Building Multivariate Longitudinal Models

Peter Norwood, NCSU, Janssen R&D

1 Introduction

This writeup discusses a how to fit multivariate longitudinal models in R, the limitations with the common packages, and then results from a monte carlo study where we evaluate how well models estimate covariance parameters at different sample sizes and under correct model specification and varying levels of misspecification. The outline is below:

- Overview of potential models to use

- Overview of fitting models using the `nlme` package

- Simulation Setup

- Simulation Results

- Discussion of Results and Model Building Suggestions

- Real Data Example

2 Models to Fit

2.1 Multivariate Mixed Model One, Continuous Time

The first model we fit involves two responses with multiple time points each:

$$y_{i1j} = \beta_{10} + b_{i10} + \beta_{11} * A_i + \beta_{12} * t_{ij} + b_{i11} * t_{ij} + \beta_{13} * A_i * t_{ij} + \epsilon_{i1j}$$

and

$$y_{i2j} = \beta_{20} + b_{i20} + \beta_{21} * A_i + \beta_{22} * t_{ij} + b_{i21} * t_{ij} + \beta_{23} * A_i * t_{ij} + \epsilon_{i2j}$$

.

Here y_{i1j} is the response type 1 for individual i at time point j. y_{i2j} is the response type 2 for individual i at time point j. $A_i \in \{0, 1\}$ is the treatment indicator for individual i and t_{ij} is the time variable (hour, week, month, etc) at time point j for individual i. We will refer to $t_{i1} = 1, t_{i2} = 2, t_{i3} = 3, t_{i4} = 4$ and treat this as continuous. b_{i10}, b_{i20} are the random intercepts for y_{1ij} and y_{2ij} respectively. Likewise, b_{i11}, b_{i21} are the random slopes. We take:

$$\begin{pmatrix} b_{i10} \\ b_{i11} \\ b_{i20} \\ b_{i21} \end{pmatrix} \sim MVN\left(\mathbf{0}, \mathbf{G}\right)$$

where

$$\mathbf{G} = \begin{pmatrix} \sigma_{b_{10}}^2 & \sigma_{b_{10},b_{11}} & \sigma_{b_{10},b_{20}} & \sigma_{b_{10},b_{21}} \\ . & \sigma_{b_{11}}^2 & \sigma_{b_{11},b_{20}} & \sigma_{b_{11},b_{21}} \\ . & . & \sigma_{b_{20}}^2 & \sigma_{b_{20},b_{21}} \\ . & . & . & \sigma_{b_{21}}^2 \end{pmatrix}$$

Additionally, we let

$$\begin{pmatrix} \epsilon_{i1j} \\ \epsilon_{i2j} \end{pmatrix} \sim MVN\left(\mathbf{0}, \mathbf{R}\right)$$

where

$$\mathbf{R} = \begin{pmatrix} \sigma_{\epsilon_1}^2 & \sigma_{\epsilon_1,\epsilon_2} \\ . & \sigma_{\epsilon_1}^2 \end{pmatrix} = \begin{pmatrix} \sigma_{\epsilon_1} & 0 \\ 0 & \sigma_{\epsilon_2} \end{pmatrix} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \begin{pmatrix} \sigma_{\epsilon_1} & 0 \\ 0 & \sigma_{\epsilon_2} \end{pmatrix}$$

where ρ is the correlation between errors at the same time point. The overall covariance matrix for an individual is

$$\mathbf{V}_i = Cov \begin{pmatrix} y_{i11} \\ y_{i21} \\ y_{i12} \\ y_{i22} \\ y_{i13} \\ y_{i23} \\ y_{i14} \\ y_{i24} \end{pmatrix} = \mathbf{Z}_i \mathbf{G} \mathbf{Z}_i^T + \begin{pmatrix} \mathbf{R} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} \\ . & \mathbf{R} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} \\ . & . & \mathbf{R} & \mathbf{0}_{2\times2} \\ . & . & . & \mathbf{R} \end{pmatrix}$$

2.2 Multivariate Mixed Model Two, Continuous Time

This model is the same as before, but with a different covariance structure for the random error.

$$y_{i1j} = \beta_{10} + b_{i10} + \beta_{11} * A_i + \beta_{12} * t_{ij} + b_{i11} * t_{ij} + \beta_{13} * A_i * t_{ij} + \epsilon_{i1j}$$

and

$$y_{i2j} = \beta_{20} + b_{i20} + \beta_{21} * A_i + \beta_{22} * t_{ij} + b_{i21} * t_{ij} + \beta_{23} * A_i * t_{ij} + \epsilon_{i2j}$$

.

Here y_{i1j} is the response type 1 for individual i at time point j. y_{i2j} is the response type 2 for individual i at time point j. $A_i \in \{0,1\}$ is the treatment indicator for individual i and t_{ij} is the time variable (hour, week, month, etc) at time point j for individual i. We will refer to $t_{i1} = 1, t_{i2} = 2, t_{i3} = 3, t_{i4} = 4$ and treat this as continuous. b_{i10}, b_{i20} are the random intercepts for y_{1ij} and y_{2ij} respectively. Likewise, b_{i11}, b_{i21} are the random slopes. We take:

$$\begin{pmatrix} b_{i10} \\ b_{i11} \\ b_{i20} \\ b_{i21} \end{pmatrix} \sim MVN\left(\mathbf{0}, \mathbf{G}\right)$$

2

where

$$\mathbf{G} = \begin{pmatrix} \sigma_{b_{10}}^2 & \sigma_{b_{10},b_{11}} & \sigma_{b_{10},b_{20}} & \sigma_{b_{10},b_{21}} \\ \cdot & \sigma_{b_{11}}^2 & \sigma_{b_{11},b_{20}} & \sigma_{b_{11},b_{21}} \\ \cdot & \cdot & \sigma_{b_{20}}^2 & \sigma_{b_{20},b_{21}} \\ \cdot & \cdot & \cdot & \sigma_{b_{21}}^2 \end{pmatrix}$$

Additionally, we let

$$\begin{pmatrix} \epsilon_{i11} \\ \epsilon_{i21} \\ \epsilon_{i12} \\ \epsilon_{i22} \\ \epsilon_{i13} \\ \epsilon_{i23} \\ \epsilon_{i14} \\ \epsilon_{i24} \end{pmatrix} \sim MVN\left(\mathbf{0}, \Sigma\right)$$

where Σ is an unstructred, symmetric, and positive definite matrix. The overall covariance for the responses is

$$\mathbf{V}_i = Cov \begin{pmatrix} y_{i11} \\ y_{i21} \\ y_{i12} \\ y_{i22} \\ y_{i13} \\ y_{i23} \\ y_{i14} \\ y_{i24} \end{pmatrix} = \mathbf{Z}_i \mathbf{G} \mathbf{Z}_i^T + \Sigma$$

2.3 Generalized Least Squares, Continuous Time

The next model is a generalized least squares model. Using the same notation and indices as before, the model is:

$$y_{i1j} = \beta_{10} + \beta_{11} * A_i + \beta_{12} * t_{ij} + \beta_{13} A_i * t_{ij} + \epsilon_{i1j}$$

and

$$y_{i2j} = \beta_{20} + \beta_{21} * A_i \beta_{22} * t_{ij} + \beta_{23} * A_i * t_{ij} + \epsilon_{i2j}$$

.

3

All covariance is estimated through the ϵ terms. We have

$$\mathbf{V}_i = Cov \begin{pmatrix} y_{i11} \\ y_{i21} \\ y_{i12} \\ y_{i22} \\ y_{i13} \\ y_{i23} \\ y_{i14} \\ y_{i24} \end{pmatrix} = Cov \begin{pmatrix} \epsilon_{i11} \\ \epsilon_{i21} \\ \epsilon_{i12} \\ \epsilon_{i22} \\ \epsilon_{i13} \\ \epsilon_{i23} \\ \epsilon_{i14} \\ \epsilon_{i24} \end{pmatrix}$$

where V_i is an unstructured, symmetric, and positive definite matrix.

2.4 Generalized Least Squares, Categorical Time

The next model we fit involves two responses with two time points each, here the time points are treated as categorical however:

$$y_{i1j} = \beta_{10} + \beta_{11}*A_i + \beta_{12}*I(t_{ij}=1) + \beta_{13}I(t_{ij}=2) + \beta_{14}*A_i*I(t_{ij}=1) + \beta_{15}*A_i*I(t_{ij}=2) + \epsilon_{i1j}$$

and

$$y_{i2j} = \beta_{20} + \beta_{21}*A_i + \beta_{22}*I(t_{ij}=1) + \beta_{23}*I(t_{ij}=2) + \beta_{24}*A_i*I(t_{ij}=1) + \beta_{25}*A_i*I(t_{ij}=2) + \epsilon_{i2j}$$

.

$I()$ is the indicator function, that is:

$$I(x=1) = \begin{cases} 1, & \text{if } x=1 \\ 0, & \text{otherwise} \end{cases}$$

Here y_{i1j} is the response type 1 for individual i at time point j. y_{i2j} is the response type 2 for individual i at time point j. $A_i \in \{0,1\}$ is the treatment indicator for individual i and $t_{ij} \in \{1,2\}$ is the time variable (hour, week, month, etc) at time point j for individual i. We will refer to $t_{i1}=1, t_{i2}=2$ and treat this as categorical. Additionally, we let

$$\begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{22} \end{pmatrix} \sim MVN\left(\mathbf{0}, \Sigma\right)$$

where Σ is an unstructured, symmetric, positive definite matrix:

$$\Sigma = \begin{pmatrix} \sigma_{11}^2 & \sigma_{11,12} & \sigma_{11,21} & \sigma_{11,22} \\ . & \sigma_{12}^2 & \sigma_{12,21} & \sigma_{12,22} \\ . & . & \sigma_{21}^2 & \sigma_{21,22} \\ . & . & . & \sigma_{22}^2 \end{pmatrix}$$

4

.

So our overall, multivariate response is the following:

$$\mathbf{y}_i = \begin{pmatrix} y_{i11} \\ y_{i12} \\ y_{i21} \\ y_{i22} \end{pmatrix} \sim MVN\left(\begin{pmatrix} \mu_{i11} \\ \mu_{i12} \\ \mu_{i21} \\ \mu_{i22} \end{pmatrix}, \begin{pmatrix} \sigma_{11}^2 & \sigma_{11,12} & \sigma_{11,21} & \sigma_{11,22} \\ \cdot & \sigma_{12}^2 & \sigma_{12,21} & \sigma_{12,22} \\ \cdot & \cdot & \sigma_{21}^2 & \sigma_{21,22} \\ \cdot & \cdot & \cdot & \sigma_{22}^2 \end{pmatrix} \right)$$

with the mean vector:

$$\begin{pmatrix} \mu_{i11} \\ \mu_{i12} \\ \mu_{i21} \\ \mu_{i22} \end{pmatrix} = \begin{pmatrix} \beta_{10} + \beta_{11} * A_i + \beta_{12} + \beta_{14} * A_i \\ \beta_{10} + \beta_{11} * A_i + \beta_{13} + \beta_{15} * A_i \\ \beta_{20} + \beta_{21} * A_i + \beta_{22} + \beta_{24} * A_i \\ \beta_{20} + \beta_{21} * A_i + \beta_{23} + \beta_{25} * A_i \end{pmatrix}$$

Since we have categorical time points, the original model is overparameterized. Since we use week one as the reference group, the rest of the parameters are then change from week one. Below is the parameterization used to fit the model:

$$y_{i1j} = \gamma_{10} + \gamma_{11} * A_i + \gamma_{12} * I(t_{ij} = 2) + \gamma_{13} * A_i * I(t_{ij} = 2) + \epsilon_{i1j}$$
$$= (\beta_{10} + \beta_{12}) + (\beta_{11} + \beta_{14}) * A_1 + (\beta_{13} - \beta_{12}) * I(t_{ij} = 2) + (\beta_{15} - \beta_{14}) * A_i * I(t_{ij} = 2) + \epsilon_{i1j}$$

and

$$y_{i2j} = \gamma_{20} + \gamma_{21} * A_i + \gamma_{22} * I(t_{ij} = 2) + \gamma_{23} * A_i * I(t_{ij} = 2) + \epsilon_{i2j}$$
$$= (\beta_{20} + \beta_{22}) + (\beta_{21} + \beta_{24}) * A_1 + (\beta_{23} - \beta_{22}) * I(t_{ij} = 2) + (\beta_{25} - \beta_{24}) * A_i * I(t_{ij} = 2) + \epsilon_{i2j}$$

Since we have categorical time points, the original model is overparameterized. Since we use week one as the reference group, the rest of the parameters are then change from week one. Below is the parameterization used to fit the model:

$$y_{i1j} = \gamma_{10} + \gamma_{11} * A_i + \gamma_{12} * I(t_{ij} = 2) + \gamma_{13} * A_i * I(t_{ij} = 2) + \epsilon_{i1j}$$
$$= (\beta_{10} + \beta_{12}) + (\beta_{11} + \beta_{14}) * A_1 + (\beta_{13} - \beta_{12}) * I(t_{ij} = 2) + (\beta_{15} - \beta_{14}) * A_i * I(t_{ij} = 2) + \epsilon_{i1j}$$

and

$$y_{i2j} = \gamma_{20} + \gamma_{21} * A_i + \gamma_{22} * I(t_{ij} = 2) + \gamma_{23} * A_i * I(t_{ij} = 2) + \epsilon_{i2j}$$
$$= (\beta_{20} + \beta_{22}) + (\beta_{21} + \beta_{24}) * A_1 + (\beta_{23} - \beta_{22}) * I(t_{ij} = 2) + (\beta_{25} - \beta_{24}) * A_i * I(t_{ij} = 2) + \epsilon_{i2j}$$

3 Fitting Models

3.1 Data Preparation

Before we fit this model, we need to format the data where we have an ID, response, treatment, time, response type, and response-time type variables. An example is below:

ID	y	trt	time	y_type	y_type_time
1	14.98	0.00	1.00	y1	y1_1
2	13.31	0.00	1.00	y1	y1_1
3	16.04	0.00	1.00	y1	y1_1
4	11.99	0.00	1.00	y1	y1_1
5	14.64	0.00	1.00	y1	y1_1
6	17.96	0.00	1.00	y1	y1_1
7	13.54	0.00	1.00	y1	y1_1
8	12.96	0.00	1.00	y1	y1_1
9	13.20	0.00	1.00	y1	y1_1
10	12.64	0.00	1.00	y1	y1_1

3.2 Main Commands

Using the ready-made software in R, we are able to fit multivariate mixed models, however we need to be care about model specification and how many parameters we choose to estimate. We use the `lme` function from the `nlme` package.

