

Within-Unit Covariance Models

See DHLZ, Chapter 5; Verbeke and Molenberghs, Chapter 10; Pinheiro and Bates, Chapter 5.

Again consider the model

$$\mathbf{Y}_i = \mathbf{x}_i \boldsymbol{\beta} + \mathbf{z}_i \mathbf{b}_i + \boldsymbol{\delta}_i + \boldsymbol{\epsilon}_i,$$

where $\mathbf{b}_i \sim_{ind} \mathbf{N}(\mathbf{0}, \mathbf{D})$ and $\boldsymbol{\delta}_i \sim \mathbf{N}(\mathbf{0}, \sigma_\delta^2 \mathbf{R}_i)$ represent error terms with serial dependence (on the same individual), and $\boldsymbol{\epsilon}_i \sim \mathbf{N}(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}_{n_i})$ represent “measurement error”.

Autoregressive errors

A widely-used time series model is the autoregressive process

$$\delta_{ij} = \rho \delta_{i,j-1} + u_{ij}, \quad (9)$$

for $j \geq 2$, $|\rho| \leq 1$ where $u_{ij} \sim_{iid} \mathbf{N}(0, \sigma_u^2)$ and are independent of δ_{ik} , $k > 0$.

Repeated application of (9) gives, for $k > 0$,

$$\delta_{ij} = u_{ij} + \rho u_{i,j-1} + \rho^2 u_{i,j-2} + \dots + \rho^{k-1} u_{i,j-k+1} + \rho^k \delta_{i,j-k}. \quad (10)$$

Assume the process has been running since $j = -\infty$ and that it is ‘stable’ so that $|\rho| < 1$ and the δ_{ij} all have the same distribution. Then, from (10)

$$\text{var}(\delta_{ij}) = \sigma_u^2 (1 + \rho^2 + \rho^4 + \dots + \rho^{2(k-1)}) + \rho^{2k} \text{var}(\delta_{i,j-k}),$$

and so as $k \rightarrow \infty$, since $\sum_{l=1}^{\infty} x^{l-1} = 1/(1-x)$,

$$\text{var}(\delta_{ij}) = \frac{\sigma_u^2}{(1-\rho^2)} = \sigma_\delta^2,$$

and, by substitution of (10),

$$\text{cov}(\delta_{ij}, \delta_{i,j-k}) = \mathbf{E}[\delta_{ij} \delta_{i,j-k}] = \frac{\sigma_u^2 \rho^k}{(1-\rho^2)} = \sigma_\delta^2 \rho^k.$$

Hence under this model we have

$$\mathbf{R}_i = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n_i-1} \\ \rho & 1 & \rho & \dots & \rho^{n_i-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n_i-3} \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{n_i-1} & \rho^{n_i-2} & \rho^{n_i-3} & \dots & 1 \end{bmatrix}$$

as the correlation matrix for $\boldsymbol{\delta}_i$.

Often this model is written in the form

$$\text{cov}(Y_{ij}, Y_{ik}) = \sigma_\delta^2 \exp(-\phi d_{ijk}),$$

with $d_{ijk} = |t_{ij} - t_{ik}|$ which is valid for unequally-spaced times also.

Covariance Models

In general there are many choices for the covariance model.

Exchangeable: We have already come across this model (also known as compound symmetry):

$$\text{var}(\mathbf{Y}_i) = \sigma^2 \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$$

Sometimes written as

$$\text{var}(\mathbf{Y}_i) = (\sigma_e^2)\mathbf{I}_{n_i} + \sigma_0^2\mathbf{J}_{n_i}$$

where \mathbf{J}_{n_i} is the $n_i \times n_i$ matrix of 1's.

This model is particularly appropriate for clustered data with no time ordering (e.g. ANOVA).

