Hierarchical Models - random variance

02424 - Advanced Dataanalysis and Statistical Modelling
Assingment 3, Part 2

Tymoteusz Barcinski - s221937
Soren Skjernaa - s223316

Technical University of Denmark

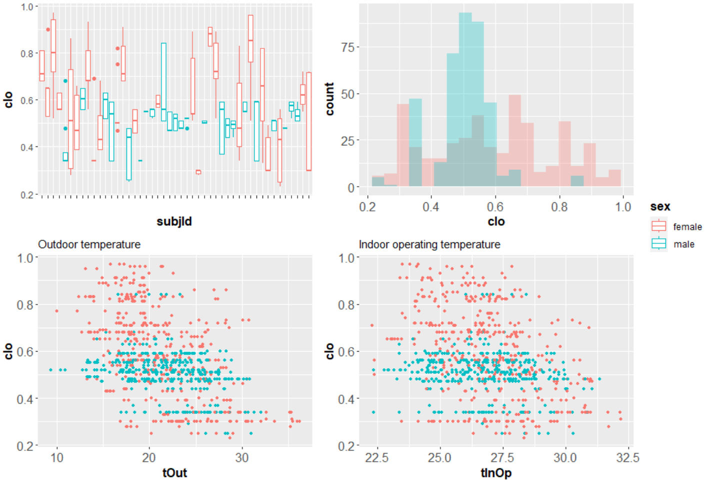
May 14, 2023

Outline

Overview of the experiment

Variable	Type	Explanation
clo	Clothing insulation level	Positive variable, with higher values implying higher insulation.
tOut	Outdoors air temperature	Measured in C^o
tInOp	Indoor operating temperature	Measured in C^o
sex	Sex	Female/male.
subjId	Subject ID	Uniqiue ID for each subject.
time	Time	Time difference since last observation for the subject (continues variable, but unit not given).
day	Day	Number of experimentation day for the subject.
subDay	Subject $\times$ Day ID	Unique ID for each combination of subject and day.

Overview of the experiment



First model - Formulation

$$Y_{i,j,k} = \log(\text{clo}_{i,j,k}) = \mu + \beta\,(\text{sex}_i) + u_i + \varepsilon_{ijk};$$
$$u_i \sim N\left(0, \sigma_u^2\right) \quad \varepsilon_{ijk} \sim \left(0, \sigma^2\right)$$

where $Y_{i,j,k}$ is the logarithm of the clothing insulation level for subject i on day j and k refer to the observation number within the day. We have that

$$E\left[Y_{ijk}\right] = \mu + \beta(\text{sex}_i);$$
$$\text{Cov}\left[Y_{ijk}, Y_{hlm}\right] = \begin{cases} \sigma_u^2 + \sigma^2 & \text{for } (i,j) = (h,l) \quad \text{(subject, obs)} \\ \sigma_u^2 & \text{for } i = h, \quad j \neq l \quad \text{(subject)} \\ 0 & \text{for } i \neq h \end{cases}$$

First model - Formulation

We can further write the model in the following form for subject i:

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{U} + \varepsilon_i; \quad \mathbf{U} \sim N(0, \sigma_u^2), \quad \varepsilon_i \sim N_{n_i}(0, \sigma^2 I)$$

We can write the model as a multivariate normal distribution

$$\mathbf{Y}_i \sim N_{n_i}\left(\mathbf{X}_i\boldsymbol{\beta}, \sigma_u^2\mathbf{Z}\mathbf{Z}^T + \sigma^2 I\right)$$

The model is parametrized by $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma_u^2, \sigma^2)$. We can write the log likelihood for the model

$$l(\boldsymbol{\theta}; \boldsymbol{y}) = \log L(\boldsymbol{\theta}; \boldsymbol{y}) = \sum_{i=1}^{n} l_i(\boldsymbol{\theta}; \boldsymbol{y}_i)$$

where n is the number of subjects. We implement the likelihood function and perform the optimisation with respect to the parameters.

First model - Likelihood implementation

The detailed implementation in R is presented:

```
n = dim(df)[1]
X = cbind(rep(1, n), df$sex_optimization)
p = dim(X)[2]
Z = cbind(rep(1, n))
u_number = dim(Z)[2]
y = log(clo)

obj_1 = function(beta){
  result = 0
  for(subject_i in unique(df$subjId)){
    X_i = X[df$subjId == subject_i, , drop = F]
    n_i = dim(X_i)[1]
    y_i = y[df$subjId == subject_i]
    y_hat_i = X_i %*% beta[1:p]
    Psi_i = exp(beta[p+1])*diag(u_number)
    Z_i = Z[df$subjId == subject_i, , drop = F]
    Sigma_full_i = exp(beta[p+2])*diag(n_i) + Z_i %*%
    Psi_i %*% t(Z_i)
    result = result +
      sum(dmvnorm(y_i, mean = y_hat_i,
      sigma = Sigma_full_i, log = TRUE))
  }
  return(-result)
```

First model - Results

We verify our results with the model estimated by the library lme4.

	own implementation		lme4	
Parameter	Estimate	Std. Error	Estimate	Std. Error
(Intercept)	-0.584	0.047	-0.584	0.047
sexmale	-0.109	0.067	-0.109	0.067
σ_u^2	0.050	-	0.050	0.224
σ^2	0.035	-	0.035	0.186

Table: Model parameters and their uncertainties for the final model.