To fit the 2.1 model with

$$\mathbf{R} = \begin{pmatrix} \sigma_{\epsilon_1}^2 & 0 \\ . & \sigma_{\epsilon_2}^2 \end{pmatrix}$$

we use the following code:

```
## increase maximum number of iterations
cntrl <- lmeControl(maxIter=1000,
                 msMaxIter=1000,
                 niterEM=1000)


## y1 ~ int r-int trt time r-time trt*time
## y2 ~ int r-int trt time r-time trt*time
fit_lme <- lme(y~-1+y_type+y_type:(trt+time+trt:time),
        ## random slope and intercept for each y_type
        random=~-1+(y_type+y_type:time)|id,
        ## different error variances for different y_type
        weights=varIdent(form=~1|y_type),
        control=cntrl,
        data=dat)
```

If we allow for a fully unstructured Σ as in 2.2, we add the `correlation=` command and base the error variances on `y_type_time`.

```
fit_lme_big <- lme(y~-1+y_type+y_type:(trt+time+trt:time),
            ## random slope and intercept for each y_type
            random=~-1+(y_type+y_type:time)|id,
```

```
## unstructured correlation matrix
correlation=corSymm(form=~1|id),
## different error variances for different y_type
weights=varIdent(form=~1|y_type_time),
control=cntrl,
data=dat)
```

if we wished to keep $Var(\epsilon_{1ij}) = Var(\epsilon_{1ik}), j \neq k$, and likewise for $Var(\epsilon_{2ij})$, then we would use `weights=varIdent(form=~1|y_type)` instead.

Unfortunately, if we want to take the 2.1 model where

$$\mathbf{R} = \begin{pmatrix} \sigma^2_{\epsilon_1} & \sigma_{\epsilon_1,\epsilon_2} \\ . & \sigma^2_{\epsilon_1} \end{pmatrix}$$

then we must fit the bigger model (2.2). This forces us to estimate many additional, unnecessary parameters, but ultimately will capture the correct covariance structure.

The generalized least squares model has similar syntax and specifies a totally unstructured $\mathbf{V}_i$ matrix.

```
fit_gls <- gls(y~-1+y_type+y_type:(trt+time+trt:time),
        ## correlation between all responses for an individual
        correlation=corSymm(form=~1|id),
        ## different error variances for different y_type_time
        weights=varIdent(form=~1|y_type_time),
        control=cntrl,
        data=dat)
```

Next we will get into the details of mean specification, random effects specification, and error specification.

For the categorical time generalized least squares model, we change the command a bit and force the time variable to be a factor variable.

```
## factor the time variable
dat$time <- factor(dat$time,levels=c("1","2"))

## fit the model
fit_gls_cat <- gls(y~-1+y_type+y_type:(trt+time+trt:time),
                ## correlation between all responses for an individual
                correlation=corSymm(form = ~ 1|id),
                ## different error variances for different y_type_time
                weights=varIdent(form = ~ 1|y_type_time),
                control=cntrl,
                data=dat)
```

3.3 Fixed Effects Specification

This section is applicable to both `lme` and `gls` models.

We specify the mean model using `y~-1+y_type+y_type:(trt+time+trt:time)`. The `-1+y_type` takes away a common intercept for the response and includes an intercept for each `y_type`. Next we want to include a different treatment effect, time effect, and treatment-time interaction for each `y_type`. This is done by adding `y_type:(trt+time+trt:time)`. The output from `sum_fit <- summary(fit)` for the fixed effects is:

```
> sum_fit$tTable[,1:2]
                          Value    Std.Error
y_typey1           0.515393744 0.012833735
y_typey2          -1.089249454 0.012994978
y_typey1:trt      -0.031988342 0.018214877
y_typey2:trt      -0.011597337 0.018443728
y_typey1:time      2.088942726 0.008654293
y_typey2:time      1.491853770 0.007712711
y_typey1:trt:time  0.003681536 0.012283010
y_typey2:trt:time  0.007253904 0.010946625
```

where you can see there is an intercept for each `y_type` and other appropriate fixed effects and their estimates.

For the categorical time function, we must remember the special parameterizaiton that R enforces using the reference time point.

3.4 Random Effects Specification

This section is only relevant to `lme`.

Similar to fixed effects, we specify a formula for the random effects we are interested in: `random=~-1+(y_type+y_type:time)|id`. The `-1` takes away a common random intercept and the `(y_type+type:time)` adds a random slope and intercept for each `y_type`. We specify we want these random effects for each individual by `|id` at the end. To look at the elements of **G**, use the below code. Be careful, since the structure may be different than what you originally thought. Luckily R gives you row and column names to indicate what estimates you are looking at.

```
> getVarCov(fit_lme,type="random.effects")
Random effects variance covariance matrix
              y_typey1 y_typey2 y_typey1:time y_typey2:time
y_typey1       1.51160  0.40183       1.01690       0.59324
y_typey2       0.40183  1.13830       0.89321       0.80225
y_typey1:time  1.01690  0.89321       1.26780       0.64216
y_typey2:time  0.59324  0.80225       0.64216       0.89645
```

3.5   Error Specification

For the `lme` model, when we ask for random slopes and intercepts for each individual, R automatically generates a covariance matrix for all error terms, that is $(\epsilon_{i11}, \epsilon_{i21}, \epsilon_{i12}, \epsilon_{i22}, \epsilon_{i13}, \epsilon_{i23}, \epsilon_{i14}, \epsilon_{i24})$, hence an 8×8 covariance matrix, rather then the block-diagonal matrix made up of $\mathbf{R}$. The default for this matrix is to constant down the diagonal and 0 on the off diagonal. Since we specified `weights=varIdent(form=~1|y_type)` we ask for different variances for different `y_type`, but still 0 on the off diagonal. Below is the output, showing the variances for all of the ϵ_{i1j} then ϵ_{i2j}. Depending on how your data is sorted, this output may be different. The data used for the below output was sorted $y_{i11}, y_{i12}, y_{i13}, y_{i14}, y_{i21}, y_{i22}, y_{i23}, y_{i24}$. If we have $y_{i11}, y_{i21}, y_{i12}, y_{i22}, y_{i13}, y_{i23}, y_{i14}, y_{i24}$ instead then we will have alternating terms down the diagonal equating to the small $\mathbf{R}$ blocks down the diagonal.

```
> getVarCov(fit_lme,type="conditional",individual=1)$'1'
          1        2        3        4        5        6        7        8
1 1.204063 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
2 0.000000 1.204063 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
3 0.000000 0.000000 1.204063 0.000000 0.000000 0.000000 0.000000 0.000000
4 0.000000 0.000000 0.000000 1.204063 0.000000 0.000000 0.000000 0.000000
5 0.000000 0.000000 0.000000 0.000000 1.508858 0.000000 0.000000 0.000000
6 0.000000 0.000000 0.000000 0.000000 0.000000 1.508858 0.000000 0.000000
7 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 1.508858 0.000000
8 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 1.508858
```

As previously discussed, if we wish to fit the 2.1 model with

$$\mathbf{R} = \begin{pmatrix} \sigma^2_{\epsilon_1} & \sigma_{\epsilon_1, \epsilon_2} \\ . & \sigma^2_{\epsilon_1} \end{pmatrix}$$

we must fit a much larger `gls` or `lme` model, or fit the smaller `lme` model knowing that we have misspecified the model.

This brings up one of the questions that we examine in this simulation study. If we misspecify the error structure: if $\sigma_{\epsilon_1, \epsilon_2} > 0$ but we force it to be 0, what are the ramifications? Do we have trouble estimating the random effects? Do we have trouble estimating the covariance between responses? How much correlation between $\epsilon_{i1j}, \epsilon_{i2j}$ does there need to be before we start running into issues.

3.6   Marginal Covariance

For the overall covariance $\mathbf{V}_i$ we can extract the estimated matrix for both models as follows.

```
V_lme <- getVarCov(fit_lme,type="marginal",individual=1)$'1'
V_gls <- getVarCov(fit_gls,type="marginal",individual=1)
```

Sometimes the generalized least squares model can be trickier when obtaining the variance estimates. Below is an alternative way to gain the estimated $\mathbf{V}_i$ matrix.

```
## vector of estimated variances
weights_cat <-  unique(attributes(fit_cat$modelStruct$varStruct)$weights)
est_sigmas_cat <- matrix( ((1/weights_cat)*fit_cat$sigma)**2,
                          nrow=1,byrow=TRUE)

## estimated correlation matrix
est_cor_cat <- corMatrix(fit_cat$modelStruct$corStruct)[1]$'1'

## estimated covariance matrix
sd_diag_cat <- (diag(c(est_sigmas_cat))**0.5)
est_cov_cat <- sd_diag_cat %*% est_cor_cat %*% sd_diag_cat
```

4 Simulation Setup

4.1 Questions to Address

For this simulation, we generate data from the model in 2.1 with varying levels of error correlation. When $\rho = 0$, we are able to correctly specify a parsimonious mixed model. The larger mixed model and the generalized least squares model are also still correctly specified in this case, but need far more parameters to estimate the same parameters.

When $\rho > 0$, the smaller mixed model is incorrectly specified and the others are correctly specified. The goal of the simulation is to understand, given certain sample sizes and levels of ρ, which models are appropriate. Some specific questions we wish to address are:

- If $\rho = 0$, what do we gain from using the smallest model?

- If $\rho > 0$, what issues do we have with the smallest model?

- How large does ρ need to be before it is clear we should use a much larger, but correctly specified model?

- Since we don't know ρ and it's difficult to directly estimate it, how effective are other methods in telling us the larger model is necessary?

4.2 Data Generation and Simulation Inputs

Using data from a relevant clinical trial, we simulate data with the following parameters.

$$
\mathbf{G} = \begin{pmatrix} \sigma_{b_{10}}^2 & \sigma_{b_{10},b_{11}} & \sigma_{b_{10},b_{20}} & \sigma_{b_{10},b_{21}} \\ . & \sigma_{b_{11}}^2 & \sigma_{b_{11},b_{20}} & \sigma_{b_{11},b_{21}} \\ . & . & \sigma_{b_{20}}^2 & \sigma_{b_{20},b_{21}} \\ . & . & . & \sigma_{b_{21}}^2 \end{pmatrix} = \begin{pmatrix} 0.959 & 0.049 & 0.730 & -0.043 \\ . & 0.031 & 0.033 & 0.010 \\ . & . & 1.826 & -0.025 \\ . & . & . & 0.008 \end{pmatrix}
$$

and

$$\mathbf{R} = \begin{pmatrix} \sigma_{\epsilon_1}^2 & \sigma_{\epsilon_1,\epsilon_2} \\ \cdot & \sigma_{\epsilon_1}^2 \end{pmatrix} = \begin{pmatrix} \sqrt{1.55} & 0 \\ 0 & \sqrt{0.571} \end{pmatrix} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \begin{pmatrix} \sqrt{1.55} & 0 \\ 0 & \sqrt{0.571} \end{pmatrix}$$

and we vary $\rho = \{0, 0.25, 0.50, 0.75\}$ where 0 indicates correct specification (no correlation), and then we increase the correlation to a low setting (0.25), medium setting (0.50), and a high setting (0.75). Since we are forcing $\rho = 0$ in the software, the higher correlation indicates more severe misspecification.