Toeplitz: Unstructured correlation:

$$\text{var}(\mathbf{Y}_i) = \sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_1 \\ \rho_3 & \rho_2 & \rho_1 & 1 \end{bmatrix}$$

AR(1): We have also seen this model:

$$\text{var}(\mathbf{Y}_i) = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

Heterogeneous versions also exist, for example, the heterogeneous exchangeable model is given by:

$$\text{var}(\mathbf{Y}_i) = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho\sigma_1\sigma_3 & \rho\sigma_1\sigma_4 \\ \rho\sigma_2\sigma_1 & \sigma_2^2 & \rho\sigma_2\sigma_3 & \rho\sigma_2\sigma_4 \\ \rho\sigma_3\sigma_1 & \rho\sigma_3\sigma_2 & \sigma_3^2 & \rho\sigma_3\sigma_4 \\ \rho\sigma_4\sigma_1 & \rho\sigma_4\sigma_2 & \rho\sigma_4\sigma_3 & \sigma_4^2 \end{bmatrix}$$

Note that we should be careful when specifying the covariance structure – identifiability problems may arise if we try to be too flexible.

Inference in the LMEM

Consider again the model

$$\mathbf{y}_i = \mathbf{x}_i\boldsymbol{\beta} + \mathbf{z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i,$$

with

- $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$,
- $\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}_{n_i})$,
- \mathbf{b}_i and $\boldsymbol{\epsilon}_i$ independent.

This leads to the marginal model

$$E[\mathbf{Y}_i] = \mathbf{x}_i\boldsymbol{\beta},$$

$$\text{var}(\mathbf{Y}_i) = \mathbf{V}_i(\boldsymbol{\alpha}) = \mathbf{z}_i\mathbf{D}\mathbf{z}_i^T + \sigma_\epsilon^2\mathbf{I}_{n_i}.$$

Fixed Effects $\boldsymbol{\beta}$

The ML/GLS estimator is given by

$$\hat{\boldsymbol{\beta}} = \sum_{i=1}^m (\mathbf{x}_i\mathbf{V}_i^{-1}\mathbf{x}_i)^{-1} \mathbf{x}_i^T \mathbf{V}_i^{-1} \mathbf{Y}_i.$$

Under correct specification of the model we have

$$\hat{\boldsymbol{\beta}} \sim N_{k+1} \left\{ \boldsymbol{\beta}, \left(\sum_{i=1}^m \mathbf{x}_i\mathbf{V}_i^{-1}\mathbf{x}_i \right)^{-1} \right\},$$

if $\mathbf{V}_i = \mathbf{V}_i(\boldsymbol{\alpha})$ is known, which corresponds to knowing $\boldsymbol{\alpha}$.

In practice, $\boldsymbol{\alpha}$ is never known, but asymptotically (as $m \rightarrow \infty$):

$$\hat{\boldsymbol{\beta}} \dot{\sim} N_{k+1} \left\{ \boldsymbol{\beta}, \left(\sum_{i=1}^m \mathbf{x}_i\mathbf{V}_i(\hat{\boldsymbol{\alpha}})^{-1}\mathbf{x}_i \right)^{-1} \right\},$$

where $\hat{\boldsymbol{\alpha}}$ is a consistent estimator of $\boldsymbol{\alpha}$.

Various t and F -like approaches have been suggested for correcting for the estimation of $\boldsymbol{\alpha}$, see Verbeke and Molenberghs (2000, Chapter 6).

So far as the MLE is concerned, the expected information matrix is partitioned as

$$\mathbf{I}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \begin{bmatrix} \mathbf{I}_{\beta\beta} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{\alpha\alpha} \end{bmatrix}.$$

Standard ML theory gives the asymptotic distribution for the MLE $\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}}$, as

$$\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\alpha}} \end{bmatrix} \sim N_{k+1+r+1} \left(\begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{bmatrix}, \begin{bmatrix} \mathbf{I}_{\beta\beta}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{\alpha\alpha}^{-1} \end{bmatrix} \right),$$

where r is the number of distinct elements in \mathbf{D} .

We have already seen the form of $\mathbf{I}_{\beta\beta}^{-1}$; the form of $\mathbf{I}_{\alpha\alpha}^{-1}$ is not pleasant (but produces standard errors in software such as `lme`).

