Forest Biometrics with Examples in R
# Contents

## I Introduction

### 1 Random variables

1.1 Introduction to random variables ........................................ 3
1.2 Univariate random variables ............................................... 6
   1.2.1 Distribution function and density ............................... 6
   1.2.2 Transformations of random variables ........................... 15
   1.2.3 Expected value .................................................. 18
   1.2.4 Variance and standard deviation ............................... 20
   1.2.5 Quantile function .............................................. 21
   1.2.6 Distribution of sums of independent random variables 26
   1.2.7 Weighted and censored distributions .......................... 28
   1.2.8 Transformations of a standard normal variable ............. 33
      1.2.8.1 $\chi^2$-distribution .................................. 33
      1.2.8.2 Student’s $t$-distribution .............................. 33
      1.2.8.3 $F$-distribution ......................................... 34
1.3 Multivariate random variables ......................................... 35
   1.3.1 Joint, conditional and marginal distributions ............... 35
   1.3.2 Independence, covariance and correlation ..................... 39
   1.3.3 Conditional expected value and variance ...................... 41
   1.3.4 Multivariate normal distribution ............................. 42
   1.3.5 Compound distributions ....................................... 50
   1.3.6 Computing using variance and covariance .................... 51

## 2 Statistical modeling and prediction

2.1 Concepts ............................................................... 53
2.2 Comparing alternative estimators .................................... 56
2.3 Prediction ............................................................. 57
   2.3.1 Linear prediction .............................................. 59
2.4 Testing and confidence intervals .................................... 60
   2.4.1 Multiple tests ................................................ 62
   2.4.2 Confidence interval ........................................... 62
3 Estimation using observed data

3.1 Estimating expected value ........................................... 65
3.2 Estimating variance and covariance ............................... 69
3.3 Estimating parameters of an assumed distribution: Maximum likelihood ........................................... 70

II Regression

4 Linear Model

4.1 Model formulation .................................................. 84
4.1.1 Model for an individual observation .......................... 84
4.1.2 The model for all data ........................................... 88
4.1.3 The systematic part .............................................. 89
4.1.4 The random part ................................................... 103
        4.1.4.1 Modeling non-constant error variance using a variance function .......................... 104
        4.1.4.2 Modeling dependence using autocorrelation functions ......................................... 105
        4.1.4.3 Implementing the variance-covariance structures in the model for all data .................. 106
4.2 Estimation ............................................................. 111
4.2.1 Ordinary least squares (OLS) .................................. 112
4.2.2 Generalized least squares (GLS) ............................... 117
4.2.3 Maximum Likelihood (ML) ........................................ 120
4.2.4 Restricted maximum likelihood (REML) ....................... 121
4.3 Does the model fit? .................................................. 126
4.3.1 Model residuals ................................................... 127
4.3.2 Graphical evaluation of the fit ................................ 129
4.3.3 Numerical criteria for model fit ............................... 138
4.4 Inference ............................................................. 143
4.4.1 Hypothesis tests of nested models ............................ 143
4.4.2 $F$-tests on regression coefficients $\beta$ ......................... 144
4.4.3 $t$-tests of regression coefficients $\beta$ ......................... 155
4.4.4 Likelihood Ratio - A general large-sample test procedure for nested models .......................... 158
4.4.5 Comparing non-nested models ................................ 160
4.4.6 Confidence intervals for model parameters ................... 161
4.5 Prediction ............................................................. 162

5 Linear mixed-effects models ............................................. 167

5.1 Model formulation ................................................... 171
5.1.1 The variance component model .................................. 171
5.1.2 Mixed-effects model with a random intercept 175  
5.1.3 Multiple random effects 179  
5.1.4 Random effects for categorical predictors 183  
5.1.5 Matrix formulation for single group 187  
5.1.6 Relaxing the assumptions on residual variance 191  
5.1.7 Matrix formulation for all data 193  

5.2 Estimation of model parameters and prediction of random effects 197  
5.2.1 Analysis of variance 198  
5.2.2 Maximum Likelihood 199  
5.2.3 Restricted maximum likelihood 204  
5.2.4 More about \( R \) 206  
5.2.5 Bayesian estimation 207  
5.2.6 Prediction of random effects 209  
5.2.7 Fixed or random group effect? 217  

5.3 Evaluation of model fit 220  
5.3.1 The systematic part 221  
5.3.2 The random part 222  
5.3.3 Coefficient of determination for mixed-effect models 229  

5.4 Tests and confidence intervals 230  
5.4.1 Hypothesis tests 230  
5.4.2 Confidence intervals 237  

5.5 Reporting a fitted mixed-effects model 239  

6 Nonlinear models 241  
6.1 Nonlinear fixed-effects model 241  
6.1.1 Model formulation 241  
6.1.2 Explaining variability in \( \phi \) through second-order predictors 246  
6.1.3 Parameter estimation and inference when \( \text{var}(e) = \sigma^2 I \) 249  
6.1.4 Parameter estimation and inference when \( \text{var}(e) = \sigma^2 V \) 256  

6.2 Nonlinear mixed-effects models 258  
6.2.1 Model formulation 258  
6.2.2 Parameter estimation and inference 263  
6.2.3 Convergence problems 276  
6.2.4 Multiple levels of grouping 276  
6.2.5 Prediction of random effects 278  
6.2.6 Principal components of random parameters 285  

6.3 Exercises 287  

7 GLM's and GLMM's 291  
7.1 Common parent distributions 293  
7.1.1 The normal distribution 293
8 Multivariate (Mixed-Effects) Models 345

8.1 Simultaneous equations 345
8.2 Why model systems 347
8.3 Seemingly unrelated regression models 350
  8.3.1 Model formulation 350
  8.3.2 Estimation 351
  8.3.3 Inference 357
  8.3.4 Prediction 357
8.4 Multivariate mixed-effect models 360
  8.4.1 Model formulation 360
  8.4.2 Estimation and inference 362
  8.4.3 Prediction 368
8.5 Exercises 373

9 Special topics on regression 375

9.1 Random regressors 382
  9.1.1 Connection between OLS linear regression and BLP 382
  9.1.2 Random regressors in a linear model 385
9.2 Modeling nonlinear responses with linear models 387
  9.2.1 Logarithmic regression 387
  9.2.2 Correcting back transformations bias using Taylor series 389
  9.2.3 Correcting back transformations bias using two-point distribution 390
  9.2.4 Regression splines 391
Part I

Introduction
Part II

Regression
The models discussed until now include only one response variable. However, sometimes several response variables are measured from the same sampling units. In many cases, it is sufficient just to formulate univariate models for each response variable separately. However, it is also possible to formulate the model jointly for all the response variables. The resulting model is called *model system, multivariate model* or *joint model*. Whether a multivariate model has some benefits over a separate univariate models depends largely on the research question at hand.

The individual equations of a model system may be *directly related, seemingly unrelated, or unrelated*. In directly related models, response of one model may be the predictor in another model. In seemingly unrelated model, direct relationships do not exist between the individual models. However, the models may be related through the dependence of residual errors and random effects. In unrelated models, no direct relationship exists between the models, and the residual errors and random effects are independent across models.

### 8.1 Simultaneous equations

Multivariate models where the equations are directly related are called *simultaneous equation models* or *structural equation models*. They are useful in econometrics or in social sciences for studying causal paths among several dependent variables. For example, the so called Klein’s small macroeconomic model includes equations for consumption, demand and investment (Greene, 2003). In forestry, model systems are sometimes used, for example, in developing a forest growth model (e.g. Hasenauer et al., 1998).

In simultaneous equations the dependent variables are called *endogenous variables* which are determined within the model. *Exogenous variables* correspond to the regressor or explanatory variables in ordinary regression models. In addition the model contains *disturbance variables* or *equation errors*.

A simultaneous equation model for observation \( i \) can be presented in a matrix equation:

\[
\mathbf{By}_i + \Gamma \mathbf{x}_i = \mathbf{e}
\]
Ones appear on the diagonal of $B$. This means that $j^{th}$ row describes how $y_{ij}$ depends on other $y$ variables and $x$ variables. There are no intercepts in the model, i.e., both $x$ and $y$ variables are centered. Generally the disturbances of different equations can be correlated. The disturbances are assumed to be uncorrelated with $x$ variables. If parameters of $j^{th}$ row are estimated with OLS, the estimates are not consistent because the error term is correlated with other $y$ variables which contain error which is correlated with the error term of $j^{th}$ row. The estimation can be done using such $x$ variables which do not appear on the $j^{th}$ row as instrumental variables (see Greene, 2003). The identification of the model requires that there are enough zeros in matrix $B$ and in the covariance matrix of $e$.

We are in this book interested in models which can be used for prediction. If $y$-variables are not known in an application, all the predictive power in a simultaneous equation models comes from the $x$ variables. It is then much more straightforward to do predictions using ordinary (seemingly unrelated, see Section 8.3) regression models. Also even if some $y$-variables are known in a multivariate model, the prediction can utilize these values, as presented in Sections 8.3.4 and 8.4.3. That is why we do not present simultaneous equation models more closely.

However, the simultaneous equation idea can be applied when specifying an ordinary regression equation. Assume that we know that

$$ y = f(x,z) + e $$

is an reasonable model for predicting $y$ when $x$ and $z$ are known. If $z$ is not known, assume that it can be predicted using equation

$$ z = g(x) + u $$

After estimating parameters of $f$ and $g$, then $y$ can be predicted as

$$ \hat{y} = \hat{f}(x, \hat{g}(x)) $$

However, it is better to estimate directly $y$ as a function of $x$ using model

$$ y = f(x, g(x)) + e. $$

No advantage can be obtained using the two-stage procedure. The simultaneous equation idea can just used to specify a reasonable model in terms of the exogenous variables.

Example 8.1. Let $H$ be tree height, $V$ tree volume and $D$ tree diameter. Assume that one wants to model the standing tree volume.

If the model is to be used for prediction in a situation where both height and diameter are known for the target trees, then a univariate model of form

$$ V = f(D, H) + e $$

for the volume is sufficient, where the true height and diameter are used as predictors.
Multivariate (Mixed-Effects) Models

If the model is to be used for prediction in a situation where only the diameter of the target tree is known, then one might be tempted to fit a model system of form

\[ V = f_1(D, H) + e_1 \]  \hspace{1cm} (8.1)
\[ H = f_2(D) + e_2 \]  \hspace{1cm} (8.2)

In the prediction situation, the unknown \( H \) in (8.1) would be replaced by prediction \( \hat{H} \) based on the fitted model (8.2). However, \( \hat{H} \) is just a transformation of \( D \), an estimate of \( E(H|D) \). Especially, it is a different random variable than \( H \), with same expected value (if estimation errors are ignored) but smaller variance. Therefore, the prediction of \( V \) using \( \hat{H} \) has less variability, leading to larger prediction variance than the estimated residual standard error of model (8.1) would suggest. Furthermore, the prediction would also be biased even if the systematic part of model (8.1) would give the correct expectation of \( E(V|D, H) \), unless the volume model (8.1) is unrealistically a linear function of \( H \).