Since the variance parameters were estimated in the log domain constructing the Wald confidence interval for them would require taking the transformation into account as the Wald confidence intervals are not invariant to parameter transformations.

Second model - Formulation

$$Y_{i,j,k} = \log(\text{clo}_{i,j,k}) = \mu + \beta\,(\text{sex}_i) + u_i + v_{ij} + \varepsilon_{ijk};$$

$$u_i \sim N\left(0, \sigma_u^2\right) \quad v_{ij} \sim N\left(0, \sigma_v^2\right) \quad \varepsilon_{ijk} \sim \left(0, \sigma^2\right)$$

We have that

$$\mathrm{E}\left[Y_{ijk}\right] = \mu + \beta(\text{sex}_i);$$

$$\text{Cov}\left[Y_{ijk}, Y_{hlm}\right] = \begin{cases} \sigma_u^2 + \sigma_v^2 + \sigma^2 & \text{for } (i,j,k) = (h,l,m) \\ \sigma_u^2 + \sigma_v^2 & \text{for } (i,j) = (h,l), \quad k \neq m \\ \sigma_u^2 & \text{for } i = h, \quad j \neq l \\ 0 & \text{for } i \neq h \end{cases}$$

Second model - Formulation

We can further write the model in the following form for subject i and day j:

$$\mathbf{Y}_i = \mathbf{X}_i\beta + \mathbf{Z}_{1,i}\mathbf{U} + \mathbf{Z}_{2,i}\mathbf{V} + \varepsilon_i;$$
$$\mathbf{U} \sim N(0, \sigma_u^2), \quad \mathbf{V} \sim N_{days_i}(0, \sigma_v^2 I), \quad \varepsilon_i \sim N_{n_i}(0, \sigma^2 I)$$

Therefore we can write the model as a multivariate normal distribution

$$\mathbf{Y}_i \sim N_{n_i}\left(\mathbf{X}_i\beta, \sigma_u^2\mathbf{Z_{1,i}}\mathbf{Z_{1,i}}^T + \sigma_v^2\mathbf{Z_{2,i}}\mathbf{Z_{2,i}}^T + \sigma^2 I\right)$$

The model is parametrized by $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma_u^2, \sigma_v^2, \sigma^2)$.

Second model - Results

We verify our results with the model estimated by the library lme4. The comment from the section about model 1 about the Wald confidence intervals for variance parameters still applies.

	own implementation		lme4	
Parameter	Estimate	Std. Error	Estimate	Std. Error
(Intercept)	-0.583	0.047	-0.583	0.047
sexmale	-0.111	0.067	-0.111	0.067
σ_u^2	0.039	-	0.039	0.197
σ_v^2	0.038	-	0.038	0.195
σ^2	0.0079	-	0.0079	0.089

Table: Model parameters and their uncertainties

Third model - Formulation

$$
\begin{aligned}
Y_{i,j,k} = \log(\mathrm{clo}_{i,j,k}) &= \mu + \beta\left(\mathrm{sex}_i\right) + u_i + v_{ij} + \varepsilon_{ijk};\\
u_i &\sim N\left(0, \sigma_u^2 \alpha(\mathrm{sex}_i)\right) \quad v_{ij} \sim N\left(0, \sigma_v^2(\alpha\,\mathrm{sex}_i)\right)\\
\varepsilon_{ijk} &\sim N\left(0, \sigma^2 \alpha(\mathrm{sex}_i)\right)
\end{aligned}
$$

We have that

$$
\begin{aligned}
\mathrm{E}\left[Y_{ijk}\right] &= \mu + \beta(\mathrm{sex}_i);\\
\mathrm{Cov}\left[Y_{ijk}, Y_{hlm}\right] &= \begin{cases}
(\sigma_u^2 + \sigma_v^2 + \sigma^2)\alpha(\mathrm{sex}_i) & \text{for } (i,j,k) = (h,l,m)\\
(\sigma_u^2 + \sigma_v^2)\alpha(\mathrm{sex}_i) & \text{for } (i,j) = (h,l), \quad k \neq m\\
\sigma_u^2 \alpha(\mathrm{sex}_i) & \text{for } i = h, \quad j \neq l\\
0 & \text{for } i \neq h
\end{cases}
\end{aligned}
$$

