

# Overview of the package LMMstar

Brice Ozenne

September 6, 2022

This vignette describes the main functionalities of the **LMMstar** package. This package implements specific types of linear mixed models, mainly useful when having repeated observations over a discrete variable (e.g. time, brain region, ...). Key assumptions are that at the cluster level, observations are independent and that the mean and variance are independent (conditionally on covariates). In particular, in large samples the residuals do not have to be normally distributed.

The user interface of the **LMMstar** package is made of the following functions:

- the function `lmm` is the main function of the package which fits linear mixed models. The user can interact with *lmm* objects using:
  - `anova` to test linear combinations of coefficients (Wald test or Likelihood ratio tests). The output be combined via `rbind`.
  - `coef` to extract the estimates.
  - `confint` to extract the estimates with their confidence intervals.
  - `dummy.coef` to extract the estimated (marginal) mean for each combination of categorical covariate.
  - `estimate` to test non-linear combinations of coefficients (Wald test via a first order delta method).
  - `levels` to extract the reference level for the mean structure. (i.e. what (**Intercept**) refers to in presence of categorical covariates).
  - `logLik` to output the log-likelihood of the estimated model.
  - `model.tables` to extract the estimates, standard errors, p-value, and confidence intervals.
  - `plot` to obtain a diagnostic plots, partial residual plots, or a graphical display of the fitted values.
  - `predict` to compute the conditional mean for new observations.
  - `profile` to display the likelihood or profile likelihood of the model.
  - `residuals` to extract the observed residuals of the fitted model.
  - `sigma` to extract the modeled residual variance covariance matrix.
  - `summary` to obtain a summary of the input, model fit, and estimated values.
- the `mlmm` function to fit (distinct) linear mixed models on different outcome, and gather the estimated coefficients.

- the `summarize` function to compute summary statistics stratified on a categorical variable.
- the `partialCor` function to compute partial correlation between two variables.
- the `sampleRem` function to simulate longitudinal data.
- the `LMMstar.options` function enables the user to display the default values used in the **LMMstar** package. The function can also change the default values to better match the user needs.

Before going further we need to load the **LMMstar** package in the R session:

```
library(LMMstar)
```

To illustrate the functionalities of the package, we will use the `gastricbypassL` dataset:

```
data(gastricbypassL, package = "LMMstar")
head(gastricbypassL)
```

	id	visit	time	weight	glucagonAUC
1	1	1	3monthsBefore	127.2	5032.50
2	2	1	3monthsBefore	165.2	12142.50
3	3	1	3monthsBefore	109.7	10321.35
4	4	1	3monthsBefore	146.2	6693.00
5	5	1	3monthsBefore	113.1	7090.50
6	6	1	3monthsBefore	158.8	10386.00

See `?gastricbypassL` for a presentation of the dataset. We will shorten the values of the time variable:

```
gastricbypassL$time <- factor(gastricbypassL$time,
  levels = c("3monthsBefore", "1weekBefore",
    "1weekAfter", "3monthsAfter" ),
  labels = c("B3m", "B1w", "A1w", "A3m"))
gastricbypassL$visit <- as.numeric(gastricbypassL$time) ## convert to numeric
gastricbypassL$baseline <- gastricbypassL$visit<=2
```

rescale the glucagon values

```
gastricbypassL$glucagon <- as.double(scale(gastricbypassL$glucagonAUC))+5
```

and add a group variable:

```
gastricbypassL$group <- as.numeric(gastricbypassL$id)%%2
```

Note: the **LMMstar** package is under active development. Newer package versions may include additional functionalities and fix previous bugs. The version of the package that is being used is:

```
utils::packageVersion("LMMstar")
```

```
[1] '0.8.0'
```

# 1 Descriptive statistics

Mean, standard deviation, and other summary statistic can be computed with respect to a categorical variable (typically time) using the `summarize` function:

```
sss <- summarize(weight+glucagon ~ time, data = gastricbypassL, na.rm = TRUE)
print(sss, digits = 3)
```

	outcome	time	observed	missing	mean	sd	min	q1	median	q3	max
1	weight	B3m	20	0	128.97	20.269	100.90	115.30	123.10	139.82	173.00
2	weight	B1w	20	0	121.24	18.910	95.70	107.78	114.50	134.53	162.20
3	weight	A1w	20	0	115.70	18.275	89.90	102.22	110.60	128.38	155.00
4	weight	A3m	20	0	102.36	17.054	78.80	90.40	98.50	108.25	148.00
5	glucagon	B3m	20	0	4.51	0.641	3.61	4.06	4.33	4.93	6.03
6	glucagon	B1w	19	1	4.39	0.558	3.58	4.05	4.23	4.55	5.95
7	glucagon	A1w	19	1	6.06	1.044	4.52	5.30	5.94	6.62	8.27
8	glucagon	A3m	20	0	5.06	0.760	3.95	4.52	5.03	5.27	7.12

Correlation matrices are also output when a cluster and ordering variable have been specified (here respectively `id` and `time`):

```
sss <- summarize(weight ~ time|id, data = gastricbypassL, na.rm = TRUE)
print(sss, digits = 3)
```

	outcome	time	observed	missing	mean	sd	min	q1	median	q3	max
1	weight	B3m	20	0	129	20.3	100.9	115.3	123.1	140	173
2	weight	B1w	20	0	121	18.9	95.7	107.8	114.5	135	162
3	weight	A1w	20	0	116	18.3	89.9	102.2	110.6	128	155
4	weight	A3m	20	0	102	17.1	78.8	90.4	98.5	108	148

Pearson's correlation:

	weight.B3m	weight.B1w	weight.A1w	weight.A3m
weight.B3m	1.000	0.990	0.986	0.946
weight.B1w	0.990	1.000	0.997	0.959
weight.A1w	0.986	0.997	1.000	0.966
weight.A3m	0.946	0.959	0.966	1.000

Alternatively, correlation and partial correlations can be computed using the `partialCor` function:

```
data(gastricbypassW, package = "LMMstar")
partialCor(weight1 + weight3 ~ 1, data = gastricbypassW)
```

	estimate	se	df	lower	upper	p.value
rho(weight1,weight3)	0.986	0.0984	36	0.966	0.995	6.57e-13

```
partialCor(weight + glucagonAUC ~ group,
  data = gastricbypassL[gastricbypassL$time=="B3m",])
```

	estimate	se	df	lower	upper	p.value
rho(weight,glucagonAUC)	-0.124	0.232	9.14	-0.576	0.386	0.61

## 2 Linear mixed model

### 2.1 Classical covariance patterns

Fit a linear model with **identity** structure:

```
eId.lmm <- lmm(weight ~ time + glucagon,
  repetition = ~time|id, structure = "ID",
  data = gastricbypassL)
eId.lmm
cat(" covariance structure: \n");sigma(eId.lmm)
```

Linear regression

```
outcome/cluster/time: weight/id/time
data                  : 78 observations and distributed in 20 clusters
parameters            : 5 mean ((Intercept) timeB1w timeA1w timeA3m glucagon)
                      1 variance (sigma)
log-restr.likelihood: -323.086426918519
convergence           : TRUE (0 iterations)
covariance structure:
      B3m    B1w    A1w    A3m
B3m 330.0427  0.0000  0.0000  0.0000
B1w  0.0000 330.0427  0.0000  0.0000
A1w  0.0000  0.0000 330.0427  0.0000
A3m  0.0000  0.0000  0.0000 330.0427
```

Fit a linear model with **independence** structure:

```
eInd.lmm <- lmm(weight ~ time + glucagon,
  repetition = ~time|id, structure = "IND",
  data = gastricbypassL)
eInd.lmm
cat(" covariance structure: \n");sigma(eInd.lmm)
```

Linear regression with heterogeneous residual variance

```
outcome/cluster/time: weight/id/time
data                  : 78 observations and distributed in 20 clusters
parameters            : 5 mean ((Intercept) timeB1w timeA1w timeA3m glucagon)
                      4 variance (sigma k.B1w k.A1w k.A3m)
log-restr.likelihood: -321.457830361849
convergence           : TRUE (8 iterations)
covariance structure:
      B3m    B1w    A1w    A3m
B3m 442.6475  0.0000  0.0000  0.0000
B1w  0.0000 418.9934  0.0000  0.0000
A1w  0.0000  0.0000 222.8463  0.0000
A3m  0.0000  0.0000  0.0000 237.2049
```

Fit a linear mixed model with **compound symmetry** structure:

```
eCS.lmm <- lmm(weight ~ time + glucagon,  
  repetition = ~time|id, structure = "CS",  
  data = gastricbypassL)  
eCS.lmm  
cat(" covariance structure: \n");sigma(eCS.lmm)
```

Linear Mixed Model with a compound symmetry covariance matrix

```
outcome/cluster/time: weight/id/time  
data                  : 78 observations and distributed in 20 clusters  
parameters            : 5 mean ((Intercept) timeB1w timeA1w timeA3m glucagon)  
                      1 variance (sigma)  
                      1 correlation (rho)  
log-restr.likelihood: -243.600523870252  
convergence           : TRUE (9 iterations)  
covariance structure:  
      B3m      B1w      A1w      A3m  
B3m 355.3062 344.6236 344.6236 344.6236  
B1w 344.6236 355.3062 344.6236 344.6236  
A1w 344.6236 344.6236 355.3062 344.6236  
A3m 344.6236 344.6236 344.6236 355.3062
```

Fit a linear mixed model with **stratified compound symmetry** covariance matrix:

```
eSCS.lmm <- lmm(weight ~ time*group,
  repetition = ~time|id, structure = CS(group~1),
  data = gastricbypassL)
eSCS.lmm
cat(" covariance structure: \n");sigma(eSCS.lmm)
```

Linear Mixed Model with a stratified compound symmetry covariance matrix

```
outcome/cluster/time: weight/id/time
data                  : 80 observations and distributed in 20 clusters
parameters            : 8 mean ((Intercept) timeB1w timeA1w timeA3m group timeB1w:group timeA1w:grou
                        2 variance (sigma:0 sigma:1)
                        2 correlation (rho:0 rho:1)
log-restr.likelihood: -233.141302306302
convergence            : TRUE (6 iterations)
covariance structure:
$`0`
      B3m      B1w      A1w      A3m
B3m 348.0783 334.7404 334.7404 334.7404
B1w 334.7404 348.0783 334.7404 334.7404
A1w 334.7404 334.7404 348.0783 334.7404
A3m 334.7404 334.7404 334.7404 348.0783

$`1`
      B3m      B1w      A1w      A3m
B3m 345.1388 340.0877 340.0877 340.0877
B1w 340.0877 345.1388 340.0877 340.0877
A1w 340.0877 340.0877 345.1388 340.0877
A3m 340.0877 340.0877 340.0877 345.1388
```

Fit a linear mixed model with **block compound symmetry** covariance matrix<sup>1</sup>:

```
eBCS.lmm <- lmm(weight ~ time*group,
  repetition = ~time|id, structure = CS(~baseline, heterogeneous = FALSE),
  data = gastricbypassL)
eBCS.lmm
cat(" covariance structure: \n");sigma(eBCS.lmm)
```

#### Linear Mixed Model with a block compound symmetry covariance matrix

```
outcome/cluster/time: weight/id/time
data                  : 80 observations and distributed in 20 clusters
parameters           : 8 mean ((Intercept) timeB1w timeA1w timeA3m group timeB1w:group timeA1w:grou
                      1 variance (sigma)
                      2 correlation (rho(TRUE) rho(TRUE,FALSE))
log-restr.likelihood: -234.971305082514
convergence           : TRUE (6 iterations)
covariance structure:
      B3m      B1w      A1w      A3m
B3m 346.6085 339.4747 336.3836 336.3836
B1w 339.4747 346.6085 336.3836 336.3836
A1w 336.3836 336.3836 346.6085 339.4747
A3m 336.3836 336.3836 339.4747 346.6085
```

Fit a linear mixed model with **toeplitz** covariance matrix:

```
eTOE.lmm <- lmm(weight ~ time*group,
  repetition = ~time|id, structure = "TOEPLITZ",
  data = gastricbypassL)
eTOE.lmm
cat(" correlation structure: \n");cov2cor(sigma(eTOE.lmm))
```

```
outcome/cluster/time: weight/id/time
data                  : 80 observations and distributed in 20 clusters
parameters           : 8 mean ((Intercept) timeB1w timeA1w timeA3m group timeB1w:group timeA1w:grou
                      4 variance (sigma k.B1w k.A1w k.A3m)
                      3 correlation (rho(B3m,B1w) rho(B3m,A1w) rho(B3m,A3m))
log-restr.likelihood: -224.790790046711
convergence           : TRUE (19 iterations)
correlation structure:
      B3m      B1w      A1w      A3m
B3m 1.0000000 0.9857538 0.9675323 0.9481027
B1w 0.9857538 1.0000000 0.9857538 0.9675323
A1w 0.9675323 0.9857538 1.0000000 0.9857538
A3m 0.9481027 0.9675323 0.9857538 1.0000000
```

---

<sup>1</sup>same as nested random effects

Fit a linear mixed model with **block unstructured** covariance matrix:

```
eBUN.lmm <- lmm(weight ~ time*group,
  repetition = ~time|id, structure = CS(~baseline),
  data = gastricbypassL)
eBUN.lmm
cat(" covariance structure: \n");sigma(eBUN.lmm)
```

#### Linear Mixed Model with a block unstructured covariance matrix

```
outcome/cluster/time: weight/id/time
data                  : 80 observations and distributed in 20 clusters
parameters           : 8 mean ((Intercept) timeB1w timeA1w timeA3m group timeB1w:group timeA1w:grou
                      2 variance (sigma k.TRUE)
                      3 correlation (rho(TRUE) rho(TRUE,FALSE) rho(FALSE))
log-restr.likelihood: -231.80588606934
convergence           : TRUE (6 iterations)
covariance structure:
      B3m      B1w      A1w      A3m
B3m 377.4267 372.4602 336.3836 336.3836
B1w 372.4602 377.4267 336.3836 336.3836
A1w 336.3836 336.3836 315.7904 306.4892
A3m 336.3836 336.3836 306.4892 315.7904
```

