Class Notes: Linear Models in Animal Breeding

Contents

Chapter 1

Classification of Linear Models

1.1 Introductory concepts

Models are used to describe complex relatioships between variables in simple mathematical terms. A linear model gives a simple description of the distribution of the data or observations $\boldsymbol{y}$. It only describes the expected value (E($\boldsymbol{y}$)) and the variance (Var($\boldsymbol{y}$)) of the data.

Definition 1.1.1 *The model for $E(\boldsymbol{y})$ is said to be linear if*

$$E(\boldsymbol{y}) = \boldsymbol{X\beta} \tag{1.1}$$

where $\boldsymbol{X}$ is a known matrix and $\boldsymbol{\beta}$ is a vector of unknown parameters.The above model for $E(\boldsymbol{y})$ is linear in the parameters $\boldsymbol{\beta}$.

Definition 1.1.2 *The general linear model for $\boldsymbol{y}$ has the form:*

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon} \tag{1.2}$$

where $E(\boldsymbol{\epsilon}) = \mathbf{0}$, $Var(\boldsymbol{\epsilon}) = \boldsymbol{V}$ or

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e} \tag{1.3}$$

where $\boldsymbol{X}$ and $\boldsymbol{Z}$ are known matrices,

$$\begin{aligned}
&E(\boldsymbol{u}) = \mathbf{0}, E(\boldsymbol{e}) = \mathbf{0} \\
&Var(\boldsymbol{u}) = \boldsymbol{G},\ Var(\boldsymbol{e}) = \boldsymbol{R}\sigma_e^2, Cov(\boldsymbol{u}, \boldsymbol{e}') = \mathbf{0} \\
&Var(\boldsymbol{y}) = \boldsymbol{ZGZ}' + \boldsymbol{R}\sigma_e^2 = \boldsymbol{V}
\end{aligned}$$

$\boldsymbol{G}$ and $\boldsymbol{R}$ will often be simplified further, eg. $\boldsymbol{R} = \boldsymbol{I}$.

1.2 Classification based on the variable type

Linear models can be classified based on the type of variables used in the models as:

1. Regression models that contain only quantitative independent variables;
2. ANOVA models that contain only qualitiative independent variables; or
3. Analysis of covariance models (ANCOVA) that contain qualitative and quantitative independent variables.

1.3 Clasification based on the distribution of effects in the model

Linear models are also classified based on the distribution of the effects in the model

1. Fixed linear models contain only fixed effects apart from a random residual.
2. Random linear models contain only random effects.
3. Mixed linear models contain fixed and random effects.

Chapter 2

Estimation of Fixed Effects and Hypothesis Testing

2.1 Least Squares Estimation

Let us consider a fixed linear model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e} \tag{2.1}$$

where $\boldsymbol{X}$ is a $n \times p$ known matrix with rank $r \leq p$, with $\mathrm{E}(\boldsymbol{e}) = \boldsymbol{0}$ and $\mathrm{Var}(\boldsymbol{y}) = \mathrm{Var}(\boldsymbol{e}) = \boldsymbol{I}\sigma_e^2$. To estimate $\mathrm{E}(\boldsymbol{y}) = \boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta^*}$ by least squares, where $\boldsymbol{\beta^*}$ is the unique not known $\boldsymbol{\beta}$ from the whole parameter space, minimize

$$Q = \sum_{i=1}^{n} (y_i - \tilde{\eta}_i)^2 \tag{2.2}$$

with respect to $\tilde{\boldsymbol{\beta}}_j$, where $\tilde{\eta}_i = \sum_j x_{ij}\tilde{\beta}_j$. Here $\tilde{\beta}_j$ are the possible values β_j can take. In matrix notation we can write

$$\begin{aligned} Q &= (\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}) \\ &= \boldsymbol{y'y} - 2\boldsymbol{X'y}\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\beta}}'(\boldsymbol{X'X})\tilde{\boldsymbol{\beta}} \end{aligned} \tag{2.3}$$

Using (6.28) and (6.30) to obtain the derivative of Q with respect to $\tilde{\boldsymbol{\beta}}$ and equating the derivative to 0 gives:

$$\frac{\partial Q}{\partial \tilde{\boldsymbol{\beta}}} = -2\boldsymbol{X'y} + 2(\boldsymbol{X'X})\widehat{\boldsymbol{\beta}} = \boldsymbol{0} \tag{2.4}$$

where $\widehat{\boldsymbol{\beta}}$ is the estimate that minimizes Q. Rearranging (2.4) gives the ordinary least square equations (OLS) or the normal equations

$$(\boldsymbol{X'X})\widehat{\boldsymbol{\beta}} = \boldsymbol{X'y} \tag{2.5}$$

In a geometric interpretation of least square theory we want to minimize $Q = \| \boldsymbol{y} - \tilde{\boldsymbol{\eta}} \|^2$ with respect to $\tilde{\boldsymbol{\eta}} \in V_r$. Note that V_r is the space spanned by the columns of $\boldsymbol{X}$. Theorem 6.1.2 shows that $\widehat{\boldsymbol{\eta}}$, the projection of $\boldsymbol{y}$ on V_r, minimizes Q. Note that because $\widehat{\boldsymbol{\eta}} \in V_r$ we can write $\widehat{\boldsymbol{\eta}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$. Also because $(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})$ is orthogonal to the columns of $\boldsymbol{X}$ we can write

$$\boldsymbol{X'}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) = \mathbf{0} \Rightarrow \boldsymbol{X'y} = (\boldsymbol{X'X})\widehat{\boldsymbol{\beta}} \tag{2.6}$$

So, the normal equations, can be derived by using this approach too.

2.2 Estimable functions

Definition 2.2.1 *Let $\boldsymbol{\psi} = \boldsymbol{k'\beta}$ be a linear function of the unknown parameters $\boldsymbol{\beta}$. $\boldsymbol{\psi}$ is estimable if it has an unbiased linear estimate i.e. if there exists $\boldsymbol{a'}$ such that $\mathrm{E}(\boldsymbol{a'y}) = \boldsymbol{\psi}$ for all β.*

For example if $\mathrm{E}(\boldsymbol{y}) = \boldsymbol{\mu} + \boldsymbol{\alpha}_i$ where

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_{11} \\ \boldsymbol{y}_{12} \\ \boldsymbol{y}_{21} \\ \boldsymbol{y}_{22} \end{bmatrix}; \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \end{bmatrix}$$

and if $\boldsymbol{\psi} = \boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2 = \begin{bmatrix} 0 & 1 & -1 \end{bmatrix} \boldsymbol{\beta}$, can show that $\boldsymbol{\psi}$ is estimable. For $\boldsymbol{a'} = \begin{bmatrix} 1 & 0 & -1 & 0 \end{bmatrix}$, we can write

$$\mathrm{E}(\boldsymbol{a'y}) = \mathrm{E}(\boldsymbol{y}_{11} - \boldsymbol{y}_{21}) = \boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2 \tag{2.7}$$

and based on the definition (2.2.1) $\boldsymbol{\psi}$ is estimable.