We take $t_{i1} = 0, t_{i2} = 2, t_{i3} = 4, t_{i4} = 8$ and treat this as a continuous feature. Thus our overall covariance matrix for all responses at all time points is the following:

$$\mathbf{V}_i = Cov \begin{pmatrix} y_{i11} \\ y_{i21} \\ y_{i12} \\ y_{i22} \\ y_{i13} \\ y_{i23} \\ y_{i14} \\ y_{i24} \end{pmatrix} = \mathbf{Z}_i \mathbf{G} \mathbf{Z}_i^T + \begin{pmatrix} \mathbf{R} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} \\ \cdot & \mathbf{R} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} \\ \cdot & \cdot & \mathbf{R} & \mathbf{0}_{2\times 2} \\ \cdot & \cdot & \cdot & \mathbf{R} \end{pmatrix}$$

$$= \begin{pmatrix} 2.509 & 0.730 + \sigma_{\epsilon_1,\epsilon_2} & 1.057 & 0.645 & 1.155 & 0.559 & 1.351 & 0.387 \\ \cdot & 2.398 & 0.797 & 1.777 & 0.864 & 1.728 & 0.997 & 1.629 \\ \cdot & \cdot & 2.827 & 0.752 + \sigma_{\epsilon_1,\epsilon_2} & 1.498 & 0.707 & 1.938 & 0.616 \\ \cdot & \cdot & \cdot & 2.331 & 0.859 & 1.743 & 1.074 & 1.708 \\ \cdot & \cdot & \cdot & \cdot & 3.391 & 0.855 + \sigma_{\epsilon_1,\epsilon_2} & 2.526 & 0.845 \\ \cdot & \cdot & \cdot & \cdot & \cdot & 2.329 & 1.150 & 1.787 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 5.251 & 1.303 + \sigma_{\epsilon_1,\epsilon_2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 2.518 \end{pmatrix}$$

where

$$\mathbf{Z}_i = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 2 \\ 1 & 4 & 0 & 0 \\ 0 & 0 & 1 & 4 \\ 1 & 8 & 0 & 0 \\ 0 & 0 & 1 & 8 \end{pmatrix}$$

We simulate out of the following mean model:

11

$$\begin{pmatrix} \beta_{10} \\ \beta_{20} \\ \beta_{11} \\ \beta_{21} \\ \beta_{12} \\ \beta_{22} \\ \beta_{13} \\ \beta_{23} \end{pmatrix} = \begin{pmatrix} 7.044 \\ 1.243 \\ 0.068 \\ 0.081 \\ -0.056 \\ -0.004 \\ -0.107 \\ -0.081 \end{pmatrix}$$

4.3   Models to Fit

We fit a small `lme` model as discussed in 2.1 with a diagonal R matrix. The code is below.

```
fit_lme <- lme(y~-1+y_type+y_type:(trt+time+trt:time),
               ## random slope and intercept for each y_type
               random=~-1+(y_type+y_type:time)|id,
               ## different error variances for different y_type
               weights=varIdent(form=~1|y_type),
               control=cntrl,
               data=dat)
```

We also fit a larger `lme` model as discussed in 2.2 with unstructured covariances between all error terms and variances based on the response type.

```
fit_lme2 <- lme(y~-1+y_type+y_type:(trt+time+trt:time),
                ## random slope and intercept for each y_type
                random=~-1+(y_type+y_type:time)|id,
                ## unstructured correlation matrix
                correlation=corSymm(form=~1|id),
                ## different error variances for different y_type
                weights=varIdent(form=~1|y_type),
                control=cntrl,
                data=dat)
```

The last model we fit is `gls` model with a totally unstructured covariance matrix for the errors.

```
fit_gls <- gls(y~-1+y_type+y_type:(trt+time+trt:time),
               ## different error variances for different y_type_time
               correlation=corSymm(form=~1|id),
               weights=varIdent(form=~1|y_type_time),
               control=cntrl,
               data=dat)
```

4.4   Simulation Inputs

We vary the following factors:

- Sample size: $N = 40, 100, 200, 300, 400$ with $N/2$ patients randomized to $A_i = 1$, $N/2$ to $A_i = 0$.

- Correlation of errors: $\rho = 0.0, 0.25, 0.5, 0.75$

and replicate each scenario 1000 times. We are interested in the monte carlo mean, the monte carlo standard deviation, and the sampling distribution of parameter estimates for both fixed and random effects.

5   Simulation Results

First we will look at the effects of misspecification for mixed models. We wish to look at the estimates of covariances parameters for marginal variance, random effects, and random error.

5.1   Marginal Variance

When estimating variances of certain responses and covariances between different responses, we want to compare the mixed models versus a generalized least squares model with a totally unstructured covariance matrix.

Generally, we find that when $\rho = 0$ all models consistently estimate the covariance parameters. In this case, the smallest mixed model's estimates are more precise. Intuitively, since we have fewer parameters to estimate in the mixed model, it should have more precision.

When $\rho > 0$, the bias of the smaller mixed model's estimates increase and the larger mixed model and the generalized least squares model still consistently estimates the parameters. This bias is exaggerated when estimating covariance parameters that depend on $\sigma_{\epsilon_1, \epsilon_2} = \rho \sigma_{\epsilon_1} \sigma_{\epsilon_2}$. The next pages display these results graphically.

Below is a plot of $\sigma_{y_{11},y_{22}}$, which does not directly depend on ρ.

Estimates of Cov(y11,y22)
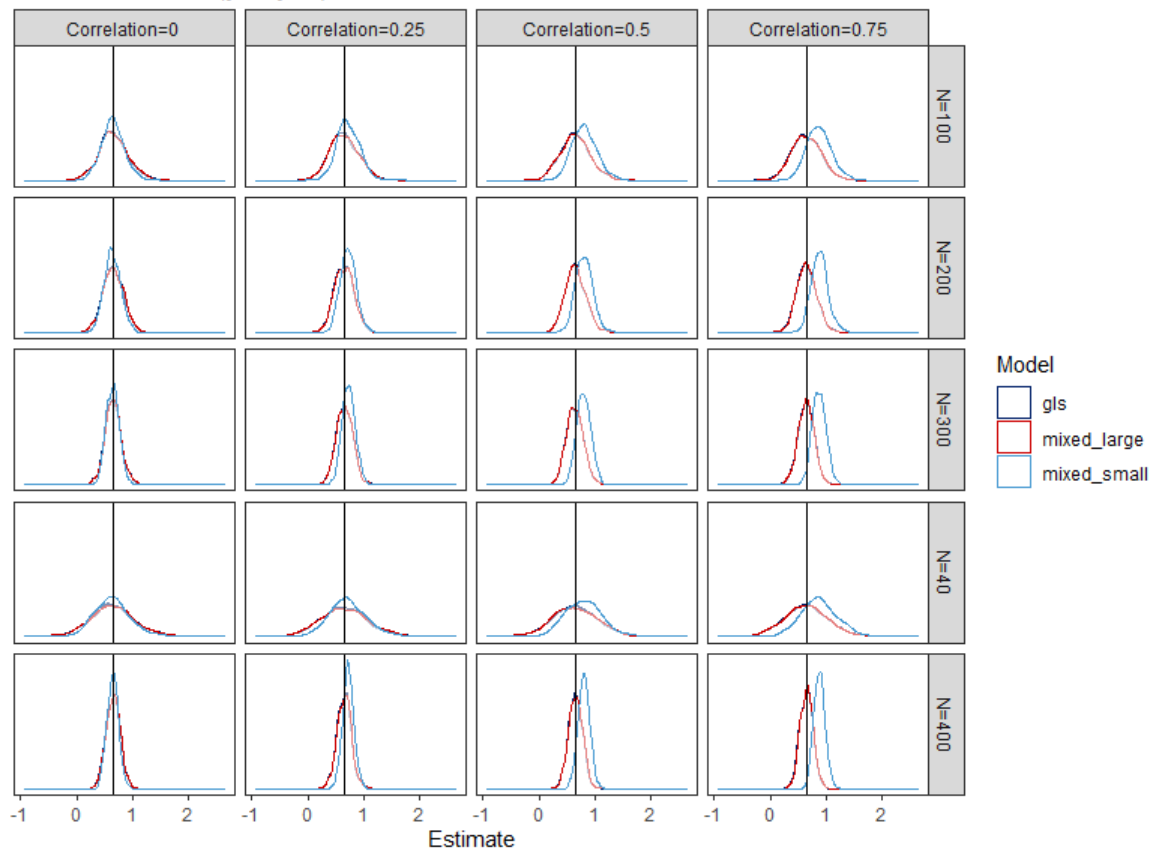
And now a plot of $\sigma_{y_{11}, y_{21}}$, which does directly depend on ρ.

Estimates of Cov(y11,y21)
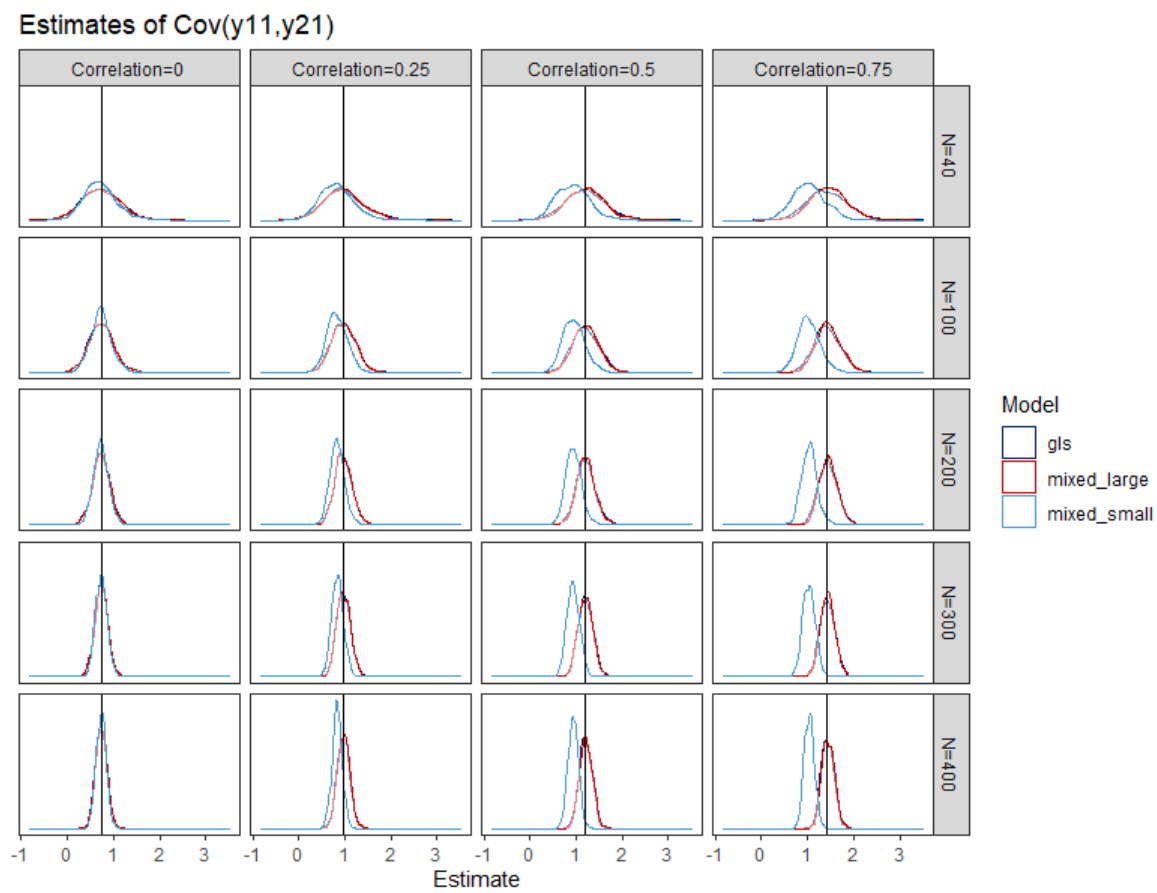
To zero in on the precision of both models' estimators, we examine plots and summary statistics for $\sigma^2_{y_{i11}}$.
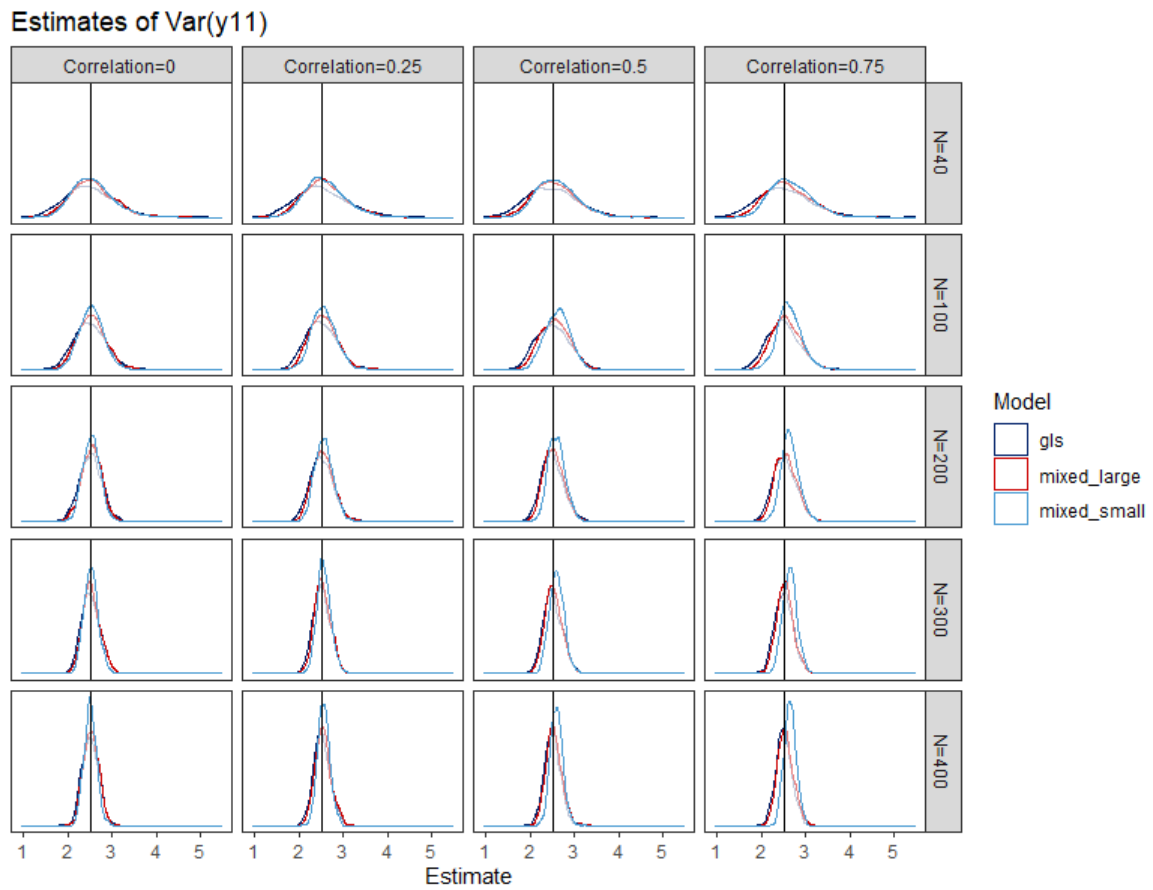
Estimates of Var(y11)