Fit a linear mixed model with **unstructured** covariance matrix:

```
eUN.lmm <- lmm(weight ~ time + glucagon,
  repetition = ~time|id, structure = "UN",
  data = gastricbypassL)
eUN.lmm
cat(" covariance structure: \n");sigma(eUN.lmm)
```

#### Linear Mixed Model with an unstructured covariance matrix

```
outcome/cluster/time: weight/id/time
data                  : 78 observations and distributed in 20 clusters
parameters           : 5 mean ((Intercept) timeB1w timeA1w timeA3m glucagon)
                      4 variance (sigma k.B1w k.A1w k.A3m)
                      6 correlation (rho(B3m,B1w) rho(B3m,A1w) rho(B3m,A3m) rho(B1w,A1w) rho(B1w,A
log-restr.likelihood: -216.318937004306
convergence           : TRUE (22 iterations)
covariance structure:
      B3m      B1w      A1w      A3m
B3m 411.3114 381.9734 352.6400 318.8573
B1w 381.9734 362.7326 335.4649 304.6314
A1w 352.6400 335.4649 311.6921 285.8077
A3m 318.8573 304.6314 285.8077 280.9323
```



Fit a linear mixed model with **stratified unstructured** covariance matrix:

```
eSUN.lmm <- lmm(weight ~ time*group + glucagon,
  repetition = ~time|id, structure = UN(~group),
  data = gastricbypassL)
eSUN.lmm
cat(" covariance structure: \n");sigma(eSUN.lmm)
```

#### Linear Mixed Model with a stratified unstructured covariance matrix

```
outcome/cluster/time: weight/id/time
data                  : 78 observations and distributed in 20 clusters
parameters            : 9 mean ((Intercept) timeB1w timeA1w timeA3m group glucagon timeB1w:group tim
                        8 variance (sigma:0 sigma:1 k.B1w:0 k.A1w:0 k.A3m:0 k.B1w:1 k.A1w:1 k.A3m:1)
                        12 correlation (rho(B3m,B1w):0 rho(B3m,A1w):0 rho(B3m,A3m):0 rho(B1w,A1w):0
log-restr.likelihood: -197.171312062213
convergence           : TRUE (50 iterations)
covariance structure:
$`0`
      B3m      B1w      A1w      A3m
B3m 417.3374 382.8829 362.5674 301.7430
B1w 382.8829 364.4515 346.4039 292.7507
A1w 362.5674 346.4039 331.1789 282.9301
A3m 301.7430 292.7507 282.9301 253.3324

$`1`
      B3m      B1w      A1w      A3m
B3m 383.8877 363.6405 336.5771 350.0416
B1w 363.6405 347.9898 321.5908 331.5182
A1w 336.5771 321.5908 297.5329 308.1345
A3m 350.0416 331.5182 308.1345 334.8267
```

## 2.2 Model output

The `summary` method can be used to display the main information relative to the model fit:

```
summary(eUN.lmm)
```

### Linear Mixed Model

Dataset: `gastricbypassL`

- 20 clusters
- 78 observations were analyzed, 2 were excluded because of missing values
- between 3 and 4 observations per cluster

Summary of the outcome and covariates:

```
$ weight : num  127 165 110 146 113 ...
$ time    : Factor w/ 4 levels "B3m","B1w","A1w",...: 1 1 1 1 1 1 1 1 1 1 ...
$ glucagon: num   4.03 5.24 4.93 4.32 4.38 ...
reference level: time=B3m
```

Estimation procedure

- Restricted Maximum Likelihood (REML)
- log-likelihood : -216.3189
- parameters: mean = 5, variance = 4, correlation = 6
- convergence: TRUE (22 iterations)
- largest `|score|` = 7.034659e-05 for `k.A1w`  
|change| = 1.09738491005373e-06 for (Intercept)

Residual variance-covariance: unstructured

- correlation structure: `~time - 1`

	B3m	B1w	A1w	A3m
B3m	1.000	0.989	0.985	0.938
B1w	0.989	1.000	0.998	0.954
A1w	0.985	0.998	1.000	0.966
A3m	0.938	0.954	0.966	1.000
- variance structure: `~time`

	standard.deviation	ratio
sigma.B3m	20.3	1.000
sigma.B1w	19.0	0.939
sigma.A1w	17.7	0.871
sigma.A3m	16.8	0.826

Fixed effects: weight ~ time + glucagon

	estimate	se	df	lower	upper	p.value
(Intercept)	132.98	4.664	19.8	123.243	142.717	< 0.001 ***
timeB1w	-7.882	0.713	19.2	-9.374	-6.39	< 0.001 ***
timeA1w	-11.788	1.018	21.6	-13.9	-9.676	< 0.001 ***
timeA3m	-26.122	1.656	18.8	-29.591	-22.654	< 0.001 ***
glucagon	-0.888	0.242	13.7	-1.408	-0.369	0.00257 **

-----  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1.

Columns lower and upper contain 95% pointwise confidence intervals for each coefficient.

Model-based standard errors are derived from the observed information (column se).

Degrees of freedom were computed using a Satterthwaite approximation (column df).

Note: the calculation of the degrees of freedom, especially when using the observed information can be quite slow. Setting the arguments `df` to `FALSE` and `type.information` to `"expected"` when calling `lmm` should lead to a more reasonable computation time.

## 2.3 Extract estimated coefficients

The value of the estimated coefficients can be output using `coef`:

```
coef(eUN.lmm)
```

(Intercept)	timeB1w	timeA1w	timeA3m	glucagon
132.9801355	-7.8822331	-11.7879545	-26.1223908	-0.8883081

Variance coefficients can be output by specifying the `effects` argument:

```
coef(eUN.lmm, effects = "variance")
```

sigma	k.B1w	k.A1w	k.A3m
20.2808131	0.9390916	0.8705176	0.8264480

It is possible to apply specific transformation on the variance coefficients, for instance to obtain the residual variance relative to each outcome:

```
coef(eUN.lmm, effects = "variance", transform.k = "sd")
```

sigma.B3m	sigma.B1w	sigma.A1w	sigma.A3m
20.28081	19.04554	17.65480	16.76104

The marginal means at each timepoint can be obtained using `dummy.coef`:

```
dummy.coef(eUN.lmm)
```

	time	estimate	se	df	lower	upper
1	B3m	128.5386	4.536445	18.97584	119.04289	138.0343
2	B1w	120.6564	4.261691	19.04078	111.73783	129.5749
3	A1w	116.7506	3.956964	19.04925	108.47007	125.0312
4	A3m	102.4162	3.747908	19.05531	94.57328	110.2591

## 2.4 Extract estimated coefficient and associated uncertainty

The uncertainty about the mean coefficients can be obtained using the `model.tables` method <sup>2</sup>:

```
model.tables(eUN.lmm)
```

	estimate	se	df	lower	upper	p.value
(Intercept)	132.9801355	4.6642475	19.75815	123.243045	142.7172256	0.000000e+00
timeB1w	-7.8822331	0.7131797	19.17147	-9.374032	-6.3904339	9.273644e-10
timeA1w	-11.7879545	1.0175135	21.64404	-13.900162	-9.6757467	9.552470e-11
timeA3m	-26.1223908	1.6564077	18.84049	-29.591280	-22.6535021	2.617462e-12
glucagon	-0.8883081	0.2416081	13.70759	-1.407545	-0.3690712	2.571605e-03

Values for the all correlation parameters can be displayed too, by specifying `effect="all"`:

```
model.tables(eUN.lmm, effect = "all") ## not shown
```

Because these parameters are constrained (e.g. strictly positive), they uncertainty is by default computed after transformation (e.g. `log`) and then backtransformed.

## 2.5 Extract estimated residual variance-covariance structure

The method `sigma` can be used to output the covariance structure of the residuals:

```
Sigma <- sigma(eUN.lmm)
Sigma
```

	B3m	B1w	A1w	A3m
B3m	411.3114	381.9734	352.6400	318.8573
B1w	381.9734	362.7326	335.4649	304.6314
A1w	352.6400	335.4649	311.6921	285.8077
A3m	318.8573	304.6314	285.8077	280.9323

and then converted to a correlation matrix using `cov2cor`:

```
cov2cor(Sigma)
```

	B3m	B1w	A1w	A3m
B3m	1.0000000	0.9889048	0.9848800	0.9380157
B1w	0.9889048	1.0000000	0.9976791	0.9542904
A1w	0.9848800	0.9976791	1.0000000	0.9658511
A3m	0.9380157	0.9542904	0.9658511	1.0000000

The method can also be used to extract the residual covariance relative to a "known" individual:

```
sigma(eUN.lmm, cluster = 5)
```

---

<sup>2</sup>it is equivalent to `confint` method except that by default it also outputs `se` and `p.value`

	B3m	A1w	A3m
B3m	411.3114	352.6400	318.8573
A1w	352.6400	311.6921	285.8077
A3m	318.8573	285.8077	280.9323

or for a new individual:

```
newdata <- data.frame(id = "X", time = c("B3m", "B1w", "A1w", "A3m"))
sigma(eUN.lmm, cluster = newdata)
```

	B3m	B1w	A1w	A3m
B3m	411.3114	381.9734	352.6400	318.8573
B1w	381.9734	362.7326	335.4649	304.6314
A1w	352.6400	335.4649	311.6921	285.8077
A3m	318.8573	304.6314	285.8077	280.9323

## 2.6 Random effects

Mixed model having a compound symmetry structure with positive correlation parameters are equivalent to random intercept models, possibly with nested random effects. Indeed the residual variance-covariance matrix can then be decomposed as  $\Omega = Z\Omega_1 Z^\top + \Omega_2$  where  $Z$  is the design matrix associated to the possibly nested clustering factors,  $\Omega_1$  is the variance-covariance of the random effects and  $\Omega_2$  the residual-variance covariance conditional to the random effects. Denoting by  $\varepsilon_i$  the vector of marginal residuals relative to individual  $i$  with variance-covariance matrix  $\Omega_i$ ,  $\eta_{ij} = \omega_{2j} Z_{ij} \Omega_i^{-1} \varepsilon_i$  is its  $j$ -th random effect with design matrix  $Z_{ij}$  and variance  $\omega_{2j}$ . This is what the `coef` method returns when setting the argument `effects` to `"ranef"`:

```
head(coef(eCS.lmm, effects = "ranef"))
```

	id
1	0.9036038
2	32.5542378
3	-18.3099658
4	20.2561307
5	-15.4258816
6	19.3751847

```
head(coef(eBCS.lmm, effects = "ranef"))
```

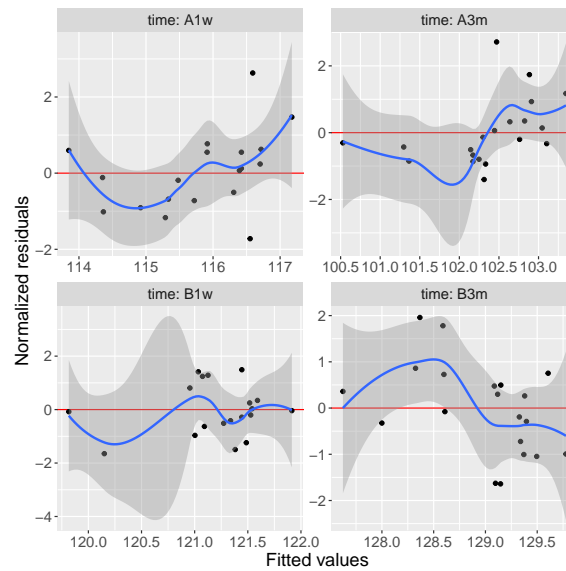
	id	baseline1	baseline2
1	4.958429	0.55088599	-0.5053222
2	28.398952	-0.09700981	0.3579722
3	-13.706851	0.20977987	-0.3357343
4	15.650120	0.83098280	-0.6871714
5	-11.181840	-0.31252621	0.2097745
6	15.006490	-2.67719285	2.8150898

## 2.7 Model diagnostic

The method `plot` can be used to display diagnostic plots about:

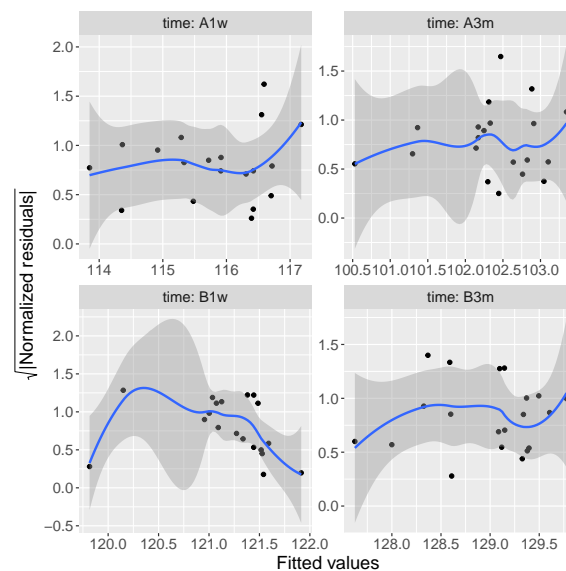
- misspecification of the mean structure

```
plot(eUN.lmm, type = "scatterplot")
```



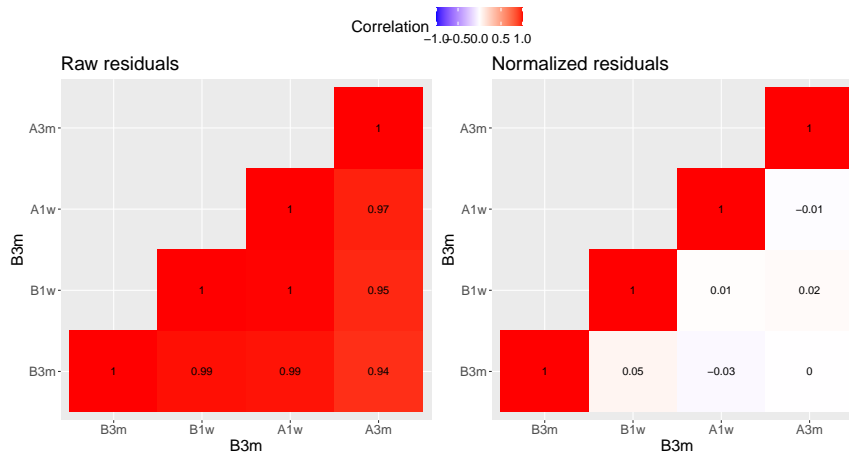
- misspecification of the variance structure

```
plot(eUN.lmm, type = "scatterplot2")
```