We require $m \rightarrow \infty$, it is not sufficient to have m fixed and $n_i \rightarrow \infty$ for $i = 1, \dots, m$.

The diagonal form of the expected information has a number of implications.

1. As we have shown above, we may carry out separate maximization of the log-likelihood with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$.
2. Asymptotically we have independence between $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\alpha}}$, so any consistent estimator of $\boldsymbol{\alpha}$ will give an asymptotically efficient estimator for $\boldsymbol{\beta}$.

Example: Ovarian Follicles

Pinheiro and Bates (2000) use data from a study of the number of large ovarian follicles detected in different mares at several times in their menstrual cycle.

The response is the number of ovarian follicles greater than 10mm in diameter, as a function of time in the estrus cycle (the times during which the mare is “receptive”). The data were recorded daily from 3 days before ovulation until days after the next ovulation – the measurements are scaled so that the ovulations for each mare occur at times 0 and 1.

Figure 7 shows the number of follicles versus time for each of the 11 mares, created using:

```
> library(nlme)
> data(Ovary)
> trelldat<-groupedData(follicles~Time|Mare,data=Ovary)
> plot(trelldat)
```

Note: count data but counts not too close to zero, so maybe not too bad to assume that response is continuous.

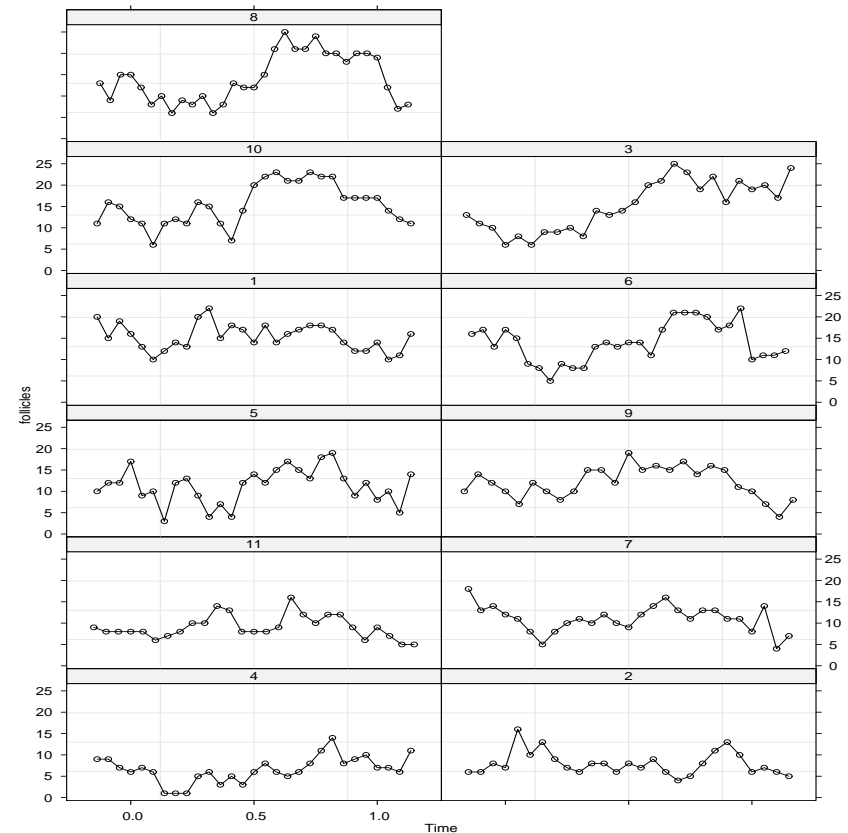


Figure 7: Follicle counts versus standardized time for 11 mares.