Theorem 2.2.1 *$\psi = \boldsymbol{k'\beta}$ is estimable if and only if $\boldsymbol{k'}$ is a linear function of the rows of $\boldsymbol{X}$. That is ψ is estimable if and only if there exists $\boldsymbol{a'}$ such that $\boldsymbol{k'} = \boldsymbol{a'X}$*

Proof: If $\boldsymbol{\psi}$ is estimable, from definition (2.2.1), $\mathrm{E}(\boldsymbol{a'y}) = \boldsymbol{\psi}$ and as a result

$$\mathrm{E}(\boldsymbol{a'y}) = \boldsymbol{a'}\mathrm{E}(\boldsymbol{y}) = \boldsymbol{a'X\beta} = \boldsymbol{k'\beta} = \psi \quad \forall\, \boldsymbol{\beta} \Rightarrow \boldsymbol{a'X} = \boldsymbol{k'} \tag{2.8}$$

Now if $\boldsymbol{k'} = \boldsymbol{a'X}$

$$\psi = \boldsymbol{k'\beta} = \boldsymbol{a'X\beta} = \boldsymbol{a'}\mathrm{E}(\boldsymbol{y}) = \mathrm{E}(\boldsymbol{a'y}) \quad \forall\, \boldsymbol{\beta} \tag{2.9}$$

and as a result $\boldsymbol{\psi} = \boldsymbol{k'\beta}$ is estimable.

Lemma 2.2.1 *If $\psi = \boldsymbol{k'\beta}$ is estimable, then there exists a unique linear unbiased estimate of $\boldsymbol{\psi}$, $\boldsymbol{a^{*'}y}$ for $\boldsymbol{a^*} \in V_r$. Further, $\boldsymbol{a^*}$ is the projection of $\boldsymbol{a}$ on V_r where $E(\boldsymbol{a'y}) = \boldsymbol{\psi}$.*

Proof: There exists $\boldsymbol{a'}$ such that $\mathrm{E}(\boldsymbol{a'y}) = \boldsymbol{\psi}$. Otherwise, $\boldsymbol{\psi}$ is not estimable. Let $\boldsymbol{a} = \boldsymbol{a}^* + \boldsymbol{b}$ where $\boldsymbol{a}^* \in V_r$ and $\boldsymbol{b} \perp V_r$. Then

$$\begin{aligned}\boldsymbol{\psi} &= \mathrm{E}(\boldsymbol{a'y}) \\ &= \mathrm{E}(\boldsymbol{a^{*'}y}) + \mathrm{E}(\boldsymbol{b'y}) \\ &= \mathrm{E}(\boldsymbol{a^{*'}y}) + \boldsymbol{b'X\beta} \\ &= \mathrm{E}(\boldsymbol{a^{*'}y})\end{aligned} \tag{2.10}$$

because $\boldsymbol{b'X} = \boldsymbol{0}$. We have shown that there exists an $\boldsymbol{a}^* \in V_r$, now we need to show the uniqueness of $\boldsymbol{a}^*$. Suppose α' is also in V_r and $\mathrm{E}(\boldsymbol{\alpha' y}) = \boldsymbol{\psi}$. Then

$$\begin{aligned}\boldsymbol{0} &= \mathrm{E}(\boldsymbol{a^{*'}y}) - \mathrm{E}(\boldsymbol{\alpha' y}) \\ &= (\boldsymbol{a}^* - \boldsymbol{\alpha})'\boldsymbol{X\beta} \quad \forall\, \boldsymbol{\beta}\end{aligned} \tag{2.11}$$

Because this holds for all $\boldsymbol{\beta}$-s, this implies that

$$(\boldsymbol{a}^* - \boldsymbol{\alpha})'\boldsymbol{X} = \boldsymbol{0} \Rightarrow (\boldsymbol{a}^* - \boldsymbol{\alpha})' \perp V_r$$

But we know also that $(\boldsymbol{a}^* - \boldsymbol{\alpha})' \in V_r$. As a result $(\boldsymbol{a}^* - \boldsymbol{\alpha})' = \boldsymbol{0}$, so $\boldsymbol{a}^*$ is unique.

2.3 Gauss-Markoff theorem

Theorem 2.3.1 (Gauss-Markoff Theorem) *Given* $E(\boldsymbol{y}) = \boldsymbol{X\beta}$ *and* $Var(\boldsymbol{y}) = \boldsymbol{I}\sigma_e^2$ *every estimable function* ψ *has a unique unbiased linear estimator* $\widehat{\psi}$ *which has minimum variance in the class of all linear unbiased estimators. This is given by* $\widehat{\psi} = \boldsymbol{k'}\widehat{\boldsymbol{\beta}}$ *where* $\widehat{\boldsymbol{\beta}}$ *is a solution to the normal equations.*

Proof: Let $\boldsymbol{a^{*'}y}$ with $\boldsymbol{a}^* \in V_r$ be the unbiased linear estimator of $\boldsymbol{\psi}$, and let $\boldsymbol{a'y}$ be any unbiased estimator. Note that $\boldsymbol{a}^*$ is the projection of $\boldsymbol{a}$ on V_r. Because $(\boldsymbol{a} - \boldsymbol{a}^*)$ is orthogonal on $\boldsymbol{a}$ we can write

$$||\, \boldsymbol{a}\, ||^2 = ||\, \boldsymbol{a}^*\, ||^2 + ||\, \boldsymbol{a} - \boldsymbol{a}^*\, ||^2 \tag{2.12}$$

Consider now

$$\begin{aligned}\mathrm{Var}(\boldsymbol{a'y}) &= \boldsymbol{a'}\mathrm{Var}(\boldsymbol{y})\boldsymbol{a} \\ &= \boldsymbol{a'I}\sigma_e^2\boldsymbol{a} \\ &= ||\, \boldsymbol{a}\, ||^2\ \sigma_e^2\end{aligned} \tag{2.13}$$

then using (2.12) we can write

$$\mathrm{Var}(\boldsymbol{a'y}) = ||\, \boldsymbol{a}^*\, ||^2\ \sigma_e^2 + ||\, \boldsymbol{a} - \boldsymbol{a}^*\, ||^2\ \sigma_e^2 \tag{2.14}$$

Note the fact that

$$\mathrm{Var}(\boldsymbol{a^{*\prime}y}) = \boldsymbol{a^{*\prime}I}\sigma_e^2\boldsymbol{a}^* = ||\, \boldsymbol{a}^*\, ||^2\ \sigma_e^2 \tag{2.15}$$

Taking in account the results of (2.14) and (2.15) we conclude that

$$\mathrm{Var}(\boldsymbol{a'y}) \geq \mathrm{Var}(\boldsymbol{a^{*\prime}y}) \tag{2.16}$$

with equality at $\boldsymbol{a^*} = \boldsymbol{a}$. As a result $\boldsymbol{a^{*\prime}y}$ is the unique linear unbiased estimator with minimum variance. This estimator will be called the best linear unbiased estimator (BLUE) of $\boldsymbol{\psi}$.