A better multivariate solution is to replace model (8.1) by

\[ V = f_1(D, \hat{H}) + e_1 \]  \hspace{1cm} (8.3)

where predicted height from model (8.2) is now used as predictor in (8.3) in addition to \( D \). If model (8.3) is formulated and estimated so that the systematic part gives the correct expectation of \( E(V|D, \hat{H}) \), the model would be unbiased and give lower prediction variance than model (8.1).

However, the role of the height model is only to specify such a transformation of the diameter that describes the height-diameter relationship well. Therefore, the model system could be replaced with a univariate model

\[ V = f_3(D) + e_3 \]

where the function \( f_3(D) \) is formulated so that it fits well to the volume-diameter relationship. The knowledge about the well-fitting height-diameter relationship could be used in formulation of the model, but no parameters of that model need not to be fixed based on a fitted height-diameter model. This approach would lead to even better fitting prediction error variance than model system based on equations (8.3) and (8.2).

### 8.2 Why model systems

In the next two subsections, we will discuss the seemingly unrelated linear models and linear mixed-effect models. In those models, the most important difference to separately fitted models is the estimation of the cross-model correlation of residual errors and random effects. The estimated cross-model correlation may improve the analysis because

1. The fixed parameters of the model may be better estimated.
2. Prediction of the models will be more efficient, if the response of one of the component models have been observed.

3. A more realistic simulations could be obtained from the models system.

4. Hypotheses about the effects of a predictor on several response variables can be tested.

The first item seldom causes any big improvement to the efficiency, as we will discuss later in Sections 8.3.2 and 8.4.2.

Even though the information on between-model correlation would not change the estimates of fixed effects at all, utilizing a known correlation would lead to much better and more realistic results in prediction. The prediction from SUR model may be highly more efficient than from separately estimated models if the residuals of the models are correlated, and if observed response of one component model can be used in prediction. One of the earliest applications is Lappi (1991), which developed a bivariate mixed-effect model system for tree volumes and heights as a function of diameter. He proposed that the volume model could be calibrated using measured heights for improved prediction of standing tree volume.

An even more early applications can be found in Lappi (1986) in the context of taper curves; the method was further developed and simplified in Lappi (2006). The system included a total of 13 equations: 12 equations for the stem diameter at fixed absolute or relative heights and an additional equation for tree height. The model expressed the expected value for each model conditional on tree diameter. Furthermore, the variance-covariance structure among the model residuals was modeled as a function of tree diameter. In applications, the cross-model correlations of the models are utilized to predict the residuals (and stand-level random effects) of all models by using typically one or two measurements of one tree, such as DBH, height and upper stem diameter. Interpolation is used to generalize the expected values, variances and covariances from fixed relative heights to the relative heights corresponding to the actual measurements.

A similar idea was utilized also in Mehtätalo (2005) for diameter distributions. A system of 12 models for the percentiles of stand-specific diameter distribution for fixed percentages was available from Kangas and Maltamo (2000b). Sample quantiles, such as the smallest tree of a plot, or e.g. 3rd smallest tree of a plot with a total of 7 trees, were interpreted as a measured sample percentile from the stand. The expected values, variances and covariances of the model system were interpolated for the percentage corresponding to the measurement, and the measured diameter was used to predict the residual errors for all all models of the system in the target stand.

Siipilehto (2011) developed a system of models for 20 stand-level variables, some of which were important stand characteristics such as basal area, mean diameter and dominant height, whereas others were parameters of models that describe tree allometry and stand structure. Measurements of some
stand variables, typically the basal area and mean diameter, can then be used to improve the prediction of all variables of the system, which is utilized in a stand-level forest planning tool. The benefit of the system over a set of predictive regression models is flexibility in the use of measurements: any set of measurements (including the empty set of no measurements) can be used to produce the prediction. The more variables have been measured, the more accurate are the predictions. The approach illustrates that the estimates of mean and the variance-covariance matrix of the variables includes all information that is needed to produce OLS-estimates of the regression coefficients of any regression model based on the available variables. In old days the regression coefficients were usually estimated this way (Yule and Kendall, 1958). Once the means and the variance-covariance matrix of the variables had been computed, all possible regression models could be estimated using minimal computing effort, as we described in Section 9.1.1.

Other prediction applications with similar ideas are Eerikäinen (2009) for generalizing tree heights; the different responses were tree species in mixture stands. de Souza Vismara et al. (2016) developed a bivariate model for individual tree volumes in eucalyptus plantations; they had two responses which were associated with two rotations of the plantation. We will demonstrate the cross-model calibration in Examples 8.4 and 8.5.

A realistic description of the dependence structure is also important if models are used for simulation (Leskinen et al., 2009). Simulation is used, for example, in growth simulators, which commonly comprise of several models for different natural processes (such as ingrowth, diameter growth and mortality) or different structural components (such as stemwood, branch, needle and root biomass). In such a situation, it is extremely important to take into account the dependencies of model predictions to avoid unrealistic simulation results and to estimate the prediction errors of the output. Thus, knowledge of the interrelationships between-models is extremely important in simulation. It may be possible that an assumed constant correlation is not sufficiently realistic and much more detailed assumptions are needed. In such situations, a realistic model for the dependency structure may be far more important than the effective estimation of regression coefficients. Furthermore, it is not only the variance-covariance structure that matters, but whole distribution of errors may have an effect on the results.

The fourth benefit from joint estimation is the possibility to test hypotheses of form “Does predictor $x$ have an effect in some of the responses” against the null hypothesis “Predictor $x$ does not have an effect on any of the response variables”. We will discuss this issue in more detail in the context of fixed-effect models in Section 8.3.3. An example will be presented in Example 8.3. Similar benefits occur in the case of seemingly unrelated mixed-effects models as well. For more discussion about hypothesis tests of model systems, see Greene (2003).
8.3 Seemingly unrelated regression models

8.3.1 Model formulation

The seemingly unrelated models do not have such a direct relationship as the directly related models do. However, they are related to each other because the residuals are correlated. Let us consider an example where tree height and volume are both modeled on tree diameter:

\[
\begin{align*}
\ln h_i &= \beta_1^{(h)} + \beta_2^{(h)} \ln d_i + e_i^{(h)} \quad (8.4) \\
\ln v_i &= \beta_1^{(v)} + \beta_2^{(v)} \ln d_i + e_i^{(v)},
\end{align*}
\]

where \(d_i\) is tree diameter, \(\beta_1^{(h)}, \beta_2^{(h)}, \beta_1^{(v)},\) and \(\beta_2^{(v)}\) are parameters to be estimated, and \(e_i^{(h)}\) and \(e_i^{(v)}\) are the residuals of the models. It is assumed that the residuals have mean zero and variances \(\text{var}(e_i^{(h)}) = \sigma_h^2\) and \(\text{var}(e_i^{(v)}) = \sigma_v^2\).

The residuals within both models are uncorrelated: \(\text{cov}(e_i^{(h)}, e_i^{(v)}) = 0\) and \(\text{cov}(e_i^{(v)}, e_{i'}^{(v)}) = 0\) for \(i \neq i'\). These assumptions would also be made if the two models were estimated separately.

The difference to separate models is in the assumption about covariance across models, where we assume \(\text{cov}(e_i^{(v)}, e_i^{(h)}) = \varphi\), where \(\varphi\) is the cross-model covariance to be estimated. For example, it might be realistic to assume that if height is higher than expected by the systematic part of the model, then also the volume is higher than predicted by the systematic part. In a SUR model, we allow correlation among responses. This fixed, constant correlation is assumed to be common to all observations.

To formulate the model in a matrix form, let us consider the following \(M\) individual linear models,

\[
\begin{align*}
y_1 &= X_1 \beta_1 + e_1 \\
y_2 &= X_2 \beta_2 + e_2 \\
&\vdots \\
y_M &= X_M \beta_M + e_M .
\end{align*}
\]

where \(E(e_m) = 0\), \(\text{var}(e_m) = \sigma_m^2 I\) for \(m = 1, \ldots, M\) and \(\text{cov}(e_m, e_{m'}) = \varphi_{mm'} I\) for every pair \(m \neq m'\).
By defining

\[ y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix}, \quad X = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_M \end{bmatrix} \]

\[ e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_M \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_M \end{bmatrix} \]

we can write the model system in the form of univariate linear model (4.6) as

\[ y = X \beta + e, \]

where the variance-covariance matrix of residual errors is

\[ \text{var}(e) = \begin{bmatrix} \sigma_1^2 I & \varphi_{12} I & \cdots & \varphi_{1M} I \\ \varphi_{12} I & \sigma_2^2 I & \cdots & \varphi_{2M} I \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{1M} I & \varphi_{2M} I & \cdots & \sigma_M^2 I \end{bmatrix} \]

\[ = \begin{bmatrix} \sigma_1^2 & \varphi_{12} & \cdots & \varphi_{1M} \\ \varphi_{12} & \sigma_2^2 & \cdots & \varphi_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{1M} & \varphi_{2M} & \cdots & \sigma_M^2 \end{bmatrix} \otimes I = \Sigma \otimes I. \]

Here terms \( \sigma_1^2, \ldots, \sigma_M^2 \) are the error variances of each component and \( \varphi_{mm'} \) is the cross-model covariance between models \( m \) and \( m' \).

It is straightforward to extend the model formulation to situation where the diagonal elements of the response-specific variance-covariance matrices are used to model the heteroscedastic error variance. In that case it is common to allow the cross-model covariances vary in such a way that the cross-model correlation for each pair of models is constant.

### 8.3.2 Estimation

As the above specification showed that the SUR model is a special case of the linear model, the methods presented in Section 4 for estimation and inference apply. Thus, the estimation involves first estimating the parameters specifying matrix \( W = \text{var}(e) \) using REML, and then \( \beta \) using GLS. Note that even though matrix \( W \) may be large, its inverse is easy to find as \( \Sigma^{-1} \otimes I \).

It is also possible to use the iteratively reweighted least squares:

1. Estimate \( \beta \) by fitting each model separately using OLS.
2. Estimate $W$ as the empirical variance-covariance matrix of the observed residuals of the current model.

3. Re-estimate the model parameters using GLS.

4. Repeat steps (2) and (3) until the estimates do not change.

Usually the estimation is done without step 4, as suggested in the classical paper of Zellner (1962).