Model for Ovarian Data

Figure 7 shows that the counts appear to vary cyclically over time. Hence we start with a model

$$Y_{ij} = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1}) \sin(2\pi t_{ij}) + \beta_2 \cos(2\pi t_{ij}) + \epsilon_{ij},$$

where Y_{ij} is the no. of follicles on occasion t_{ij} for mare i ,

$$\begin{bmatrix} b_{i0} \\ b_{i1} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & 0 \\ 0 & \sigma_1^2 \end{bmatrix} \right),$$

and $\epsilon_{ij} \sim N(0, \sigma^2)$ with \mathbf{b}_i and ϵ_{ij} independent.

```
> x1 <- sin(2*pi*Ovary$Time)
> x2 <- cos(2*pi*Ovary$Time)
> mod1 <- lme( follicles ~ x1+x2,data=Ovary,random=pdDiag(~x1) )
> mod1
```

Linear mixed-effects model fit by REML

```
(Intercept)          x1          x2
 12.1820241  -3.2985373  -0.8623725
```

Random effects:

Formula: ~x1 | Mare

Structure: Diagonal

```
(Intercept)          x1 Residual
```

```
StdDev:    3.052136  2.079312  3.112854
```

Number of Observations: 308

Number of Groups: 11

We examine modeling assumptions.

Figure 8 was created using the following sequence:

```
> par(mfrow=c(2,2))
> empBayes <- ranef(mod1)
> plot(empBayes[,1],empBayes[,2])
> qqnorm(empBayes[,1])
> qqnorm(empBayes[,2])
> temp <- ACF(mod1,maxLag=10)
> plot(temp[,1],temp[,2],type="h",ylab="Autocorelation",
      xlab="Lag")
> abline(2/sqrt(27*11),0,lty=2)
```

Here we see that there is no evidence of non-normality for either of the random effects (though caution for such plots, here probably OK as many measurements per mare), and there does not appear to be a strong correlation between \hat{b}_{i0} and \hat{b}_{i1} .

However, there does appear to be serial correlation.

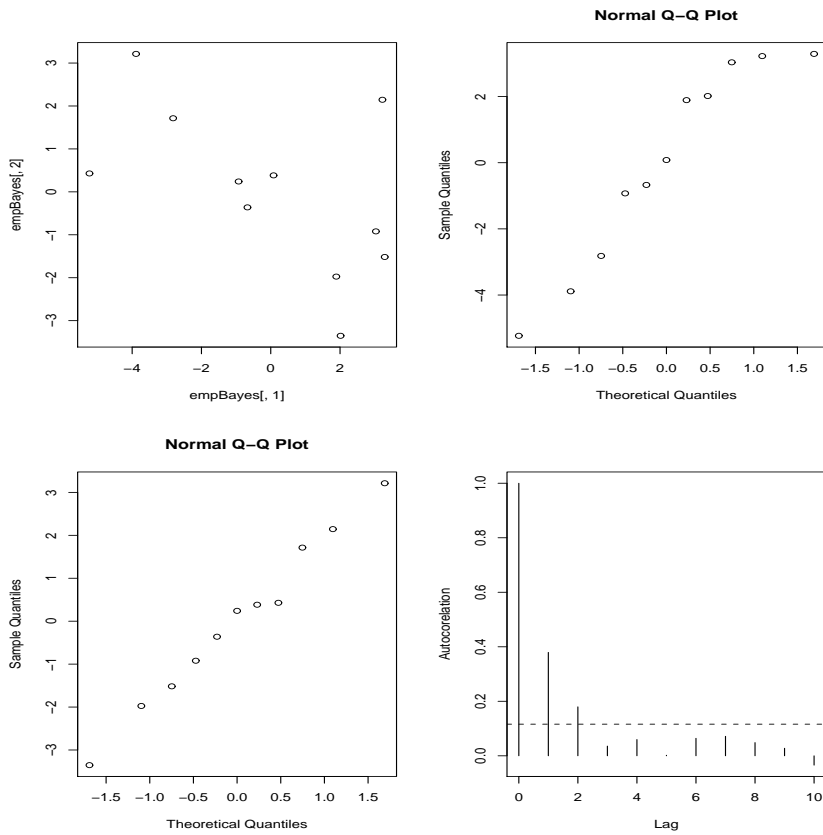


Figure 8: Diagnostics for model fit.