We need to show also that $\boldsymbol{a^{*\prime}y} = \boldsymbol{k'}\widehat{\boldsymbol{\beta}}$. Recall that $\boldsymbol{a^*} \in V_r$ and $(\boldsymbol{y} - \widehat{\boldsymbol{\eta}}) \perp V_r$where $\widehat{\boldsymbol{\eta}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$ is the projection of $\boldsymbol{y}$ on V_r. So

$$\boldsymbol{a^{*}}'(\boldsymbol{y} - \widehat{\boldsymbol{\eta}}) = \boldsymbol{0} \Rightarrow \boldsymbol{a^{*}}'\boldsymbol{y} = \boldsymbol{a^{*}}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} \tag{2.17}$$

But

$$\boldsymbol{k'\beta} = \mathrm{E}(\boldsymbol{a^{*\prime}y}) = \boldsymbol{a^{*}}'\boldsymbol{X\beta} \quad \forall\, \boldsymbol{\beta} \Rightarrow \boldsymbol{k'} = \boldsymbol{a^{*}}'\boldsymbol{X} \tag{2.18}$$

and as a result (2.17) and (2.18) imply that $\boldsymbol{a^{*\prime}y} = \boldsymbol{k'}\widehat{\boldsymbol{\beta}}$.

2.4 Solving normal equations

Gauss-Markoff theorem shows that the solutions of the normal equations provide the unique unbiased linear estimator with minimum variance. Now we need to solve the normal equations to obtain this solution. Let $\boldsymbol{X}$ be a $n \times p$ matrix with rank $r < p$. Then there are infinitely many solutions to $(\boldsymbol{X'X})\widehat{\boldsymbol{\beta}} = \boldsymbol{X'y}$. We need to consider the following situations:

1. Let $\boldsymbol{X}_r$ be a set of linearly independent columns of $\boldsymbol{X}$. Then

$$\mathrm{E}(\boldsymbol{y}) = \boldsymbol{X}_r\boldsymbol{\beta}_r = \boldsymbol{\eta} \tag{2.19}$$

where $\boldsymbol{\eta}$ is unique and $\widehat{\boldsymbol{\eta}} = \boldsymbol{X}_r\widehat{\boldsymbol{\beta}_r}$ where $\widehat{\boldsymbol{\beta}_r}$ is the solution to $(\boldsymbol{X'}_r\boldsymbol{X}_r)\widehat{\boldsymbol{\beta}_r} = \boldsymbol{X'}_r\boldsymbol{y}$.

2. Consider now $t = p - r$ side conditions. The side conditions are $\boldsymbol{H\beta} = \boldsymbol{0}$.

Definition 2.4.1 *$\boldsymbol{H}^{t\times p}$ is defined as a matrix of rank t with rows linearly independent of the rows of $\boldsymbol{X}$.*

Note that $\boldsymbol{H\beta}$ is not estimable. That is because $\boldsymbol{H}$ is linearly independent of $\boldsymbol{X}$. Note also that $\boldsymbol{X\beta} = \boldsymbol{\eta}$ has an infinite number of solutions. Here is where we will use the side conditions to obtain the desired solution.

Lemma 2.4.1 *There will be a unique value for $\boldsymbol{\beta}$ that satisfies*

$$(\boldsymbol{X'X})\boldsymbol{\beta} = \boldsymbol{\eta} \quad \textit{and} \quad \boldsymbol{H\beta} = \boldsymbol{0} \tag{2.20}$$

and a unique $\widehat{\boldsymbol{\beta}}$ for

$$\begin{aligned} &(\boldsymbol{X'X})\widehat{\boldsymbol{\beta}} = \boldsymbol{X'y} \quad \textit{and} \quad \boldsymbol{H}\widehat{\boldsymbol{\beta}} = \boldsymbol{0} \quad \textit{or} \\ &\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\eta}} \quad \textit{and} \quad \boldsymbol{H}\widehat{\boldsymbol{\beta}} = \boldsymbol{0} \end{aligned} \tag{2.21}$$

Proof: From the definition of $\boldsymbol{H}$, it follows that

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{H} \end{bmatrix}_{(n+t)\times p} \tag{2.22}$$

has full column rank p. Therefore if

$$\boldsymbol{G}\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \widehat{\boldsymbol{\eta}} \\ \boldsymbol{0} \end{bmatrix} \tag{2.23}$$

has a solution, it will be unique. We still have to show that there is a solution. Now if $\begin{bmatrix} \boldsymbol{G} & \widehat{\boldsymbol{\eta}} \\ & \boldsymbol{0} \end{bmatrix}$ and $\boldsymbol{G}$ have the same rank then there is a solution to (2.23).We can write

$$\begin{bmatrix} \boldsymbol{G} & \widehat{\boldsymbol{\eta}} \\ & \boldsymbol{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X} & \widehat{\boldsymbol{\eta}} \\ \boldsymbol{H} & \boldsymbol{0} \end{bmatrix} \Rightarrow \text{rank of} \quad \begin{bmatrix} \boldsymbol{G} & \widehat{\boldsymbol{\eta}} \\ & \boldsymbol{0} \end{bmatrix} \quad \text{is} \quad p \tag{2.24}$$

due to the fact that the rows of $\boldsymbol{H}$ are linearly independent of the rows of $\boldsymbol{X}$. Based on the results derived in (2.23), (2.22) and (2.24) we conclude that there is a solution and because $\boldsymbol{G}$ is a full rank matrix the solution is unique. In order to obtain $\widehat{\boldsymbol{\beta}}$ premultiply

$$\boldsymbol{G}\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \boldsymbol{X}\widehat{\boldsymbol{\beta}} \\ \boldsymbol{0} \end{bmatrix}$$

by $\boldsymbol{G'} = \begin{bmatrix} \boldsymbol{X'} & \boldsymbol{H'} \end{bmatrix}$ and as a result

$$\boldsymbol{G'G}\widehat{\boldsymbol{\beta}} = \boldsymbol{X'X}\widehat{\boldsymbol{\beta}} \tag{2.25}$$

Note that $\boldsymbol{G'G} = \boldsymbol{X'X} + \boldsymbol{H'H}$ and as a result we can write

$$\begin{aligned} &(\boldsymbol{X'X} + \boldsymbol{H'H})\widehat{\boldsymbol{\beta}} = \boldsymbol{X'X}\widehat{\boldsymbol{\beta}} \quad \text{but also} \\ &\boldsymbol{X'X}\widehat{\boldsymbol{\beta}} = \boldsymbol{X'y} \quad \text{and as a result} \\ &(\boldsymbol{X'X} + \boldsymbol{H'H})\widehat{\boldsymbol{\beta}} = \boldsymbol{X'y} \end{aligned} \tag{2.26}$$

which has a unique solution because $\boldsymbol{G'G}$ is a full rank matrix. The unique solution is given by

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X'X} + \boldsymbol{H'H})^{-1}\boldsymbol{X'y} \tag{2.27}$$

3. From $(\boldsymbol{X'X})\widehat{\boldsymbol{\beta}} = \boldsymbol{X'y}$ using the generalized inverse concept we obtain $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X'X})^{-}\boldsymbol{X'y}$.