Summary of Estimates for $\hat{\sigma}^2_{y_{i11}} = 2.509$ When N=40

N	ρ	Model	Mean($\hat{\sigma}^2_{y_{i11}}$)	SD($\hat{\sigma}^2_{y_{i11}}$)	Mean Abs Error	Mean Sq Error
40	0.00	gls	2.52	0.56	0.44	0.31
	0.00	mixed_large	2.57	0.49	0.38	0.24
	0.00	mixed_small	2.56	0.44	0.35	0.20
	0.25	gls	2.52	0.57	0.45	0.33
	0.25	mixed_large	2.58	0.47	0.37	0.23
	0.25	mixed_small	2.58	0.44	0.35	0.20
	0.50	gls	2.50	0.56	0.45	0.32
	0.50	mixed_large	2.55	0.47	0.38	0.22
	0.50	mixed_small	2.60	0.44	0.35	0.20
	0.75	gls	2.53	0.58	0.46	0.34
	0.75	mixed_large	2.59	0.48	0.39	0.24
	0.75	mixed_small	2.66	0.44	0.36	0.21

Summary of Estimates for $\hat{\sigma}^2_{y_{i11}} = 2.509$ When N=100

N	ρ	Model	Mean($\hat{\sigma}^2_{y_{i11}}$)	SD($\hat{\sigma}^2_{y_{i11}}$)	Mean Abs Error	Mean Sq Error
100	0.00	gls	2.50	0.36	0.29	0.13
	0.00	mixed_large	2.54	0.31	0.25	0.10
	0.00	mixed_small	2.54	0.27	0.21	0.07
	0.25	gls	2.50	0.34	0.27	0.12
	0.25	mixed_large	2.54	0.30	0.24	0.09
	0.25	mixed_small	2.56	0.27	0.21	0.07
	0.50	gls	2.50	0.35	0.29	0.13
	0.50	mixed_large	2.55	0.31	0.26	0.10
	0.50	mixed_small	2.61	0.27	0.24	0.08
	0.75	gls	2.49	0.35	0.28	0.12
	0.75	mixed_large	2.55	0.31	0.25	0.10
	0.75	mixed_small	2.63	0.27	0.23	0.09

Summary of Estimates for $\hat{\sigma}^2_{y_{i11}} = 2.509$ When N=200

N	ρ	Model	Mean($\hat{\sigma}^2_{y_{i11}}$)	SD($\hat{\sigma}^2_{y_{i11}}$)	Mean Abs Error	Mean Sq Error
200	0.00	gls	2.52	0.25	0.20	0.06
	0.00	mixed_large	2.54	0.22	0.18	0.05
	0.00	mixed_small	2.53	0.19	0.15	0.04
	0.25	gls	2.51	0.26	0.20	0.07
	0.25	mixed_large	2.54	0.24	0.19	0.06
	0.25	mixed_small	2.56	0.20	0.16	0.04
	0.50	gls	2.50	0.24	0.19	0.06
	0.50	mixed_large	2.53	0.22	0.18	0.05
	0.50	mixed_small	2.60	0.19	0.16	0.04
	0.75	gls	2.52	0.26	0.21	0.07
	0.75	mixed_large	2.55	0.24	0.20	0.06
	0.75	mixed_small	2.64	0.20	0.19	0.06

Summary of Estimates for $\hat{\sigma}^2_{y_{i11}} = 2.509$ When N=300

N	ρ	Model	Mean($\hat{\sigma}^2_{y_{i11}}$)	SD($\hat{\sigma}^2_{y_{i11}}$)	Mean Abs Error	Mean Sq Error
300	0.00	gls	2.52	0.21	0.17	0.04
	0.00	mixed_large	2.54	0.19	0.15	0.04
	0.00	mixed_small	2.53	0.16	0.13	0.02
	0.25	gls	2.51	0.19	0.15	0.04
	0.25	mixed_large	2.53	0.18	0.14	0.03
	0.25	mixed_small	2.56	0.15	0.13	0.03
	0.50	gls	2.51	0.21	0.17	0.04
	0.50	mixed_large	2.53	0.19	0.15	0.04
	0.50	mixed_small	2.60	0.16	0.15	0.03
	0.75	gls	2.51	0.20	0.16	0.04
	0.75	mixed_large	2.53	0.19	0.15	0.04
	0.75	mixed_small	2.64	0.16	0.17	0.04

Summary of Estimates for $\hat{\sigma}^2_{y_{i11}} = 2.509$ When N=400

N	ρ	Model	Mean($\hat{\sigma}^2_{y_{i11}}$)	SD($\hat{\sigma}^2_{y_{i11}}$)	Mean Abs Error	Mean Sq Error
400	0.00	gls	2.51	0.18	0.15	0.03
	0.00	mixed_large	2.53	0.17	0.14	0.03
	0.00	mixed_small	2.51	0.14	0.11	0.02
	0.25	gls	2.51	0.18	0.14	0.03
	0.25	mixed_large	2.54	0.18	0.14	0.03
	0.25	mixed_small	2.56	0.14	0.11	0.02
	0.50	gls	2.50	0.18	0.14	0.03
	0.50	mixed_large	2.53	0.18	0.14	0.03
	0.50	mixed_small	2.59	0.14	0.13	0.03
	0.75	gls	2.51	0.18	0.14	0.03
	0.75	mixed_large	2.54	0.18	0.14	0.03
	0.75	mixed_small	2.65	0.14	0.16	0.04

Finally, we look at overall difference between the true marginal covariance matrix and the estimated. We measure difference by the frobenius norm

$$||\mathbf{V}_i - \hat{\mathbf{V}}_i||_F = \left(\sum_{i=1}^{8} \sum_{j=1}^{8} |v_{ij} - \hat{v}_{ij}|^2 \right)^{1/2}$$

where v_{ij} is the component of $\mathbf{V}_i$ taking the i-th row and j-th column. $\hat{v}_{ij}$ is similarly defined.

As expected, when the mixed model is correctly specified, we see that the small mixed model provides better estimation. When the model is slightly off ($\rho = 0.25$), then the smaller mixed model still provides a more accurate estimate of the overall covariance. Once we have a higher correlation (0.5 and above), the `gls` and larger `lme` models provide a better fit.

Model Performance on Overall Covariance Matrix V
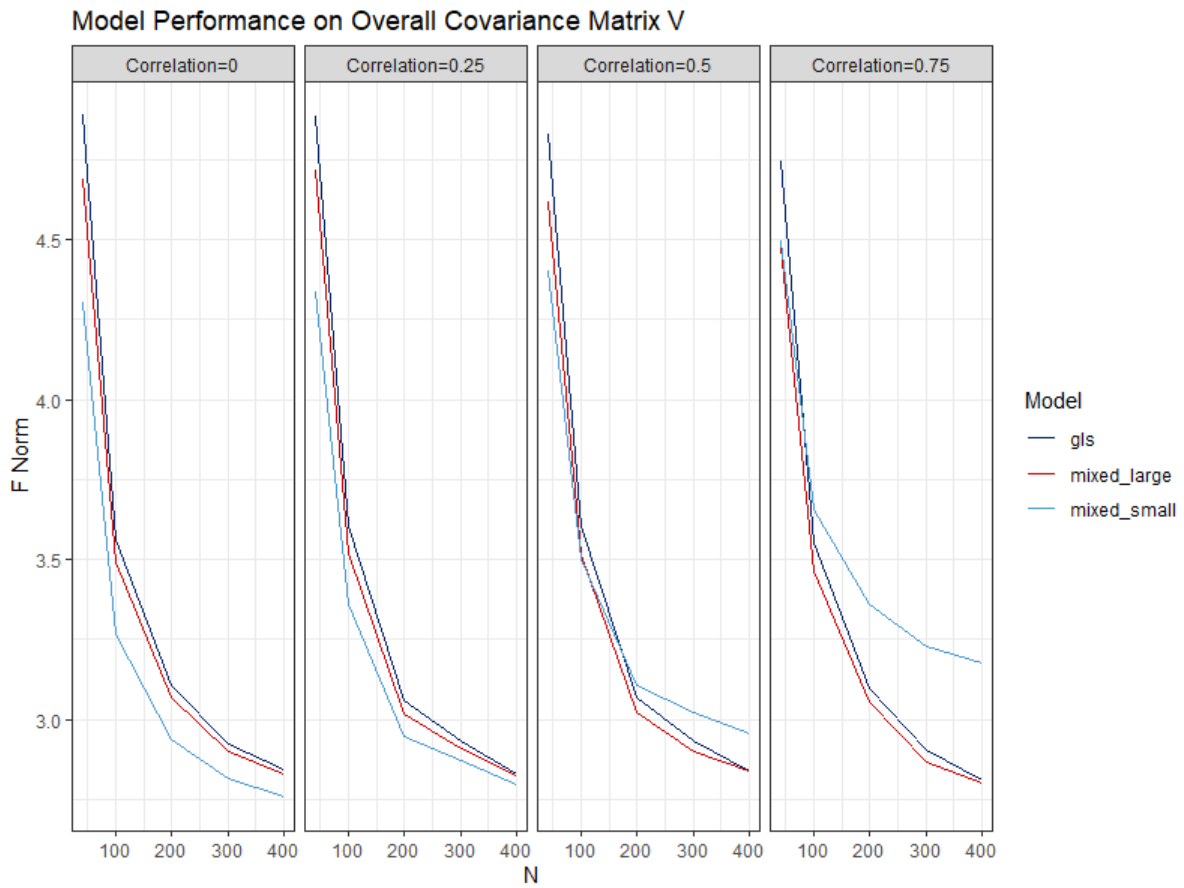
5.2   Random Effects

As ρ increases, both mixed models have a harder time estimating the random effects. When $\rho \leq 0.5$, we see that the smaller model uniformly outperforms the larger model in estimating the random effects. When ρ is very high, then the larger model is better. This is likely due to the fact that the larger model has to estimate so many more covariance parameters that are actually 0, so it has a harder time discerning if covariance is coming from random effects or from residual error.

Model Performance on Random Effects Covariance Matrix G
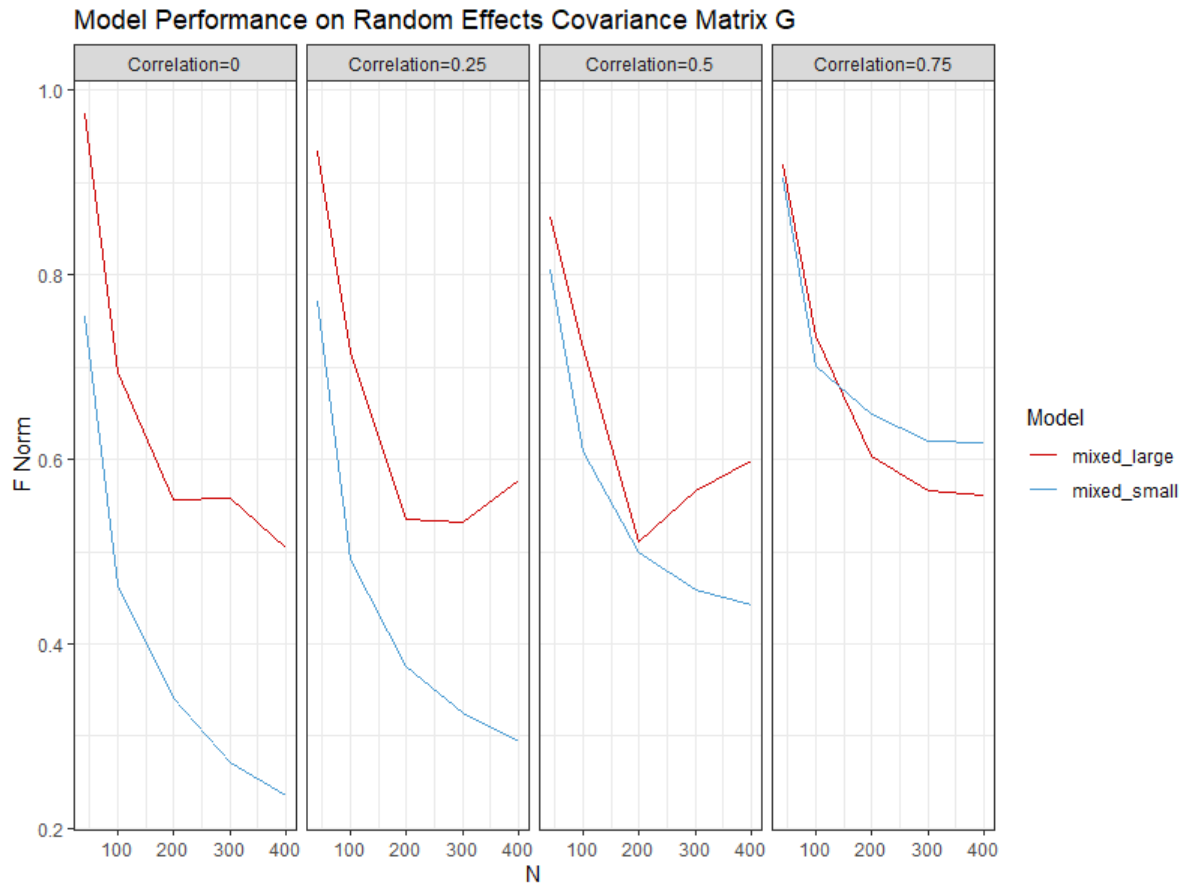
5.3   Random Error