If the same predictor variables are used in all equations of the system, or if the cross-model correlations are zero, then the parameter estimates obtained by SUR fit are exactly same as one would bet from separate fits. Their standard errors are the same as well. If different predictor variables are used and cross-model correlation exists, then the estimates differ and the estimates from SUR fit have lower variance than the estimates from separate fits. However, if the models are formulated well and all important predictors are included in all models, the differences in the estimates of regression coefficients between jointly and separately fitted models should not be very large. If the differences are large, it means that some of the equations includes a predictor that can explain the variation in also some of those responses where that predictor is not used. In such cases, the formulation of the fixed part should be reconsidered and updated. Because the improvement in the estimates of $\beta$ from fitting models jointly is not large, an acceptable strategy for model fitting is to estimate $\beta$ from separate fits and thereafter estimate the cross-model correlation.

The correlation can be estimated as the empirical correlation of the model residuals, as it was done in step 2 of the iterative procedure above. It can also be estimated by utilizing a model for the sum of the two variables. Consider a model with two response variables, $y_1$ and $y_2$. Start by fitting the univariate models for both responses, and denote the residual standard errors from these models as $\hat{\sigma}_1$ and $\hat{\sigma}_2$. Then fit also a model for the sum $y_1 + y_2$, resulting to estimated residual standard error $\hat{\sigma}_{12}$. The cross-model covariance can now be estimated using equations $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2 \text{cov}(X + Y)$ as

$$\text{cov}(X,Y) = \frac{\hat{\sigma}_{12}^2 - \hat{\sigma}_1^2 - \hat{\sigma}_2^2}{2}.$$ 

In R, estimation of seemingly unrelated regression models using the iterative procedure presented above can be done by using function `systemfit::systemfit`; by default the function does only steps 1-3, but iterative procedure can be conducted by argument `maxit`. Function `nlme::gls` allows such specification of the variance-covariance matrix of the multivariate model through arguments `weights` and `cor`.

**Example 8.2.** Consider data set `stumplift`. The processing time $t_i$ of a stump was previously explained by machine and stump diameter in the examples of Chapter 4. The data includes also variable "productivity", denoted by $p_i$. Let us
consider the following models for these two variables:

\[ t_i = \beta_1^{(1)} + \beta_2^{(1)} d_i + \beta_3^{(1)} M2_i + \beta_4^{(1)} M2_i d_i + \beta_5^{(1)} M3_i d_i + e_i^{(1)} \]

\[ p_i = \beta_1^{(p)} + \beta_2^{(p)} d_i + \beta_3^{(p)} M2_i + \beta_4^{(p)} M2_i d_i + \beta_5^{(p)} M3_i d_i + e_i^{(p)} \]

These models are fitted first separately using OLS. The empirical correlation among the residuals is negative and rather high.

```r
modT<-gls(Time~Machine*Diameter,data=stumplift)
modP<-gls(Productivity~Machine*Diameter,data=stumplift)
cor(resid(modP),resid(modT))
[1] -0.7345888
```

Exactly same estimate of cross-model correlation is obtained by utilizing the model for the sum of processing time and productivity:

```r
modTplusP<-gls(I(Productivity+Time)~Machine*Diameter,data=stumplift)
covTP<-(modTplusP$sigma^2-modT$sigma^2-modP$sigma^2)/2
covTP/(modT$sigma*modP$sigma)
[1] -0.7345888
```

The joint model can be fitted using `systemfit`:

```r
library(systemfit)
eqT <- Time~Machine*Diameter
eqP <- Productivity~Machine*Diameter
system <- list( Time = eqT, Productivity = eqP)
modSUR<-systemfit(system,"SUR",data=stumplift,maxit=10)
summary(modSUR)
```

The correlations of the residuals:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>Productivity</td>
</tr>
<tr>
<td>1.0000</td>
<td>-0.734589</td>
</tr>
</tbody>
</table>

SUR estimates for 'Time' (equation 1):

|                | Estimate | Std. Error | t value | Pr(>|t|) |
|----------------|----------|------------|---------|---------|
| (Intercept)    | -9.509293| 7.723040   | -1.2313 | 0.2188 |
| Machine2       | -44.931988| 10.771147  | -4.1715 | 3.5930e-05*** |
| Machine3       | -43.777689| 13.110970  | -3.3390 | 0.00090634*** |
| Diameter       | 1.621759  | 0.218281   | 7.4297  | 5.0360e-13*** |
| Machine2:Diameter | 2.334425 | 0.300030   | 7.7806  | 4.4631e-14*** |
| Machine3:Diameter | 1.907349 | 0.367602   | 5.1863  | 3.1340e-07*** |

Residual standard error: 26.203231 on 479 degrees of freedom
Number of observations: 485 Degrees of Freedom: 479

SUR estimates for 'Productivity' (equation 2):

|                | Estimate | Std. Error | t value | Pr(>|t|) |
|----------------|----------|------------|---------|---------|
| (Intercept)    | -0.7488020| 0.934094   | -0.8017 | 0.4231 |
| Machine2       | 2.2378383 | 1.306645   | 1.7179  | 0.0845 |
| Machine3       | 2.6487354 | 1.585615   | 1.6704  | 0.0954 |
| Diameter       | 0.3360912 | 0.026395   | 12.7314 | < 2.2e-16*** |
| Machine2:Diameter | -0.1941110| 0.0362850  | -5.4961 | 1.3660e-07*** |
| Machine3:Diameter | -0.1707520| 0.0444570  | -3.8063 | 0.001391 *** |

Residual standard error: 3.168968 on 479 degrees of freedom
Number of observations: 485 Degrees of Freedom: 479

Multiple R-Squared: 0.61 Adjusted R-Squared: 0.605299

SUR estimates for 'Productivity' (equation 2):

|                | Estimate | Std. Error | t value | Pr(>|t|) |
|----------------|----------|------------|---------|---------|
| (Intercept)    | -0.7488020| 0.934094   | -0.8017 | 0.4231 |
| Machine2       | 2.2378383 | 1.306645   | 1.7179  | 0.0845 |
| Machine3       | 2.6487354 | 1.585615   | 1.6704  | 0.0954 |
| Diameter       | 0.3360912 | 0.026395   | 12.7314 | < 2.2e-16*** |
| Machine2:Diameter | -0.1941110| 0.0362850  | -5.4961 | 1.3660e-07*** |
| Machine3:Diameter | -0.1707520| 0.0444570  | -3.8063 | 0.001391 *** |

Residual standard error: 3.168968 on 479 degrees of freedom
Number of observations: 485 Degrees of Freedom: 479

Multiple R-Squared: 0.428163 Adjusted R-Squared: 0.422194

```
Also `systemfit` gave the same estimate of cross-model correlation that was computed earlier by using residuals of separate OLS fits. Also the parameter estimates are the same as in separate OLS fit, as one can see by comparing the coefficients for processing time below to those from the fit using `systemfit`.

```r
> summary(modT)

Value Std.Error  t-value p-value
(Intercept) -9.50929 7.723040 -1.231289 0.2188
Machine2 -44.93199 10.771147 -4.171514 0.0000
Machine3 -43.77769 13.110970 -3.339012 0.0009
Diameter  2.33442  0.300030  7.780640 0.0000
Machine2:Diameter  1.90735  0.367602  5.188633 0.0000
Machine3:Diameter  1.90735  0.367602  5.188633 0.0000
```

We could also fit the model by using function `nlme::gls`. The script below shows the necessary scripts.

```r
> library(nlme)
> mvmod1<-gls(y~-1+IT+M2T+M3T+DT+M2DT+M3DT+IP+M2P+M3P+DP+M2DP+M3DP,
+ weights=varIdent(form = ~1|response),
+ corr=corSymm(form = ~1 | obs),
+ data=mvdat)
> summary(mvmod1)
```

Correlation Structure: General
Formula: ~1 | obs
Parameter estimate(s):
Correlation:
```

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.735</td>
</tr>
</tbody>
</table>
```

Variance function:
Structure: Different standard deviations per stratum
Formula: ~1 | response
Parameter estimates:
The model output is practically identical to that of modSUR. The residual variance for the processing time is \( \hat{\sigma}_t^2 = 26.20^2 \). For productivity, it is \( \hat{\sigma}_p^2 = (26.20 \times 0.1209)^2 = 3.169^2 \). In our example, the fixed parts of both models included the same terms. If, for example, the model for processing time would also include second-order terms, then the parameter estimates from SUR would differ from those based on separate fits.

We already know from Example 4.10 that the variance of processing time increases as a function of fitted value. This is also the case with productivity. Therefore, it might be realistic to assume that

\[
\text{var}(e_i^{(r)}) = \sigma^2_t |\hat{y}_i|^{2\delta_t}
\]

where \( \delta_t \) is the power parameter specified separately for the two responses \( r \in \{t, p\} \) and \( \sigma^2_t \) is correspondingly the scaling factor of the variance function. To fit a model with this assumption, we first fit separate models for both responses to
estimate the power parameters. The estimates were $\hat{\delta}_t = 0.933$ and $\hat{\delta}_p = 1.096$. We use these estimates and predicted values from model `mvmod1` to compute the value of $\hat{y}_i^{\hat{\delta}_r}$ to variable z of data set `mvdat`. Thereafter we specify the variance function as a product of variance functions `varIdent` and `varFixed` using `varComb` to re-estimate the multivariate model, see details about these functions in Pinheiro and Bates (2000, Section 5.2). Figure 8.1 shows that the variance is well modeled by using this approach. The parameter estimates of the fitted model now differ from the previous estimates because of modeling the residual variance. For example, the estimated cross-model correlation is now even stronger than previously.