Third model - Formulation

We can further write the model in the following form:

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_{1,i}\mathbf{U} + \mathbf{Z}_{2,i}\mathbf{V} + \varepsilon_i;$$
$$\mathbf{U} \sim N(0, \sigma_u^2\alpha(\text{sex}_i)), \quad \mathbf{V} \sim N_{days_i}(0, \sigma_v^2\alpha(\text{sex}_i)I),$$
$$\varepsilon_i \sim N_{n_i}(0, \sigma^2\alpha(\text{sex}_i)I)$$

Therefore we can write the model as a multivariate normal distribution

$$\mathbf{Y}_i \sim N_{n_i}\left(\mathbf{X}_i\boldsymbol{\beta}, \left(\sigma_u^2\mathbf{Z_{1,i}}\mathbf{Z_{1,i}}^T + \sigma_v^2\mathbf{Z_{2,i}}\mathbf{Z_{2,i}}^T + \sigma^2 I\right)\alpha(\text{sex}_i)\right)$$

The model is parametrized by $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma_u^2, \sigma_v^2, \sigma^2, \alpha)$.
Here α gives the weight for the variance for males, i.e. if $\tilde{\sigma^2}_{female}$ is the estimated variance parameter for females then the corresponding variance parameter for males is given by $\tilde{\sigma^2}_{male} = \alpha\tilde{\sigma^2}_{female}$.

Third model - Results

Par.	Estimate log domain	Std. Error log domain	Estimate	Std. Error
μ	-	-	-0.5833	0.0797
β	-	-	-0.1117	0.0830
σ_u^2	-2.4366	0.3886	0.0874	-
σ_v^2	-1.680	0.1775	0.186	-
σ^2	-4.365	0.0696	0.0127	-
α	-2.499	0.1049	0.0821	-

Table: Model parameters and their uncertainties

Theorem formulation

Given

$$Y_i \mid \gamma_i \sim N\left(\mu, \frac{\sigma^2}{\gamma_i}\right); \quad \gamma_i \sim \text{Gamma}(1, \lambda); \quad E[\gamma_i] = 1 \quad Var[\gamma_i] = \frac{1}{\lambda}$$

We want to show that the marginal distribution of Y_i is

$$f_{Y_i} \sim \frac{1}{\sigma} f_0\left(\frac{y-\mu}{\sigma}; 2\lambda\right)$$

where f_0 is the pdf of a student t-distributed random variable with 2λ degrees of freedom.

Gamma distribution

Canonical parametrization of the Gamma distribution is

$$Y \sim \mathrm{G}(\alpha, \beta); \quad E[Y] = \alpha\beta \quad Var[Y] = \alpha\beta^2$$

Reparametrizion in a way such that the expected value is the mean value parameter μ_{G} and the variation is characterized by the precision parameter λ as it is done on page 96.

$$\alpha = \lambda \quad \beta = \frac{\mu_{\mathrm{G}}}{\lambda}$$

We obtain the following density function after reparametrization.

$$f_{\gamma_i}(\gamma_i, \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \gamma_i^{\alpha-1} \exp\left(-\frac{\gamma_i}{\beta}\right) = \frac{\lambda^\lambda}{\mu_{\mathrm{G}}^\lambda \Gamma(\lambda)} \gamma_i^{\lambda-1} \exp\left(-\frac{\gamma_i \lambda}{\mu_{\mathrm{G}}}\right)$$

In the Theorem we have $\mu_{\mathrm{G}} = 1$, hence we obtain

$$f_{\gamma_i}(\gamma_i, \lambda) = f_{\gamma_i}(\gamma_i, \mu_{\mathrm{G}} = 1, \lambda) = \frac{\lambda^\lambda}{\Gamma(\lambda)} \gamma_i^{\lambda-1} \exp\left(-\gamma_i \lambda\right)$$

Normal and t-student distributions

The conditional density of $Y_i|\gamma_i$ is the normal distribution

$$f_{Y|\gamma_i}(y;\gamma_i) = \frac{\sqrt{\gamma_i}}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\gamma_i}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right)$$

The density of the student t distribution with 2λ degrees of freedom is

$$f_0(y) = \frac{\Gamma\left(\frac{2\lambda+1}{2}\right)}{\sqrt{2\lambda\pi}\Gamma\left(\lambda\right)}\left(1+\frac{y^2}{2\lambda}\right)^{-(2\lambda+1)/2}$$