The results are less clear for random error. When $\rho = 0$, then the smaller mixed model provides a consistent and more efficient estimate of both $Var(\epsilon_{i1j})$ and $Var(\epsilon_{i2j})$. When ρ increases, the smaller model's estimates become biased and the larger model's become biased for ϵ_{i1j} but not ϵ_{i2j}. Ultimately we are not terribly interested in the estimates of residual variance, but rather the overall variance. The next pages display these results graphically.
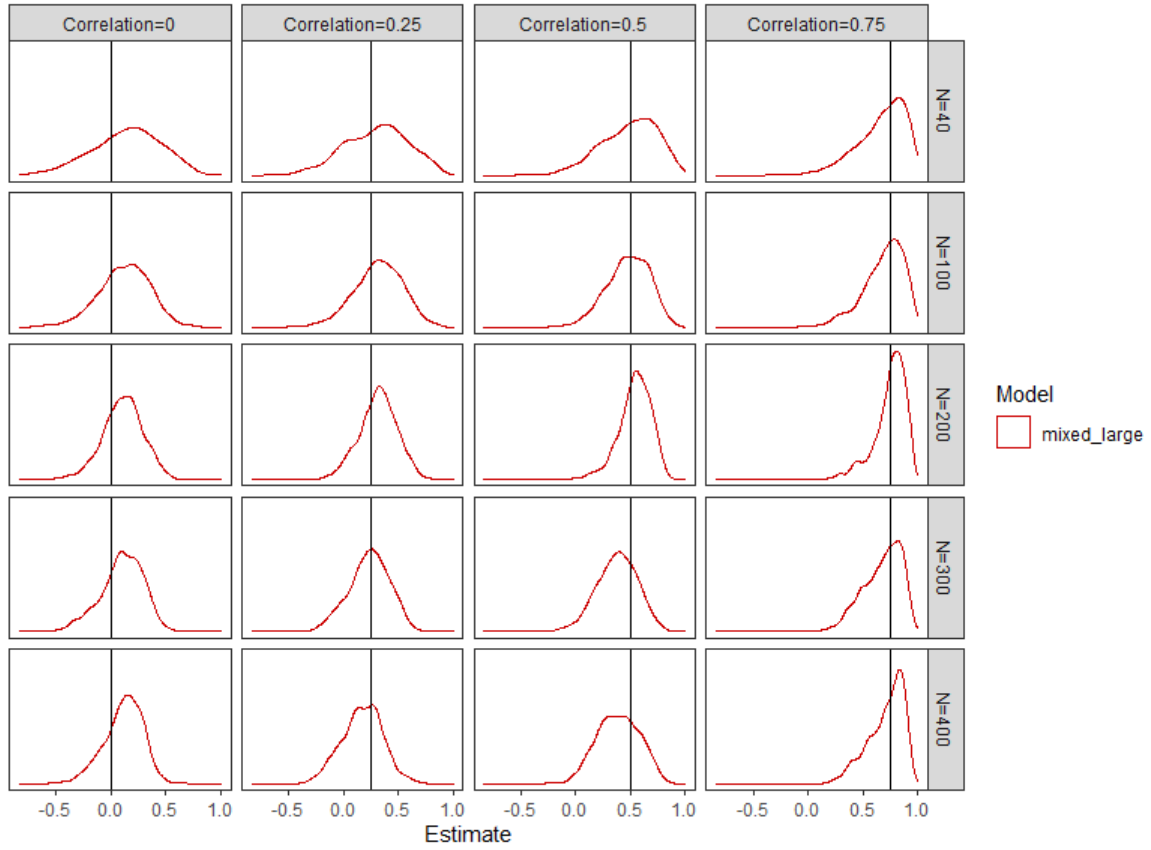
Estimates of Var(e1)
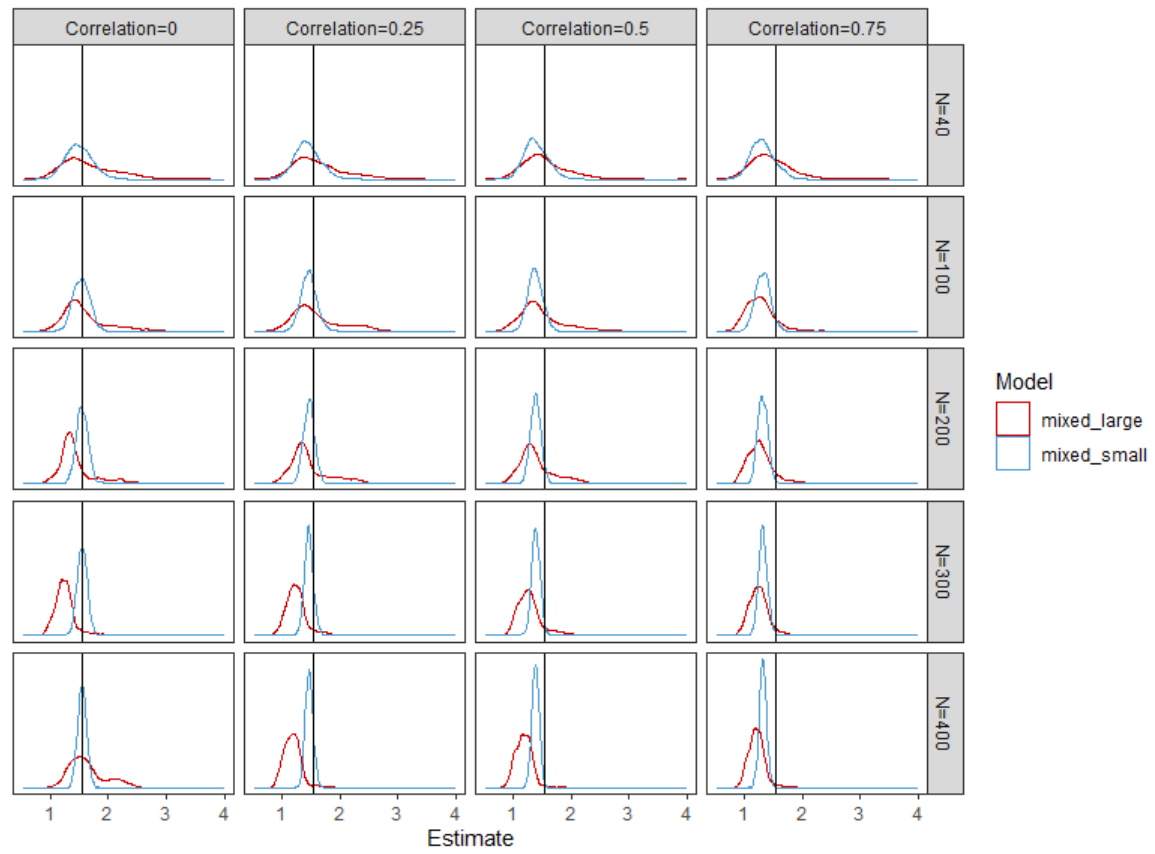
Estimates of Var(e2)
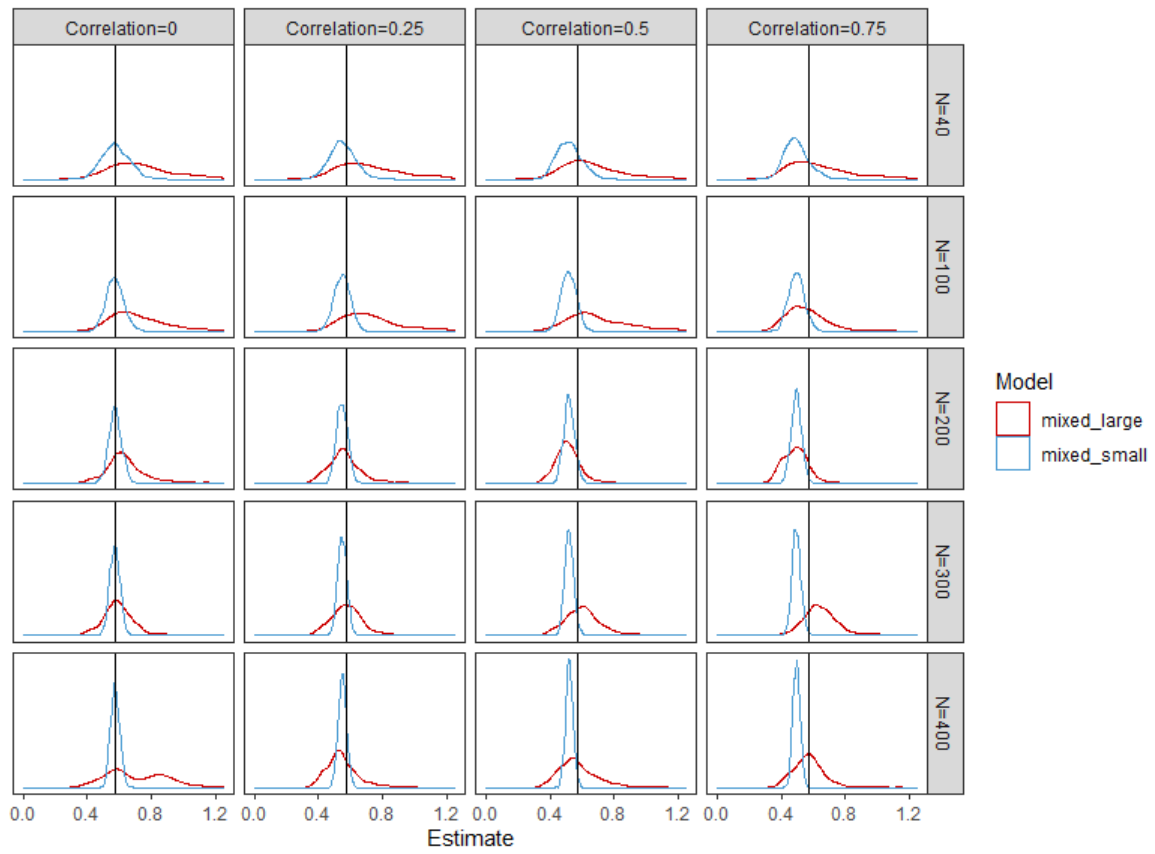
Finally, we see biased and not-precise but "in the ballpark" estimates of ρ by looking at the correlation between $\epsilon_{i11}, \epsilon_{i21}$ from the biggest model.

Estimates of Cor(e11,e21)

5.4 Fixed Effects

For all three models, the fixed effects estimation is virtually the same. All models produce consistent results with similar precision.


Model Performance on Fixed Effects
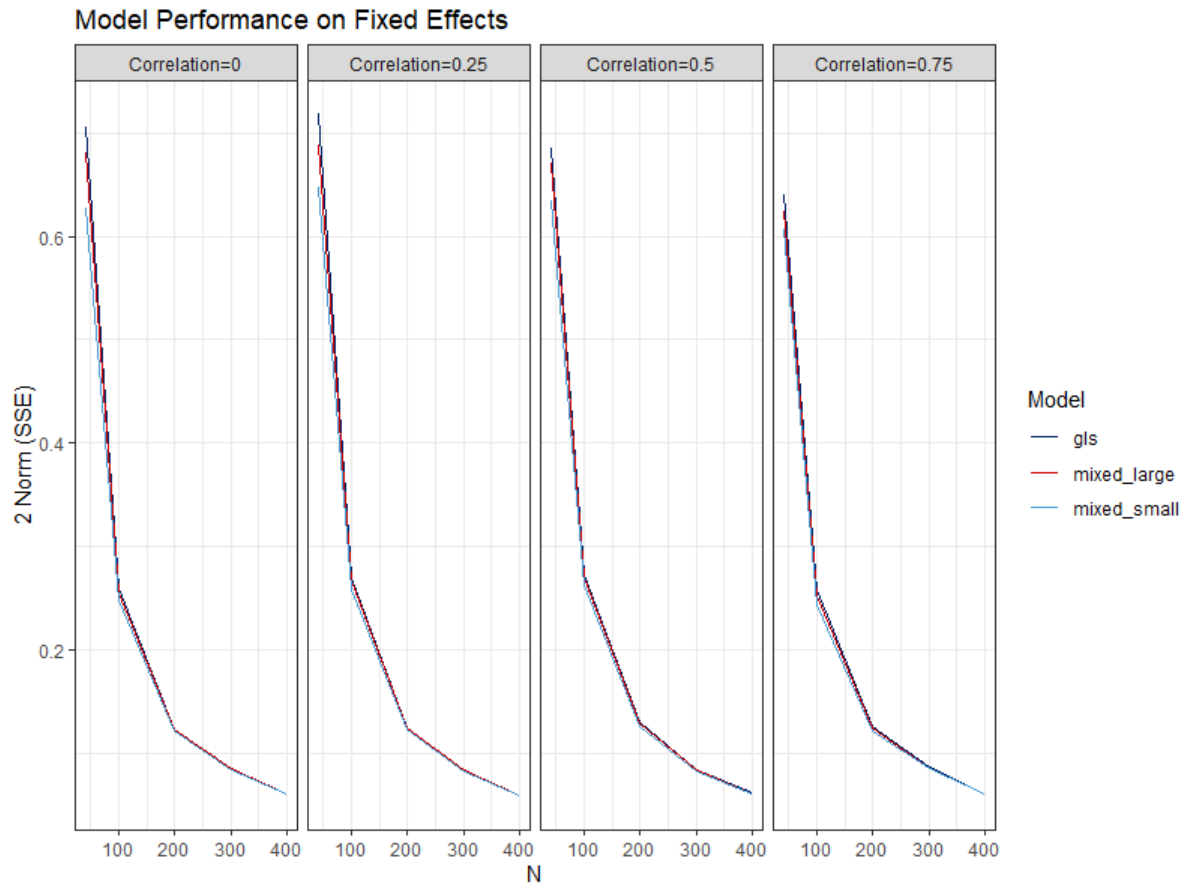
For illustration, here are the monte carlo means and standard deviations for the estimate of β_{12}. Similar trends (equal estimates and standard deviations) are see across sample sizes and other parameters.