```r
> # Fit univariate models using varPower
> modt<-gls(Time~Machine*Diameter,
+ weights=varPower(),
+ data=stumplift)
> modp<-gls(Productivity~Machine*Diameter,
+ weights=varPower(),
+ data=stumplift)
> # Save the estimated response-specific variance function power to mvdat
> mvdat$delta<-attributes(modt$apVar)$Pars[1]
> mvdat$delta[mvdat$response==2]<-attributes(modp$apVar)$Pars[1]
> mvdat$z<-fitted(mvmod1)^mvdat$delta
+ weight=varComb(varIdent(form = ~1|response),varFixed(~z)),
+ corr=corSymm(form = ~1 | obs),
+ data=mvdat)
> # left subplot of the residual plot
> plot(fitted(mvmod2)[mvdat$response==1],resid(mvmod2,type="p")[mvdat$response==1],
+ xlab="Predicted processing time, s",ylab="Pearson residual")
> mywhiskers(fitted(mvmod2)[mvdat$response==1],resid(mvmod2,type="p")[mvdat$response==1],
+ add=TRUE,se=FALSE)

> summary(mvmod2)

Call: gls(formula = y ~ -1 + IT + M2T + M3T + DT + M2DT + M3DT + IP + M2P + M3P + DP + M2DP + M3DP,

weights = varComb(varIdent(form = ~1|response),
varFixed(~z)),

corr = corSymm(form = ~1 | obs),

data = mvdat)

Coefficients:
                  Value Std.Error t-value p-value
IT -13.921357  4.200570 -3.314159 0.0010
MGT -12.552604  7.511373 -1.671146 0.0950
MGT -22.174553  7.463255 -2.971164 0.0030
DT  1.756717  0.144440 12.162267 0.0000
M2DT  1.370425  0.262268  5.225289 0.0000
M3DT  1.272717  0.253834  5.013972 0.0000
IP -1.072897  1.006550 -1.065916 0.2867
M2P  1.289585  1.167601  1.104475 0.2697
M3P  1.667803  1.382803  1.206104 0.2281
DP  0.340526  0.037138  9.262470 0.0000
M2DP -0.167217  0.039239 -4.261541 0.0000

Correlation Structure: General
Formula: ~1 | obs
Parameter estimate(s):
Correlation:
   1
   2 -0.808
Combination of variance functions:
Structure: Different standard deviations per stratum
Formula: ~1 | response
Parameter estimates:
   1
   2
1.000000 0.623679
Variance function:
Structure: fixed weights
Formula: ~z

Coefficients:
                  Value Std.Error t-value p-value
IT -13.921357  4.200570 -3.314159 0.0010
MGT -12.552604  7.511373 -1.671146 0.0950
MGT -22.174553  7.463255 -2.971164 0.0030
DT  1.756717  0.144440 12.162267 0.0000
M2DT  1.370425  0.262268  5.225289 0.0000
M3DT  1.272717  0.253834  5.013972 0.0000
IP -1.072897  1.006550 -1.065916 0.2867
M2P  1.289585  1.167601  1.104475 0.2697
M3P  1.667803  1.382803  1.206104 0.2281
DP  0.340526  0.037138  9.262470 0.0000
M2DP -0.167217  0.039239 -4.261541 0.0000
```
8.3.3 Inference

The inference on a seemingly unrelated system of equations is based on the same procedures than in the case of univariate model. However, an important difference is that the number of observations in the model where multivariate response is modeled using univariate model is \( nM \). Therefore, the denominator degrees of freedom become much higher than in univariate model and the test results for individual coefficient are different compared to univariate models even though the parameter estimates and their standard errors are the same. The difference may be substantial if \( M \) is large and \( n \) is small. We suggest that if inference is made on coefficients of one equation only, it is better to use the univariate model. However, the model system allows also inference on parameters of several models simultaneously. This allows e.g. hypothesis tests on whether a certain predictor has significant effect on some of the response variables. If that hypothesis shows significant effect, then one could continue to test the effect separately for each response variable.

**Example 8.3.** To test whether there are differences between the machines in either processing time or productivity, we could consider hypothesis:

\[
H_0 : \beta_3^{(i)} = \beta_4^{(i)} = \beta_5^{(i)} = \beta_3^{(p)} = \beta_4^{(p)} = \beta_5^{(p)} = 0
\]

Such a hypothesis is tested below by using the conditional F-test for model `mvmod2`.