Lemma 2.4.2 *If $\psi = \boldsymbol{K'\beta}$ is estimable, then by Gauss-Markoff theorem BLUE of $(\boldsymbol{K'\beta}) = \boldsymbol{K'}\widehat{\boldsymbol{\beta}}$.*

Proof:

$$\begin{aligned}\mathrm{E}(\boldsymbol{K'}\widehat{\boldsymbol{\beta}}) &= \mathrm{E}(\boldsymbol{K'}(\boldsymbol{X'X})^{-}\boldsymbol{X'y}) \\ &= \boldsymbol{K'}(\boldsymbol{X'X})^{-}\boldsymbol{X'}\mathrm{E}(\boldsymbol{y}) \\ &= \boldsymbol{K'}(\boldsymbol{X'X})^{-}\boldsymbol{X'X\beta}\end{aligned} \tag{2.28}$$

But because $\boldsymbol{\psi} = \boldsymbol{K'\beta}$ is estimable, we have also

$$\mathrm{E}(\boldsymbol{K'}\widehat{\boldsymbol{\beta}}) = \boldsymbol{K'\beta} \quad \forall\, \boldsymbol{\beta} \tag{2.29}$$

As a result

$$\boldsymbol{K'}(\boldsymbol{X'X})^{-}\boldsymbol{X'X\beta} = \boldsymbol{K'\beta} \quad \forall\, \boldsymbol{\beta} \tag{2.30}$$

and this implies $\boldsymbol{K'}(\boldsymbol{X'X})^{-}\boldsymbol{X'X} = \boldsymbol{K'}$ if $\boldsymbol{\psi}$ is estimable. This is a necessary condition, now we need also to show that it is sufficient. We need to show that if $\boldsymbol{K'}(\boldsymbol{X'X})^{-}\boldsymbol{X'X} = \boldsymbol{K'}$ then $\boldsymbol{\psi}$ is estimable. As a result we need to find $\boldsymbol{A'}$ such that $\mathrm{E}(\boldsymbol{A'y}) = \boldsymbol{K'\beta}$. Take $\boldsymbol{A'} = \boldsymbol{K'}(\boldsymbol{X'X})^{-}\boldsymbol{X'}$ then we can write

$$\begin{aligned}\mathrm{E}(\boldsymbol{A'y}) &= \mathrm{E}(\boldsymbol{K'}(\boldsymbol{X'X})^{-}\boldsymbol{X'y}) \\ &= \boldsymbol{K'}(\boldsymbol{X'X})^{-}\boldsymbol{X'}\mathrm{E}(\boldsymbol{y}) \\ &= \boldsymbol{K'}(\boldsymbol{X'X})^{-}\boldsymbol{X'X\beta} \\ &= \boldsymbol{K'\beta}\end{aligned} \tag{2.31}$$

Now we consider

$$\begin{aligned}\mathrm{Var}(\boldsymbol{K'}\widehat{\boldsymbol{\beta}}) &= \boldsymbol{K'}\mathrm{Var}(\widehat{\boldsymbol{\beta}})\boldsymbol{K} \\ &= \boldsymbol{K'}(\boldsymbol{X'X})^{-}\boldsymbol{X'I}\sigma_e^2\boldsymbol{X}(\boldsymbol{X'X})^{-}\boldsymbol{K} \\ &= \boldsymbol{K'}(\boldsymbol{X'X})^{-}\boldsymbol{X'X}(\boldsymbol{X'X})^{-}\boldsymbol{K}\sigma_e^2\end{aligned} \tag{2.32}$$

where if $\boldsymbol{K'\beta}$ is estimable then $\boldsymbol{K'}(\boldsymbol{X'X})^{-}\boldsymbol{X'X} = \boldsymbol{K'}$ and as a result we can write

$$\mathrm{Var}(\boldsymbol{K'}\widehat{\boldsymbol{\beta}}) = \boldsymbol{K'}(\boldsymbol{X'X})^{-}\boldsymbol{K}\sigma_e^2 \tag{2.33}$$

2.5 Generalized least squares

Consider the case where $\mathrm{Var}(\boldsymbol{y}) = \boldsymbol{R}\sigma_e^2 \neq \boldsymbol{I}\sigma_e^2$ and $\boldsymbol{R}$ is known. The underlying assumptions are: $\mathrm{E}(\boldsymbol{y}) = \boldsymbol{X\beta}, \mathrm{Var}(\boldsymbol{y}) = \boldsymbol{R}\sigma_e^2$. There is a $n \times n$ matrix $\boldsymbol{P}$ with $\boldsymbol{P'RP} = \boldsymbol{I}$. Let

$$\begin{aligned}&\tilde{\boldsymbol{y}} = \boldsymbol{P'y} \\ &\mathrm{E}(\tilde{\boldsymbol{y}}) = \boldsymbol{P'X\beta} = \tilde{\boldsymbol{X}}\boldsymbol{\beta} \\ &\mathrm{Var}(\tilde{\boldsymbol{y}}) = \boldsymbol{P'RP}\sigma_e^2 = \boldsymbol{I}\sigma_e^2\end{aligned} \tag{2.34}$$

so, $\widehat{\boldsymbol{\beta}}$ can be obtained from:

$$\begin{aligned}&\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}\widehat{\boldsymbol{\beta}} = \tilde{\boldsymbol{X}}'\tilde{\boldsymbol{y}} \\ &(\boldsymbol{X'PP'X})\widehat{\boldsymbol{\beta}} = \boldsymbol{X'PP'y}\end{aligned} \tag{2.35}$$

and using the fact that $\boldsymbol{PP'} = \boldsymbol{R}^{-1}$

$$(\boldsymbol{X'R}^{-1}\boldsymbol{X})\widehat{\boldsymbol{\beta}} = \boldsymbol{X'R}^{-1}\boldsymbol{y} \tag{2.36}$$

These are the generalized least squares (*GLS*) equations and by Gauss-Markoff theorem, *BLUE*($\boldsymbol{K'\beta}$)= $\boldsymbol{K'}\widehat{\boldsymbol{\beta}}$. Note also that

$$\mathrm{E}(\boldsymbol{K'}\widehat{\boldsymbol{\beta}}) = \boldsymbol{K'}(\boldsymbol{X'R}^{-1}\boldsymbol{X})^{-}\boldsymbol{X'R}^{-1}\boldsymbol{X\beta} = \boldsymbol{K'\beta} \tag{2.37}$$

for all $\boldsymbol{\beta}$. So $\boldsymbol{K'}(\boldsymbol{X'R}^{-1}\boldsymbol{X})^{-}\boldsymbol{X'R}^{-1}\boldsymbol{X} = \boldsymbol{K'}$ if $\boldsymbol{K'\beta}$ is estimable. The reverse is also true, that is if $\boldsymbol{K'}(\boldsymbol{X'R}^{-1}\boldsymbol{X})^{-}\boldsymbol{X'R}^{-1}\boldsymbol{X} = \boldsymbol{K'}$ then $\boldsymbol{K'\beta}$ is estimable. That is due to the fact that there is a linear function $\boldsymbol{A'y}$ that has $\mathrm{E}(\boldsymbol{A'y}) = \boldsymbol{K'\beta}$. Note that the above result is true for any covariance matrix $\boldsymbol{R}$.