Summary of Estimates for $\hat{\beta}_{12}$ When N=100

| N	ρ	Model	Mean($\hat{\beta}_{12}$)	SD($\hat{\beta}_{12}$)	Mean Abs Error	Mean Sq Error
100	0.00	gls	-0.05	0.04	0.03	0.00
	0.00	mixed_large	-0.06	0.04	0.03	0.00
	0.00	mixed_small	-0.06	0.04	0.03	0.00
	0.25	gls	-0.06	0.04	0.03	0.00
	0.25	mixed_large	-0.06	0.04	0.03	0.00
	0.25	mixed_small	-0.06	0.04	0.03	0.00
	0.50	gls	-0.06	0.04	0.03	0.00
	0.50	mixed_large	-0.06	0.04	0.03	0.00
	0.50	mixed_small	-0.06	0.04	0.03	0.00
	0.75	gls	-0.06	0.04	0.03	0.00
	0.75	mixed_large	-0.06	0.04	0.03	0.00
	0.75	mixed_small	-0.06	0.04	0.03	0.00

5.5 Model Selection

Results up until this point indicate that the smaller mixed model is certainly a better fit when $\rho = 0$ and likely a better fit when ρ is small (≤ 0.25). Since our estimates of ρ from the larger mixed model are likely to be off and fitting the larger mixed model may pose computational difficulties, we examine other methods to gauge if we have misspecified the model and if we should use the larger model.

The first item we examine is the average BIC across monte carlo replicates. When $\rho = 0$, BIC unsurprisingly considers the smallest model – which is correctly specified – as the best fit. When the sample size is large, BIC is quick to indicate that the smaller model is misspecified.

BIC for Different Models
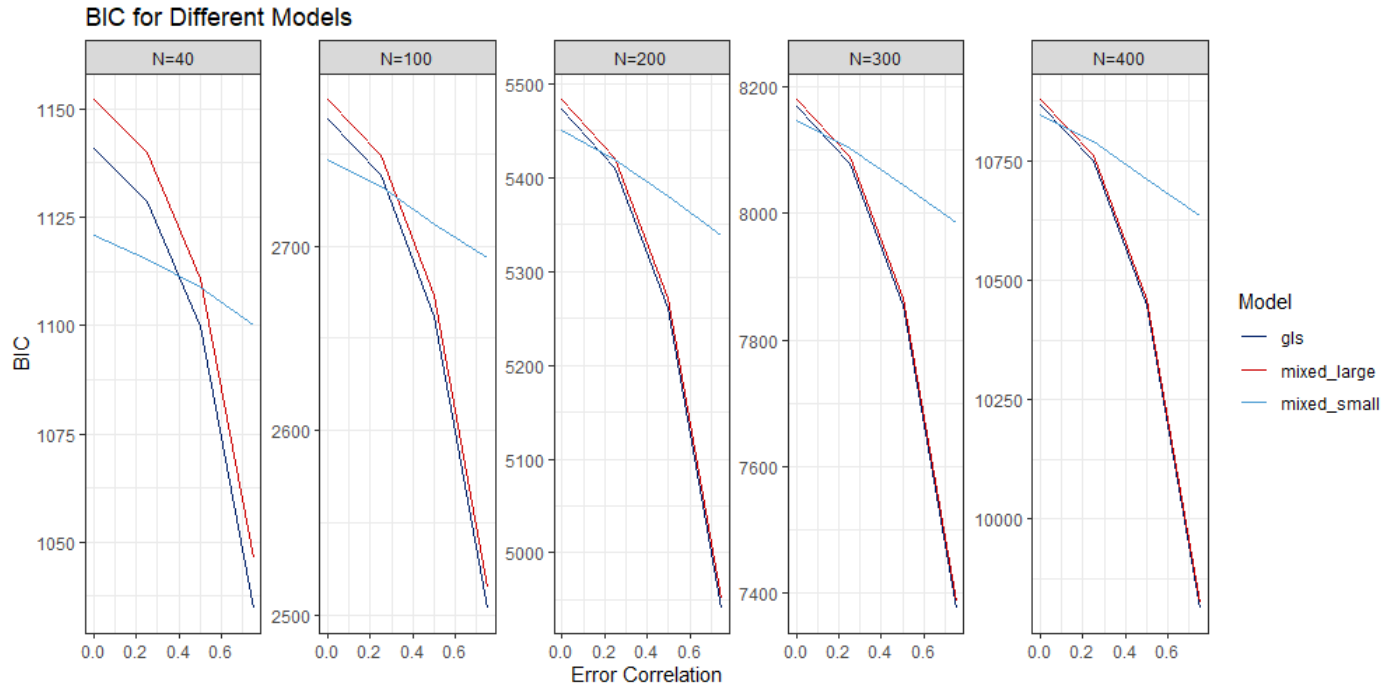
The next item are likelihood ratio tests between the smaller model and the larger models. We find that when comparing the smaller and larger mixed models, for all sample sizes, the likelihood ratio test will have a p-value less than 0.05 with high probability if $\rho \geq 0.5$. For larger sample sizes, we see that we will likely reject if $\rho \geq 0.25$. When $\rho = 0$, we can see that we rarely reject and type one error is held.

Probability of Rejecting H0 in LRT Small vs Large Mixed Models
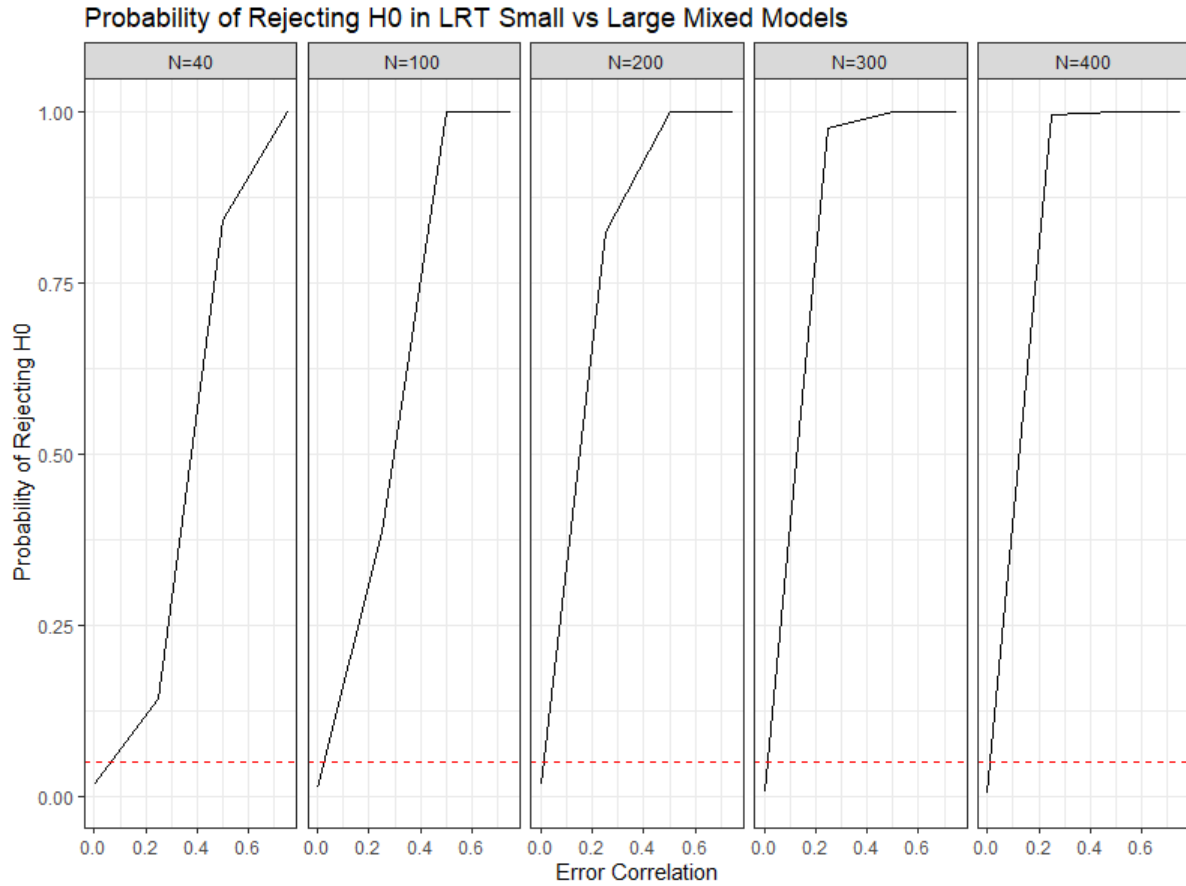
When testing the smaller mixed model to the generalized least squares model, we see similar results, but the chance for a type one error is inflated, especially when the sample size is small.

Probability of Rejecting H0 in LRT Small Mixed vs Generalized Least Squares
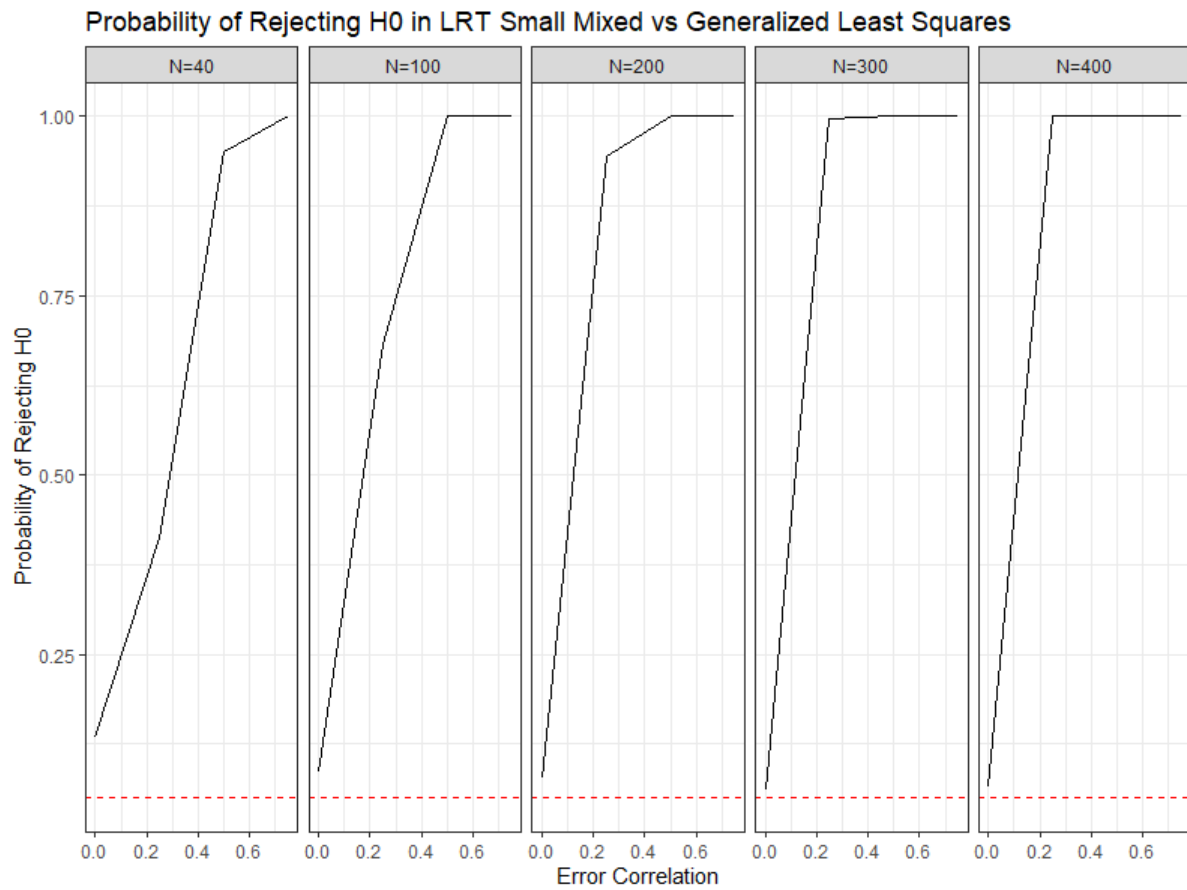
Additionally, the likelihood ratio test rarely suggests the larger mixed model over the generalized least squares model.