```r
> anova(mvmod2, Terms=c("M2T","M3T","M2DT","M3DT","M2P","M3P","M2DP","M3DP"))
Denom. DF: 958
F-test for: M2T, M3T, M2DT, M3DT, M2P, M3P, M2DP, M3DP
numDF F-value p-value
1  8  27.23083  <.0001
```

The low p-value indicates that there are differences between machines in either processing time or productivity. We could now progress further to test separately whether the Machine has effect on processing time or productivity by using separately fitted models for these responses. If the p-value from the above test were so high that we would not reject the null hypothesis, then we would not make any further tests about the effect of Machine. This way we could dramatically reduce the number of tests needed, thus protecting ourselves against the multiple testing problem. This approach was used e.g. in ? to test the effect of different treatments on 20 chemical compounds in birch leaves.

8.3.4 Prediction

Prediction from a SUR model that utilizes only the systematic part of the model is very similar to prediction from any other linear model. We just write the values of the predictors to the model and make prediction. The prediction intervals are calculated in very similar manner than with linear models.
However, the prediction can be improved if the value of one or more responses of the model system have been observed from a sampling unit, and other responses are being predicted for that unit. In that case, the estimated cross-model correlation can be used to carry information from one model to another to predict the residual error for the unobserved response. The prediction is a straightforward application of the linear predictor of section 2.3.1.

Assume that the response variables of models $m, \ldots, M$ have been observed from sampling unit, and the other responses are to be predicted for that sampling unit. Let $y_1, \ldots, y_M$ be the response vector for the sampling unit in question and $x_1, \ldots, x_M$ the vectors of predictors for models $1, \ldots, M$. We define

$$h_1 = \begin{bmatrix} y_1 \\ \vdots \\ y_{m-1} \end{bmatrix}, \quad h_2 = \begin{bmatrix} y_m \\ \vdots \\ y_M \end{bmatrix}$$

So that

$$\mu_1 = \begin{bmatrix} x_1' \beta_1 \\ x_2' \beta_2 \\ \vdots \\ x_{m-1}' \beta_{m-1} \end{bmatrix}, \quad \mu_2 = \begin{bmatrix} x_m' \beta_m \\ x_{m+1}' \beta_{m+1} \\ \vdots \\ x_M' \beta_M \end{bmatrix}.$$

The required variance-covariance-matrices are blocks of matrix $\Sigma$ of Equation (??):

$$V_1 = \begin{bmatrix} \sigma_1^2 & \varphi_{1,2} & \cdots & \varphi_{1,m-1} \\ \varphi_{1,2} & \sigma_2^2 & \cdots & \varphi_{2,m-1} \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{1,m-1} & \varphi_{2,m-1} & \cdots & \sigma_{m-1}^2 \end{bmatrix},$$

$$V_2 = \begin{bmatrix} \sigma_m^2 & \varphi_{m,m+1} & \cdots & \varphi_{m,M} \\ \varphi_{m,m+1} & \sigma_{m+1}^2 & \cdots & \varphi_{m+1,M} \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{m,M} & \varphi_{m+1,M} & \cdots & \sigma_M^2 \end{bmatrix},$$

$$V_{12} = \begin{bmatrix} \varphi_{1,m} & \varphi_{1,m+1} & \cdots & \varphi_{1,M} \\ \varphi_{2,m} & \varphi_{2,m+1} & \cdots & \varphi_{m+1,M} \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{m-1,m} & \varphi_{m-1,m+1} & \cdots & \varphi_{m-1,M} \end{bmatrix}.$$

By replacing the parameters in the above equations by their estimates, the EBLUP of $h_1$ can be computed using Formula (2.15) in Section 2.3.1. The prediction variance that ignores the estimation errors of parameters can be estimated using Equation (2.16). The estimation errors of fixed effects can be taken into account through formulas presented in Section 4.5. A complete
solution would also take into account the correlation between the estimation errors of fixed effect and prediction errors of residuals. However, the effect of estimation errors of fixed effects is usually small compared to the prediction error of residuals if the modeling data is sufficiently large.

**Example 8.4.** We illustrate the prediction using the system `mvmod2` of two models, which was previously fitted in Example 8.2. Assume that a prediction of processing time is made for a 40-cm stump processed using machine 1. The conventional prediction utilizes only the information about the fixed effects. The predicted processing time is 56 seconds, and the 95% prediction interval, based on normality, is 32 – 80 seconds. The required computations are shown below.

Let us then consider a hypothetical case where the productivity is known to be 17.27 m$^3$/h, but the processing time is unknown. This might be interesting e.g. in some imputation situation. However, this example should be taken more as a technical illustration of the procedure, which was applied e.g. by Lappi (2006), Miettitalo (2005) and Siipilehto (2011) in more realistic situations.

The observed productivity is higher than the expected productivity based on the systematic part of the model (12.79 m$^3$/h). Because processing time is negatively correlated with productivity, this means that the processing time should therefore be shorter than expected by the fixed part only (56.35 s). The script below defines the required components for prediction; we use general matrix definitions for wider applicability of the script, even though objects $\mu_1$, $\mu_2$, $V_1$, $V_2$, and $V_{12}$ are all scalars for this two-model system. The predicted value of the processing time is 41.06 seconds, and the 95% prediction interval is much shorter than previously: 17 – 65 seconds. Our data includes this example tree (tree id 51) for which the true processing time was 41.4 seconds. This value falls well within our prediction interval.
8.4 Multivariate mixed-effect models

8.4.1 Model formulation

Seemingly unrelated mixed-effects are such extension of the SUR model to grouped data, where group effects are modeled as random effects. Assume that we have $M$ individual mixed-effects models for group $i$:

$$
y_{i}^{(1)} = X_{i}^{(1)}\beta_{1} + Z_{i}^{(1)}b_{i}^{(1)} + \epsilon_{i}^{(1)}
$$

$$
y_{i}^{(2)} = X_{i}^{(2)}\beta_{2} + Z_{i}^{(2)}b_{i}^{(2)} + \epsilon_{i}^{(2)}
$$

$$
\vdots
$$

$$
y_{i}^{(M)} = X_{i}^{(M)}\beta_{M} + Z_{i}^{(M)}b_{i}^{(M)} + \epsilon_{i}^{(M)}
$$

where $\text{var}(\epsilon_{i}^{(m)}) = \sigma_{m}^{2}I_{n_{i} \times n_{i}}$ and $\text{var}(b_{i}^{(m)}) = D_{s}^{(m)}$ for $m = 1, \ldots, M$, recall Section 5.1.5. The random effects may be correlated across the responses, i.e., $\text{cov}(b_{i}^{(m)}, b_{i}^{(m')}) = C_{i}^{(mm')}$ may be nonzero. We assume that the cross-model correlation is the same for all groups, and denote therefore $C_{i}^{(mm')} = C_{i}^{(mm')}$.

The cross-model covariance of the random effects models the joint behaviour of group-specific curves. For example, if the models include random intercepts, so that $Z_{i}^{(m)}b_{i}^{(m)} = 1b_{i}^{(m)}$ and $C_{i}^{(mm')}$ is scalar, a positive cross-model covariance between the models $m$ and $m'$ means that for groups where the group-specific curve is above the curve of a typical group for response $m$, the group-specific curve for response $m'$ is usually above the curve of a typical group, too.

In addition to the cross-model correlation of random effects, also the residual errors may be correlated across models. That is, we may assume that $\text{cov}(\epsilon_{ij}^{(m)}, \epsilon_{ij}^{(m')}) = \varphi_{mm'}$ is nonzero. For example, a positive correlation between the residual errors of models $m$ and $m'$ means that for a individual sampling unit $j$ within group $i$, a positive residual error of model $m$ is usually associated with a positive residual error of model $m'$. However, the same individual sampling units are not necessarily observed for all responses, and the cross-model correlation for the residual errors is therefore not identifiable. For example, de Souza Vismara et al. (2016) analyzed the individual tree volumes of eucalyptus plantation data using a system of two mixed-effect models, with individual models for two rotations. The first-rotation trees were planted trees of the same clone, which were cut at the end of rotation. The second rotation trees were sprouts from the stumps of second-rotation trees. The grouping was associated with farm and the residuals with individual trees. It was justified that the trees on a certain plantation are similar to each other across the rotations. However, the individual trees are not the same in both rotations, therefore the cross-model correlation between residual errors was not estimated.

For estimation, the model system can also be written as a univariate model.
To do that for a single group, we make the following definitions:

\[ y_i = \begin{bmatrix} y^{(1)}_i \\ y^{(2)}_i \\ \vdots \\ y^{(M)}_i \end{bmatrix}, \quad X_i = \begin{bmatrix} X^{(1)}_i \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \epsilon_i = \begin{bmatrix} \epsilon^{(1)}_i \\ \epsilon^{(2)}_i \\ \vdots \\ \epsilon^{(M)}_i \end{bmatrix} \]

\[ \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_M \end{bmatrix}, \quad Z_i = \begin{bmatrix} Z^{(1)}_i \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad b_i = \begin{bmatrix} b^{(1)}_i \\ b^{(2)}_i \\ \vdots \\ b^{(M)}_i \end{bmatrix} \]

\[ R_i = \Sigma \otimes I \quad D_* = \begin{bmatrix} D^{(1)}_* & C^{(12)}_* & \ldots & C^{(1M)}_* \\ C^{(12)*} & D^{(2)}_* & \ldots & C^{(2M)}_* \\ \vdots & \vdots & \ddots & \vdots \\ C^{(1M)*} & C^{(2M)*} & \ldots & D^{(M)}_* \end{bmatrix} \]

where \( R_i = \text{var}(\epsilon_i), \ D_* = \text{var}(b_i) \) as previously in Section 5.1.5, and

\[ \Sigma_* = \begin{bmatrix} \sigma_1 & \varphi_{1,2} & \ldots & \varphi_{1,M} \\ \varphi_{1,2} & \sigma_2 & \ldots & \varphi_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{1,M} & \varphi_{2,M} & \ldots & \sigma_M \end{bmatrix} \]

if observations of all response variables are taken from the same sampling units. If the observations are taken from different sampling units of the group, then the off-diagonal elements of \( \Sigma_* \) are zeroes. We can now write the model in the familiar formula

\[ y_i = X_i \beta + Z_i b_i + \epsilon_i, \quad (8.6) \]

where \( \text{var}(b_i) = D_* \) and \( \text{var}(\epsilon_i) = R_i \). Once the model for a single group is defined, the model for all data is defined in the same way as for the univariate model, see Section 5.1.7. It should be noticed that the number of parameters increases rapidly when the the number of responses increases, especially if random slopes are included in addition to random intercepts.

An alternative to the above-specified formulation for single group is to define \( y_i \) so that the different responses of each sampling unit are first stacked on each other, and these vectors are thereafter stacked to form vector \( y_i \). That would change the order of observations so that matrix \( R_i \) is block-diagonal. Parameter estimation of the model formulated in this way is computationally less intensive than the formulation presented above.

Definitions to multiple levels of grouping is possible by combining the ideas presented above for single level of grouping and the ideas from Sections ?? and ??, For discussion on such models, see e.g. Goldstein (2003) and Hox et al. (2018).
8.4.2 Estimation and inference

The model is a special case of the univariate linear mixed-effect model, therefore the parameter estimation and inference can be based on REML and GLS, as described in Section 5.2. However, if the number of responses is high, convergence problems frequently occur due to high number of parameters in matrices $D_*$ and $\Sigma_*$. As a solution, Fieuws and Verbeke (2006) proposed a pairwise fitting approach. Instead of fitting the large system of $M$ mixed-effect models jointly, one could formulate all the $M(M-1)/2$ pairs of mixed-effect models and fit each of them jointly using REML. Each of these bivariate models would provide estimates of the cross-model covariances of random effects and residuals for the specific pair in question. In addition, one would get $M-1$ different unbiased estimates of $\beta$, $D^{(m)}$ and $\sigma_m$ for $m = 1, \ldots, M$, which are averaged over the fitted model pairs to get the final estimates. They also provided means to estimate the variance of the averaged estimates. A forestry application of their procedure is in Mehtätalo et al. (2008).

Often joint estimation of the model system does not provide much improvement in the estimates of fixed effects. In addition, the main motivation may be in estimation of the cross-model covariances or correlations of the random effects and residual errors to be used in prediction or simulation. One might be tempted to estimate the individual models separately, and thereafter estimate the cross-model covariances using the predicted random effects and residuals. However, because of the shrinkage of the predicted random effects (see Section 5.2), the variances and covariances estimated in this way are downward biased, and this method is not acceptable.

An acceptable strategy is to estimate the model so that the group effects as fixed, and thereafter estimate the variances based on these fixed group effects. Another alternative for models with random intercept is to use the method proposed by Searle and Rounsaville (1974). They suggested that the univariate mixed-effect model is estimated for all response variables separately. In addition, similar model is formulated for all possible sums of two response variables. For example, consider a system of two variance component models with random intercept:

\[
\begin{align*}
y_{ij}^{(1)} &= \mu_1 + b_i^{(1)} + \epsilon_{ij}^{(1)} \\
y_{ij}^{(2)} &= \mu_2 + b_i^{(2)} + \epsilon_{ij}^{(2)}
\end{align*}
\]

Summing the left and right sides separately and reorganizing we see that the model for the sum $z_{ij} = y_{ij}^{(1)} + y_{ij}^{(2)}$ can be written as

\[
z_{ij} = \mu_z + b_i^{(z)} + \epsilon_{ij}^{(z)}
\]

where $\mu_z = \mu_1 + \mu_2$, $b_i^{(z)} = b_i^{(1)} + b_i^{(2)}$ and $\epsilon_{ij}^{(z)} = \epsilon_{ij}^{(1)} + \epsilon_{ij}^{(2)}$. Fitting these three model separately gives estimates of $\text{var}(b_i^{(1)})$, $\text{var}(b_i^{(2)})$ and $\text{var}(b_i^{(1)} + b_i^{(2)}) = \text{var}(b_i^{(z)})$. 

The estimated variances can be used to estimate the cross-model covariance of random effects as
\[ \hat{\text{cov}}(b_i^{(1)}, b_i^{(2)}) = \hat{\text{var}}(b_i^{(2)}) - \hat{\text{var}}(b_i^{(1)}) - \hat{\text{var}}(b_i^{(2)}) \]

two

Correspondingly, the cross-model covariance of residual errors can be estimated as
\[ \hat{\text{cov}}(\epsilon_{ij}^{(1)}, \epsilon_{ij}^{(2)}) = \frac{\hat{\text{var}}(\epsilon_{ij}^{(2)}) - \hat{\text{var}}(\epsilon_{ij}^{(1)}) - \hat{\text{var}}(\epsilon_{ij}^{(2)})}{2} \]

This idea extends also to multiple levels of grouping. If the fixed predictors of the two models differ, then the sum model should include all predictors that are included in either of the models. If both models fit well, the regression coefficients of the sum model should approximately equal to the sum of the estimates from the two component models. If that is not the case, it may indicate problems in the systematic part in similar ways as in the case of SUR model. An application of this approach in forestry context is in Lappi (1986).

Sometimes the assumption of constant correlation between the models may be too restrictive. However, estimation of the model parameters with a more flexible covariance structure may be too demanding. An example of such situation is in Lappi (2006) (see page 348) in the context of taper curves. As a solution, he modeled the means, variances and covariances of the 13-equation system as a function of tree DBH to produce a realistic description of the variance-covariance structure for trees of different size. The estimation started by fitting nonparametrically models for the response variable by using smoothing splines. Thereafter the residuals of these models were extracted. The estimation of variance function was based on modeling the squared residuals (because \( \text{var}(X) = E(X^2) \) if \( E(X) = 0 \)) and the covariance was based on modeling the cross-products of the residuals of all possible pairs of observations (because \( \text{cov}(X, Y) = E(XY) \) if \( E(X) = E(Y) = 0 \)). The cost of the more flexible modeling of the means, variances and covariances was that the properties of the estimates are not that well known. However, this was only a small problem compared to the benefit gained from a more realistic model, which allowed realistic predictions of taper curves through cross-model calibration.

**Example 8.5.** Maltamo et al. (2012) developed a system of seemingly unrelated mixed-effects models for five tree specific characteristics: Tree diameter at breast height (D, cm), tree height (H, m), tree stem volume (V, m\(^2\)), height of dead branch (HDB, m), and crown base height (HCB, m). The predictors were tree-specific quantities based on airborne laser scanning. The data included a total of 1510 individual Scots Pine trees from 56 sample plots in a study area in Eastern
The ultimate aim is to use the models to cross-calibrate the models (see Example 8.7). The component models for tree \( j \) on plot \( i \) were as follows

\[
D_{ij} = \beta_{1}^{(1)} + \beta_{2}^{(1)} h100_{ij} + \beta_{3}^{(1)} h_{ij}^{*} + \beta_{4}^{(1)} \sqrt{h30}_{ij} + b_{1}^{(1)} + b_{2}^{(1)} h100_{ij} + \epsilon_{ij}^{(1)}
\]

\[
H_{ij} = \beta_{1}^{(2)} + \beta_{2}^{(2)} h100_{ij} + \beta_{3}^{(2)} h80_{ij} + b_{1}^{(2)} + b_{2}^{(2)} h100_{ij} + \epsilon_{ij}^{(2)}
\]

\[
\ln V_{ij} = \beta_{1}^{(3)} + \beta_{2}^{(3)} h_{ij}^{*} + \beta_{3}^{(3)} \ln h30^{*}_{ij} + b_{1}^{(3)} + b_{2}^{(3)} h100_{ij} + \epsilon_{ij}^{(3)}
\]

\[
\sqrt{HDB_{ij}} = \beta_{1}^{(4)} + \beta_{2}^{(4)} h100_{ij} + \beta_{3}^{(4)} h70_{ij} + b_{1}^{(4)} + b_{2}^{(4)} h100_{ij} + \epsilon_{ij}^{(4)}
\]

\[
HCB_{ij} = \beta_{1}^{(5)} + \beta_{2}^{(5)} h20_{ij} + \beta_{3}^{(5)} h_{ij}^{*} + \beta_{4}^{(5)} \text{veg}_{ij} + b_{1}^{(5)} + b_{2}^{(5)} h20_{ij} + \epsilon_{ij}^{(5)}
\]

where \( h_{xx} \) refers to the \( xx \)th quantile and \( \bar{h} \) to the mean of the ALS return height; \( \text{veg} \) is the proportion of returns from vegetation. An asterisk is used for predictors that are based on a 250 m\(^2\) neighbourhood around the tree, the other predictors are based on the returns from the crown of the tree in question. All models include both random intercept and slope, therefore the variance-covariance matrix of random effects is a 10 by 10 matrix with 10 variances and 45 covariances. The matrix \( \Sigma \), is an unstructured, positive definite 5 by 5 matrix.

There are few missing values of \( HDB \) and \( HCB \) in the data, therefore the residual variance-covariance matrix for the data where missing values are removed is obtained from \( \Sigma^{*}I \) by removing the rows and columns corresponding to the missing values.

The modeling data set is available in data set \( \text{alsTree} \). To fit the model, we reorganize the data so that the five responses are all in variable \( y \) and variable \( \text{index} \) specifies the type of response for each row. In addition, the predictors of each of the response are in separate blocks so that columns 5, . . ., 22 includes the model matrix of the model system.