2.6 Canonical form of linear model

Let $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_r\}$ be an orthonormal basis for $V_r \subseteq V_n$, the space spanned by the columns of $\boldsymbol{X}$. Let the basis be extended to $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_n\}$ an orthonormal basis for for V_n. Note that this can be done for example by applying the Gramm-Schmidt process to $[\boldsymbol{X}\ \boldsymbol{I}]$ and taking the non-zero columns. Let $\boldsymbol{P'}_1 = [\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r]$, $\boldsymbol{P'}_2 = [\boldsymbol{\alpha}_{r+1}, \ldots, \boldsymbol{\alpha}_n]$, $\boldsymbol{P'} = [\boldsymbol{P'}_1, \boldsymbol{P'}_2]$. Also, $\boldsymbol{P'P} = \boldsymbol{I} = \boldsymbol{PP'}$. Then

$$\begin{aligned}\boldsymbol{y} &= \boldsymbol{P'z} = \boldsymbol{P'}_1\boldsymbol{z}_1 + \boldsymbol{P'}_2\boldsymbol{z}_2 \\ \boldsymbol{z} &= \boldsymbol{P'y} = \begin{bmatrix}\boldsymbol{P}_1\\ \boldsymbol{P}_2\end{bmatrix}\boldsymbol{y} = \begin{bmatrix}\boldsymbol{z}_1\\ \boldsymbol{z}_2\end{bmatrix}\end{aligned} \tag{2.38}$$

where $z's$ are the coordinates of y in this new basis. Also

$$\mathrm{E}(\boldsymbol{z}) = \begin{bmatrix}\boldsymbol{P}_1\\ \boldsymbol{P}_2\end{bmatrix}\boldsymbol{X\beta} = \begin{bmatrix}\boldsymbol{P}_1\boldsymbol{X\beta}\\ \boldsymbol{P}_2\boldsymbol{X\beta}\end{bmatrix} = \begin{bmatrix}\boldsymbol{\xi}_1\\ \boldsymbol{\xi}_2\end{bmatrix}$$

and because the rows of $\boldsymbol{P}_2$ are perpendicular to V_r

$$\mathrm{E}(\boldsymbol{z}) = \begin{bmatrix}\boldsymbol{\xi}_1\\ \boldsymbol{0}\end{bmatrix}$$

Also, $\mathrm{Var}(\boldsymbol{z}) = \boldsymbol{P}\mathrm{Var}(\boldsymbol{y})\boldsymbol{P'} = \boldsymbol{PP'}\sigma_e^2 = \boldsymbol{I}\sigma_e^2$. We can write $\boldsymbol{y} = \boldsymbol{P'}_1\boldsymbol{z}_1 + \boldsymbol{P'}_2\boldsymbol{z}_2$. Now recall that columns of $\boldsymbol{P'}_1 \in V_r$ and columns of $\boldsymbol{P'}_2 \perp V_r$. So, $\boldsymbol{P'}_1\boldsymbol{z}_1 = \widehat{\boldsymbol{\eta}}$, the projection of $\boldsymbol{y}$ onto V_r. So,

$$\begin{aligned}\boldsymbol{y} - \widehat{\boldsymbol{\eta}} &= (\boldsymbol{P'}_1\boldsymbol{z}_1 + \boldsymbol{P'}_2\boldsymbol{z}_2) - \boldsymbol{P'}_1\boldsymbol{z}_1 \\ &= \boldsymbol{P'}_2\boldsymbol{z}_2\end{aligned}$$

and

$$\begin{aligned}\boldsymbol{Q} &= (\boldsymbol{y} - \widehat{\boldsymbol{\eta}})'(\boldsymbol{y} - \widehat{\boldsymbol{\eta}}) \\ &= \boldsymbol{z'}_2\boldsymbol{P}_2\boldsymbol{P'}_2\boldsymbol{z}_2 \\ &= \boldsymbol{z'}_2\boldsymbol{z}_2 \\ &= \sum_{i=r+1}^{n} \boldsymbol{z}_i^2\end{aligned}$$

represents the residual sum of squares. Because $\mathrm{E}(\boldsymbol{z}_2) = 0, \mathrm{Var}(\boldsymbol{z}_2) = \boldsymbol{I}\sigma_e^2 \Rightarrow \mathrm{E}(\boldsymbol{z}_i^2) = \sigma_e^2$. Here r is the rank of the X matrix. The estimate used for σ_e^2 is given by

$$\begin{aligned}\widehat{\sigma}_e^2 &= \frac{\sum_{i=r+1}^{n} \boldsymbol{z}_i^2}{n-r} = \frac{(\boldsymbol{y}-\widehat{\boldsymbol{\eta}})'(\boldsymbol{y}-\widehat{\boldsymbol{\eta}})}{n-r} \\ \mathrm{E}(\widehat{\sigma}_e^2) &= \sigma_e^2\end{aligned}$$

Theorem 2.6.1 *$\boldsymbol{z}_1$ is called the estimation space, if for any estimable function $\boldsymbol{\psi}$, it's least square (BLUE) estimate $\widehat{\boldsymbol{\psi}}$, is a linear function of $\boldsymbol{z}_1$.*

Proof: Recall that $\widehat{\boldsymbol{\psi}} = \boldsymbol{a}^{*\prime}\boldsymbol{y}$ where $\boldsymbol{a}^* \in V_r$ and we can also write

$$\begin{aligned}\widehat{\boldsymbol{\psi}} &= \boldsymbol{a}^{*\prime}\boldsymbol{P}'\boldsymbol{z} \\ &= \boldsymbol{a}^{*\prime}(\boldsymbol{P}'_1\boldsymbol{z}_1 + \boldsymbol{P}'_2\boldsymbol{z}_2) \\ &= \boldsymbol{a}^{*\prime}\boldsymbol{P}'_1\boldsymbol{z}_1\end{aligned} \tag{2.39}$$

because $\boldsymbol{a}^* \perp$ on the columns of $\boldsymbol{P}'_2$. Also $\boldsymbol{z}_2$ is called the *error space* because $\mathrm{E}(\boldsymbol{z}_2) = \boldsymbol{0}$ and σ_e^2 is estimated for $\boldsymbol{z}_2$.

2.7 Hypothesis Testing

The following assumption is made

$$\Omega : \boldsymbol{y}^{n\times 1} \sim \aleph\left(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{I}\sigma_e^2\right) \tag{2.40}$$

where the rank of $\boldsymbol{X}^{n\times p} = r$.