Derivation

Consider

$$
\begin{aligned}
f_{Y_i}(y) &= \int_0^\infty f_{Y_i|\gamma_i}(y_i;\gamma_i) f_{\gamma_i}(\gamma_i,\lambda)\, d\gamma_i \\
&= \int_0^\infty \frac{\sqrt{\gamma_i}}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\gamma_i}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right) \frac{\lambda^\lambda}{\Gamma(\lambda)} \gamma_i^{\lambda-1} \exp\left(-\gamma_i\lambda\right) d\gamma_i \\
&= \frac{\lambda^\lambda}{\sigma\sqrt{2\pi}\Gamma(\lambda)} \int_0^\infty \gamma_i^{(\lambda+\frac{1}{2})-1} \exp\left(-\gamma_i\left(\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2 + \lambda\right)\right) d\gamma_i
\end{aligned}
$$

The integrand is seen as the kernel of a Gamma distribution: $\mathrm{G}\left(\alpha = \lambda + \frac{1}{2}, \beta = 1/\left(\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2 + \lambda\right)\right)$. Therefore we need to adjust the integrating constant to obtain the true distribution which will integrate to 1.

Derivation

After doing so we obtain the following:

$$f_{Y_i}(y) = \frac{1}{\sqrt{2\pi}\sigma}\frac{\lambda^\lambda}{\Gamma(\lambda)}\frac{\Gamma\left(\lambda+\frac{1}{2}\right)}{\left(\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2+\lambda\right)^{\lambda+\frac{1}{2}}}$$

Thus substituting $t = \frac{y-\mu}{\sigma}$ and $v = 2\lambda$ we get

$$\begin{aligned}
f_{Y_i}(y) &= \frac{1}{\sigma}\frac{\sigma\left(\frac{v+1}{2}\right)}{\sqrt{2\pi}\Gamma(v/2)}\lambda^{-\frac{1}{2}}\lambda^{\lambda+\frac{1}{2}}\left(\frac{t^2+v}{2}\right)^{-\frac{v+1}{2}} \\
&= \frac{1}{\sigma}\frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\Gamma(v/2)}\left(1+\frac{t^2}{v}\right)^{-\frac{v+1}{2}} \\
&= \frac{1}{\sigma}f_0\left(\frac{y-\mu}{\sigma};2\lambda\right)
\end{aligned}$$

which was to be proven.

Laplace Approximation in 4 steps

1) The likelihood of the General Mixed Effect model is

$$L_M(\boldsymbol{\theta}, \boldsymbol{y}) = \int_{\mathbb{R}^q} L(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{y}) d\boldsymbol{u}$$

2) Quadratic approximation of the log-likelihood is

$$\ell(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{y}) \approx \ell\left(\boldsymbol{\theta}, \hat{\boldsymbol{u}}_{\boldsymbol{\theta}}, \boldsymbol{y}\right) - \frac{1}{2}\left(\boldsymbol{u} - \hat{\boldsymbol{u}}_{\boldsymbol{\theta}}\right)^T \left(- \ell''_{uu}(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{y})\big|_{\boldsymbol{u}=\hat{\boldsymbol{u}}_{\boldsymbol{\theta}}}\right)\left(\boldsymbol{u} - \hat{\boldsymbol{u}}_{\boldsymbol{\theta}}\right)$$

3) Inner optimization is

$$\hat{\boldsymbol{u}}_{\boldsymbol{\theta}} = \underset{\boldsymbol{u}}{\operatorname{argmax}} L(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{y})$$

4) After simplifications the Laplace approximation emerges

$$\ell_{M,LA}(\boldsymbol{\theta}, \boldsymbol{y}) \approx \ell\left(\boldsymbol{\theta}, \hat{\boldsymbol{u}}_{\boldsymbol{\theta}}, \boldsymbol{y}\right) - \frac{1}{2}\log\left(\det\left(- \ell''_{uu}(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{y})\big|_{\boldsymbol{u}=\hat{\boldsymbol{u}}_{\boldsymbol{\theta}}}\right)\right) + \frac{q}{2}\log(2\pi)$$

Hierarchical model - Formulation

$$
\begin{aligned}
Y_{i,j,k}|u_i, v_{ij}, \gamma_i &\sim N\left(\mu + \beta\,(\text{sex}_i) + u_i + v_{ij},\, \sigma^2\alpha(\text{sex}_i)/\gamma_i\right) \\
u_i|\gamma_i &\sim N\left(0, \sigma_u^2\alpha(\text{sex}_i)/\gamma_i\right) \\
v_{ij}|\gamma_i &\sim N\left(0, \sigma_v^2\alpha(\text{sex}_i)/\gamma_i\right) \\
\gamma_i &\sim G\left(1, \lambda\right)
\end{aligned}
$$