```r
> data(alsTree)
> dat<-cbind(  
> cbind(plot=alsTree$plot,tree=alsTree$tree,index=1,y=alsTree$DBH,  
> conD=1,hmaxD=alsTree$hmax,a_hmeanD=alsTree$a_hmean,sqh30D=sqrt(alsTree$h30),  
> conH=0,hmaxH=alsTree$hmax,h80H=alsTree$h80,  
> conV=0,hmaxV=alsTree$hmax,lna_h30V=log(alsTree$a_h30),lnh70V=log(alsTree$h70),  
> conHDB=0,hmaxHDB=alsTree$hmax,a_hmeanHDB=alsTree$a_hmean,a_h70HDB=alsTree$a_hmean,a_hmeanHDB=alsTree$a_hmean,  
> conHCB=0,h20HCB=alsTree$h20,a_hmeanHCB=alsTree$a_hmean,lna_vegHCB=log(alsTree$a_veg)),  
> cbind(plot=alsTree$plot,tree=alsTree$tree,index=2,y=alsTree$H,  
> conV=1,hmaxV=alsTree$hmax,a_hmeanV=alsTree$a_hmean,a_hmeanH=alsTree$a_hmean,  
> conH=1,hmaxH=alsTree$hmax,h80H=alsTree$h80,h20H=alsTree$h20,a_hmeanH=alsTree$a_hmean,a_hmeanH=alsTree$a_hmean,  
> conHDB=1,hmaxHDB=alsTree$hmax,a_hmeanHDB=alsTree$a_hmean,a_hmeanHDB=alsTree$a_hmean,  
> conHCB=1,h20HCB=alsTree$h20,a_hmeanHCB=alsTree$a_hmean,lna_vegHCB=log(alsTree$a_veg)),  
> cbind(plot=alsTree$plot,tree=alsTree$tree,index=3,y=log(alsTree$V),  
> conH=1,hmaxH=alsTree$hmax,h80H=alsTree$h80,h20H=alsTree$h20,a_hmeanH=alsTree$a_hmean,  
> conHDB=1,hmaxHDB=alsTree$hmax,a_hmeanHDB=alsTree$a_hmean,a_hmeanHDB=alsTree$a_hmean,lna_vegHDB=log(alsTree$a_veg)),  
> cbind(plot=alsTree$plot,tree=alsTree$tree,index=4,y=sqrt(alsTree$HDB),  
> conH=1,hmaxH=alsTree$hmax,h80H=alsTree$h80,h20H=alsTree$h20,a_hmeanH=alsTree$a_hmean,  
> conHDB=1,hmaxHDB=alsTree$hmax,a_hmeanHDB=alsTree$a_hmean,a_hmeanHDB=alsTree$a_hmean,  
> conHCB=1,h20HCB=alsTree$h20,a_hmeanHCB=alsTree$a_hmean,lna_vegHCB=log(alsTree$a_veg)),  
> cbind(plot=alsTree$plot,tree=alsTree$tree,index=5,y=alsTree$HCB,  
> conH=1,hmaxH=alsTree$hmax,h80H=alsTree$h80,h20H=alsTree$h20,a_hmeanH=alsTree$a_hmean,  
> conHDB=1,hmaxHDB=alsTree$hmax,a_hmeanHDB=alsTree$a_hmean,a_hmeanHDB=alsTree$a_hmean,  
> conHCB=1,h20HCB=alsTree$h20,a_hmeanHCB=alsTree$a_hmean,lna_vegHCB=log(alsTree$a_veg)),  
> cbind(plot=alsTree$plot,tree=alsTree$tree,index=6,y=alsTree$HDB,  
> conH=1,hmaxH=alsTree$hmax,h80H=alsTree$h80,h20H=alsTree$h20,a_hmeanH=alsTree$a_hmean,  
> conHDB=1,hmaxHDB=alsTree$hmax,a_hmeanHDB=alsTree$a_hmean,a_hmeanHDB=alsTree$a_hmean,  
> conHCB=1,h20HCB=alsTree$h20,a_hmeanHCB=alsTree$a_hmean,lna_vegHCB=log(alsTree$a_veg)))
> dat<-as.data.frame(dat)
> dat<-dat[!is.na(dat$y),]
> head(dat)

<table>
<thead>
<tr>
<th>plot</th>
<th>tree</th>
<th>index</th>
<th>y</th>
<th>conD</th>
<th>hmaxD</th>
<th>a_hmeanD</th>
<th>sqh30D</th>
<th>hmaxH</th>
<th>h80H</th>
<th>conV</th>
<th>hmaxV</th>
<th>a_hmeanV</th>
<th>sqh30V</th>
<th>hmaxH</th>
<th>h80H</th>
<th>conV</th>
<th>hmaxV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1148</td>
<td>109</td>
<td>4</td>
<td>1.24</td>
<td>1</td>
<td>20.4</td>
<td>14.19</td>
<td>3.76</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
The model is fitted below with function \texttt{nlme::lme} using REML. The common random intercept is dropped and the response-specific random intercepts are specified through the random slopes for the \texttt{con}-variables. The response-specific random slopes are defined by the other terms in the argument \texttt{random}. The weight argument specifies different variances for each response, i.e., the structure for the diagonal of matrix $\Sigma$. The off-diagonal structure is determined by the argument \texttt{corr}. Because of the large number of parameters and observations, the estimation took about 30 minutes using a modern PC at the time of writing. To monitor the fitting progress, we have requested function \texttt{lme} to produce some intermediate results by using argument \texttt{control}.

```r
> syssur <- lme(y ~ -1 + conD + hmaxD + a_hmeanD + sqh30D + conH + hmaxH + h80H + ...
+ conV + hmaxV + lna_h30V + lnh70V + conHDB + hmaxHDB + a_h70HDB + conHCB
+ h20HCB + a_hmeanHCB + lna_vegHCB,
+ data = dat,
+ random = ~ -1 + conD + hmaxD + conH + hmaxH + conV + hmaxV + ...
+ conHDB + hmaxHDB + conHCB + h20HCB | plot,
+ weights=varIdent(form = ~1|index),
+ corr=corSymm(form = ~1|plot/tree), control=list(msVerbose=TRUE,
+ maxIter=5000,msMaxIter=5000,
+ msMaxEval=5000,niterEM=50))

0: 34302.509: 1.27227 2.55406 2.04874 3.96844 2.55755 4.39270 1.55443 3.86243 ...
< 625 rows have been removed >
626: 28782.401: 2.29269 4.00359 2.91645 4.37326 3.19984 4.85242 1.88012 3.93374 ...
```

The estimated model is printed below. The output shows first the estimates of $\beta$. Thereafter there are the standard deviations and correlations based on the estimate of $D_\nu$. Term \texttt{Residual} specifies the scaling factor $\sigma$ of the residual variance-covariance matrix. Thereafter the output includes the estimated correlation matrix corresponding to the matrix $\Sigma$. Finally, there are the parameters of the variance function based on the diagonal elements of $\Sigma$. The response-specific residual standard errors are obtained by multiplying these by $\hat{\sigma} = 2.781$.

```r
> syssur
Linear mixed-effects model fit by REML
Data: dat
Log-restricted-likelihood: -7203.678
Fixed: y ~ -1 + conD + hmaxD + a_hmeanD + sqh30D + conH + hmaxH + h80H + ...
conD hmaxD a_hmeanD sqh30D conH hmaxH ...
-2.57056979 1.29651029 -0.85796990 0.18906254 1.28243491 0.68741940
h80H conV hmaxV lna_h30V lnh70V conHDB ...
0.36073188 -4.77798341 0.1966693 -1.34523931 1.31988229 -0.72958999

Random effects:
```

```
```
Formula: `-1 + conD + hmaxD + conH + hmaxH + conV + hmaxV + conHDB + ...
Structure: General positive-definite, Log-Cholesky parametrization

<table>
<thead>
<tr>
<th>StdDev Corr</th>
<th>conD</th>
<th>hmaxD</th>
<th>conH</th>
<th>hmaxH</th>
<th>conV</th>
<th>hmaxV</th>
<th>conHDB</th>
<th>hmaxHDB</th>
<th>conHCB</th>
<th>hmaxHCB</th>
<th>h20HCB</th>
</tr>
</thead>
<tbody>
<tr>
<td>conD</td>
<td>4.61917617</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hmaxD</td>
<td>0.28368158</td>
<td>-0.934</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>conH</td>
<td>0.74496709</td>
<td>-0.519</td>
<td>0.456</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hmaxH</td>
<td>0.05064401</td>
<td>0.409</td>
<td>-0.313</td>
<td>-0.894</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>conV</td>
<td>0.40132448</td>
<td>-0.157</td>
<td>0.215</td>
<td>-0.033</td>
<td>-0.033</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hmaxV</td>
<td>0.02651238</td>
<td>0.226</td>
<td>-0.165</td>
<td>0.008</td>
<td>-0.003</td>
<td>-0.897</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>conHDB</td>
<td>1.06330907</td>
<td>0.905</td>
<td>-0.896</td>
<td>-0.571</td>
<td>0.425</td>
<td>-0.453</td>
<td>0.444</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hmaxHDB</td>
<td>0.06745221</td>
<td>-0.727</td>
<td>0.786</td>
<td>0.321</td>
<td>-0.225</td>
<td>0.452</td>
<td>-0.437</td>
<td>-0.904</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>conHCB</td>
<td>0.84668486</td>
<td>-0.509</td>
<td>0.476</td>
<td>0.000</td>
<td>-0.101</td>
<td>-0.346</td>
<td>0.404</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hmaxHCB</td>
<td>0.07866226</td>
<td>0.239</td>
<td>-0.239</td>
<td>0.151</td>
<td>-0.197</td>
<td>0.072</td>
<td>-0.123</td>
<td>0.086</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>2.78109728</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Correlation Structure: General
Formula: `~1 | plot/tree`
Parameter estimate(s):
Correlation:

1 2 3 4 5
2 0.116
3 0.916 0.229
4 -0.202 0.072 -0.197
5 -0.249 0.159 -0.223 0.086

Variance function:
Structure: Different standard deviations per stratum
Formula: `~1 | index`
Parameter estimates:

1 2 3 4 5
1.0000000 0.2474477 0.1144037 0.1252440 0.3792651

Number of Observations: 7167
Number of Groups: 56

If the estimation had failed to converge, we could have fitted the model using the pairwise fitting approach. With 5 responses, we would get a total of 10 pairs of models, each of which would be fitted separately. Thereafter the results of the 10 models would have been pooled to a single model through averaging whenever several alternative estimates for the same parameter are available. Implementing the pairwise fitting approach is left as an exercise.