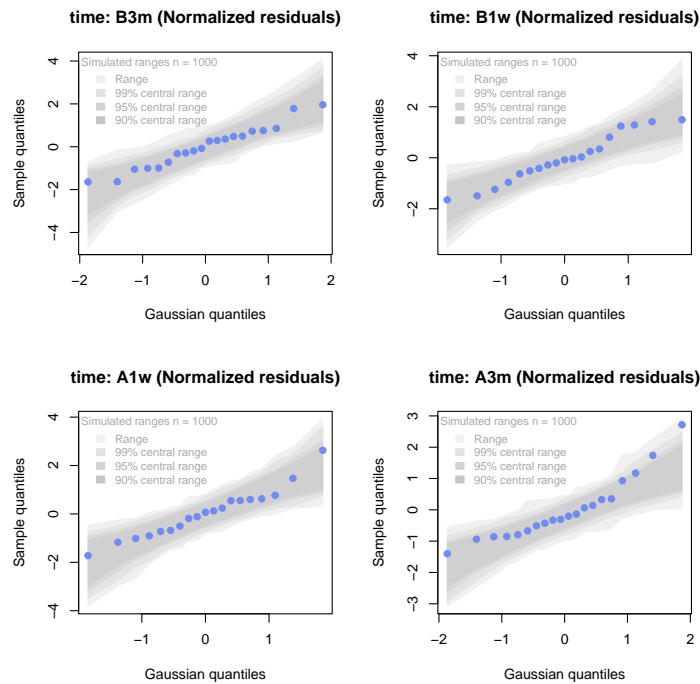
- misspecification of the correlation structure

```
plot(eUN.lmm, type = "correlation", type.residual = "response")
plot(eUN.lmm, type = "correlation", type.residual = "normalized")
```



- residual distribution vs. normal distribution <sup>3</sup>:

```
plot(eUN.lmm, type = "qqplot", engine.qqplot = "qqtest")
## Note: the qqtest package to be installed to use the argument engine.plot = "qqtest"
```



<sup>3</sup>see Oldford (2016) for guidance about how to read quantile-quantile plots.

The method `residuals` returns the residulas in the wide format:

```
eUN.diagW <- residuals(eUN.lmm, type = "normalized", format = "wide")
colnames(eUN.diagW) <- gsub("normalized.", "", colnames(eUN.diagW))
head(eUN.diagW)
```

	cluster	r.B3m	r.B1w	r.A1w	r.A3m
1	1	-0.2897365	-0.2027622	-1.16864038	0.3258573
2	2	0.8603117	-1.6492164	0.62578801	1.7370660
3	3	0.7273066	-0.4155171	-0.68266741	-0.8510316
4	4	-1.6403082	-0.5128368	0.06806206	1.1725813
5	5	0.4755409	NA	-0.18736415	-0.8634200
6	6	1.7801675	1.2847703	2.63004812	0.3505542

or in the long format:

```
eUN.diagL <- residuals(eUN.lmm, type = "normalized", format = "long")
head(eUN.diagL)
```

```
[1] -0.2897365  0.8603117  0.7273066 -1.6403082  0.4755409  1.7801675
```

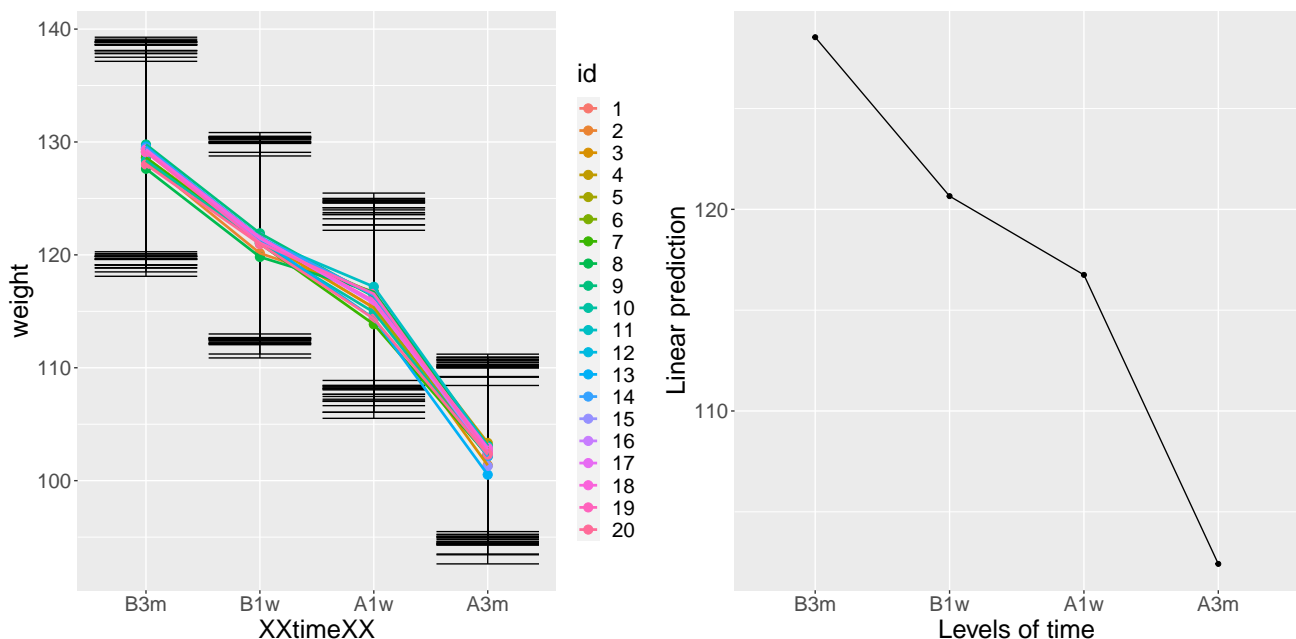
Various type of residuals can be extract but the normalized one are recommanded when doing model checking.

## 2.8 Model fit

The fitted values can be displayed via the `plot` method or using the `emmeans` package:

```
library(ggplot2) ## left panel
plot(eUN.lmm, type = "fit", color = "id", ci.alpha = NA, size.text = 20)
```

```
library(emmeans) ## right panel
emmip(eUN.lmm, ~time) + theme(text = element_text(size=20))
```



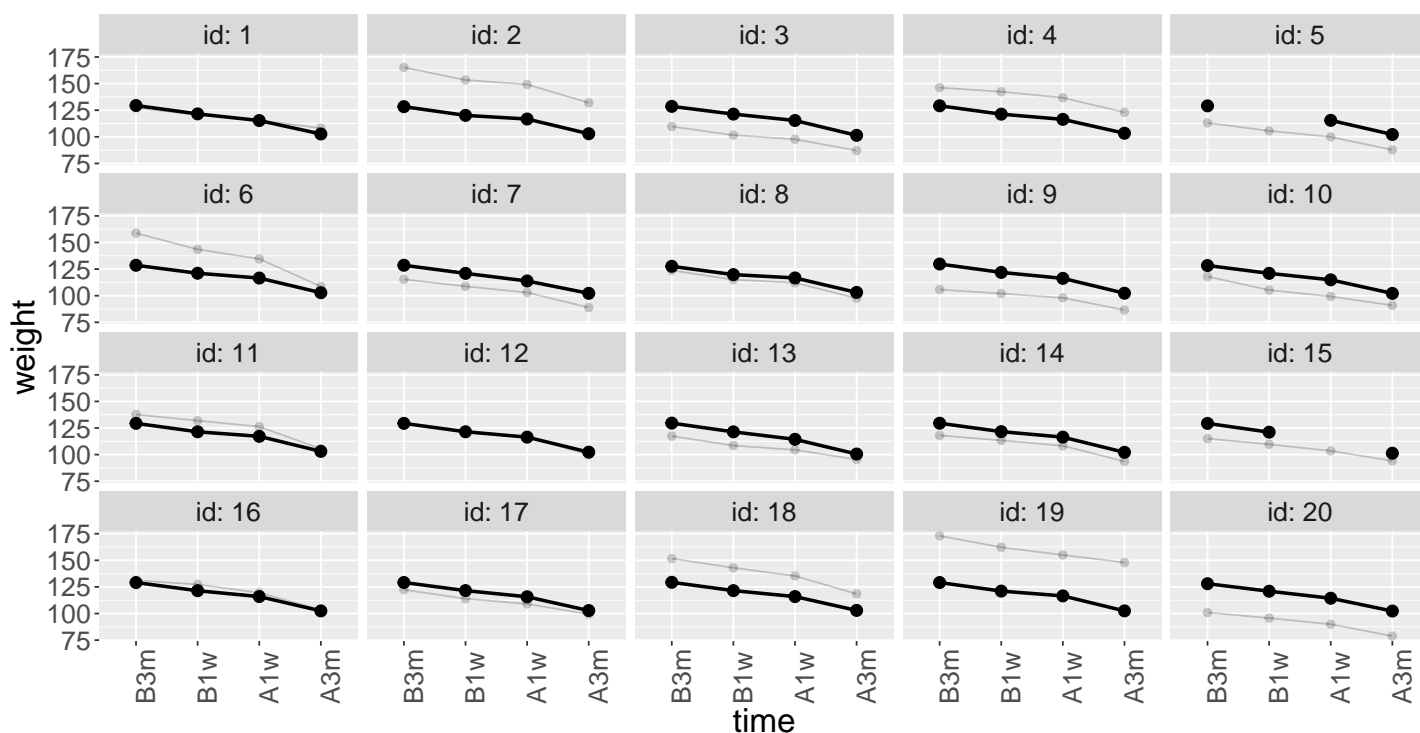


In the first case each possible curve is displayed while in the latter the average curve (over glucagon values). With the `plot` method, it is possible to display a curve specific to a glucagon value via the argument `at`:

```
plot(eUN.lmm, type = "fit", at = data.frame(glucagon = 10), color = "glucagon")
## result not shown
```

It is also possible to display the observed values along with the fitted values by setting the argument `obs.alpha` to a strictly positive value below or equal to 1. This argument controls the transparency of the color used to display the observed values:

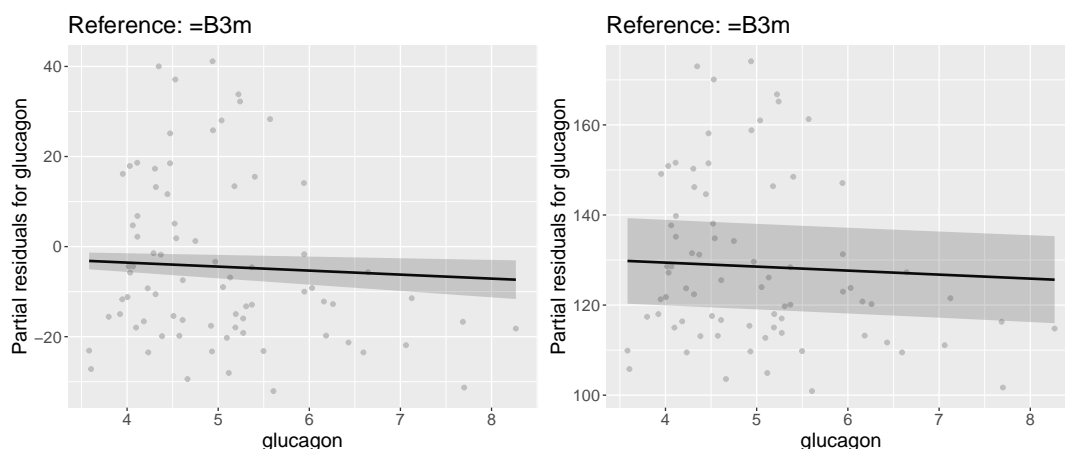
```
gg <- plot(eUN.lmm, type = "fit", obs.alpha = 0.2, ci = FALSE, plot = FALSE)$plot
gg <- gg + facet_wrap(~id, labeller = label_both)
gg <- gg + theme(axis.text.x=element_text(angle = 90, hjust = 0))
gg
```



## 2.9 Partial residuals

Partial residuals can also be displayed via the `plot` method:

```
gg1 <- plot(eUN.lmm, type = "partial", var = "glucagon", plot = FALSE)$plot
gg2 <- plot(eUN.lmm, type = "partial", var = c("(Intercept)", "glucagon"), plot = FALSE)$plot
ggarrange(gg1, gg2)
```



Their value can be extracted via the `residuals` method, e.g.:

```
df.pres <- residuals(eUN.lmm, type = "partial", var = "glucagon", keep.data = TRUE)
head(df.pres)
```

	id	visit	time	weight	glucagonAUC	baseline	glucagon	group	r.partial
1	1	1	B3m	127.2	5032.50	TRUE	4.034616	1	-5.780135
2	2	1	B3m	165.2	12142.50	TRUE	5.240766	0	32.219865
3	3	1	B3m	109.7	10321.35	TRUE	4.931824	1	-23.280135
4	4	1	B3m	146.2	6693.00	TRUE	4.316306	0	13.219865
5	5	1	B3m	113.1	7090.50	TRUE	4.383738	1	-19.880135
6	6	1	B3m	158.8	10386.00	TRUE	4.942791	0	25.819865

This matches manual calculation:

```
m.pres <- gastricbypassL$weight - model.matrix(~time, gastricbypassL) %*% coef(eUN.lmm)[1:4]
range(df.pres$r.partial - m.pres, na.rm = TRUE)
```

```
[1] -1.065814e-14 1.421085e-14
```

Note: to match the partial residuals obtained from `lm`:

```
eIID.lm <- lm(glucagon ~ time + weight, data = gastricbypassL)
pRes.lm <- residuals(eIID.lm, type = "partial")[, "weight"]
```

one should use `type` equal to `"partial-center"` which also removes the average effect of the covariate:

```
eIID.lmm <- lmm(glucagon ~ time + weight, data = gastricbypassL)
pRes.lmm <- residuals(eIID.lmm, type = "partial-center", var = "weight")
range(pRes.lm - na.omit(pRes.lmm))
```

```
[1] -6.883383e-15 8.881784e-15
```

## 2.10 Statistical inference (linear)

The `anova` method can be used to test one or several linear combinations of the model coefficients using Wald tests. By default, it will simultaneously test all parameters associated to a variable:

```
anova(eUN.lmm)
```

Multivariate Wald test

	F-statistic	df	p.value
mean: time	86.743 (3,19.0)	2.84e-11	***
: glucagon	13.518 (1,13.7)	0.00257	**

Note that here the p-values are not adjusted for multiple comparisons over variables. It is possible to specify a null hypothesis to be tested: e.g. is there a change in average weight just after taking the treatment:

```
anova(eUN.lmm, effects = c("timeA1w-timeB1w=0"))
```

Multivariate Wald test

	F-statistic	df	p.value
all: 1	43.141 (1,17.9)	3.72e-06	***

One can also simultaneously test several null hypotheses:

```
e.anova <- anova(eUN.lmm, effects = c("timeA1w-timeB1w=0", "timeA3m-timeB1w=0"))
summary(e.anova)
```

Multivariate Wald test

	F-statistic	df	p.value
all: 1	98.651 (2,18.6)	1.2338e-10	***

-----  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1.