Theorem 2.7.1 *Under Ω*

$$\widehat{\boldsymbol{\psi}} = \boldsymbol{K}'\widehat{\boldsymbol{\beta}} \sim \aleph\left(\boldsymbol{K}'\boldsymbol{\beta}, \boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{K}\boldsymbol{I}\sigma_e^2\right) \tag{2.41}$$

and is independent of $\frac{\boldsymbol{Q}_\Omega}{\sigma_e^2} \sim \boldsymbol{\chi}^2_{n-r}$.

Proof: From Gauss-Markoff we know that $\mathrm{E}(\boldsymbol{K}'\widehat{\boldsymbol{\beta}}) = \boldsymbol{K}'\boldsymbol{\beta}$ and that $\mathrm{Var}(\boldsymbol{K}'\widehat{\boldsymbol{\beta}}) = \boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{K}\boldsymbol{I}\sigma_e^2$. Also a linear function of normal random variables is normal. To show the independence recall that $\boldsymbol{K}'\boldsymbol{\beta}$ is a linear function of $\boldsymbol{z}_1$ and $\boldsymbol{Q}$ is a function of $\boldsymbol{z}_2$. Under Ω, $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ are independent (under normality, null covariance implies independence). Also $\frac{\boldsymbol{z}_i}{\sigma_e^2} \sim \aleph(0,1)$ for $i > r$. So

$$\frac{\boldsymbol{Q}_\Omega}{\sigma_e^2} = \sum_{i=r+1}^{n}\left(\frac{\boldsymbol{z}_i}{\sigma_e}\right)^2 \sim \boldsymbol{\chi}^2_{n-r} \tag{2.42}$$

Notation 1 *$\omega = H \cap \Omega$ is the set of assumptions obtained by imposing the assumptions of hypothesis H in addition to the assumptions Ω.*

Definition 2.7.1 *Let $f(\boldsymbol{y};\boldsymbol{\theta})$ denote the probability density function of $\boldsymbol{y}$. Then the likelihood ratio statistic for testing H is*

$$\lambda = \frac{max_{\omega} f(\boldsymbol{y};\boldsymbol{\theta})}{max_{\Omega} f(\boldsymbol{y};\boldsymbol{\theta})} \tag{2.43}$$

Two forms of Ω and ω assumptions

1. $\Omega_1 : \boldsymbol{y} \sim \boldsymbol{N}\left(\boldsymbol{X\beta}, \boldsymbol{I}\sigma_e^2\right)$ rank $\boldsymbol{X}^{n\times p} = r$

   $H_1 : \psi_1 = \psi_2 = \ldots = \psi_q = 0$ where ψ_i are linearly independent estimable functions, or there exists $\boldsymbol{K'}_{q\times p}$ such that $\psi_i = \boldsymbol{K'\beta} = 0$, meanining that, $\boldsymbol{K'}$ has q linearly independent rows.

2. $\Omega_2 : \boldsymbol{y} \sim \boldsymbol{N}\left(\boldsymbol{\eta}, \boldsymbol{I}\sigma_e^2\right)$ $\boldsymbol{\eta} \in V_r \subset V_n$

   $H_2 : \boldsymbol{\eta} \in V_{r-q} \subset V_r \subset V_n$

   V_r is the space spanned by the columns of $\boldsymbol{X}$ and V_{r-q} is the subspace to which $\boldsymbol{\eta}$ is restricted to by $H : \psi_1 = \psi_2 = \ldots = \psi_q = 0$

It is obvious that $\Omega_1 = \Omega_2 = \Omega$. Need to show that $\omega_1 = H_1 \cap \Omega$ is equivalent to $\omega_2 = H_2 \cap \Omega$. Let $BLUE(\psi = \boldsymbol{K'\beta} = \boldsymbol{A'y}$. Note that $BLUE(\boldsymbol{K'\beta}) = \boldsymbol{K'}\widehat{\boldsymbol{\beta}} = \boldsymbol{K'}(\boldsymbol{X'X})^{-}\boldsymbol{X'y}$, so $\boldsymbol{A'} = \boldsymbol{K'}(\boldsymbol{X'X})^{-}\boldsymbol{X'}$. Also

$$\mathrm{E}(\boldsymbol{A'y}) = \boldsymbol{A'\eta} = \boldsymbol{A'X\beta} = \boldsymbol{K'\beta}$$

and because this holds for all $\boldsymbol{\beta}, \Rightarrow \boldsymbol{A'X} = \boldsymbol{K'}$. Note that rank($\boldsymbol{K'}$)= q = rank($\boldsymbol{A'X}$) $\leq$ rank($\boldsymbol{A}^{'q\times n}$) $\leq q \Rightarrow$ rank($\boldsymbol{A'}$) $= q$. Also from the Gauss-Markoff Theorem, the i^{th} row of $\boldsymbol{A'}, \boldsymbol{a}_i \in V_r$. Under $H_1, \boldsymbol{\psi} = \boldsymbol{A'\eta} = \boldsymbol{0} \Rightarrow \boldsymbol{\eta} \perp V_A$, where V_A is the q dimensional subspace of V_r spanned by the rows of $\boldsymbol{A'}$. The set of all vectors in V_r that are orthogonal to V_A is an $(r-q)$ dimensional subspace $V^*_{(r-q)} \subset V_r$. Thus, $\omega_1 \Rightarrow \boldsymbol{\eta} \in V^*_{(r-q)}$ or ω_1 implies ω_2 with $V_{(r-q)} = V^*_{(r-q)}$. Now to show that ω_2 implies ω_1, let V_A^* be the orthocomplement of $V_{(r-q)}$ in V_r. Let $\{\boldsymbol{a}_1^*, \ldots, \boldsymbol{a}_q^*\}$ be a set of vectors that span V_A^* and

$$\boldsymbol{A}^{*'} = \begin{bmatrix} \boldsymbol{a}_1^{*'} \\ \vdots \\ \boldsymbol{a}_q^{*'} \end{bmatrix}$$

Let $\boldsymbol{K}^{*'} = \boldsymbol{A}^{*'}\boldsymbol{X}$, then

$$\psi^* = \boldsymbol{K}^{*'}\boldsymbol{\beta} = \boldsymbol{A}^{*'}\boldsymbol{X\beta} = \boldsymbol{A}^{*'}\boldsymbol{\eta} = \boldsymbol{0}$$

because ω_2 specifies $\boldsymbol{\eta} \perp V_A^*$. Need to show that there are q linearly independent ψ^* or that rank($\boldsymbol{K}^{*,q\times n}$) $= q$. Suppose there is some $\boldsymbol{c'}$ such that

$$\boldsymbol{c'K}^* = \boldsymbol{c'A}^{*'}\boldsymbol{X} = \boldsymbol{u'X} = \boldsymbol{0}$$

where $\boldsymbol{u} = \boldsymbol{A}^*\boldsymbol{c}$. This implies $\boldsymbol{u} \perp V_r$ but also $\boldsymbol{u} \in V_r$ so $\Rightarrow \boldsymbol{u} = \boldsymbol{0} \Rightarrow \boldsymbol{A}^*\boldsymbol{c} = \boldsymbol{0}$. This fact implies that $\boldsymbol{c} = \boldsymbol{0}$ and that means that $\boldsymbol{K}$ has q linearly independent columns and so ω_2 implies ω_1. Now we can conclude that $\omega_1 = \omega_2$.