Example 8.6. To illustrate the approach of Searle and Rounsaville (1974) in the estimation of cross-model covariances of random effects and residuals, consider the model system for tree diameter and height in data `alsTree`. The univariate models for `DBH`, `H` and `DBH + H` are as follows

\[
D_{ij} = x_{ij}^{(1)} \beta_{1} + b_{1i}^{(1)} + \epsilon_{ij}^{(1)} \\
H_{ij} = x_{ij}^{(2)} \beta_{2} + b_{1i}^{(2)} + \epsilon_{ij}^{(2)} \\
D_{ij} + H_{ij} = x_{ij}^{(3)} \beta_{3} + b_{1i}^{(3)} + \epsilon_{ij}^{(3)}
\]

we use random intercept models because the method cannot be applied for the random slope models. The estimates of fixed effects for the sum model are approximately equal to the sum of fixed effects for the two component models. The cross-model correlations estimated by using the method of Searle and Rounsaville are \(\text{corr}_{SR}(b_{1i}^{(1)}, b_{1i}^{(2)}) = 0.120\) and \(\text{corr}_{SR}(\epsilon_{ij}^{(1)}, \epsilon_{ij}^{(2)}) = 0.080\). Estimation using REML gives \(\text{corr}_{R}(b_{1i}^{(1)}, b_{1i}^{(2)}) = 0.164\) and \(\text{corr}_{R}(\epsilon_{ij}^{(1)}, \epsilon_{ij}^{(2)}) = 0.079\).

```r
> modD<-lme(DBH~hmax,random=~1|plot,data=alsTree)
> modH<-lme(H~hmax,random=~1|plot,data=alsTree)
```
Multivariate (Mixed-Effects) Models

```r
> modDplusH <- lme(I(DBH+H) ~ hmax, random = ~1 | plot, data = alsTree)
> fixef(modDplusH)
(Intercept)       hmax
-7.401627        2.810454
> fixef(modD) + fixef(modH)
(Intercept)       hmax
-7.446093        2.813016
> covb <- (getVarCov(modDplusH)[1] - getVarCov(modD)[1] - getVarCov(modH)[1])/2
> corb <- covb/(getVarCov(modD)[1] * getVarCov(modH)[1])
> cove <- (modDplusH$sigma^2 - modD$sigma^2 - modH$sigma^2)/2
> core <- cove/(modD$sigma * modH$sigma)
> corb
[1] 0.1199896
> core
[1] 0.08030711
> datDH <- rbind(cbind(plot = alsTree$plot, tree = alsTree$tree, index = 1, y = alsTree$DBH, conD = 1, hmaxD = alsTree$hmax, a_hmeanD = alsTree$a_hmean, sqh30D = sqrt(alsTree$h30)),
  cbind(plot = alsTree$plot, tree = alsTree$tree, index = 2, y = alsTree$H, conH = 0, hmaxH = alsTree$hmax, h80H = alsTree$h80))
> sysDH <- lme(y ~ -1 + conD + hmaxD + conH + hmaxH,
  data = datDH,
  random = ~ -1 + conD + conH | plot,
  weights = varIdent(form = ~ 1 | index),
  corr = corSymm(form = ~ 1 | plot/tree))
> sysDH
Linear mixed-effects model fit by REML
Data: datDH
Log-restricted-likelihood: -5499.01
Fixed: y ~ -1 + conD + hmaxD + conH + hmaxH
  conD hmaxD conH hmaxH
-8.606148 1.796589 1.128128 1.018543
Random effects:
Formula: ~ -1 + conD + conH | plot
Structure: General positive-definite, Log-Cholesky parametrization
StdDev Corr
conD      2.7797620 conD
conH     0.3661644 0.164
Residual 2.8612900

Correlation Structure: General
Formula: ~ 1 | plot/tree
Parameter estimates:
Correlation:
1
2 0.079
Variance function:
Structure: Different standard deviations per stratum
Formula: ~ 1 | index
Parameter estimates:
1 2
1.0000000 0.2488368
Number of Observations: 3020
Number of Groups: 56
```
8.4.3 Prediction

Prediction based on the fixed part and group-specific prediction using an individual components of the system follows the principles we have presented earlier in Section ??.

The model system allows utilizing also the cross-model correlations of random effects to predict the group effects for all response variables using measurements of some of the responses of the target group. In addition, the model system allows utilizing the cross-model correlation of residual errors to predict the residual errors for all response variables using measurements of some responses of the target observation. These predictions utilize the cross-model correlations and are beneficial only if the correlations exist. If the correlations are low, the predictions are very close to the prediction based on fixed part only. The prediction is based on the EBLUP (see Section 2.3.1), and the required expectations, variances and covariances are based on the fitted model as described below.

For simplicity, consider first the situation where only the random effects are being predicted and all response variables have been observed for the sampling units $1, \ldots, n$ that are used for prediction. Start by defining the unobserved ($h_1$) and observed ($h_2$) parts of vector $h$ of (2.15). We consider a single group $i$ but drop the group index $i$ from the following presentation to simplify notations. The unobserved part includes the random effects $h_1 = b$ of all models. The observed part includes $y_{ij}$ for all responses for sampling units $1, \ldots, n$ of group $i$, i.e., $h_2 = y$, the vector of all responses for the observations of the target group, i.e., $y = (y^{(1)}_i, y^{(2)}_i, \ldots, y^{(M)}_i)'$. Let $X$, $Z$, $D$ and $R$ be the corresponding model matrices for the fixed and random parts and the variance-covariance matrices of random effect and residuals, with the structure defined before Equation (8.6). Now $\text{var}(y) = ZDZ' + R$ and $\text{cov}(b, y') = DZ'$, and the prediction of random effect can be done using the procedures presented earlier in Section 5.2.6.

To consider the case where where not all responses have been observed, we adjust the procedure as follows. Denote the vector of observations by $y_o$. It is constructed from $y$ by removing the elements that have not been observed. For example, if responses 1 and 3 have not been observed, then $y_o = (y^{(2)}_i, y^{(4)}_i, \ldots, y^{(M)}_i)'$. We define the corresponding matrices and vectors $X_o$, $Z_o$, $D_o$, $R_o$ and $\beta_o$ by removing from $X$, $Z$, $D$, $R$ and $\beta$ the rows and columns that correspond to the unobserved responses. Furthermore, we define matrix $C$ by removing from $D$ the columns that correspond to the unobserved response variables but keeping all rows. Now

$$
\begin{pmatrix}
  b \\
  y_o
\end{pmatrix} \sim \left[ \begin{pmatrix}
  0 \\
  X_o \beta_o
\end{pmatrix}, \begin{pmatrix}
  D & CZ_o' \\
  Z_oC' & Z_oD_oZ_o' + R_o
\end{pmatrix} \right].
$$
The BLP of random effect vector $\mathbf{b}$ and its prediction variance become

$$\text{BLP}(\mathbf{b}) = C\mathbf{Z}'_o (\mathbf{Z}_o \mathbf{D}_o \mathbf{Z}'_o + \mathbf{R}_o)^{-1} (\mathbf{y}_o - \mathbf{X}_o \mathbf{\beta}_o)$$

$$\text{var} (\text{BLP}(\mathbf{b}) - \mathbf{b}) = \mathbf{D} - C\mathbf{Z}'_o (\mathbf{Z}_o \mathbf{D}_o \mathbf{Z}'_o + \mathbf{R}_o)^{-1} \mathbf{Z}_o \mathbf{C}'.$$

In practice, $\mathbf{\beta}$ and variance-covariance matrices are replaced by their estimates, which leads to EBLUP. The variance formula above is underestimate because it ignores the related estimation errors.