We have that

$$
\mathrm{E}\left[Y_{ijk}|\gamma_i\right] = \mu + \beta(\text{sex}_i);
$$

$$
\text{Cov}\left[Y_{ijk}|\gamma_i, Y_{hlm}|\gamma_h\right] = \begin{cases} (\sigma_u^2 + \sigma_v^2 + \sigma^2)\alpha(\text{sex}_i)\gamma_i^{-1} & (i,j,k) = (h,l,m) \\ (\sigma_u^2 + \sigma_v^2)\alpha(\text{sex}_i)\gamma_i^{-1} & (i,j) = (h,l) \\ \sigma_u^2\alpha(\text{sex}_i)\gamma_i^{-1} & i = h, \quad j \neq l \\ 0 & i \neq h \end{cases}
$$

Hierarchical model - Formulation

We can further write the model in the following form:

$$\begin{aligned}
\mathbf{Y}_i &= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_{1,i}\mathbf{U} + \mathbf{Z}_{2,i}\mathbf{V} + \varepsilon_i; \\
\mathbf{U}|\gamma_i &\sim N(0, \sigma_u^2\alpha(\mathrm{sex}_i)\gamma_i^{-1}), \\
\mathbf{V}|\gamma_i &\sim N_{days_i}(0, \sigma_v^2\alpha(\mathrm{sex}_i)\gamma_i^{-1}I), \\
\varepsilon_i|\gamma_i &\sim N_{n_i}(0, \sigma^2\alpha(\mathrm{sex}_i)\gamma_i^{-1}I)
\end{aligned}$$

Therefore we can write the model as a multivariate normal distribution

$$\mathbf{Y}_i|\gamma_i \sim N_{n_i}\left(\mathbf{X}_i\boldsymbol{\beta}, \left(\sigma_u^2\mathbf{Z_{1,i}}\mathbf{Z_{1,i}}^T + \sigma_v^2\mathbf{Z_{2,i}}\mathbf{Z_{2,i}}^T + \sigma^2 I\right)\alpha(\mathrm{sex}_i)\gamma_i^{-1}\right)$$

$$\gamma_i \sim G(1, \lambda)$$

Hierarchical model - Formulation

Theorem 6.7 generalizes to the multivariate setting as changing the proof to account for the dispersion matrix Σ in the multivariate setting is straightforward.

The extended Theorem 6.7 tells us that the marginal distribution for Y_i is

$$Y_i \sim t_{n_i}\left(\mathbf{X}_i\beta, \left(\sigma_u^2 \mathbf{Z_{1,i}}\mathbf{Z_{1,i}}^T + \sigma_v^2 \mathbf{Z_{2,i}}\mathbf{Z_{2,i}}^T + \sigma^2 I\right)\alpha(\text{sex}_i), 2\lambda\right)$$

where t_{n_i} is the multivariate t-distribution with mean, scale and degrees of freedom as specified.

We write the likelihood and parametrize the model by

$$\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma_u^2, \sigma_v^2, \sigma^2, \alpha, \lambda)$$

Hierarchical model - C++ TMB file

We estimate the model with TMB because it estimates the random effects. The likelihood function in C++ is presented:

```
for(int i=0; i < nsubjects; i++){
    f -= dnorm(u[i], mean_random_subject,
               sqrt(exp(sigma2_u_log)*exp(alpha*sex[index])/
               gamma[i]), true);
}
for(int j=0; j < ndays; j++){
    i = subjectId_day_factor_gamma[j];
    f -= dnorm(v[j], mean_random_day,
               sqrt(exp(sigma2_v_log)*exp(alpha*sex[index])/
               gamma[i]), true);
}
for(int i=0; i < nsubjects; i++){
    f -= dgamma(gamma[i], exp(lambda),(1/exp(lambda)),true);
}
for(int index = 0; index < nobs; index++){
i = subjectId_factor[index];
j = subjectId_day_factor[index];
f -= dnorm(y[index], (beta[0] + beta[1]*sex[index]
           + u[i] + v[j]),
          sqrt(exp(sigma2_log) *exp(alpha * sex[index]) /
          gamma[i]), true);
}
```

Hierarchical model - Results

The parameter estimates for the TMB are slightly different than for the regular optimization which is explained by the fact that TMB uses the Laplace approximation which introduces some level of approximation.

	marginal t-student		TMB			
	log	regular	log	std.error log	regular	std.error
μ	-	-0.475	-	-	-0.484	0.043
β	-	-0.199	-	-	-0.189	0.053
σ_u^2	-3.846	0.0214	-4.035	0.409	0.018	-
σ_v^2	-3.845	0.0214	-4.246	0.304	0.014	-
σ^2	-8.8047	0.00015	-8.741	0.333	0.00015	-
α	-0.7178	0.4878	-0.762	0.335	0.467	-
λ	-0.984	0.3738	-0.959	0.172	0.383	-