Degrees of freedom were computed using a Satterthwaite approximation (column df).

Univariate Wald test

	estimate	se	df	lower	upper	p.value
timeA1w - timeB1w	-3.906	0.595	17.9	-5.323	-2.489	< 1e-05 ***
timeA3m - timeB1w	-18.24	1.323	19	-21.392	-15.088	< 1e-05 ***

-----  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1.

Columns lower/upper/p.value adjusted for multiple comparisons -- max-test.

(1e+05 samples have been used)

Model-based standard errors are derived from the observed information (column se).

Degrees of freedom were computed using a Satterthwaite approximation (column df).

or return all pairwise comparisons for a given factor using the `mcp` function of the `multcomp` package:

```
library(multcomp)
summary(anova(eUN.lmm, effects = mcp(time = "Tukey")))
```

Singular contrast matrix: contrasts "A1w - B1w" "A3m - B1w" "A3m - A1w" have been removed.

Multivariate Wald test

	F-statistic	df	p.value
all: time	86.743	(3,19.0)	2.84e-11 ***

-----  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1.

Degrees of freedom were computed using a Satterthwaite approximation (column df).

Univariate Wald test

	estimate	se	df	lower	upper	p.value
B1w - B3m	-7.882	0.713	19.2	-9.825	-5.94	<1e-05 ***
A1w - B3m	-11.788	1.018	21.6	-14.559	-9.017	<1e-05 ***
A3m - B3m	-26.122	1.656	18.8	-30.633	-21.611	<1e-05 ***
A1w - B1w	-3.906	0.595	17.9	-5.525	-2.286	<1e-05 ***
A3m - B1w	-18.24	1.323	19	-21.843	-14.638	<1e-05 ***
A3m - A1w	-14.334	1.057	20.3	-17.212	-11.457	<1e-05 ***

-----  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1.

Columns lower/upper/p.value adjusted for multiple comparisons -- max-test.

(1e+05 samples have been used)

Model-based standard errors are derived from the observed information (column se).

Degrees of freedom were computed using a Satterthwaite approximation (column df).

Here the `summary` method prints not only the global test but also the result associated to each hypothesis. When testing transformed variance or correlation parameters, parentheses (as in `log(k).B1w`) cause problem for recognizing parameters:

```
try(
  anova(eUN.lmm,
  effects = c("log(k).B1w=0","log(k).A1w=0","log(k).A3m=0"))
)
```

Error in `.anova_Wald(object, effects = effects, robust = robust, rhs = rhs, :`

Possible misspecification of the argument 'effects' as running `mulcomp::glht` lead to the following

Error in `parse(text = ex[i])` : <text>:1:7: uventet symbol

1: log(k).B1w

^

It is then advised to build a contrast matrix, e.g.:

```
name.coef <- rownames(confint(eUN.lmm, effects = "all"))
name.varcoef <- grep("^k",name.coef, value = TRUE)
C <- matrix(0, nrow = 3, ncol = length(name.coef), dimnames = list(name.varcoef, name.coef))
diag(C[name.varcoef,name.varcoef]) <- 1
C[,1:9]
```

```
      (Intercept) timeB1w timeA1w timeA3m glucagon sigma k.B1w k.A1w k.A3m
k.B1w           0         0         0         0         0         0         1         0         0
k.A1w           0         0         0         0         0         0         0         1         0
k.A3m           0         0         0         0         0         0         0         0         1
```

And then call the `anova` method specifying the null hypothesis via the contrast matrix:

```
anova(eUN.lmm, effects = C)
```

Multivariate Wald test

```
      F-statistic      df p.value
all: 1          6.203 (3,18.0) 0.00442 **
```

Note that using the approach of [Pipper et al. \(2012\)](#) it is also possible to adjust for multiple testing across several `lmm` objects. To do so, one first fit the mixed models, then use the `anova` method to indicate which hypotheses are being tested, and combine them using `rbind`. Here is an (artificial) example:

```
Manova <- rbind(anova(eInd.lmm, effects = "glucagon = 0"),
  anova(eCS.lmm, effects = "glucagon = 0"),
  anova(eUN.lmm, effects = "glucagon = 0"),
  name = c("Ind","CS","UN"))
summary(Manova)
```

Multivariate Wald test

```
      Chi2-statistic      df p.value
all: 1          6.393 (3,Inf) 0.000251 ***
```

-----  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1.

Univariate Wald test

```
      estimate      se      df      lower      upper p.value
Ind: glucagon    -8.27  2.574  34.2  -14.869  -1.671  0.0122 *
CS: glucagon      0.822  0.59  53.8   -0.691   2.335  0.4325
UN: glucagon    -0.888  0.353  13.7   -1.793   0.017  0.0557 .
```

-----  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1.

Columns lower/upper/p.value adjusted for multiple comparisons -- max-test.

(1e+05 samples have been used)

Robust standard errors are derived from the observed information (column se).

## 2.11 Statistical inference (non-linear)

The `estimate` function can be used to test one or several non-linear combinations of model coefficients, using a first order delta method to quantify uncertainty. The combination has to be specified via a function (argument `f`). To illustrate its use consider an ANCOVA analysis:

$$Y_{i1} = \alpha + \beta Y_{i,0} + \gamma X_i + e_i$$

```
gastricbypassW <- reshape(gastricbypassL[,c("id","time","weight","group")],
  direction = "wide",
  timevar = "time", idvar = c("id","group"))
e.ANCOVA <- lm(weight.A1w ~ weight.B1w + group, data = gastricbypassW)
summary(e.ANCOVA)$coef
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-1.4823022	2.31781138	-0.6395267	5.310047e-01
weight.B1w	0.9654917	0.01803988	53.5198489	2.156258e-20
group	0.2521714	0.66499945	0.3792054	7.092302e-01

We can replicate this analysis by first fitting a mixed model:

$$Y_{ij} = \alpha_j + \gamma_j X_i + \varepsilon_{i,j} \text{ where } \varepsilon_i \sim \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \right)$$

```
gastricbypassL23 <- gastricbypassL[gastricbypassL$visit %in% 2:3,]
gastricbypassL23$time <- droplevels(gastricbypassL23$time)
e.lmmANCOVA <- lmm(weight ~ time+time:group, repetition = ~time|id,
  data = gastricbypassL23)
```

and then perform a delta-method:

```
lava::estimate(e.lmmANCOVA, f = function(p){
  c(Y1 = as.double(p["rho(B1w,A1w)"]*p["k.A1w"]),
    X1 = as.double(p["timeA1w:group"]-p["rho(B1w,A1w)"]*p["k.A1w"]*p["timeB1w:group"]))
})
```

	estimate	se	df	lower	upper	p.value
Y1	0.9654917	0.01753161	15.96758	0.9283202	1.002663	0.0000000
X1	0.2521714	0.64626331	15.00340	-1.1252790	1.629622	0.7018732

Indeed:

$$\begin{aligned} \mathbb{E}[Y_{i2}|Y_{i1}, X_i] &= \alpha_2 + \gamma_2 X_i + \rho \frac{\sigma_2}{\sigma_1} (Y_{i1} - \alpha_1 - \gamma_1 X_i) \\ &= \alpha_2 - \rho \frac{\sigma_2}{\sigma_1} \alpha_1 + \rho \frac{\sigma_2}{\sigma_1} Y_{i1} + \left( \gamma_2 - \rho \frac{\sigma_2}{\sigma_1} \gamma_1 \right) X_i \end{aligned}$$

We obtain identical estimate but different standard-errors/degrees of freedom compared to the univariate linear model approach. The later is to be prefer as it does not rely on approximation. The former is nevertheless useful as it can handle missing data in the outcome variable.

## 2.12 Baseline adjustment

In clinical trial the group and intervention variable often do not coincide, e.g., in presence of baseline measurement. In our running example, the first two measurement are pre-treatment (i.e. treatment should be "none") while the last two measurements are post-treatment (i.e. treatment should be 1 or 2). The `baselineAdjustment` function can be helpful to:

- define the treatment variable from the time and allocation variable, where baseline has its specific value

```
gastricbypassL$treat <- baselineAdjustment(gastricbypassL, variable = "group",
  repetition = ~time|id, constrain = c("B3m","B1w"),
  new.level = "none")
table(treat = gastricbypassL$treat, time = gastricbypassL$time, group = gastricbypassL$group)
```

```
, , group = 0
```

```
      time
treat  B3m B1w A1w A3m
  none   10  10   0   0
    0     0   0  10  10
    1     0   0   0   0
```

```
, , group = 1
```

```
      time
treat  B3m B1w A1w A3m
  none   10  10   0   0
    0     0   0   0   0
    1     0   0  10  10
```

- define the treatment variable from the time and allocation variable, where baseline corresponds to the reference group

```
gastricbypassL$treat2 <- baselineAdjustment(gastricbypassL, variable = "group",
  repetition = ~time|id, constrain = c("B3m","B1w"))
table(treat = gastricbypassL$treat2, time = gastricbypassL$time, group = gastricbypassL$group)
```

```
, , group = 0
```

```
      time
treat  B3m B1w A1w A3m
    1   10  10   0   0
    0    0   0  10  10
```

```
, , group = 1
```

```

      time
treat B3m B1w A1w A3m
      1  10  10  10  10
      0   0   0   0   0

```

- define a time varying treatment variable from the time and allocation variable

```

gastricbypassL$timeXtreat <- baselineAdjustment(gastricbypassL, variable = "group",
  repetition = ~time|id, constrain = c("B3m", "B1w"),
  collapse.time = ".")

table(treat = gastricbypassL$timeXtreat, time = gastricbypassL$time, group = gastricbypassL$
  group)

```

```

, , group = 0

```

```

      time
treat   B3m B1w A1w A3m
   B3m    10  0  0  0
   B1w     0 10  0  0
   A1w.0   0  0 10  0
   A3m.0   0  0  0 10
   A1w.1   0  0  0  0
   A3m.1   0  0  0  0

```

```

, , group = 1

```

```

      time
treat   B3m B1w A1w A3m
   B3m    10  0  0  0
   B1w     0 10  0  0
   A1w.0   0  0  0  0
   A3m.0   0  0  0  0
   A1w.1   0  0 10  0
   A3m.1   0  0  0 10

```

We would then typically like to model group differences only after baseline (i.e. only at 1 week and 3 months after). This can be performed using the time varying treatment variable, e.g.:

```

eC.lmm <- lmm(weight ~ timeXtreat, data = gastricbypassL,
  repetition = ~time|id, structure = "UN")
coef(eC.lmm) ## change from baseline

```

```

(Intercept)      timeXtreatB1w timeXtreatA1w.0 timeXtreatA3m.0 timeXtreatA1w.1 timeXtreatA3m.1
  128.97000         -7.73000         -13.38978         -28.52130         -13.15022         -24.68870

```



or

```
eC2.lmm <- lmm(weight ~ 0 + timeXtreat, data = gastricbypassL,  
  repetition = ~time|id, structure = "UN")  
coef(eC2.lmm) ## absolute value
```

```
timeXtreatB3m    timeXtreatB1w timeXtreatA1w.0 timeXtreatA3m.0 timeXtreatA1w.1 timeXtreatA3m.1  
128.9700        121.2400        115.5802        100.4487        115.8198        104.2813
```

The parametrization however does not (directly) output treatment effects. Instead one may be tempted to use a formula like `treatment*time`. However this will lead to a non-identifiable model. Indeed we are only able to estimate a total of 6 means when constraining the expected baseline value between the two groups to be the same. Therefore can at most identify 6 effects. However the design matrix for the interaction model:

```
colnames(model.matrix(weight ~ treat*time, data = gastricbypassL))
```

```
[1] "(Intercept)"    "treat0"          "treat1"          "timeB1w"         "timeA1w"  
[6] "timeA3m"        "treat0:timeB1w"  "treat1:timeB1w"  "treat0:timeA1w"  "treat1:timeA1w"  
[11] "treat0:timeA3m" "treat1:timeA3m"
```

contains 12 parameters (i.e. 6 too many). Fortunately, the `lmm` will drop non-identifiable effects from the model and fit the resulting simplified model:

```
eC3.lmm <- lmm(weight ~ treat2*time, data = gastricbypassL,  
  repetition = ~time|id, structure = "UN")
```

Constant values in the design matrix for the mean structure.

Coefficients "treat20" "treat20:timeB1w" relative to interactions "treat2:time" have been removed.

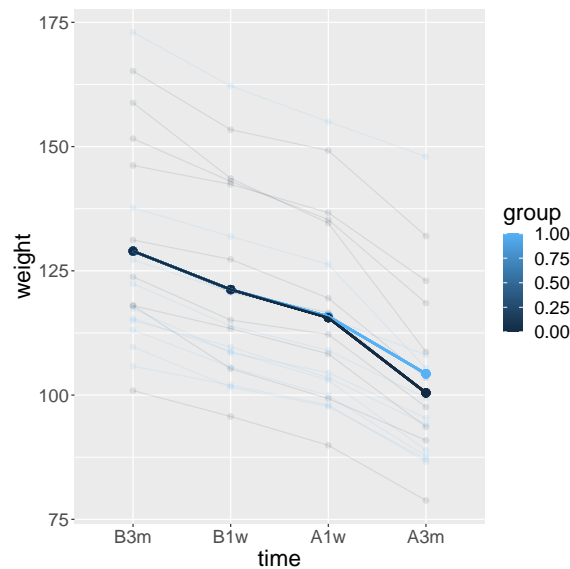
with the following coefficients:

```
model.tables(eC3.lmm)
```

	estimate	se	df	lower	upper	p.value
(Intercept)	128.9700000	4.5323695	18.98130	119.483009	138.4569912	0.000000e+00
timeB1w	-7.7300000	0.6974427	18.97552	-9.189892	-6.2701082	9.938186e-10
timeA1w	-13.1502219	0.8970429	22.87334	-15.006465	-11.2939786	4.058975e-13
timeA3m	-24.6886957	1.7751662	22.25061	-28.367762	-21.0096290	1.863398e-12
treat20:timeA1w	-0.2395562	0.6484895	17.66860	-1.603816	1.1247037	7.162149e-01
treat20:timeA3m	-3.8326086	2.1066817	17.60613	-8.265691	0.6004734	8.592047e-02

One can visualize the baseline adjustment via the `autoplot` function:

```
autoplot(eC3.lmm, color = "group", ci = FALSE, size.text = 20, obs.alpha = 0.1)
```



## 2.13 Marginal means

The `emmeans` package can be used to output marginal means. Consider the following model:

```
gastricbypassL$group2 <- as.numeric(gastricbypassL$id) %% 3 == 0
e.group <- lmm(glucagon ~ time*group2, data = gastricbypassL,
  repetition = ~time|id, structure = "UN")
```

We can for instance compute the average value over time *assuming balanced groups*:

```
emmeans(e.group, specs=~time)
```

NOTE: Results may be misleading due to involvement in interactions

time	emmean	SE	df	lower.CL	upper.CL
B3m	4.45	0.156	18.0	4.12	4.78
B1w	4.32	0.131	18.0	4.05	4.60
A1w	5.95	0.262	18.4	5.40	6.50
A3m	5.12	0.187	18.0	4.73	5.51

Results are averaged over the levels of: group2

Confidence level used: 0.95

This differs from the average value over time over the whole sample:

```
df.pred <- cbind(gastricbypassL, predict(e.group, newdata = gastricbypassL))
summarize(formula = estimate~time, data = df.pred)
```

	outcome	time	observed	missing	mean	sd	min	q1	median	q3	max
1	estimate	B3m	20	0	4.514352	0.1502565	4.290643	4.290643	4.610227	4.610227	4.610227
2	estimate	B1w	20	0	4.390071	0.1617778	4.149209	4.149209	4.493298	4.493298	4.493298
3	estimate	A1w	20	0	6.044056	0.2109650	5.729961	5.729961	6.178668	6.178668	6.178668
4	estimate	A3m	20	0	5.057642	0.1465315	4.964144	4.964144	4.964144	5.275805	5.275805

as the groups are not balanced:

```
table(group = gastricbypassL$group2, time = gastricbypassL$time)
```

```
      time
group  B3m B1w A1w A3m
FALSE  14  14  14  14
TRUE   6   6   6   6
```

The "emmeans" approach gives equal "weight" to the expected value of both group:

```
mu.group1 <- as.double(coef(e.group)["(Intercept)"])
mu.group2 <- as.double(coef(e.group)["(Intercept)"] + coef(e.group)["group2TRUE"])
p.group1 <- 14/20 ; p.group2 <- 6/20
c(emmeans = (mu.group1+mu.group2)/2, predict = mu.group1 * p.group1 + mu.group2 * p.group2)
```

```
emmeans predict
4.450435 4.514352
```

Which one is relevant depends on the application. The `emmeans` function can also be used to display expected value in each group over time:

```
emmeans.group <- emmeans(e.group, specs = ~group2|time)
emmeans.group
```

```
time = B3m:
group2 emmean    SE    df lower.CL upper.CL
FALSE   4.61 0.171 18.0    4.25    4.97
TRUE    4.29 0.262 18.0    3.74    4.84
```

```
time = B1w:
group2 emmean    SE    df lower.CL upper.CL
FALSE   4.49 0.145 18.4    4.19    4.80
TRUE    4.15 0.219 17.9    3.69    4.61
```

```
time = A1w:
group2 emmean    SE    df lower.CL upper.CL
FALSE   6.18 0.277 17.8    5.60    6.76
TRUE    5.73 0.446 18.6    4.80    6.66
```

```
time = A3m:
group2 emmean    SE    df lower.CL upper.CL
FALSE   4.96 0.205 18.0    4.53    5.39
TRUE    5.28 0.313 18.0    4.62    5.93
```

Confidence level used: 0.95

Using the `pair` function displays the differences:

```
epairs.group <- pairs(emmeans.group, reverse = TRUE)
epairs.group
```

```
time = B3m:
  contrast      estimate    SE   df t.ratio p.value
TRUE - FALSE   -0.320 0.313 18.0  -1.022  0.3202
```

```
time = B1w:
  contrast      estimate    SE   df t.ratio p.value
TRUE - FALSE   -0.344 0.262 18.0  -1.311  0.2062
```

```
time = A1w:
  contrast      estimate    SE   df t.ratio p.value
TRUE - FALSE   -0.449 0.525 18.4  -0.855  0.4034
```

```
time = A3m:
  contrast      estimate    SE   df t.ratio p.value
TRUE - FALSE    0.312 0.374 18.0   0.834  0.4153
```

One can adjust for multiple comparison via the `adjust` argument and display confidence intervals setting the argument `infer` to `TRUE`:

```
summary(epairs.group, by = NULL, adjust = "mvt", infer = TRUE)
```

contrast	time	estimate	SE	df	lower.CL	upper.CL	t.ratio	p.value
TRUE - FALSE	B3m	-0.320	0.313	18.0	-1.156	0.516	-1.022	0.6927
TRUE - FALSE	B1w	-0.344	0.262	18.0	-1.046	0.357	-1.311	0.5058
TRUE - FALSE	A1w	-0.449	0.525	18.4	-1.852	0.954	-0.855	0.7961
TRUE - FALSE	A3m	0.312	0.374	18.0	-0.688	1.311	0.834	0.8084

Confidence level used: 0.95

Conf-level adjustment: mvt method for 4 estimates

P value adjustment: mvt method for 4 tests

This should also work when doing baseline adjustment (because of baseline adjustment no difference is expected at the first two timepoints):

```
summary(pairs(emmeans(eC3.lmm , specs = ~treat2|time), reverse = TRUE), by = NULL)
```

Note: `adjust = "tukey"` was changed to `"sidak"`

because `"tukey"` is only appropriate for one set of pairwise comparisons

contrast	time	estimate	SE	df	t.ratio	p.value
treat20 - treat21	B3m	0.00	0.000	Inf	NaN	NaN
treat20 - treat21	B1w	0.00	0.000	Inf	NaN	NaN
treat20 - treat21	A1w	-0.24	0.648	18	-0.369	0.9935
treat20 - treat21	A3m	-3.83	2.107	18	-1.819	0.3019

P value adjustment: sidak method for 4 tests

## 2.14 Predictions

Two types of predictions can be performed with the `predict` method:

- **static predictions** that are only conditional on the covariates:

```
news <- gastricbypassL[gastricbypassL$id==1,]  
news$glucagon <- 0  
predict(eUN.lmm, newdata = news)
```

```
      estimate      se      df      lower      upper  
1 132.9801 4.664247 19.75815 123.24305 142.7172  
2 125.0979 4.388294 19.91418 115.94155 134.2543  
3 121.1922 4.214230 20.55331 112.41660 129.9678  
4 106.8577 3.942058 20.95499  98.65871 115.0568
```

which can be computing by creating a design matrix:

```
X.12 <- model.matrix(formula(eUN.lmm), news)  
X.12
```

```
      (Intercept) timeB1w timeA1w timeA3m glucagon  
1             1         0         0         0         0  
21            1         1         0         0         0  
41            1         0         1         0         0  
61            1         0         0         1         0  
attr(,"assign")  
[1] 0 1 1 1 2  
attr(,"contrasts")  
attr(,"contrasts")$time  
[1] "contr.treatment"
```

and then multiplying it with the regression coefficients:

```
X.12 %*% coef(eUN.lmm)
```

```
      [,1]  
1 132.9801  
21 125.0979  
41 121.1922  
61 106.8577
```

- **dynamic predictions** that are conditional on the covariates and the outcome measured at other timepoints. Consider two subjects for who we would like to predict the weight 1 week before the intervention based on the weight 3 months before the intervention:

```
newd <- rbind(
  data.frame(id = 1, time = "B3m", weight = coef(eUN.lmm)["(Intercept)"], glucagon = 0),
  data.frame(id = 1, time = "B1w", weight = NA, glucagon = 0),
  data.frame(id = 2, time = "B3m", weight = 100, glucagon = 0),
  data.frame(id = 2, time = "B1w", weight = NA, glucagon = 0)
)
predict(eUN.lmm, newdata = newd, type = "dynamic", keep.newdata = TRUE)
```

	id	time	weight	glucagon	estimate	se	df	lower	upper
1	1	B3m	132.9801	0	NA	NA	NA	NA	NA
2	1	B1w	NA	0	125.09790	0.6362754	Inf	123.85083	126.3450
3	2	B3m	100.0000	0	NA	NA	NA	NA	NA
4	2	B1w	NA	0	94.47017	7.2279385	Inf	80.30367	108.6367

The first subjects has the average weight while the second has a much lower weight. The predicted weight for the first subject is then the average weight one week before while it is lower for the second subject due to the positive correlation over time. The predicted value is computed using the formula of the conditional mean for a Gaussian vector:

```
mu1 <- coef(eUN.lmm)[1]
mu2 <- sum(coef(eUN.lmm)[1:2])
Omega_11 <- sigma(eUN.lmm)["B3m","B3m"]
Omega_21 <- sigma(eUN.lmm)["B1w","B3m"]
as.double(mu2 + Omega_21 * (100 - mu1) / Omega_11)
```

```
[1] 94.47017
```

## 2.15 Missing values and imputation

We now consider the glucagon level as an outcome. The `summarize` function can be used to describe the amount of missing data at each repetition:

```
sss <- summarize(glucagon ~ time, data = gastricbypassL, na.rm = TRUE)
cbind(sss[,1:4], pc = paste0(100 * sss$missing / (sss$missing + sss$observed), "%"))
```

```
  outcome time observed missing pc
1 glucagon B3m      20      0 0%
2 glucagon B1w      19      1 5%
3 glucagon A1w      19      1 5%
4 glucagon A3m      20      0 0%
```

Further description of the missing data patterns rely on function outside the LMMstar package, e.g. appropriate call to `tapply` and `table`:

```
vec.pattern <- tapply(as.numeric(is.na(gastricbypassL$glucagon)),
  INDEX = gastricbypassL$id,
  FUN = paste, collapse=".")
table(vec.pattern)
```

```
vec.pattern
0.0.0.0 0.0.1.0 0.1.0.0
      18       1       1
```

Linear mixed model can handle missing value in the outcome variable, assuming that missigness is random conditional on the covariate and observed outcome values. The `lmm` function can be used "as usual":

```
eUN.lmmNA <- lmm(glucagon ~ time,
  repetition = ~time|id, structure = "UN",
  data = gastricbypassL)
summary(eUN.lmmNA)
```

### Linear Mixed Model

Dataset: `gastricbypassL`

- 20 clusters
- 78 observations were analyzed, 2 were excluded because of missing values
- between 3 and 4 observations per cluster

Summary of the outcome and covariates:

```
$ glucagon: num  4.03 5.24 4.93 4.32 4.38 ...
$ time    : Factor w/ 4 levels "B3m","B1w","A1w",...: 1 1 1 1 1 1 1 1 1 1 ...
reference level: time=B3m
```

The visible difference in the summary is when describing the dataset: we can see that some repetitions (here 2) have been ignored as the outcome was missing. So for some clusters only 3 values were analyzed instead of 4. It is possible to extract the most likely value for these missing observation using the `fitted` function with argument `impute=TRUE`:

```
fitted(eUN.lmmNA, impute = TRUE)
```

[1] 4.256984 6.497856

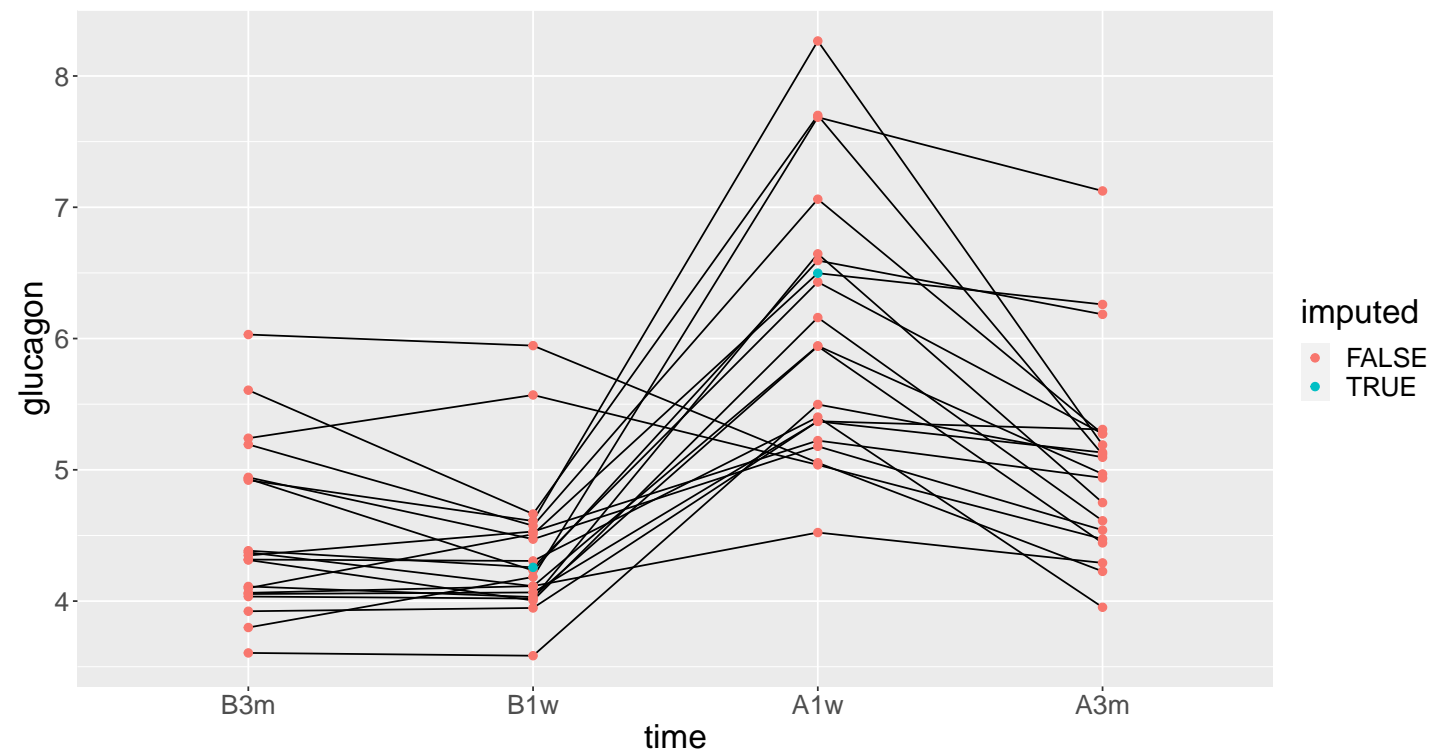
When using the argument `keep.newdata=TRUE`, the missing outcome value has been replaced by its most likely value (which is the same as the dynamic prediction, described previously):