Figure 9 was created with:

```

> par(mfrow=c(3,2))
> fitmod1 <- fitted(mod1)
> resmod1 <- residuals(mod1,type="normalized")
> empresid1<-(Ovary$follicles-fitted(mod1))/3.11
> plot(empresid1,resmod1)
> abline(0,1)
> plot(Ovary$Time,Ovary$follicles,type="n")
> text(Ovary$Time,Ovary$follicles,
      labels=as.numeric(as.character(Ovary$Mare)))
> plot(Ovary$Time,fitmod1,type="n")
> text(Ovary$Time,fitmod1,
      labels=as.numeric(as.character(Ovary$Mare)))
> plot(Ovary$Time,resmod1,type="n")
> text(Ovary$Time,resmod1,
      labels=as.numeric(as.character(Ovary$Mare)))
> abline(0,0,lty=2)
> plot(fitmod1,resmod1,type="n")
> text(fitmod1,resmod1,
      labels=as.numeric(as.character(Ovary$Mare)))
> abline(0,0,lty=2)
> qqnorm(resmod1)

```

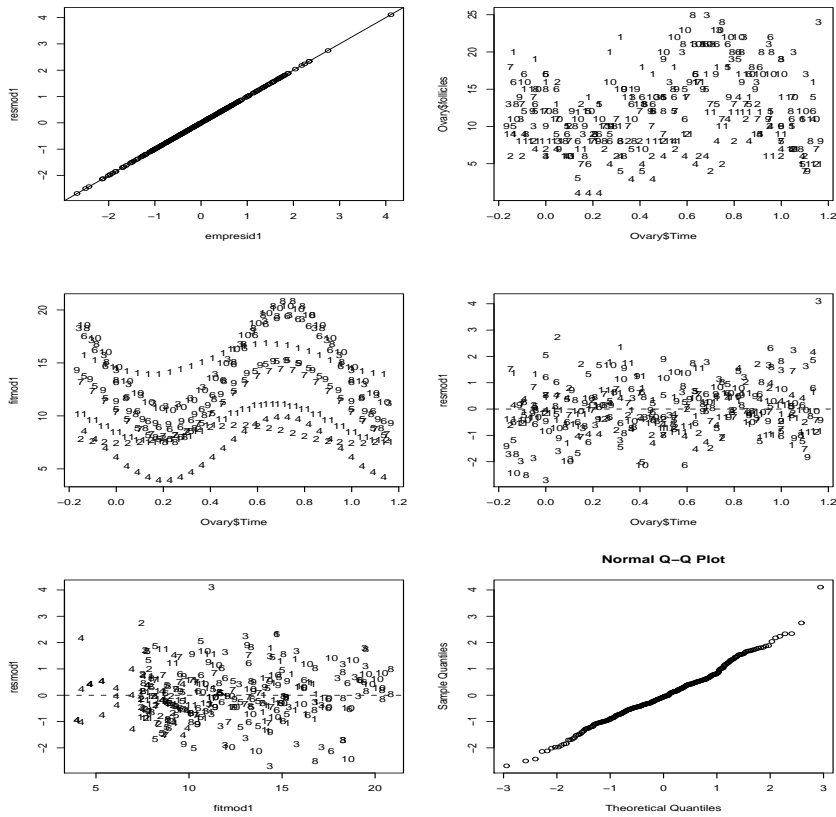


Figure 9: More diagnostics.

We attempt to fit an AR(1) correlation structure.

```
> mod3 <- lme( follicles ~ x1+x2,data=Ovary,random=pdDiag(~x1),
correlation=corAR1() )
> summary(mod3)
```

Linear mixed-effects model fit by REML

Data: Ovary
AIC BIC logLik
1564.445 1590.487 -775.2224
Random effects:
Formula: ~x1 | Mare
Structure: Diagonal
(Intercept) x1 Residual
StdDev: 2.807293 0.03630784 3.665217
Correlation Structure: AR(1)

Formula: ~1 | Mare
Parameter estimate(s):
Phi
0.6073908
Fixed effects: follicles ~ x1 + x2