Table: Model parameters and their uncertainties

Final Normal model - Formulation

$$
\begin{aligned}
Y_{i,j,k}|u_i, v_{ij}, \gamma_i &\sim N\left(\mu + \beta\left(\text{sex}_i\right) + u_i + v_{ij}, \sigma^2\alpha\left(\text{sex}_i\right)e^{-\gamma_i}\right) \\
u_i \mid \gamma_i &\sim N\left(0, \sigma_u^2\alpha\left(\text{sex}_i\right)e^{-\gamma_i}\right) \\
v_{ij} \mid \gamma_i &\sim N\left(0, \sigma_v^2\alpha\left(\text{sex}_i\right)e^{-\gamma_i}\right) \\
\gamma_i &\sim N\left(0, \sigma_G^2\right)
\end{aligned}
$$

We have that

$$
\mathrm{E}\left[Y_{ijk}|\gamma_i\right] = \mu + \beta(\text{sex}_i); \tag{1}
$$

$$
\text{Cov}\left[Y_{ijk}|\gamma_i, Y_{hlm}|\gamma_h\right] = \begin{cases} (\sigma_u^2 + \sigma_v^2 + \sigma^2)\alpha(\text{sex}_i)e^{-\gamma_i} & (i,j,k) = (h,l,m) \\ (\sigma_u^2 + \sigma_v^2)\alpha(\text{sex}_i)e^{-\gamma_i} & (i,j) = (h,l), \\ \sigma_u^2\alpha(\text{sex}_i)e^{-\gamma_i} & i = h, \quad j \neq l \\ 0 & i \neq h \end{cases} \tag{2}
$$

Final Normal model - Formulation

We can further write the model in the following form:

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_{1,i}\mathbf{U} + \mathbf{Z}_{2,i}\mathbf{V} + \varepsilon_i;$$
$$\mathbf{U}|\gamma_i \sim N(0, \sigma_u^2\alpha(\text{sex}_i)e^{-\gamma_i}), \quad \mathbf{V}|\gamma_i \sim N_{days_i}(0, \sigma_v^2\alpha(\text{sex}_i)e^{-\gamma_i}I),$$
$$\varepsilon_i|\gamma_i \sim N_{n_i}(0, \sigma^2\alpha(\text{sex}_i)e^{-\gamma_i}I)$$

Therefore we can write the model as a multivariate normal distribution

$$\mathbf{Y}_i|\gamma_i \sim N_{n_i}\left(\mathbf{X}_i\boldsymbol{\beta}, \left(\sigma_u^2\mathbf{Z_{1,i}}\mathbf{Z_{1,i}}^T + \sigma_v^2\mathbf{Z_{2,i}}\mathbf{Z_{2,i}}^T + \sigma^2 I\right)\alpha(\text{sex}_i)e^{-\gamma_i}\right)$$
$$\gamma_i \sim N(0, \sigma_G^2)$$

The model is parametrized by $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma_u^2, \sigma_v^2, \sigma^2, \alpha, \sigma_G^2)$.

Final Normal model - Likelihood implementation

The likelihood function in C++ is presented:

```
for(int i=0; i < nsubjects; i++){
    f -= dnorm(u[i], mean_random_subject,
               sqrt(exp(sigma2_u_log)*exp(alpha*sex[index])*
               exp(-gamma[i])), true);
}
for(int j=0; j < ndays; j++){
    i = subjectId_day_factor_gamma[j];
    f -= dnorm(v[j], mean_random_day,
               sqrt(exp(sigma2_v_log)*exp(alpha*sex[index])*
               exp(-gamma[i])), true);
}
for(int i=0; i < nsubjects; i++){
    f -= dnorm(gamma[i], mean_random_gamma,
    sqrt(exp(sigma2_G_log)), true);
}
for(int index = 0; index < nobs; index++){
    i = subjectId_factor[index];
    j = subjectId_day_factor[index];
    f -= dnorm(y[index], (beta[0] + beta[1]*sex[index] +
                    u[i] + v[j]),
               sqrt(exp(sigma2_log)*exp(alpha*sex[index])*
               exp(-gamma[i])), true);
}
```

Final Normal model - Results

Par.	log	Std. Error log	regular	Std. Error
μ	-	-	-0.4906	0.0407
β	-	-	-0.1810	0.0533
α	-1.1344	0.3545	0.3216	-
σ_u^2	-2.2760	0.4717	0.1027	-
σ_v^2	-2.5828	0.3761	0.0756	-
σ_G^2	1.5996	0.2225	4.9512	-
σ^2	-6.8391	0.3755	0.0011	-

Table: Model parameters and their uncertainties

Comparisons - Likelihoods, AICs

We see that the models with random effect in the dispersion for distribution of the dependent variable are significantly better than other models.

