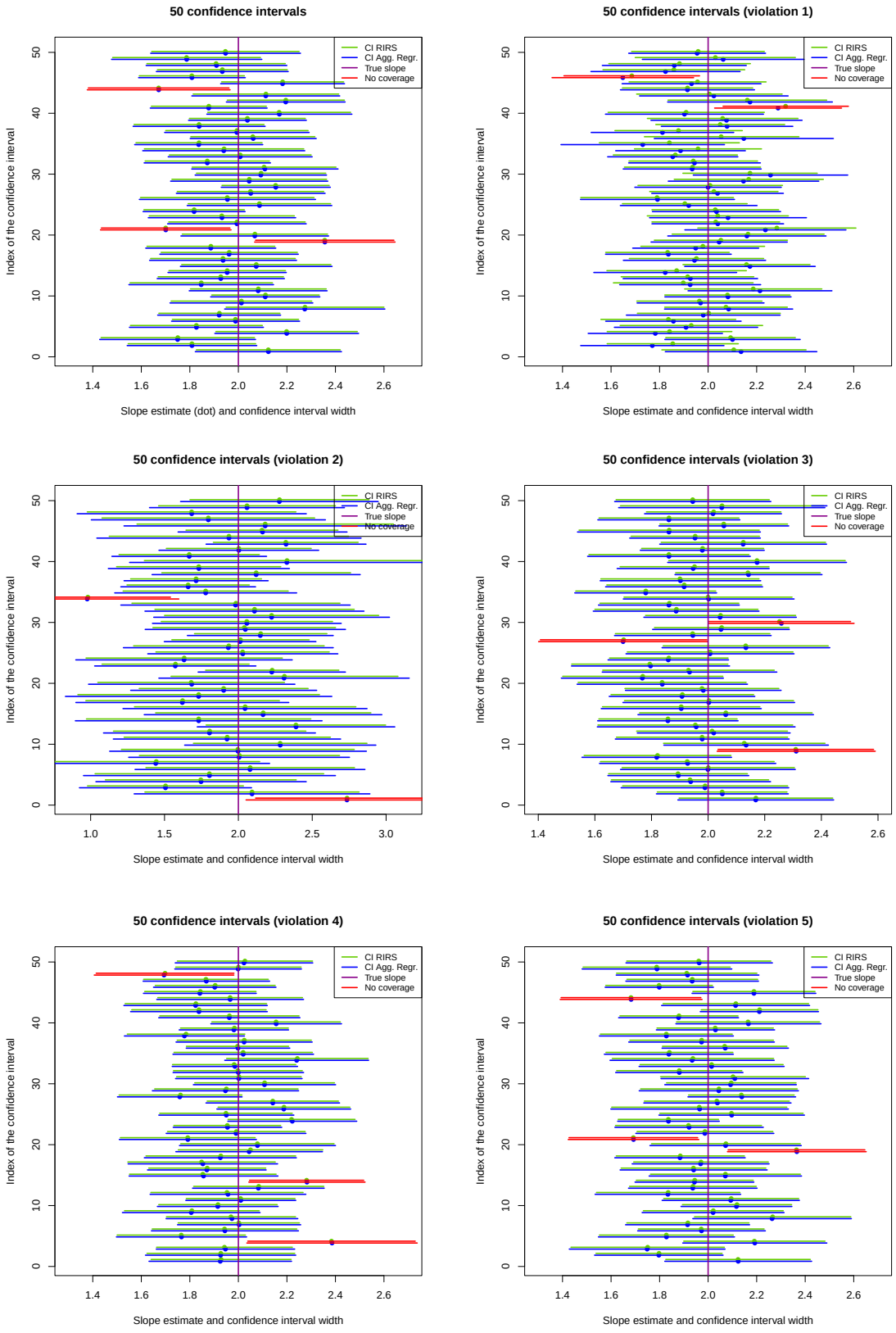


Figure A.3: 50 sample confidence intervals for all violations and the default (top left) comparing the random coefficient model (RIRS) with the aggregated regressions

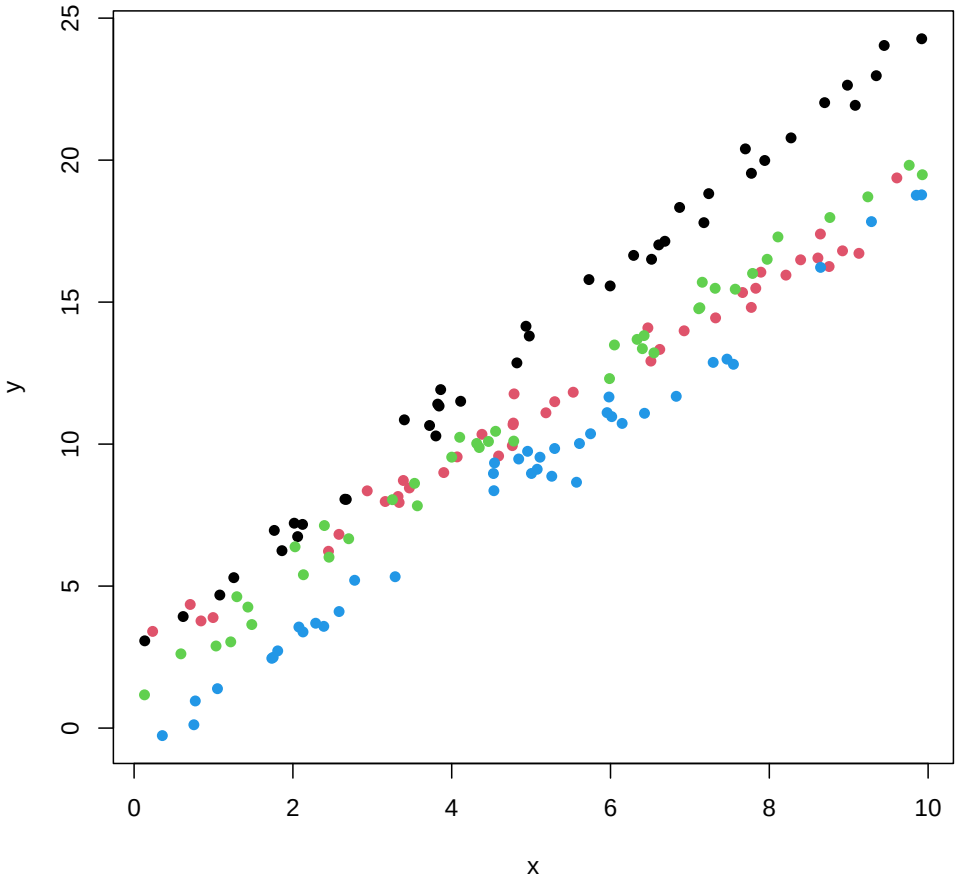


Figure A.4: Data sampled from the random coefficient model with $\tau_1 = 0.2$ and $n_{\text{ind}} = 4$

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

Aggregation as a Simple Alternative to Mixed Models

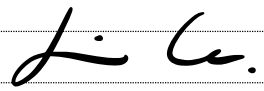
Authored by (in block letters):
For papers written by groups the names of all authors are required.

Name(s):	**First name(s):**
Kühne	Linus Anton

With my signature I confirm that

- I have committed none of the forms of plagiarism described in the 'Citation etiquette' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date	**Signature(s)**
Meisterschwanden, 15.11.2023	

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.