Consider now

$$f(\boldsymbol{y};\boldsymbol{\theta}) = (2\pi\sigma_e^2)^{-\frac{n}{2}}\exp\{-\frac{1}{2\sigma_e^2} \mid\mid \boldsymbol{y}-\boldsymbol{\eta} \mid\mid^2\} \tag{2.44}$$

Need to find the maximum of (2.44) for $0 < \sigma_e^2 < \infty$ and $\boldsymbol{\eta} \in V_r$ for Ω and $\boldsymbol{\eta} \in V_{r-q}$ for ω. First fix σ_e^2 and maximize with respect to ω. This is achived by letting $\boldsymbol{\eta}$ be the projection of $\boldsymbol{y}$ on V_r (for Ω) or V_{r-q} (for ω). Thus, maximum of (2.44) for fixed σ_e^2 is

$$(2\pi\sigma_e^2)^{-\frac{n}{2}}\exp\{-\frac{1}{2\sigma_e^2} \mid\mid \boldsymbol{y}-\widehat{\boldsymbol{\eta}} \mid\mid^2\} \tag{2.45}$$

Taking the logarithm of (2.45) and then taking the derivative with respect to σ_e^2 results in

$$\frac{\partial}{\partial\sigma_e^2}\left(-\frac{n}{2}\log(2\pi\sigma_e^2) - \frac{1}{2\sigma_e^2} \mid\mid \boldsymbol{y}-\widehat{\boldsymbol{\eta}} \mid\mid^2\right) = -\frac{n}{2\sigma_e^2} + \frac{\mid\mid \boldsymbol{y}-\widehat{\boldsymbol{\eta}} \mid\mid^2}{2\sigma_e^4} \tag{2.46}$$

Setting (2.46) to 0 and then solving it gives

$$\widehat{\sigma}_e^2 = \frac{\mid\mid \boldsymbol{y}-\widehat{\boldsymbol{\eta}} \mid\mid^2}{n} \tag{2.47}$$

So

$$\begin{aligned} \max\nolimits_{\Omega} f(\boldsymbol{y};\boldsymbol{\theta}) &= \left(\frac{2\pi \mid\mid \boldsymbol{y}-\widehat{\boldsymbol{\eta}}_\Omega \mid\mid^2}{n}\right)^{\frac{-n}{2}} \exp\left(-\frac{n}{2}\right) \\ \max\nolimits_{\omega} f(\boldsymbol{y};\boldsymbol{\theta}) &= \left(\frac{2\pi \mid\mid \boldsymbol{y}-\widehat{\boldsymbol{\eta}}_\omega \mid\mid^2}{n}\right)^{\frac{-n}{2}} \exp\left(-\frac{n}{2}\right) \end{aligned} \tag{2.48}$$

and

$$\lambda = \left(\frac{\mid\mid \boldsymbol{y}-\widehat{\boldsymbol{\eta}}_\omega \mid\mid^2}{\mid\mid \boldsymbol{y}-\widehat{\boldsymbol{\eta}}_\Omega \mid\mid^2}\right)^{-\frac{n}{2}} = \left(\frac{\boldsymbol{Q}_\omega}{\boldsymbol{Q}_\Omega}\right)^{-\frac{n}{2}} \tag{2.49}$$

In practice,

$$\begin{aligned} F(x) = F &= \frac{n-r}{q}\left(\lambda^{-\frac{2}{n}} - 1\right) \\ &= \frac{n-r}{q}\left(\frac{\boldsymbol{Q}_\omega - \boldsymbol{Q}_\Omega}{\boldsymbol{Q}_\Omega}\right) \end{aligned} \tag{2.50}$$

is used. Note

$$\lambda < \lambda_0 \quad \text{if and only if} \quad F > F_0 \quad \text{when} \quad F_0 = F_0(\lambda_0)$$

So rejecting H when $\lambda < \lambda_0$ is equivalent to rejecting H when $F > F_0$.

Definition 2.7.2 *Let $V_{r-q} \subset V_r$ be the subspace that η is restricted to by H.*

Under $\Omega, \eta \in V_r \subset V_n$ let $\{\boldsymbol{\alpha}_{q+1}, \ldots, \boldsymbol{\alpha}_r\}$ be an orthonormal basis for V_{r-q}. Extend this set to $\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_q, \boldsymbol{\alpha}_{q+1}, \ldots, \boldsymbol{\alpha}_r\}$ to be an orthonormal basis for V_r. Extend once more to $\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_q, \boldsymbol{\alpha}_{q+1}, \ldots, \boldsymbol{\alpha}_n\}$ to be an orthonormal basis for V_n.

$$\underbrace{\underbrace{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_q, \underbrace{\boldsymbol{\alpha}_{q+1} \ldots, \boldsymbol{\alpha}_r}_{O.N.B.forV_{r-q}}}_{O.N.B.forV_r}, \boldsymbol{\alpha}_{r+1} \ldots, \boldsymbol{\alpha}_n}_{O.N.B.forV_n} \tag{2.51}$$

Let

$$\begin{aligned}
\boldsymbol{P'}_1 &= \left[\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_q\right] \\
\boldsymbol{P'}_2 &= \left[\boldsymbol{\alpha}_{q+1}, \ldots, \boldsymbol{\alpha}_r\right] \\
\boldsymbol{P'}_3 &= \left[\boldsymbol{\alpha}_{r+1}, \ldots, \boldsymbol{\alpha}_n\right] \\
\boldsymbol{P'} &= \left[\boldsymbol{P'}_1 \boldsymbol{P'}_2 \boldsymbol{P'}_3\right]
\end{aligned}$$

Then,

$$\boldsymbol{y} = \boldsymbol{P'z} = \underbrace{\boldsymbol{P'}_1 \boldsymbol{z}_1 + \boldsymbol{P'}_2 \boldsymbol{z}_2}_{\in V_r} + \underbrace{\boldsymbol{P'}_3 \boldsymbol{z}_3}_{\perp V_r}$$

where $\boldsymbol{z} = \boldsymbol{Py}$, and also recall $\boldsymbol{P} = \boldsymbol{P'}^{-1}$. Also $\boldsymbol{z} \sim \boldsymbol{N}\left(\boldsymbol{\zeta}, \boldsymbol{I}\sigma_e^2\right)$ where $\boldsymbol{\zeta} = \boldsymbol{PX\beta} = \boldsymbol{P\eta}$. Note also that $\boldsymbol{y} = \widehat{\boldsymbol{\eta}}_\Omega + \boldsymbol{P'}_3 \boldsymbol{z}_3$. Note that

$$\begin{array}{llll}
\text{under} \quad \Omega: & \zeta_i = 0 & \text{for} \quad i > r & \eta \in V_r, \text{ and} \\
\text{under} \quad \omega: & \zeta_i = 0 & \text{for} \quad \begin{cases} i \le q \\ i > r \end{cases} & \eta \in V_{r-q}
\end{array}$$