Model	log-likelihood	n params	AIC
First	134.7031	4	-261.4063
Second	541.9209	5	-1073.842
Third	769.2759	6	-1526.552
Hierarchical marginal	1216.219	7	-2418.438
Hierarchical TMB	1212.806	7	-2411.612
Final Normal	1204.501	7	-2395.003

Table: Comparison of likelihood and AIC between estimated models

Comparisons - Variances

Hierarchical TMB model, Final normal model, Third TMB model

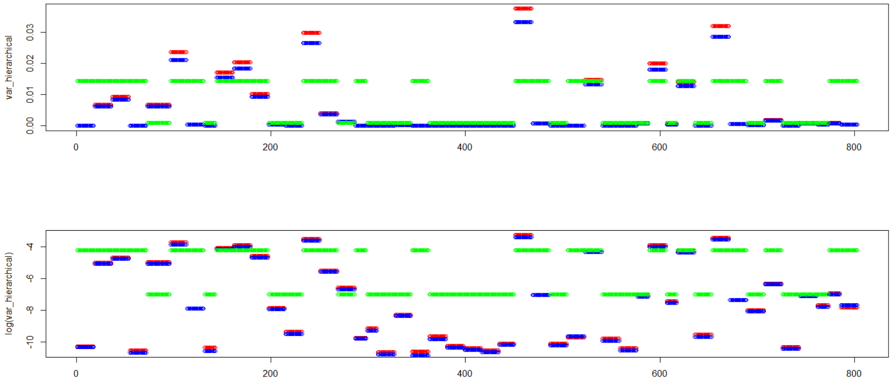


Figure: Plots of variance and log variance

Comparisons - Residuals

Hierarchical TMB model, Final normal model, Third TMB model

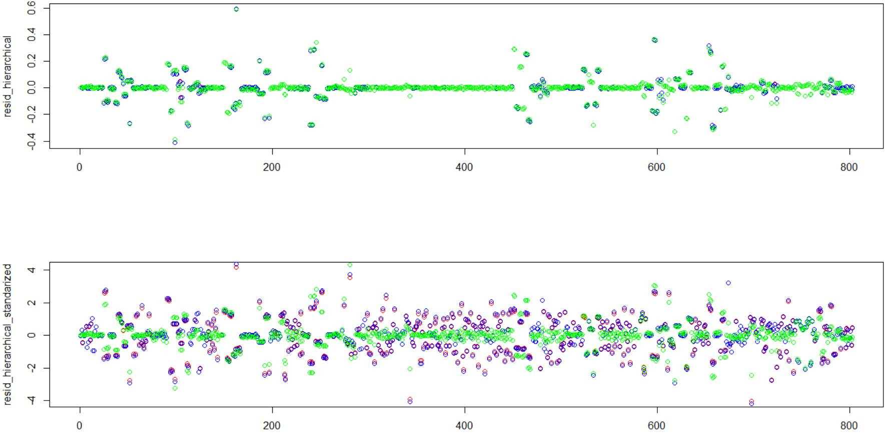


Figure: Plots of residuals and standardized residuals

Comparisons - QQplots

The following QQ plots indicate that the standardised residuals for models with random effects in the variance regression do not have heavy tails.

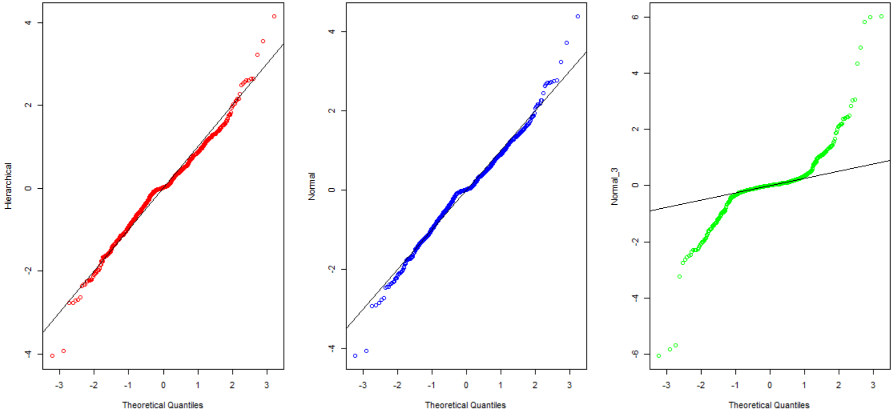
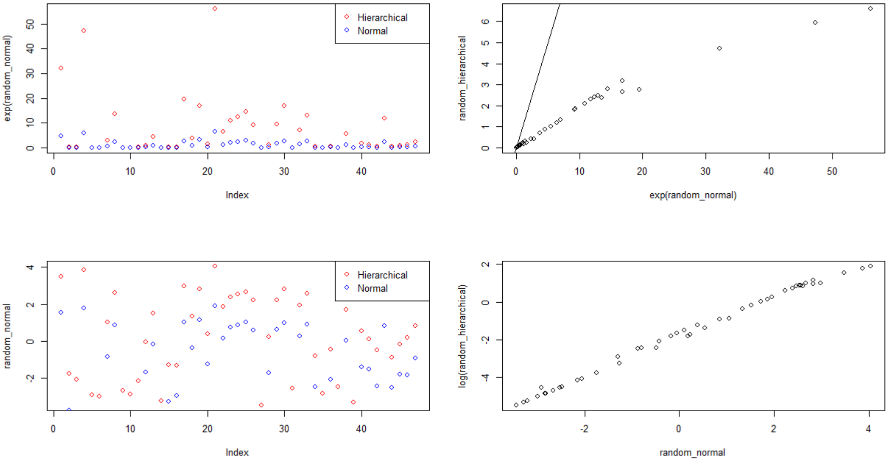