```
eData <- fitted(eUN.lmmNA, impute = TRUE, keep.newdata = TRUE)
eData$treat <- eData$treat2 <- eData$timeXtreat <- NULL
eData[eData$id %in% eData[eData$imputed,"id"],]
```

	id	visit	time	weight	glucagonAUC	baseline	glucagon	group	group2	imputed
5	5	1	B3m	113.1	7090.5	TRUE	4.383738	1	FALSE	FALSE
15	15	1	B3m	115.0	5410.5	TRUE	4.098741	1	TRUE	FALSE
25	5	2	B1w	105.6	NA	TRUE	4.256984	1	FALSE	TRUE
35	15	2	B1w	109.7	7833.0	TRUE	4.509697	1	TRUE	FALSE
45	5	3	A1w	99.9	19155.0	FALSE	6.430376	1	FALSE	FALSE
55	15	3	A1w	103.5	NA	FALSE	6.497856	1	TRUE	TRUE
65	5	4	A3m	87.7	12345.0	FALSE	5.275118	1	FALSE	FALSE
75	15	4	A3m	94.1	18148.5	FALSE	6.259632	1	TRUE	FALSE

Visually:

```
ggplot(eData, aes(x=time,y=glucagon, group=id)) + geom_line() + geom_point(aes(color=imputed))
```





It is possible to sample from the estimated distribution of the missing value instead of using the most likely value, e.g. accounting for residual variance and uncertainty related to parameter estimation:

```
set.seed(10)
fitted(eUN.lmmNA, impute = TRUE, se = "total")
fitted(eUN.lmmNA, impute = TRUE, se = "total")
fitted(eUN.lmmNA, impute = TRUE, se = "total")
```

```
[1] 4.262434 6.305287
```

```
[1] 3.858267 5.871642
```

```
[1] 4.342624 6.905246
```

### 3 User-specific covariance patterns

From version 0.6.0 and above it is possible to customize the covariance pattern under the following model for the residuals:

$$\Omega = \sigma^\top R \sigma$$

where:

- $\sigma = f(\theta_\sigma, Z_\sigma)$  is a vector of residual standard errors depending on a vector of parameters  $\theta_\sigma$  and possible covariates via the design matrix  $Z_\sigma$ .
- $R = g(\theta_R, Z_R)$  is a matrix of residual correlations depending on a vector of parameters  $\theta_R$  and possible covariates via the design matrix  $Z_R$ .

To be more concrete, consider the following correlation matrix

```
rho.2block <- function(p,time,...){
  n.time <- length(time)
  rho <- matrix(1, nrow = n.time, ncol = n.time)
  rho[1,2] <- rho[2,1] <- rho[4,5] <- rho[5,4] <- p["rho1"]
  rho[1,3] <- rho[3,1] <- rho[4,6] <- rho[6,4] <- p["rho2"]
  rho[2,3] <- rho[3,2] <- rho[5,6] <- rho[6,5] <- p["rho3"]
  rho[4:6,1:3] <- rho[1:3,4:6] <- p["rho4"]
  return(rho)
}
Rho <- rho.2block(p = c(rho1=0.25,rho2=0.5,rho3=0.4,rho4=0.1),
  time = 1:6)
Rho
```

```
      [,1] [,2] [,3] [,4] [,5] [,6]
[1,] 1.00 0.25  0.5 0.10 0.10  0.1
[2,] 0.25 1.00  0.4 0.10 0.10  0.1
[3,] 0.50 0.40  1.0 0.10 0.10  0.1
[4,] 0.10 0.10  0.1 1.00 0.25  0.5
[5,] 0.10 0.10  0.1 0.25 1.00  0.4
[6,] 0.10 0.10  0.1 0.50 0.40  1.0
```

and the corresponding dataset:

```
set.seed(11)
n <- 1000
Y <- rmvnorm(n, mean = rep(0,6), sigma = Rho)
dfL <- reshape2::melt(cbind(id = 1:n, as.data.frame(Y)), id.vars = "id")
dfL$time <- dfL$variable
dfL <- dfL[order(dfL$id),]
dfL[1:8,]
```

	id	variable	value	time
1	1	V1	-0.9842079	V1
1001	1	V2	-0.3681245	V2
2001	1	V3	-1.6174652	V3
3001	1	V4	-1.4994103	V4
4001	1	V5	0.7493107	V5
5001	1	V6	-1.0719657	V6
2	2	V1	1.2402726	V1
1002	2	V2	0.6494215	V2

To fit the corresponding mixed model, we call `lmm` using a (non-standard) correlation structure via the `CUSTOM` structure:

```
e.lmmCUSTOM <- lmm(value~time,
  repetition=~time|id,
  structure=CUSTOM(~variable,
    FCT.sigma = function(p,time,X){rep(p,length(time))}, ## function f
    init.sigma = c("sigma"=1),
    FCT.rho = rho.2block, ## function g
    init.rho = c("rho1"=0.25,"rho2"=0.25,"rho3"=0.25,"rho4"=0.25)),
  data=dfL, control = list(optimizer = "FS"),
  df = FALSE) ## df = FALSE to save computation time
logLik(e.lmmCUSTOM)
```

```
[1] -7962.243
```

The optimization procedure is not very fast but eventually reaches an optimum. We can then output the estimated correlation matrix:

```
cov2cor(sigma(e.lmmCUSTOM))
```

	V1	V2	V3	V4	V5	V6
V1	1.00000000	0.24898095	0.50058994	0.09053785	0.09053785	0.09053785
V2	0.24898095	1.00000000	0.36110943	0.09053785	0.09053785	0.09053785
V3	0.50058994	0.36110943	1.00000000	0.09053785	0.09053785	0.09053785
V4	0.09053785	0.09053785	0.09053785	1.00000000	0.24898095	0.50058994
V5	0.09053785	0.09053785	0.09053785	0.24898095	1.00000000	0.36110943
V6	0.09053785	0.09053785	0.09053785	0.50058994	0.36110943	1.00000000

Note that specifying a classical structure via `CUSTOM` (e.g. compound symmetry):

```
logLik(lmm(value~time,
  repetition=~time|id,
  structure=CUSTOM(~1,
    FCT.sigma = function(p,time,X){rep(p,length(time))},
    init.sigma = c("sigma"=1),
    FCT.rho = function(p,time,X){matrix(p,length(time),length(time))+diag(1-p,length(time),
length(time))},
    init.rho = c("rho"=0.5)),
  data=dfL, control = list(optimizer = "FS"),
  df = FALSE))
```

```
[1] -8186.859
```

should be the same as using the pre-specified structure (up the certain user-friendly displays):

```
logLik(lmm(value~time,  
  repetition=~time|id,  
  structure="CS",  
  data=dfL, control = list(optimizer = "FS"),  
  df = FALSE))
```

```
[1] -8186.859
```

## 4 Data generation

Simulate some data in the wide format:

```
set.seed(10) ## ensure reproducibility
n.obs <- 100
n.times <- 4
mu <- rep(0,4)
gamma <- matrix(0, nrow = n.times, ncol = 10) ## add interaction
gamma[,6] <- c(0,1,1.5,1.5)
dW <- sampleRem(n.obs, n.times = n.times, mu = mu, gamma = gamma, format = "wide")
head(round(dW,3))
```

	id	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	Y1	Y2	Y3	Y4
1	1	1	0	1	1	0	-0.367	1.534	-1.894	1.729	0.959	1.791	2.429	3.958	2.991
2	2	1	0	1	2	0	-0.410	2.065	1.766	0.761	-0.563	2.500	4.272	3.002	2.019
3	3	0	0	2	1	0	-1.720	-0.178	2.357	1.966	1.215	-3.208	-5.908	-4.277	-5.154
4	4	0	0	0	1	0	0.923	-2.089	0.233	1.307	-0.906	-2.062	0.397	1.757	-1.380
5	5	0	0	2	1	0	0.987	5.880	0.385	0.028	0.820	7.963	7.870	7.388	8.609
6	6	0	0	1	1	2	-1.075	0.479	2.202	0.900	-0.739	0.109	-1.602	-1.496	-1.841

Simulate some data in the long format:

```
set.seed(10) ## ensure reproducibility
dL <- sampleRem(n.obs, n.times = n.times, mu = mu, gamma = gamma, format = "long")
head(dL)
```

	id	visit	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
1	1	1	1.791444	1	0	1	1	0	-0.3665251	1.533815	-1.894425	1.7288665	0.9592499
2	1	2	2.428570	1	0	1	1	0	-0.3665251	1.533815	-1.894425	1.7288665	0.9592499
3	1	3	3.958350	1	0	1	1	0	-0.3665251	1.533815	-1.894425	1.7288665	0.9592499
4	1	4	2.991198	1	0	1	1	0	-0.3665251	1.533815	-1.894425	1.7288665	0.9592499
5	2	1	2.500179	1	0	1	2	0	-0.4097541	2.065413	1.765841	0.7613348	-0.5630173
6	2	2	4.272357	1	0	1	2	0	-0.4097541	2.065413	1.765841	0.7613348	-0.5630173

## 5 Modifying default options

The `LMMstar.options` method enable to get and set the default options used by the package. For instance, the default option for the information matrix is:

```
LMMstar.options("type.information")
```

```
$type.information  
[1] "observed"
```

To change the default option to "expected" (faster to compute but less accurate p-values and confidence intervals in small samples) use:

```
LMMstar.options(type.information = "expected")
```

To restore the original default options do:

```
LMMstar.options(reinitialise = TRUE)
```

## 6 R session

Details of the R session used to generate this document:

```
sessionInfo()
```

```
R version 4.2.0 (2022-04-22 ucrt)
```

```
Platform: x86_64-w64-mingw32/x64 (64-bit)
```

```
Running under: Windows 10 x64 (build 19044)
```

```
Matrix products: default
```

```
locale:
```

```
[1] LC_COLLATE=Danish_Denmark.utf8 LC_CTYPE=Danish_Denmark.utf8 LC_MONETARY=Danish_Denmark.utf8
```

```
[4] LC_NUMERIC=C LC_TIME=Danish_Denmark.utf8
```

```
attached base packages:
```

```
[1] stats      graphics  grDevices  utils      datasets  methods    base
```

```
other attached packages:
```

```
[1] lme4_1.1-29      sandwich_3.0-2      numDeriv_2016.8-1.1 Matrix_1.4-1
[5] lava_1.6.10      copula_1.1-0        LMMstar_0.8.0       nlme_3.1-158
[9] ggpubr_0.4.0      multcomp_1.4-19     TH.data_1.1-1       MASS_7.3-57
[13] survival_3.3-1    mvtnorm_1.1-3        qqtest_1.2.0        emmeans_1.7.4-1
[17] ggplot2_3.3.6
```