Sometimes it may be interesting to predict the residuals of the observed sampling units for all models, in addition to the random effects. That is the case, for example, in the taper curve model system of (Lappi, 1986), which includes random stand effects and tree-level residual errors. If measurements of tree diameters are available for some trees of a stand, they include information, on one hand, about the mean stem form in the stand and, on the other hand, about the form of that particular tree stem. The mean stem form of that stand can be estimated by using the predicted stand effects of the model, and the form of the stem with measurements can be further improved by using the predicted residual errors for all component models. To do so, we can construct the BLP using

$$\begin{pmatrix} \mathbf{e} \\ \mathbf{b} \\ \mathbf{y}_o \end{pmatrix} \sim \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{D} \\ \mathbf{X}_o \mathbf{\beta}_o \\ \mathbf{B}' \\ \mathbf{Z}_o \mathbf{C}' \\ \mathbf{Z}_o \mathbf{D}_o \mathbf{Z}'_o + \mathbf{R}_o \end{pmatrix},$$

where $\mathbf{B}$ is constructed from $\mathbf{R}$ by removing those columns for which we do not have measurements but keeping all rows. Defining $h_1 = (\mathbf{e}', \mathbf{b}')'$ and $h_2 = \mathbf{y}_o$, we see easily the structure of $V_1$, $V_2$ and $V_{12}$ which are needed to apply Equation 2.15. Replacing unknown parameters by their estimates leads, again, to EBLUP. Identical predictions of $\mathbf{e}$ are obtained by constructing the predictor using

$$\begin{pmatrix} \mathbf{e} \\ \mathbf{y}_o \end{pmatrix} \sim \begin{pmatrix} 0 & \mathbf{B} \\ \mathbf{X}_o \mathbf{\beta}_o \\ \mathbf{B}' \\ \mathbf{Z}_o \mathbf{D}_o \mathbf{Z}'_o + \mathbf{R}_o \end{pmatrix}.$$

**Example 8.7.** Consider the system of model that was estimated in Example 8.5. Assume that three trees have been measured for $\text{DBH}$, $\text{HDB}$ and $\text{HCB}$, with the following measurements and values of the predictors.

```
> obs
plot tree DBH HDB HCB hmax h20 h30 h70 h80 a_hmean a_veg a_h30 a_h70
2  1 30 24.60 0.75 12.3 20.87 13.74 14.94 18.42 18.84 14.44 0.8952 14.51 14.75
5  1 43 28.25 0.40 13.0 20.00 13.33 14.77 17.77 18.47 14.75 0.8533 14.89 14.52
7  1 18 25.40 2.30 14.1 21.50 13.43 15.58 18.63 19.50 14.51 0.9268 14.26 17.84
```

The aim is to predict the random effects of all five models of the system by using these data. We extract $\mathbf{\beta}$, $\mathbf{D}$ and $\Sigma$ from the fitted model. Thereafter we construct the required matrices and vectors and predict the random effects. The constructed matrices are printed to show their structure. The predicted values of random effects and their prediction errors are then computed to objects $\mathbf{b}$ and $\text{varbb}$.
> D <-getVarCov(syssur)
> corepsilon <- corMatrix(syssur$modelStruct[2]$corStruct)[1][1]
> sdepsilon <- syssur$sigma*c(1, exp(coef(syssur$modelStruct[3]$varStruct)))
> Sigma <- diag(sdepsilon)%*%corepsilon%*%diag(sdepsilon)
>
> library(magic)
> ntrees <- 3
> Z1234 <- cbind(1, obs[, "hmax"])
> Z5 <- cbind(1, obs[, "h20"])
> Zo <- adiag(Z1234, Z1234, Z5)
>
> C <- D[, -c(3, 4, 5, 6)]
> Do <- D[-c(3, 4, 5, 6), -c(3, 4, 5, 6)]
> Ro <- Sigma[-c(2, 3), -c(2, 3)]%x%diag(ntrees)
>
> Xo <- adiag(cbind(1, obs$hmax, obs$a_hmean, sqrt(obs$h30)),
+ cbind(1, obs$hmax, obs$a_h70),
+ cbind(1, obs$h20, obs$a_hmean, log(obs$a_veg)))
>
> beta <- fixef(syssur)
> betao <- beta[-seq(5, 11)]
> muo <- Xo%*%betao
> yo <- c(obs$DBH, sqrt(obs$HDB), obs$HCB)
>
> round(D, 3)

> round(C, 3)

> round(Zo[, 1:6])
Multivariate (Mixed-Effects) Models

> \textbf{To illustrate the benefit from using the predicted random effects, Table 8.1 shows the observed values of all five response variables for 5 new trees of the same target stand. Predictions were made using fixed part only and using fixed and random parts, with the predicted random effects from vector } b. In most cases, the predictions using fixed and random effects are closer to the observed values than the predictions based on fixed part only. The scripts for the prediction are shown below. The predictions of } V \text{ and } HDB \text{ were not adjusted for the back-transformation bias; an improved analysis should do this by utilizing the prediction variance from } varbb \text{ as we did in Example 5.19.}
TABLE 8.1
Predictions using the fixed part only, using the fixed part and the predicted random effects, and field-measured values of the five trees on the five response variables from the target stand. The random effects were predicted by using the observations of DBH, HDB and HCB for three sample trees in Example 8.7

<table>
<thead>
<tr>
<th></th>
<th>Fixed effects only</th>
<th>Fixed part + random effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DBH</td>
<td>H</td>
</tr>
<tr>
<td>31.2</td>
<td>24.4</td>
<td>0.97</td>
</tr>
<tr>
<td>27.2</td>
<td>22.8</td>
<td>0.64</td>
</tr>
<tr>
<td>25.2</td>
<td>21.7</td>
<td>0.49</td>
</tr>
<tr>
<td>30.6</td>
<td>24.9</td>
<td>1.03</td>
</tr>
<tr>
<td>30.2</td>
<td>23.8</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Field measurements

<table>
<thead>
<tr>
<th></th>
<th>DBH</th>
<th>H</th>
<th>V</th>
<th>HDB</th>
<th>HCB</th>
</tr>
</thead>
<tbody>
<tr>
<td>26.9</td>
<td>24.1</td>
<td>0.63</td>
<td>0.4</td>
<td>15.9</td>
<td></td>
</tr>
<tr>
<td>28.3</td>
<td>22.4</td>
<td>0.65</td>
<td>1.7</td>
<td>13.2</td>
<td></td>
</tr>
<tr>
<td>21.2</td>
<td>21.2</td>
<td>0.34</td>
<td>0.8</td>
<td>12.4</td>
<td></td>
</tr>
<tr>
<td>29.8</td>
<td>24.2</td>
<td>0.77</td>
<td>4.0</td>
<td>14.0</td>
<td></td>
</tr>
<tr>
<td>32.4</td>
<td>23.9</td>
<td>0.90</td>
<td>0.4</td>
<td>13.1</td>
<td></td>
</tr>
</tbody>
</table>

The script below predicts also the residual errors for all responses. For DBH, HDB and HCB, the predicted residuals are equal to the observed residuals $e_o = y_o - (X_o\beta_o - Z_o b_o)$, as shown below. For $H$ and $V$ they are also nonzero because the cross-model correlations of residual errors are nonzero.

```r
> Z1234 <- cbind(1, new[, "hmax"])
> Z5 <- cbind(1, new[, "h20"])
> Z <- adiag(Z1234, Z1234, Z1234, Z1234, Z5)
> predf <- X %*% beta
> predfr <- predf + Z %*% b
>
> B <- Sigma[, -c(2, 3)] %x% diag(nTrees)
> V1 <- adiag(8, 0)
> V12 <- rbind(B, C %*% t(Zo))
> V2 <- Z0 %*% Do %*% t(Zo) + No
> h1hat <- V12 %*% solve(V2) %*% (yo - muo)
> # The predicted residuals for DBH, H, V, HDB and HCB
> matrix(h1hat[1:15], ncol=5)

[1,]  -1.95  3.78 -2.43  -0.146 -0.448  0.461 -0.639  0.331  1.3

# For observed responses, the residuals above equal to y-(xbeta+zb)
> t(yo - muo - Zo %*% b[-c(3:6)])
```

8.5 Exercises

1. Estimate the parameters of the model system of example 8.5 by using the pairwise fitting algorithm of Fieuws and Verbeke (2006).
Special topics on regression
Part III

Special Topics
Bibliography


Sinkää, S., J. Heinonen, J. Miina, and K. Eerikäinen. 2014. Subject-specific prediction using a nonlinear mixed model: consequences of different approaches. Forest Science Early view(X):XXXX.


Index

Adaptive Gauss-Hermite Quadrature (AGQ), 131

Bayes rule, 38
Bayesian method, 135
Bernoulli trial, 9, 85
beta function, 50

canonical form, 102
certainty interval, 62
convolution, 26
correlation, 39
covariance, 39
covariance rules, 52
cumulative distribution function (cdf), 6
conditional, 26
design-based approach, 5
deviance residuals, 107
distribution, 6
F, 22, 34
\( \chi^2 \), 33
Bernoulli, 9, 85
Beta, 50
Beta-Binomial, 50
binomial, 50, 86
censored, 28
compound (see also distribution
conjugate), 50
conditional, 36
conjugate (see also distribution
compound), 50
exponential, 27
Gaussian multivariate (see also distribution normal multivariate), 42
joint, 35
lognormal, 15
marginal, 37
multivariate, 35
normal, 10, 85
bivariate, 47
multivariate, 42
standard, 10
Poisson, 14, 86
Student’s t, 33
truncated (see also distribution
censored), 28
Weibull, 14
weighted, 28

expected value, 18
Bernoulli distribution, 21, 85
binomial distribution, 86
conditional, 41
multivariate normal, 47
exponential distribution, 27
law of total expectation, 41
marginal, 41
normal distribution, 21
Poisson distribution, 21, 87
experiment, 5
exponential family, 102
gamma function, 33
Gaussian random field, 44
Generalized Linear Mixed Model (GLMM), 124
Generalized Linear Model, GLM, 83

independence
random variables, 39
random vectors, 39
  joint (see also independence random vectors mutual), 39
  mutual (see also independence random vectors joint), 39

latent variable, 3
likelihood ratio test, 119
linear association (see also linear dependence), 39
linear dependence, 39
link function, 88
  canonical, 102
  inverse (see mean function), 88
marginal probability, 26
Markov Chain Monte Carlo, MCMC, 135
mean (see also expected value), 18
mean function, 88
  canonical, 102
measurement, 3
model-based approach, 4
observation, 3
overdispersion, 111

parameter, 9
  space, 9
parsimonious model, 42
Pearson residuals, 107
Pearson statistic, 107
percentile, 23
point probability, 7
positive semidefinite, 40
powered exponential function, 44
prediction
  best predictor, 58
  conditionally unbiased, 58
  variance, 58
prediction interval, 22
probability density function (pdf), 9
  conditional, 37
  joint, 35
  marginal, 26, 37
  weighted, 30
probability function (see mean function), 88
probability mass function (pmf), 7
  conditional, 27, 36
  joint, 35
  marginal, 27, 37
quantile, 21
quantile function, 21
quasi-likelihood, 114
  penalized, PQL, 131
random process, 3
random variable, 3
  censored, 30
  continuous, 5
  discrete, 5
  multivariate, 6, 35
  realization, 3
  univariate, 6
random vector, 6, 35
randomness, 3
  complete spatial, 112
Rao’s score test, 120
residual deviance, 107
response function (see mean function), 88
RMSE, 57
sample space, 5
sampling, 3
saturated model, 107
standard deviation, 21
support (set), 5
  transformation, 15
    many-to-one, 26
    monotone, 15
    one-to-one, 15
underdispersion, 111
variable, 3
  variance, 20
    Bernoulli distribution, 21, 85
Index

binomial distribution, 86
conditional, 41
multivariate normal, 47
law of total variance, 42
marginal, 41
normal distribution, 21
Poisson distribution, 21, 87
rules, 52
variance-covariance matrix, 40

Wald $\chi^2$-test, 119
Wald $F$-test, 120

zero-inflation, 123