Figure: QQplots of standardized residuals for different models

Comparisons - $\gamma_i | Y_i$

Based on the plot in the bottom right corner we see that in the linear domain, the random effects have a clear linear relationship indicating that the model's hierarchical and final normal are not that different.



Comparisons - Parameters

We have $a \in \{\lambda, \sigma_G^2\}$. We note that $\gamma_i \sim \mathrm{G}(1, \lambda)$; $\quad Var(\gamma_i) = \frac{1}{\lambda} = 2.611$.

	Hierarchical TMB				Final normal TMB			
	log	sd log	regular	sd	log	sd log	regular	sd
μ	-	-	-0.484	0.04	-	-	-0.490	0.04
β	-	-	-0.189	0.05	-	-	-0.181	0.05
α	-0.762	0.335	0.467	-	-1.134	0.355	0.322	-
σ_u^2	-4.035	0.409	0.018	-	-2.276	0.472	0.103	-
σ_v^2	-4.246	0.304	0.014	-	-2.583	0.376	0.076	-
a	-0.959	0.172	0.383	-	1.5996	0.223	4.951	-
σ^2	-8.741	0.333	**0.00015**	-	-6.839	0.376	**0.0011**	-

Table: Comparison of parameters between models

Derivation of $\gamma_i | Y_i$ in hierarchical model

Consider

$$f_{\gamma_i|Y_i=y_i} = \frac{f_{Y_i=y_i|\gamma_i} f_{\gamma_i}}{f_{Y_i=y_i}} = \frac{\text{(Multivariate normal)} \cdot \text{(gamma)}}{\text{multivarite t-student}}$$

$$Y_i|\gamma_i \sim N\left(\mu, \frac{\sigma^2}{\gamma_i} I\right); \quad \gamma_i \sim G(1, \lambda)$$

Consider

$$f_{Y_i=y_i|\gamma_i} f_{\gamma_i} = \exp\left(-\frac{1}{2}(y_i - \mu)^T \Sigma^{-1} (y_i - \mu)\right) \frac{1}{\sqrt{(2\pi)^k \det(\Sigma)}}$$
$$\frac{\lambda^\lambda}{\Gamma(\lambda)} \gamma_i^{\lambda-1} \exp\left(-\gamma_i \lambda\right)$$

Note that

$$\Sigma^{-1} = \left(\frac{\sigma^2}{\gamma} I\right)^{-1} = \frac{\gamma}{\sigma^2} I; \quad \det(\Sigma) = \det\left(\frac{\sigma^2}{\gamma} I\right) = \left(\frac{\sigma^2}{\gamma}\right)^k \det(I) = \frac{\sigma^{2k}}{\gamma^k}$$

Derivation of $\gamma_i | Y_i$ in hierarchical model

Hence

$$f_{Y_i=y_i|\gamma_i} f_{\gamma_i} = \exp\left(-\frac{1}{2}||y_i-\mu||_2^2 \frac{\gamma_i}{\sigma^2}\right)(2\pi)^{\frac{-k}{2}} \frac{\gamma_i^{k/2}}{\sigma^k} \frac{\lambda^\lambda}{\Gamma(\lambda)} \gamma_i^{\lambda-1} \exp(-\gamma_i \lambda) =$$

$$(2\pi)^{\frac{-k}{2}} \frac{\lambda^\lambda}{\Gamma(\lambda)} \gamma_i^{(\lambda+\frac{k}{2})-1} \exp\left(-\gamma_i\left(\frac{1}{2\sigma^2}||y_i-\mu||_2^2+\lambda\right)\right)$$

We see the kernel of a Gamma distribution. Thus

$$\gamma_i | Y_i \sim \mathrm{G}\left(\alpha = \lambda + \frac{k}{2}, \beta = 1/\left(\frac{||y_i-\mu||_2^2}{2\sigma^2}+\lambda\right)\right)$$

where the constants in front cancel out with the multivariate t-student distribution in the denominator.