Probability of Rejecting H0 in LRT Generalized Least Squares vs Large Mixed
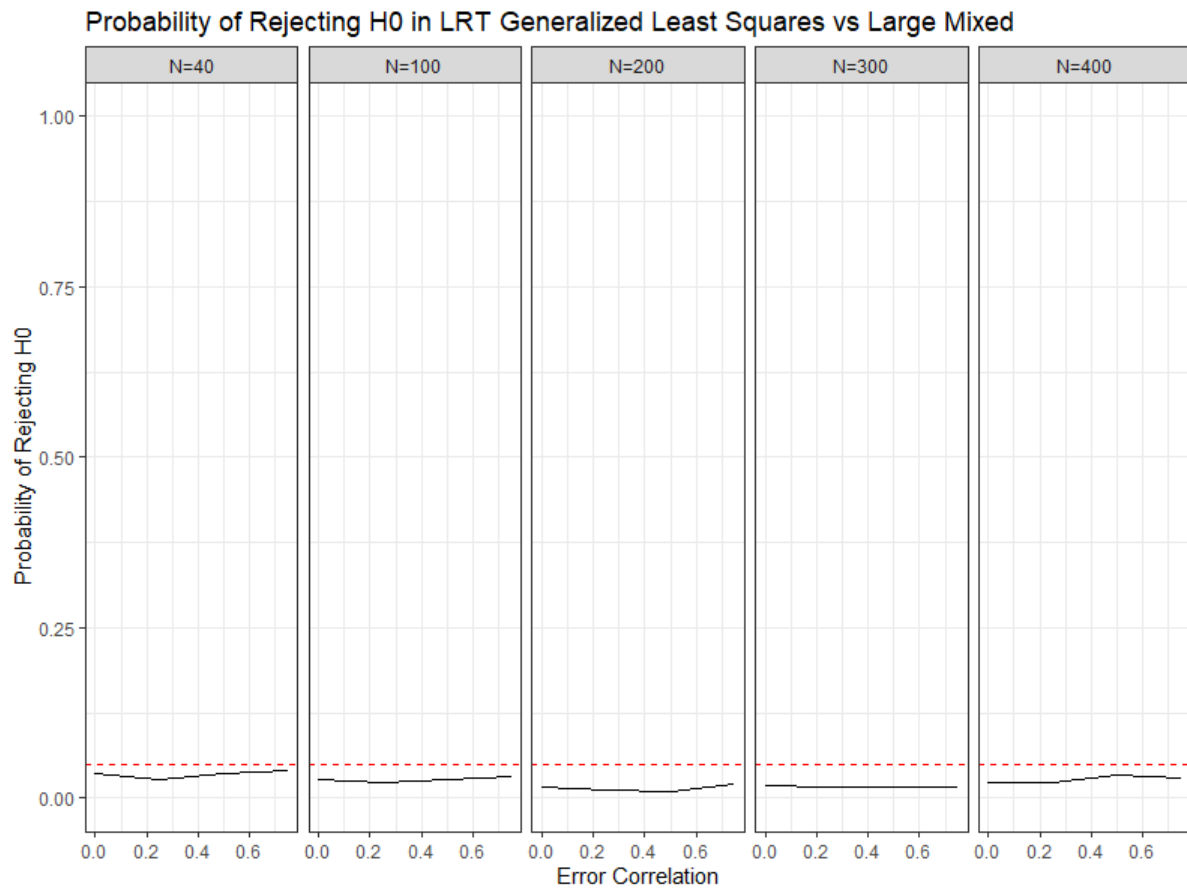
6 Discussion

6.1 Research Questions

Let us re-visit the original questions the simulation study sought to answer.

If $\rho = 0$, what do we gain from using the smallest model?

When $\rho = 0$, all models provide similar mean estimation. The smaller mixed model offers more accurate and efficient random effects estimation, random error estimation, and overall covariance estimation. When $\rho = 0$, BIC will likely show the smaller model is a better fit and it is unlikely that the a likelihood ratio test will suggest a larger model. If these criteria suggest the smaller mixed model, then there is no disadvantage to choosing it.

If $\rho > 0$, what issues do we have with the smallest model?

When $\rho > 0$, the smaller mixed model still estimates the mean parameters as well as the larger models. When $\rho > 0$ but still relatively small ($\rho \leq 0.25$) we still see the smaller mixed model estimate the overall covariance and covariance of random effects more accurately than larger models. Once ρ is relatively high ($\rho \geq 0.5$), we see that the larger models provide better covariance estimation. When N is large, we see gains from the larger models when ρ is smaller.

How large does ρ need to be before it is clear we should use a much larger, but correctly specified model?

When $\rho \leq 0.25$, the smaller mixed model provides a better estimate of the overall covariance structure. We do see, for individual covariance parameters, the smaller mixed model may have biased, but precise estimates, whereas the larger models have unbiased, but high variance estimates – leading to the classic tradeoff. Generally, if $\rho \geq 0.5$ and $N \geq 200$, the larger models will be better in estimating overall covariance.

Since we don't know ρ and it's difficult to directly estimate it, how effective are other methods in telling us the larger model is necessary?

If we are able to fit a larger mixed model, we can get a ballpark estimate of ρ by looking at $Cor(\epsilon_{i1j}, \epsilon_{i2j})$ for certain j. If this estimate is large ($|\hat{\rho}| \geq 0.5$), then it is likely we have error correlation. If the estimate is non-zero but has a small magnitude, it's hard to discern whether or not true error correlation exists based on the larger mixed model with its troubles. As expected if the sample size is large, our estimates become more accurate and more confidence can be placed on these estimates.

Using information criteria like BIC provide a reasonable way to gauge if the larger model is necessary. When the sample size is smaller, we see that the smaller model has better performance for relatively small ρ even though it is misspecified. This is reflected in BIC, where it, on average, shows smaller BIC for the smaller mixed model when N is small and $\rho \leq 0.5$. Once N is large, BIC quickly shows the larger models are better.

The likelihood ratio tests provide similar information to BIC. The likelihood ratio test between the smaller mixed model and the generalized least squares model has inflated type one error for all sample sizes, with more inflation as N decreases. It is certainly worth looking at, but in addition to the direct estimates and information criteria, not as the only

criteria.

6.2   General Process for Model Selection

Model building is an art, not a science, so there is not one process on how to build a multivariate longitudinal model. We first see what we can fit, then evaluate which models best answer our scientific questions. Based on our experience, the following offers a general sequence of steps that can help a researcher build a useful multivariate longitudinal model.

1. Fit as many relevant models as possible. With real data, convergence issues are common, so making slight adjustments to mean and covariance structures is often necessary.

2. Examine model output like $\hat{\mathbf{V}}_i, \hat{\beta}, \hat{\mathbf{R}}, \hat{\mathbf{G}}$. If a model(s) looks to have considerably different output than others, consider why that might be (certain parameter restrictions, false convergence, etc). If the outlier model(s) has an intuitive reason for being different and information criteria suggest a worse fit, then throw it out.

3. Evaluate all models using information criteria, likelihood ratio tests, and practical use (e.g. if examination of random effects' covariance structure is important scientifically, give favor to a mixed model) and choose accordingly.

4. When conducting inference on functions of mean parameters, it can be useful to conduct inference using different reasonable models. If the models agree, then the we can be more confident in the inference. If they disagree, we should be wary.

7 Real Data Example

Our goal is to jointly model two variables – MAYO score (MAYO) and partial MAYO score (pMAYO). Our main goals are to look at the effectiveness of the treatments, hoping they will reduce the MAYO score from baseline to week eight and also to look at the correlation between pMAYO at earlier weeks (two, four) and MAYO at week eight. A motivating idea is that if pMAYO at an earlier week is highly correlated with MAYO at the end of the study, then we could make inference on pMAYO earlier in the trial and get the same results at using MAYO later in the trial.

7.1 Data

Our dataset contains the following variables:

- USUBJID: Subject ID

- MAYO: MAYO score

- pMAYO: Partial MAYO score (MAYO Score without the endoscopy score, a valuable metric that depends on an invasive procedure)

- Week: Week that an observation was recorded at (0,2,4,8)

- Treatment: Treatment the individual received, either Placebo, Treatment A, or Treatment B

7.2 Models We Are Able to Fit

In the discussion, the first step to building a multivariate longitudinal model is simply to see what models we can fit. Maximizing the multivariate normal likelihood can be difficult and many models fail to converge. So, simply seeing how many reasonable models we can fit is the first step.

7.2.1 Basic Notation

The first outcomes we examine are the MAYO score and partial MAYO Score (pMAYO). We observe MAYO at baseline and at week eight. pMAYO is recorded at baseline and weeks two, four, and eight. Since MAYO is only measured at baseline and week eight, we model the scores rather than the change from baseline. Let y_{1ij} indicate MAYO score and y_{2ij} indicate pMAYO for individual i at time j.

Let A_i be the treatment given to individual i. $A_i = 0$ indicates placebo, $A_i = 1$ indicates Treatment A and $A_i = 2$ indicates Treatment B.

Also let t_{ij} be the week that is treated as a continuous variable. We let $j = 0, 2, 4, 8$ corresponding to the certain week.

7.2.2 Mixed Model

$$y_{i1j} = \beta_{10} + b_{10i} + \beta_{11} * I(A_i = 1) + \beta_{12} * I(A_i = 2) + \beta_{13} * t_{ij}$$
$$+ b_{11i} * t_{ij} + \beta_{14} * I(A_i = 1) * t_{ij} + \beta_{15} * I(A_i = 2) * t_{ij} + \epsilon_{i1j}$$
$$y_{i2j} = \beta_{20} + b_{20i} + \beta_{21} * I(A_i = 1) + \beta_{22} * I(A_i = 2) + \beta_{23} * t_{ij} + b_{21i} * t_{ij}$$
$$+ \beta_{24} * I(A_i = 1) * t_{ij} + \beta_{25} * I(A_i = 2) * t_{ij} + \epsilon_{i2j}$$

We have

$$\begin{pmatrix} b_{10i} \\ b_{11i} \\ b_{20i} \\ b_{21i} \end{pmatrix} \sim MVN\left(\mathbf{0}, \mathbf{G}\right)$$

where

$$\mathbf{G} = \begin{pmatrix} \sigma_{b_{10}}^2 & \sigma_{b_{10},b_{11}} & \sigma_{b_{10},b_{20}} & \sigma_{b_{10},b_{21}} \\ \cdot & \sigma_{b_{11}}^2 & \sigma_{b_{11},b_{20}} & \sigma_{b_{11},b_{21}} \\ \cdot & \cdot & \sigma_{b_{20}}^2 & \sigma_{b_{20},b_{21}} \\ \cdot & \cdot & \cdot & \sigma_{b_{21}}^2 \end{pmatrix}$$

Unfortunately we are only able to fit a mixed model with the same residual variance for both MAYO and pMAYO. That is,

$$\mathbf{R} = Cov\begin{pmatrix} \epsilon_{i1j} \\ \epsilon_{i2j} \end{pmatrix} = \sigma_\epsilon^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

.

This model is fit using

```
fit1_lme <- lme(y~-1+y_type+y_type:(Treatment+Week+Treatment:Week),
                ## random slope and intercept for each y_type
                random=~-1+(y_type+y_type:Week)|USUBJID,
                control=cntrl,
                na.action = "na.omit",
                data= clean_dat %>% filter(y_type_time %in%
                                   c("MAYO_0","MAYO_8",
                                     "pMAYO_0","pMAYO_2",
                                     "pMAYO_4","pMAYO_8")) %>%
                arrange(USUBJID,Week,y_type))
```

7.3 Generalized Least Squares Model

Here we have

$$
\begin{aligned}
y_{i1j} &= \beta_{10} + \beta_{11} * I(A_i = 1) + \beta_{12} * I(A_i = 2) + \beta_{13} * t_{ij} \\
&\quad + \beta_{14} * I(A_i = 1) * t_{ij} + \beta_{15} * I(A_i = 2) * t_{ij} + \epsilon_{i1j} \\
y_{i2j} &= \beta_{20} + \beta_{21} * I(A_i = 1) + \beta_{22} * I(A_i = 2) + \beta_{23} * t_{ij} \\
&\quad + \beta_{24} * I(A_i = 1) * t_{ij} + \beta_{25} * I(A_i = 2) * t_{ij} + \epsilon_{i2j}
\end{aligned}
$$

where

$$
\mathbf{V}_i = Cov \begin{pmatrix} \epsilon_{i10} \\ \epsilon_{i20} \\ \epsilon_{i12} \\ \epsilon_{i14} \\ \epsilon_{i18} \\ \epsilon_{i28} \end{pmatrix} = Cov \begin{pmatrix} y_{i10} \\ y_{i20} \\ y_{i12} \\ y_{i14} \\ y_{i18} \\ y_{i28} \end{pmatrix}
$$

is an unstructured, symmetric positive definite matrix. This model is fit using

```
fit1_gls <- gls(y~-1+y_type+y_type:(Treatment+Week+Treatment:Week),
                correlation=corSymm(form=~1|USUBJID),
                weights=varIdent(form=~1|y_type_time),
                control=cntrl,
                na.action = "na.omit",
                data= clean_dat %>% filter(y_type_time %in%
                                        c("MAYO_0","MAYO_8",
                                          "pMAYO_0","pMAYO_2",
                                          "pMAYO_4","pMAYO_8")) %>%
                arrange(USUBJID,Week,y_type))
```

7.4 Evaluating Model Output

The next step in the model building process is to examine the output of the models to see if things generally make sense and if there are any major differences between the models. In this instance, since we have to fit an overly restrictive mixed model, we see that the estimated correlation between pMAYO and MAYO at baseline is much lower in the mixed model than the generalized least squares model. This gives strong indication that the mixed model is too restrictive.