```
loaded via a namespace (and not attached):
```

```
[1] fs_1.5.2          usethis_2.1.6       devtools_2.4.3      rprojroot_2.0.3     butils_1.4.7
[6] tools_4.2.0       backports_1.4.1     utf8_1.2.2          R6_2.5.1            DBI_1.1.3
[11] mgcv_1.8-40       colorspace_2.0-3    withr_2.5.0         tidyselect_1.1.2    prettyunits_1.1.1
[16] processx_3.6.1    compiler_4.2.0      pspline_1.0-19      cli_3.3.0            desc_1.4.1
[21] labeling_0.4.2     scales_1.2.0        callr_3.7.0         pbapply_1.5-0       stringr_1.4.0
[26] digest_0.6.29     minqa_1.2.4         pkgconfig_2.0.3     parallelly_1.32.0   sessioninfo_1.2.2
[31] fastmap_1.1.0     stabledist_0.7-1    ADGofTest_0.3       rlang_1.0.4         farver_2.1.1
[36] generics_0.1.2    zoo_1.8-10          dplyr_1.0.9         car_3.1-0            magrittr_2.0.3
[41] Rcpp_1.0.8.3      munsell_0.5.0       fansi_1.0.3         abind_1.4-5          lifecycle_1.0.1
[46] stringi_1.7.6     carData_3.0-5       brio_1.1.3          plyr_1.8.7           pkgbuild_1.3.1
[51] grid_4.2.0        parallel_4.2.0      listenv_0.8.0       crayon_1.5.1         lattice_0.20-45
[56] cowplot_1.1.1     splines_4.2.0       ps_1.7.1            pillar_1.8.0         boot_1.3-28
[61] estimability_1.3  ggsignif_0.6.3      reshape2_1.4.4      future.apply_1.9.0   codetools_0.2-18
[66] stats4_4.2.0      pkgload_1.2.4       glue_1.6.2          butils.base_1.2     data.table_1.14.2
[71] remotes_2.4.2     foreach_1.5.2       vctrs_0.4.1         nloptr_2.0.3        testthat_3.1.4
[76] gtable_0.3.0      purrr_0.3.4         tidyr_1.2.0         future_1.26.1        assertthat_0.2.1
[81] cachem_1.0.6      xtable_1.8-4        broom_0.8.0         coda_0.19-4         rstatix_0.7.0
[86] pcaPP_2.0-1       gsl_2.1-7.1         tibble_3.1.7        iterators_1.0.14     memoise_2.0.1
[91] globals_0.15.1    ellipsis_0.3.2
```

## References

- Oldford, R. W. (2016). Self-calibrating quantile–quantile plots. *The American Statistician*, 70(1):74–90.
- Pipper, C. B., Ritz, C., and Bisgaard, H. (2012). A versatile method for confirmatory evaluation of the effects of a covariate in multiple models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61(2):315–326.



## Appendix A Likelihood in a linear mixed model

Denote by  $\mathbf{Y}$  a vector of  $m$  outcomes,  $\mathbf{X}$  a vector of  $p$  covariates,  $\mu(\boldsymbol{\Theta}, \mathbf{X})$  the modeled mean, and  $\Omega(\boldsymbol{\Theta}, \mathbf{X})$  the modeled residual variance-covariance. We consider  $n$  replicates (i.e.  $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ ) and  $VX_1, \dots, \mathbf{X}_n$  along with a vector of weights  $\omega = (w_1, \dots, w_n)$ , which are by default all equal to 1.

### A.1 Log-likelihood

The restricted log-likelihood in a linear mixed model can then be written:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\Theta}|\mathbf{Y}, \mathbf{X}) = & \frac{p}{2} \log(2\pi) - \frac{1}{2} \log \left( \left| \sum_{i=1}^n w_i \mathbf{X}_i \Omega_i^{-1}(\boldsymbol{\Theta}) \mathbf{X}_i^\top \right| \right) \\ & + \sum_{i=1}^n w_i \left( -\frac{m}{2} \log(2\pi) - \frac{1}{2} \log |\Omega_i(\boldsymbol{\Theta})| - \frac{1}{2} (\mathbf{Y}_i - \mu(\boldsymbol{\Theta}, \mathbf{X}_i)) \Omega_i(\boldsymbol{\Theta})^{-1} (\mathbf{Y}_i - \mu(\boldsymbol{\Theta}, \mathbf{X}_i))^\top \right) \end{aligned} \quad (\text{A})$$

This is what the `logLik` method is computing for the REML criteria. The red term is specific to the REML criteria and prevents from computing individual contributions to the likelihood<sup>4</sup>. The blue term is what `logLik` outputs for the ML criteria when setting the argument `indiv` to `TRUE`.

### A.2 Score

Using that  $\partial \log(\det(X)) = \text{tr}(X^{-1} \partial(X))$ , the score is obtained by derivating once the log-likelihood, i.e., for  $\theta \in \boldsymbol{\Theta}$ :

$$\begin{aligned} \mathcal{S}(\theta) = & \frac{\partial \mathcal{L}(\boldsymbol{\Theta}|\mathbf{Y}, \mathbf{X})}{\partial \theta} = \frac{1}{2} \text{tr} \left( \left( \sum_{i=1}^n w_i \mathbf{X}_i \Omega_i^{-1}(\boldsymbol{\Theta}) \mathbf{X}_i^\top \right)^{-1} \left( \sum_{i=1}^n w_i \mathbf{X}_i \Omega_i^{-1}(\boldsymbol{\Theta}) \frac{\partial \Omega_i(\boldsymbol{\Theta})}{\partial \theta} \Omega_i(\boldsymbol{\Theta})^{-1} \mathbf{X}_i^\top \right) \right) \\ & + \sum_{i=1}^n w_i \left( -\frac{1}{2} \text{tr} \left( \Omega_i(\boldsymbol{\Theta})^{-1} \frac{\partial \Omega_i(\boldsymbol{\Theta})}{\partial \theta} \right) + \frac{\partial \mu(\boldsymbol{\Theta}, \mathbf{X}_i)}{\partial \theta} \Omega_i(\boldsymbol{\Theta})^{-1} (\mathbf{Y}_i - \mu(\boldsymbol{\Theta}, \mathbf{X}_i))^\top \right. \\ & \quad \left. + \frac{1}{2} (\mathbf{Y}_i - \mu(\boldsymbol{\Theta}, \mathbf{X}_i)) \Omega_i(\boldsymbol{\Theta})^{-1} \frac{\partial \Omega_i(\boldsymbol{\Theta})}{\partial \theta} \Omega_i(\boldsymbol{\Theta})^{-1} (\mathbf{Y}_i - \mu(\boldsymbol{\Theta}, \mathbf{X}_i))^\top \right). \end{aligned}$$

This is what the `score` method is computing for the REML criteria. The red term is specific to the REML criteria and prevents from computing the score relative to each cluster. The blue term is what `score` outputs for the ML criteria when setting the argument `indiv` to `TRUE`.

---

<sup>4</sup>The REML is the likelihood of the observations divided by the prior on the estimated mean parameters  $\hat{\boldsymbol{\Theta}}_\mu \sim \mathcal{N}(\mu, (\mathbf{X} \Omega^{-1}(\boldsymbol{\Theta}) \mathbf{X}^\top)^{-1})$ . This corresponds to  $\frac{1}{\sqrt{2\pi^p} |(\sum_{i=1}^n \mathbf{X}_i \Omega_i^{-1}(\boldsymbol{\Theta}) \mathbf{X}_i^\top)^{-1}|} \exp \left( -(\hat{\boldsymbol{\Theta}}_\mu - \mu) (2 \sum_{i=1}^n \mathbf{X}_i \Omega_i^{-1}(\boldsymbol{\Theta}) \mathbf{X}_i^\top)^{-1} (\hat{\boldsymbol{\Theta}}_\mu - \mu)^\top \right)$ . Since  $\mu$  will be estimated to be  $\boldsymbol{\Theta}_\mu$ , the exponential term equals 1 and thus does not contribute to the log-likelihood. One divided by the other term gives  $\sqrt{2\pi^p} |(\sum_{i=1}^n \mathbf{X}_i \Omega_i^{-1}(\boldsymbol{\Theta}) \mathbf{X}_i^\top)|^{-1}$ . The log of this term equals the red term

### A.3 Hessian

Derivating a second time the log-likelihood gives the hessian,  $\mathcal{H}(\Theta)$ , with element<sup>5</sup>:

$$\begin{aligned}\mathcal{H}(\theta, \theta') &= \frac{\partial^2 \mathcal{L}(\Theta | \mathbf{Y}, \mathbf{X})}{\partial \theta \partial \theta'} = \frac{\partial \mathcal{S}(\theta)}{\partial \theta'} \\ &= \frac{1}{2} \text{tr} \left( \left( \sum_{i=1}^n w_i \mathbf{X}_i \Omega_i^{-1}(\Theta) \mathbf{X}_i^\top \right)^{-1} \left\{ \sum_{i=1}^n w_i \mathbf{X}_i \Omega_i^{-1}(\Theta) \left( \frac{\partial^2 \Omega_i(\Theta)}{\partial \theta \partial \theta'} - 2 \frac{\partial \Omega_i(\Theta)}{\partial \theta} \Omega_i^{-1}(\Theta) \frac{\partial \Omega_i(\Theta)}{\partial \theta'} \right) \Omega_i(\Theta)^{-1} \mathbf{X}_i^\top \right. \right. \\ &\quad \left. \left. + \left( \sum_{i=1}^n w_i \mathbf{X}_i \Omega_i^{-1}(\Theta) \frac{\partial \Omega_i(\Theta)}{\partial \theta} \Omega_i(\Theta)^{-1} \mathbf{X}_i^\top \right) \left( \sum_{i=1}^n w_i \mathbf{X}_i \Omega_i^{-1}(\Theta) \mathbf{X}_i^\top \right)^{-1} \left( \sum_{i=1}^n w_i \mathbf{X}_i \Omega_i^{-1}(\Theta) \frac{\partial \Omega_i(\Theta)}{\partial \theta'} \Omega_i(\Theta)^{-1} \mathbf{X}_i^\top \right) \right\} \right) \\ &\quad + \sum_{i=1}^n w_i \left( \frac{1}{2} \text{tr} \left( \Omega_i(\Theta)^{-1} \frac{\partial \Omega_i(\Theta)}{\partial \theta'} \Omega_i(\Theta)^{-1} \frac{\partial \Omega_i(\Theta)}{\partial \theta} - \Omega_i(\Theta)^{-1} \frac{\partial^2 \Omega_i(\Theta)}{\partial \theta \partial \theta'} \right) \right. \\ &\quad \left. - \frac{\partial \mu(\Theta, \mathbf{X}_i)}{\partial \theta} \Omega_i(\Theta)^{-1} \frac{\partial \Omega_i(\Theta)}{\partial \theta'} \Omega_i(\Theta)^{-1} \boldsymbol{\varepsilon}_i(\Theta)^\top - \frac{\partial \mu(\Theta, \mathbf{X}_i)}{\partial \theta} \Omega_i(\Theta)^{-1} \frac{\partial \mu(\Theta, \mathbf{X}_i)}{\partial \theta'} \right. \\ &\quad \left. + \frac{1}{2} \boldsymbol{\varepsilon}_i(\Theta) \Omega_i(\Theta)^{-1} \left( \frac{\partial^2 \Omega_i(\Theta)}{\partial \theta \partial \theta'} - \frac{\partial \Omega_i(\Theta)}{\partial \theta'} \Omega_i(\Theta)^{-1} \frac{\partial \Omega_i(\Theta)}{\partial \theta} - \frac{\partial \Omega_i(\Theta)}{\partial \theta} \Omega_i(\Theta)^{-1} \frac{\partial \Omega_i(\Theta)}{\partial \theta'} \right) \Omega_i(\Theta)^{-1} \boldsymbol{\varepsilon}_i(\Theta)^\top \right).\end{aligned}$$

where  $\boldsymbol{\varepsilon}_i(\Theta) = \mathbf{Y}_i - \mu(\Theta, \mathbf{X}_i)$ .

The `information` method will (by default) return the (observed) information which is the opposite of the hessian. So multiplying the previous formula by -1 gives what `information` output for the REML criteria. The red term is specific to the REML criteria and prevents from computing the information relative to each cluster. The blue term is what `information` outputs for the ML criteria (up to a factor -1) when setting the argument `indiv` to `TRUE`.

A possible simplification is to use the expected hessian at the maximum likelihood. Indeed for any deterministic matrix  $A$ :

- $\mathbb{E}[A(\mathbf{Y}_i - \mu(\Theta, \mathbf{X}_i))^\top | \mathbf{X}_i] = 0$
- $\mathbb{E}[(\mathbf{Y}_i - \mu(\Theta, \mathbf{X}_i))A(\mathbf{Y}_i - \mu(\Theta, \mathbf{X}_i))^\top | \mathbf{X}_i] = \text{tr}(A \text{Var}(\mathbf{Y}_i - \mu(\Theta, \mathbf{X}_i)))$

when  $\mathbb{E}[\mathbf{Y}_i - \mu(\Theta, \mathbf{X}_i)] = 0$ . This leads to:

$$\begin{aligned}\mathbb{E}[\mathcal{H}(\theta, \theta') | \mathbf{X}] &= \frac{1}{2} \text{tr} \left( \left( \sum_{i=1}^n w_i \mathbf{X}_i \Omega_i^{-1}(\Theta) \mathbf{X}_i^\top \right)^{-1} \left\{ \sum_{i=1}^n w_i \mathbf{X}_i \Omega_i^{-1}(\Theta) \left( \frac{\partial^2 \Omega_i(\Theta)}{\partial \theta \partial \theta'} - 2 \frac{\partial \Omega_i(\Theta)}{\partial \theta} \Omega_i^{-1}(\Theta) \frac{\partial \Omega_i(\Theta)}{\partial \theta'} \right) \Omega_i(\Theta)^{-1} \mathbf{X}_i^\top \right. \right. \\ &\quad \left. \left. + \left( \sum_{i=1}^n w_i \mathbf{X}_i \Omega_i^{-1}(\Theta) \frac{\partial \Omega_i(\Theta)}{\partial \theta} \Omega_i(\Theta)^{-1} \mathbf{X}_i^\top \right) \left( \sum_{i=1}^n w_i \mathbf{X}_i \Omega_i^{-1}(\Theta) \mathbf{X}_i^\top \right)^{-1} \left( \sum_{i=1}^n w_i \mathbf{X}_i \Omega_i^{-1}(\Theta) \frac{\partial \Omega_i(\Theta)}{\partial \theta'} \Omega_i(\Theta)^{-1} \mathbf{X}_i^\top \right) \right\} \right) \\ &\quad + \sum_{i=1}^n w_i \left( -\frac{1}{2} \text{tr} \left( \Omega_i(\Theta)^{-1} \frac{\partial \Omega_i(\Theta)}{\partial \theta'} \Omega_i(\Theta)^{-1} \frac{\partial \Omega_i(\Theta)}{\partial \theta} \right) - \frac{\partial \mu(\Theta, \mathbf{X}_i)}{\partial \theta} \Omega_i(\Theta)^{-1} \frac{\partial \mu(\Theta, \mathbf{X}_i)}{\partial \theta'} \right) \quad (\text{B})\end{aligned}$$

This is what `information` output when the argument `type.information` is set to "expected" (up to a factor -1).

---

<sup>5</sup>if one is relative to the mean and the other to the variance then they are respectively  $\theta$  and  $\theta'$

## A.4 Degrees of freedom

Degrees of freedom are computed using a Satterthwaite approximation, i.e. for an estimate coefficient  $\hat{\beta} \in \widehat{\Theta}$  with standard error  $\sigma_{\hat{\beta}}$ , the degree of freedom is:

$$df(\sigma_{\hat{\beta}}) = \frac{2\sigma_{\hat{\beta}}^4}{\text{Var}[\hat{\sigma}_{\hat{\beta}}]}$$

Using a first order Taylor expansion we can approximate the variance term as:

$$\begin{aligned} \text{Var}[\hat{\sigma}_{\hat{\beta}}] &\approx \frac{\partial \hat{\sigma}_{\hat{\beta}}}{\partial \Theta} \Sigma_{\Theta} \frac{\partial \hat{\sigma}_{\hat{\beta}}}{\partial \Theta}^{\top} \\ &\approx c_{\beta} (\hat{\mathcal{I}}_{\hat{\Theta}})^{-1} \frac{\partial \hat{\mathcal{I}}_{\hat{\Theta}}}{\partial \Theta} (\hat{\mathcal{I}}_{\hat{\Theta}})^{-1} c_{\beta}^{\top} \Sigma_{\Theta} c_{\beta} (\hat{\mathcal{I}}_{\hat{\Theta}})^{-1} \frac{\partial \hat{\mathcal{I}}_{\hat{\Theta}}}{\partial \Theta}^{\top} (\hat{\mathcal{I}}_{\hat{\Theta}})^{-1} c_{\beta} \end{aligned}$$

where  $\Sigma_{\Theta}$  is the variance-covariance matrix of all model coefficients,  $\mathcal{I}_{\Theta}$  the information matrix for all model coefficients,  $c_{\beta}$  a matrix used to select the element relative to  $\beta$  in the first derivative of the information matrix, and  $\frac{\partial}{\partial \Theta}$  denotes the vector of derivatives with respect to all model coefficients.

The derivative of the information matrix (i.e. negative hessian) can then be computed using numerical derivatives or using analytical formula. To obtain the later we first notice that:

$$\begin{aligned} \mathcal{H}(\theta, \theta') &= \mathbb{E}[\mathcal{H}(\theta, \theta') | \mathbf{X}] \\ &+ \sum_{i=1}^n w_i \left( \text{tr} \left( \Omega_i(\Theta)^{-1} \frac{\partial \Omega_i(\Theta)}{\partial \theta'} \Omega_i(\Theta)^{-1} \frac{\partial \Omega_i(\Theta)}{\partial \theta} - \Omega_i(\Theta)^{-1} \frac{\partial^2 \Omega_i(\Theta)}{\partial \theta \partial \theta'} \right) \right. \\ &\quad - \frac{\partial \mu(\Theta, \mathbf{X}_i)}{\partial \theta} \Omega_i(\Theta)^{-1} \frac{\partial \Omega_i(\Theta)}{\partial \theta'} \Omega_i(\Theta)^{-1} \epsilon_i(\Theta)^{\top} \\ &\quad \left. + \frac{1}{2} \epsilon_i(\Theta) \Omega_i(\Theta)^{-1} \left( \frac{\partial^2 \Omega_i(\Theta)}{\partial \theta \partial \theta'} - \frac{\partial \Omega_i(\Theta)}{\partial \theta'} \Omega_i(\Theta)^{-1} \frac{\partial \Omega_i(\Theta)}{\partial \theta} - \frac{\partial \Omega_i(\Theta)}{\partial \theta} \Omega_i(\Theta)^{-1} \frac{\partial \Omega_i(\Theta)}{\partial \theta'} \right) \Omega_i(\Theta)^{-1} \epsilon_i(\Theta)^{\top} \right) \end{aligned} \quad (\text{C})$$

where

$$\mathbb{E}[\mathcal{H}(\theta, \theta') | \mathbf{X}] = \frac{1}{2} \text{tr} \left( A(\Theta)^{-1} \left( \sum_{i=1}^n w_i b_i(\Theta) B_i(\Theta) b_i^{\top}(\Theta) + C(\Theta) A(\Theta)^{-1} C^{\top}(\Theta) \right) \right) + E(\Theta)$$

So we will first derive the derivative of  $\mathcal{H}(\theta, \theta')$  and then the one of the blue term in Equation C. To simplify the derivation of the formula we will only derive them at the maximum likelihood, i.e. when  $\mathbb{E} \left[ \frac{\partial \mathcal{H}(\theta, \theta') | \mathbf{X}}{\partial \theta''} \right] = \frac{\partial \mathbb{E}[\mathcal{H}(\theta, \theta') | \mathbf{X}]}{\partial \theta''}$  where the expectation is taken over  $\mathbf{X}$ . To find the derivative of  $\mathcal{H}(\theta, \theta')$  we can therefore take the derivative of formula (B). Its derivative with respect to the mean parameters is 0.

So we just need to compute the derivative with respect to a variance parameter  $\theta''$ :

$$\begin{aligned} & \frac{\partial A(\boldsymbol{\Theta})^{-1} (\sum_{i=1}^n w_i b_i(\boldsymbol{\Theta}) B_i(\boldsymbol{\Theta}) b_i^\top(\boldsymbol{\Theta}) + C(\boldsymbol{\Theta}) A(\boldsymbol{\Theta})^{-1} C^\top(\boldsymbol{\Theta}))}{\partial \theta''} \\ &= A(\boldsymbol{\Theta})^{-1} \frac{\partial A(\boldsymbol{\Theta})}{\partial \theta''} A(\boldsymbol{\Theta})^{-1} \left( \sum_{i=1}^n w_i b_i(\boldsymbol{\Theta}) B_i(\boldsymbol{\Theta}) b_i^\top(\boldsymbol{\Theta}) + C(\boldsymbol{\Theta}) A(\boldsymbol{\Theta})^{-1} C^\top(\boldsymbol{\Theta}) \right) \\ &+ A(\boldsymbol{\Theta})^{-1} \left( \sum_{i=1}^n w_i \left( \frac{\partial b_i(\boldsymbol{\Theta})}{\partial \theta''} B_i(\boldsymbol{\Theta}) b_i^\top(\boldsymbol{\Theta}) + b_i(\boldsymbol{\Theta}) \frac{\partial B_i(\boldsymbol{\Theta})}{\partial \theta''} b_i^\top(\boldsymbol{\Theta}) + b_i(\boldsymbol{\Theta}) B_i(\boldsymbol{\Theta}) \frac{\partial b_i^\top(\boldsymbol{\Theta})}{\partial \theta''} \right. \right. \\ &\quad \left. \left. + \frac{\partial C(\boldsymbol{\Theta})}{\partial \theta''} A^{-1}(\boldsymbol{\Theta}) C^\top(\boldsymbol{\Theta}) + C(\boldsymbol{\Theta}) A^{-1} \frac{\partial A(\boldsymbol{\Theta})}{\partial \theta''} A^{-1} C^\top(\boldsymbol{\Theta}) + C(\boldsymbol{\Theta}) A^{-1}(\boldsymbol{\Theta}) \frac{\partial C^\top(\boldsymbol{\Theta})}{\partial \theta''} \right) \right) \end{aligned}$$

and

$$\begin{aligned} \frac{\partial E(\boldsymbol{\Theta})}{\partial \theta''} &= \sum_{i=1}^n w_i \left( -\frac{1}{2} \text{tr} \left( -2 \Omega_i(\boldsymbol{\Theta})^{-1} \frac{\partial \Omega_i(\boldsymbol{\Theta})}{\partial \theta''} \Omega_i(\boldsymbol{\Theta})^{-1} \frac{\partial \Omega_i(\boldsymbol{\Theta})}{\partial \theta'} \Omega_i(\boldsymbol{\Theta})^{-1} \frac{\partial \Omega_i(\boldsymbol{\Theta})}{\partial \theta} \right. \right. \\ &\quad \left. \left. + \Omega_i(\boldsymbol{\Theta})^{-1} \frac{\partial^2 \Omega_i(\boldsymbol{\Theta})}{\partial \theta' \partial \theta''} \Omega_i(\boldsymbol{\Theta})^{-1} \frac{\partial \Omega_i(\boldsymbol{\Theta})}{\partial \theta} + \Omega_i(\boldsymbol{\Theta})^{-1} \frac{\partial \Omega_i(\boldsymbol{\Theta})}{\partial \theta'} \Omega_i(\boldsymbol{\Theta})^{-1} \frac{\partial^2 \Omega_i(\boldsymbol{\Theta})}{\partial \theta \partial \theta''} \right) \right. \\ &\quad \left. + \frac{\partial \mu(\boldsymbol{\Theta}, \mathbf{X}_i)}{\partial \theta} \Omega_i(\boldsymbol{\Theta})^{-1} \frac{\partial \Omega_i(\boldsymbol{\Theta})}{\partial \theta''} \Omega_i(\boldsymbol{\Theta})^{-1} \frac{\partial \mu(\boldsymbol{\Theta}, \mathbf{X}_i)}{\partial \theta'} \right) \end{aligned}$$

where:

$$\begin{aligned} \frac{\partial A(\boldsymbol{\Theta})}{\partial \theta''} &= \sum_{i=1}^n w_i \mathbf{X}_i \Omega_i^{-1}(\boldsymbol{\Theta}) \frac{\partial \Omega_i(\boldsymbol{\Theta})}{\partial \theta''} \Omega_i^{-1}(\boldsymbol{\Theta}) \mathbf{X}_i^\top \\ \frac{\partial b_i(\boldsymbol{\Theta})}{\partial \theta''} &= \mathbf{X}_i \Omega_i^{-1}(\boldsymbol{\Theta}) \frac{\partial \Omega_i(\boldsymbol{\Theta})}{\partial \theta''} \Omega_i^{-1}(\boldsymbol{\Theta}) \\ \frac{\partial B_i(\boldsymbol{\Theta})}{\partial \theta''} &= \frac{\partial^3 \Omega_i(\boldsymbol{\Theta})}{\partial \theta' \partial \theta''} \\ &\quad - 2 \left( \frac{\partial^2 \Omega_i(\boldsymbol{\Theta})}{\partial \theta \partial \theta''} \Omega_i^{-1}(\boldsymbol{\Theta}) \frac{\partial \Omega_i(\boldsymbol{\Theta})}{\partial \theta'} + \frac{\partial \Omega_i(\boldsymbol{\Theta})}{\partial \theta} \Omega_i^{-1}(\boldsymbol{\Theta}) \frac{\partial \Omega_i(\boldsymbol{\Theta})}{\partial \theta''} \Omega_i^{-1}(\boldsymbol{\Theta}) \frac{\partial \Omega_i(\boldsymbol{\Theta})}{\partial \theta'} + \frac{\partial \Omega_i(\boldsymbol{\Theta})}{\partial \theta} \Omega_i^{-1}(\boldsymbol{\Theta}) \frac{\partial^2 \Omega_i(\boldsymbol{\Theta})}{\partial \theta' \partial \theta''} \right) \\ \frac{\partial C(\boldsymbol{\Theta})}{\partial \theta''} &= \sum_{i=1}^n w_i \mathbf{X}_i \Omega_i^{-1}(\boldsymbol{\Theta}) \left( \frac{\partial \Omega_i(\boldsymbol{\Theta})}{\partial \theta''} \Omega_i^{-1}(\boldsymbol{\Theta}) \frac{\partial \Omega_i(\boldsymbol{\Theta})}{\partial \theta} + \frac{\partial^2 \Omega_i(\boldsymbol{\Theta})}{\partial \theta \partial \theta''} + \frac{\partial \Omega_i(\boldsymbol{\Theta})}{\partial \theta} \Omega_i^{-1}(\boldsymbol{\Theta}) \frac{\partial \Omega_i(\boldsymbol{\Theta})}{\partial \theta''} \right) \Omega_i^{-1}(\boldsymbol{\Theta}) \mathbf{X}_i^\top \end{aligned}$$

## Appendix B Likelihood ratio test with the REML criterion

The blue term of Equation A in the log-likelihood is invariant to re-parameterisation while the red term is not. This means that a re-parametrisation of  $X$  into  $\tilde{X} = BX$  with  $B$  invertible would not change the likelihood when using ML but would decrease the log-likelihood by  $\log(|B|)$  when using REML.

Let's take an example:

```
## data(gastricbypassL, package = "LMMstar")
dfTest <- gastricbypassL
dfTest$glucagon2 <- dfTest$glucagon*2
```

where we multiply one column of the design matrix by 2. As mentioned previously this does not affect the log-likelihood when using ML:

```
eML.lmmUN <- lmm(weight ~ time+glucagon, data = dfTest, repetition = ~time|id, method = "ML")
eML.lmmUN2 <- lmm(weight ~ time+glucagon2, data = dfTest, repetition = ~time|id, method = "ML")
```

```
logLik(eML.lmmUN)
logLik(eML.lmmUN2)
```

```
[1] -245.7909
[1] -245.7909
```

but it does when using REML:

```
eREML.lmmUN <- lmm(weight ~ time + glucagon, data = dfTest, repetition = ~time|id, method = "REML")
eREML.lmmUN2 <- lmm(weight ~ time + glucagon2, data = dfTest, repetition = ~time|id, method = "REML")
```

```
logLik(eREML.lmmUN)-logLik(eREML.lmmUN2)
log(2)
```

```
[1] 0.6931472
[1] 0.6931472
```

Therefore, when comparing models with different mean effects there is a risk that the difference (or part of it) in log-likelihood is due to a new parametrisation and not only to a difference in model fit. This would typically be the case when adding an interaction where we can have a smaller restricted log-likelihood when considering a more complex model:

```
set.seed(15)
dfTest$ff <- rbinom(NROW(dfTest), size = 1, prob = 0.5)
logLik(lmm(weight ~ time+glucagon, data = dfTest, repetition = ~time|id, method = "REML"))
logLik(lmm(weight ~ time+glucagon*ff, data = dfTest, repetition = ~time|id, method = "REML"))
```

```
[1] -216.3189
[1] -217.0239
```

This is quite counter-intuitive as more complex model should lead to better fit and would never happen when using ML:

```
logLik(lmm(weight ~ time + glucagon, data = dfTest, repetition = ~time|id, method = "ML"))  
logLik(lmm(weight ~ time + glucagon*ff, data = dfTest, repetition = ~time|id, method = "ML"))
```

```
[1] -218.71
```

```
[1] -217.4141
```

This is why, unless one knows what he/she is doing, it is not recommended to use likelihood ratio test to assess relevance of mean parameters in mixed models estimated with REML.