	Value	Std.Error	DF	t-value	p-value
(Intercept)	12.189581	0.9454275	295	12.893194	0.0000
x1	-2.947343	0.5026645	295	-5.863440	0.0000
x2	-0.880711	0.5139787	295	-1.713517	0.0877

Number of Observations: 308
Number of Groups: 11

GEE

We now estimate the fixed effects using GEE.

```
> summary(geese(follicles ~ x1+x2,data=Ovary,id=Mare,
  corstr="independence"))
Call:
geese(formula = follicles ~ x1 + x2, id = Mare, data = Ovary,
  corstr = "independence")
Mean Model:
Mean Link:          identity
Variance to Mean Relation: gaussian
Coefficients:
              estimate   san.se     wald      p
(Intercept) 12.2155822 1.0101245 146.244165 0.000000e+00
x1          -3.3396116 0.6971878  22.945228 1.666835e-06
x2          -0.8697358 0.4092518   4.516412 3.357116e-02
Scale Model:
Scale Link:          identity
Estimated Scale Parameters:
              estimate   san.se     wald      p
(Intercept) 19.92926 3.444361 33.47843 7.205871e-09
Correlation Model:
Correlation Structure: independence
Returned Error Value: 0
Number of clusters: 11 Maximum cluster size: 31
```

```
> summary(geese(follicles ~ x1+x2,data=Ovary,id=Mare,
  corstr="ar1"))
Call:
geese(formula = follicles ~ x1 + x2, id = Mare, data = Ovary,
  corstr = "ar1")
Coefficients:
              estimate   san.se     wald      p
(Intercept) 12.2201663 1.0927344 125.061889 0.000000e+00
x1          -2.6919481 0.5513123  23.841740 1.045897e-06
x2          -0.9270288 0.5068586   3.345128 6.740424e-02
Estimated Scale Parameters:
              estimate   san.se     wald      p
(Intercept) 20.11901 3.466789 33.67893 6.500086e-09
Correlation Model:
Correlation Structure: ar1
Correlation Link:      identity
Estimated Correlation Parameters:
              estimate   san.se     wald      p
alpha 0.8989447 0.03670874 599.6902 0
Returned Error Value: 0
Number of clusters: 11 Maximum cluster size: 31

Inference seems relatively robust – which approach we use
depends on the aim of the study.

The AR working model seems more appropriate here (more
efficient estimation).

Might question appropriateness of sandwich estimation,
given  $m = 11$ .
```

Inference for the random effects

There are two principal targets of inference in the mixed models that we have considered:

- Inference for a population of units, defined through β .
- Inference for specific unit i , which requires inference for β and \mathbf{b}_i .

In this section we consider problems of the second type.

Examples:

- Pharmacokinetics: individualization of a profile.
- Dairy herds: genetic merit of a particular bull – data are in the form of the milk yields of his daughters.
- Psychology, inference for the IQ of an individual from a set of test scores.
- Industrial applications: operating characteristics of a particular machine.

Criteria for Inference

From a frequentist perspective, inference for random effects is often viewed as *prediction* rather than estimation, since \mathbf{b} are random variables.

The usual optimality criteria for a fixed effect θ , based upon unbiasedness:

$$E[\hat{\theta}] - \theta = \mathbf{0},$$

where θ is a fixed constant, and variance of estimator

$$\text{var}(\hat{\theta}),$$

are adjusted for a random effect \mathbf{b} . Specifically, an unbiased prediction is such that

$$E[\tilde{\mathbf{b}} - \mathbf{b}] = \mathbf{0},$$

to give

$$E[\tilde{\mathbf{b}}] = E[\mathbf{b}]$$

Variance is defined with respect to a fixed number, and to account for the variance of the random effects

$$\text{var}(\tilde{\mathbf{b}} - \mathbf{b}) = \text{var}(\tilde{\mathbf{b}}) + \text{var}(\mathbf{b}) - 2\text{cov}(\tilde{\mathbf{b}}, \mathbf{b})$$

is considered.