7.4.1 Fixed Effects

First we examine the fixed effects for both models. We find that the estimates and standard errors are fairly similar across the board. All hypothesis tests arrive at the same conclusion.

```
> ## estimates for fixed effects
> ## mixed model
> round(sum1_lme$tTable[,-3],3)
                                         Value Std.Error t-value p-value
y_typeMAYO                               8.856     0.092  95.980   0.000
y_typepMAYO                              5.670     0.086  66.069   0.000
y_typeMAYO:TreatmentB                    0.069     0.130   0.533   0.594
y_typepMAYO:TreatmentB                  -0.037     0.121  -0.303   0.762
y_typeMAYO:TreatmentPlacebo              0.009     0.130   0.072   0.943
y_typepMAYO:TreatmentPlacebo             0.132     0.121   1.085   0.278
y_typeMAYO:Week                         -0.399     0.019 -20.545   0.000
y_typepMAYO:Week                        -0.298     0.016 -18.305   0.000
y_typeMAYO:TreatmentB:Week              -0.048     0.027  -1.759   0.079
y_typepMAYO:TreatmentB:Week             -0.048     0.023  -2.099   0.036
y_typeMAYO:TreatmentPlacebo:Week         0.173     0.027   6.307   0.000
y_typepMAYO:TreatmentPlacebo:Week        0.127     0.023   5.515   0.000
> ## gls model
> round(sum1_gls$tTable,3)
                                         Value Std.Error t-value p-value
y_typeMAYO                               8.637     0.075 114.678   0.000
y_typepMAYO                              5.967     0.062  96.404   0.000
y_typeMAYO:TreatmentB                   -0.008     0.106  -0.077   0.939
y_typepMAYO:TreatmentB                  -0.101     0.087  -1.155   0.248
y_typeMAYO:TreatmentPlacebo              0.091     0.115   0.795   0.426
y_typepMAYO:TreatmentPlacebo             0.080     0.099   0.810   0.418
y_typeMAYO:Week                         -0.363     0.015 -24.631   0.000
y_typepMAYO:Week                        -0.295     0.012 -23.749   0.000
y_typeMAYO:TreatmentB:Week              -0.054     0.021  -2.579   0.010
y_typepMAYO:TreatmentB:Week             -0.045     0.018  -2.579   0.010
y_typeMAYO:TreatmentPlacebo:Week         0.175     0.023   7.604   0.000
y_typepMAYO:TreatmentPlacebo:Week        0.140     0.019   7.268   0.000
```

Generally we find that the covariance matrix for fixed effects is smaller for the generalized least squares model than the mixed model model. This indicates we have more precision with the generalized least squares model.

```
> ## variance of fixed effects
> norm(fit1_lme$varFix,"F")
[1] 0.05288819
> norm(fit1_gls$varBeta,"F")
[1] 0.0492176
>
> norm(fit1_lme$varFix,"2")
```

```
[1] 0.05005047
> norm(fit1_gls$varBeta,"2")
[1] 0.04780746
```

7.4.2   Random Effects

Here we can see that the random intercepts for each response are highly, highly correlated with each other. The same result is found for the random slopes.

```
> ## covariance and correlation matricies for random effects
> fit1_lme_G
Random effects variance covariance matrix
                y_typeMAYO y_typepMAYO y_typeMAYO:Week y_typepMAYO:Week
y_typeMAYO        1.720700    1.733300       -0.010290        -0.041219
y_typepMAYO       1.733300    1.752800        0.012097        -0.023922
y_typeMAYO:Week  -0.010290    0.012097        0.089283         0.070392
y_typepMAYO:Week -0.041219   -0.023922        0.070392         0.056149
  Standard Deviations: 1.3117 1.324 0.2988 0.23696
> cov2cor(fit1_lme_G)
                y_typeMAYO y_typepMAYO y_typeMAYO:Week y_typepMAYO:Week
y_typeMAYO        1.000000    0.998080       -0.026252        -0.132610
y_typepMAYO       0.998080    1.000000        0.030578        -0.076253
y_typeMAYO:Week  -0.026252    0.030578        1.000000         0.994190
y_typepMAYO:Week -0.132610   -0.076253        0.994190         1.000000
```

7.4.3   Conditional Variance

Since the only mixed model we could fit does not change variances based on response, our **R** matrix is diagonal.

```
> ## conditional covariance
> fit1_lme_R
USUBJID CNTO1275UCO3001-100001
Conditional variance covariance matrix
        1       2       3       4       5       6
1 0.99506 0.00000 0.00000 0.00000 0.00000 0.00000
2 0.00000 0.99506 0.00000 0.00000 0.00000 0.00000
3 0.00000 0.00000 0.99506 0.00000 0.00000 0.00000
4 0.00000 0.00000 0.00000 0.99506 0.00000 0.00000
5 0.00000 0.00000 0.00000 0.00000 0.99506 0.00000
6 0.00000 0.00000 0.00000 0.00000 0.00000 0.99506
  Standard Deviations: 0.99753 0.99753 0.99753 0.99753 0.99753 0.99753
```

7.4.4 Marginal Variance

Since we cannot fit a mixed model with separate conditional variances for each response, we see some differences in the marginal covariance matricies. Notably, we see much higher correlation between y_{1i0}, y_{2i0} in `gls` model (0.958) than the `lme` (0.635). This seems curious to us, so we checked the basic sample correlation between the two responses and find it to be 0.958, the same as the `gls` model. Additionally, the sample variance is 2.478 for y_{1i0} and 2.00 for y_{2i0} whereas the estimates are 2.716 for y_{1i0} and 2.748 for y_{2i0} in the mixed model. Since we have to force $Var(\epsilon_{1ij}) = Var(\epsilon_{2ij})$ in the mixed model – which seems like an unreasonable assumption – our variance and correlations seem to be thrown off. In contrast the `gls` estimates are 2.461 for y_{1i0} and 1.984 for y_{2i0}.

```
> ## marginal covariance
> fit1_lme_V
          1        2        3        4        5        6
1 2.715729 1.733343 1.650906 1.568468 1.638353 1.403594
2 1.733343 2.747906 1.705002 1.657158 1.830115 1.561470
3 1.650906 1.705002 2.876814 2.058507 2.873955 2.412012
4 1.568468 1.657158 2.058507 3.454916 3.917795 3.262554
5 1.638353 1.830115 2.873955 3.917795 8.265176 6.005474
6 1.403594 1.561470 2.412012 3.262554 6.005474 5.958697
> fit1_gls_V
          [,1]       [,2]      [,3]      [,4]      [,5]       [,6]
[1,]  2.461870 2.1170622 1.482238 1.352929 1.649183 1.2211448
[2,]  2.117062 1.9841756 1.318737 1.153780 1.243817 0.9997635
[3,]  1.482238 1.3187371 4.394730 3.781921 3.508730 2.7984554
[4,]  1.352929 1.1537803 3.781921 5.148938 4.463454 3.6107654
[5,]  1.649183 1.2438170 3.508730 4.463454 8.029243 6.0702509
[6,]  1.221145 0.9997635 2.798455 3.610765 6.070251 4.9941420
> ## marginal correlation
> cov2cor(fit1_lme_V)
          1         2         3         4         5         6
1 1.0000000 0.6345126 0.5906403 0.5120517 0.3458105 0.3489174
2 0.6345126 1.0000000 0.6064125 0.5378292 0.3840178 0.3858843
3 0.5906403 0.6064125 1.0000000 0.6529468 0.5893835 0.5825697
4 0.5120517 0.5378292 0.6529468 1.0000000 0.7331569 0.7190570
5 0.3458105 0.3840178 0.5893835 0.7331569 1.0000000 0.8557475
6 0.3489174 0.3858843 0.5825697 0.7190570 0.8557475 1.0000000
> cov2cor(fit1_gls_V)
          [,1]       [,2]      [,3]      [,4]      [,5]       [,6]
[1,] 1.0000000 0.9578797 0.4506291 0.3800001 0.3709361 0.3482604
[2,] 0.9578797 1.0000000 0.4465829 0.3609726 0.3116227 0.3175973
[3,] 0.4506291 0.4465829 1.0000000 0.7950374 0.5906721 0.5973403
```

```
[4,] 0.3800001 0.3609726 0.7950374 1.0000000 0.6941845 0.7120492
[5,] 0.3709361 0.3116227 0.5906721 0.6941845 1.0000000 0.9586032
[6,] 0.3482604 0.3175973 0.5973403 0.7120492 0.9586032 1.0000000
```

7.5   Evaluating Information Criteria

Although the model output makes it pretty clear the generalized least squares model is better, we will look at the information criteria just to confirm. If we had many models to compare, it would be reasonable to go ahead and throw out the mixed model.

Since we are restricted in the specification of **R** in the mixed model, the generalized least squares model outperforms the mixed model pretty considerably. Below is the output looking at information criteria and a likelihood ratio test.

```
> anova(fit1_lme,fit1_gls)
         Model df      AIC      BIC    logLik   Test  L.Ratio p-value
fit1_lme     1 23 20528.25 20681.38 -10241.13
fit1_gls     2 46 17616.00 17922.25  -8762.00 1 vs 2 2958.253  <.0001
```

7.6   Answering Research Questions

Since we only have one reasonable model to use, we will conduct inference using it. We are interested in the correlation between pMAYO at earlier time points and MAYO at later time points and also inference on the treatment effects.

7.6.1   Correlation

```
> fit1_gls_cor
            MAYO_0   pMAYO_0   pMAYO_2   pMAYO_4    MAYO_8   pMAYO_8
MAYO_0   1.0000000 0.9578797 0.4506291 0.3800001 0.3709361 0.3482604
pMAYO_0  0.9578797 1.0000000 0.4465829 0.3609726 0.3116227 0.3175973
pMAYO_2  0.4506291 0.4465829 1.0000000 0.7950374 0.5906721 0.5973403
pMAYO_4  0.3800001 0.3609726 0.7950374 1.0000000 0.6941845 0.7120492
MAYO_8   0.3709361 0.3116227 0.5906721 0.6941845 1.0000000 0.9586032
pMAYO_8  0.3482604 0.3175973 0.5973403 0.7120492 0.9586032 1.0000000
```

Here we can see that $\hat{Cor}(y_{1i2}, y_{2i8}) = 0.591, \hat{Cor}(y_{1i4}, y_{2i8}) = 0.694$, so there is moderate positive correlation between pMAYO at earlier time points and the ultimate MAYO score at the end of the trial.

7.6.2   Treatment Effects

We are ultimately interested the difference in change from baseline in MAYO scores for treated and not treated individuals. Since pMAYO scores at week four and MAYO scores at

week eight at correlation, we also want to look at similar tests for pMAYO at baseline and week four.

For both MAYO and pMAYO scores, we test these hypotheses with contrasts. Our test statistic is

$$Z_{obs} = \frac{\mathbf{c}^T \hat{\theta}}{\left(\mathbf{c}^T \hat{\Sigma} \mathbf{c}\right)^{1/2}}$$

where $\hat{\Sigma}$ is the estimated covariance matrix for $\hat{\theta}$. We use the standard normal distribution to calculate p-values.

For MAYO scores, the first test looks at the difference in change from baseline at week eight for Treatment A versus placebo. The second test compares Treatment B. The third test compares the two treatments. Our formal hypotheses are:

(Test One) $H_0 : E(y_{1i8} - y_{1i0}|A_i = 1) - E(y_{1i8} - y_{1i0}|A_i = 1) = 0$
$H_A : E(y_{1i8} - y_{1i0}|A_i = 1) - E(y_{1i8} - y_{1i0}|A_i = 0) < 0$

(Test Two) $H_0 : E(y_{1i8} - y_{1i0}|A_i = 2) - E(y_{1i8} - y_{1i0}|A_i = 0) = 0$
$H_A : E(y_{1i8} - y_{1i0}|A_i = 2) - E(y_{1i8} - y_{1i0}|A_i = 0) < 0$

(Test Three) $H_0 : E(y_{1i8} - y_{1i0}|A_i = 2) - E(y_{1i8} - y_{1i0}|A_i = 1) = 0$
$H_A : E(y_{1i8} - y_{1i0}|A_i = 2) - E(y_{1i8} - y_{1i0}|A_i = 1) \neq 0$

Hypothesis Test Information on MAYO Scores

Test	Effect Size	Standard Error	p-value
One	-1.21	0.19	< 0.01
Two	-1.64	0.20	< 0.01
Three	0.43	0.19	0.02

Here we can see that both treatments are clearly effective when compared to placebo. Additionally, Treatment B looks to be more effective than Treatment A.

Next, we want to look at similar test comparing pMAYO at week four and baseline. Our

formal hypotheses are:

$$\text{(Test One) } H_0 : E(y_{2i4} - y_{2i0}|A_i = 1) - E(y_{2i4} - y_{2i0}|A_i = 1) = 0$$
$$H_A : E(y_{2i4} - y_{2i0}|A_i = 1) - E(y_{2i4} - y_{2i0}|A_i = 0) < 0$$

$$\text{(Test Two) } H_0 : E(y_{2i4} - y_{2i0}|A_i = 2) - E(y_{2i4} - y_{2i0}|A_i = 0) = 0$$
$$H_A : E(y_{2i4} - y_{2i0}|A_i = 2) - E(y_{2i4} - y_{2i0}|A_i = 0) < 0$$

$$\text{(Test Three) } H_0 : E(y_{2i4} - y_{2i0}|A_i = 2) - E(y_{2i4} - y_{2i0}|A_i = 1) = 0$$
$$H_A : E(y_{2i4} - y_{2i0}|A_i = 2) - E(y_{2i4} - y_{2i0}|A_i = 1) \neq 0$$

Hypothesis Test Information on pMAYO Scores

Test	Effect Size	Standard Error	p-value
One	-0.48	0.08	< 0.01
Two	-0.69	0.08	< 0.01
Three	0.19	0.08	0.02

Using pMAYO at week four, we gather the same results as using the MAYO score at week eight. In this instance, choosing go/no-go based on week four pMAYO information would have resulted in the correct decision.