From the Gauss-Markoff Theorem,

$$\begin{aligned}
\widehat{\boldsymbol{\eta}}_\Omega &= \boldsymbol{P'}_1 \boldsymbol{z}_1 + \boldsymbol{P'}_2 \boldsymbol{z}_2 \\
\widehat{\boldsymbol{\eta}}_\omega &= \boldsymbol{P'}_2 \boldsymbol{z}_2
\end{aligned}$$

So

$$\begin{aligned}
\boldsymbol{Q}_\Omega &= \| \boldsymbol{y} - \widehat{\boldsymbol{\eta}}_\Omega \|^2 = \| \boldsymbol{P'}_3 \boldsymbol{z}_3 \|^2 = \boldsymbol{z'}_3 \boldsymbol{P}_3 \boldsymbol{P'}_3 \boldsymbol{z}_3 = \boldsymbol{z'}_3 \boldsymbol{z}_3 \\
&= \sum_{i=r+1}^{n} \boldsymbol{z}_i^2 \\
\boldsymbol{Q}_\omega &= \| \boldsymbol{y} - \widehat{\boldsymbol{\eta}}_\omega \|^2 = \| \boldsymbol{P'}_1 \boldsymbol{z}_1 + \boldsymbol{P'}_3 \boldsymbol{z}_3 \|^2 = \boldsymbol{z'}_1 \boldsymbol{P}_1 \boldsymbol{P'}_1 \boldsymbol{z}_1 + \boldsymbol{z'}_3 \boldsymbol{P}_3 \boldsymbol{P'}_3 \boldsymbol{z}_3 \\
&= \boldsymbol{z'}_1 \boldsymbol{z}_1 + \boldsymbol{z'}_3 \boldsymbol{z}_3 \\
&= \sum_{i=1}^{q} \boldsymbol{z}_i^2 + \sum_{i=r+1}^{n} \boldsymbol{z}_i^2
\end{aligned}$$

and

$$\boldsymbol{Q}_\omega - \boldsymbol{Q}_\Omega = \sum_{i=1}^{q} \boldsymbol{z}_i^2$$

Note that $\boldsymbol{z}_1$ and $\boldsymbol{z}_3$ are statistically independent. As a result $\boldsymbol{Q}_\omega - \boldsymbol{Q}_\Omega$ and $\boldsymbol{Q}_\Omega$ are statistically independent.

$$\text{Under } \omega\text{:} \quad \frac{z_i}{\sigma_e} \quad \text{iid} \quad \boldsymbol{N}(0,1) \quad \begin{cases} i < q \\ i > r \end{cases}$$

So,

$$\frac{\boldsymbol{Q}_\omega - \boldsymbol{Q}_\Omega}{\sigma_e^2} = \sum_{i=1}^{q} \left(\frac{z_i}{\sigma_e}\right)^2 \sim \boldsymbol{\chi}_q^2 \tag{2.52}$$

and

$$\frac{\boldsymbol{Q}_\Omega}{\sigma_e^2} = \sum_{i=r+1}^{n} \left(\frac{z_i}{\sigma_e}\right)^2 \sim \boldsymbol{\chi}_{n-r}^2 \tag{2.53}$$

So,

$$\begin{aligned} \boldsymbol{F} &= \frac{n-r}{q}\left(\frac{\boldsymbol{Q}_\omega - \boldsymbol{Q}_\Omega}{\boldsymbol{Q}_\Omega}\right) = \frac{(\boldsymbol{Q}_\omega - \boldsymbol{Q}_\Omega)/q}{\sigma_e^2} \\ &\sim \frac{\boldsymbol{\chi}_q^2/q}{\boldsymbol{\chi}_{n-r}^2/(n-r)} \\ &= \boldsymbol{F}_{q,n-r} \end{aligned} \tag{2.54}$$

$$\text{Under } \omega\text{:} \quad \Pr(\boldsymbol{F} > \boldsymbol{F}_{\alpha;q,n-r}) = \alpha \tag{2.55}$$

$$\begin{aligned} \text{Under } \Omega\text{:} \quad \frac{z_i}{\sigma_e} &\sim \boldsymbol{N}(\zeta_i, \sigma_e) \quad i < r \\ &\sim \boldsymbol{N}(0,1) \quad i > r \end{aligned}$$

So,

$$\frac{\boldsymbol{Q}_\omega - \boldsymbol{Q}_\Omega}{\sigma_e^2} = \sum_{i=1}^{q} \left(\frac{z_i}{\sigma_e}\right)^2 \sim \boldsymbol{\chi}_{q,\delta}^2 \tag{2.56}$$

where δ represents a non-centrality parameter and is equal to $\delta = \sum_{i=1}^{q} \frac{\zeta_i^2}{\sigma_e^2}$. So under $\Omega : \boldsymbol{F} \sim \boldsymbol{F}_{q,n-r;\delta}$.

Consider now another derivation of $\boldsymbol{F}$ and it's distribution.

$$\begin{aligned} \text{Under } \omega: \quad & \boldsymbol{K'\beta} = \boldsymbol{0} \\ \text{and} \quad & \boldsymbol{K'}\widehat{\boldsymbol{\beta}} \sim \boldsymbol{N}\left(\boldsymbol{0}, \boldsymbol{K'}(\boldsymbol{X'X})^{-}\boldsymbol{K}\sigma_e^2\right) \end{aligned}$$

So

$$\underbrace{\left(\boldsymbol{K'}\widehat{\boldsymbol{\beta}}\right)' \left[\boldsymbol{K'}(\boldsymbol{X'X})^{-}\boldsymbol{K}\sigma_e^2\right]^{-1} \left(\boldsymbol{K'}\widehat{\boldsymbol{\beta}}\right)}_{SS_H} \sim \boldsymbol{\chi}_q^2$$

Also we saw from cannonical form for Ω that any $\widehat{\boldsymbol{\psi}}$ and $\boldsymbol{Q}_\Omega$ are statistically independent. Also we know that $\frac{\boldsymbol{Q}_\Omega}{\sigma_e^2} \sim \boldsymbol{\chi}^2_{n-r}$. So

$$\begin{aligned}
\boldsymbol{F} &= \frac{\left(\boldsymbol{K'}\widehat{\boldsymbol{\beta}}\right)' \left[\boldsymbol{K'}\left(\boldsymbol{X'X}\right)^{-}\boldsymbol{K}\sigma_e^2\right]^{-1}\left(\boldsymbol{K'}\widehat{\boldsymbol{\beta}}\right)/q}{\boldsymbol{Q}_\Omega/\sigma_e^2/(n-r)} \\
&= \frac{\left(\boldsymbol{K'}\widehat{\boldsymbol{\beta}}\right)' \left[\boldsymbol{K'}\left(\boldsymbol{X'X}\right)^{-}\boldsymbol{K}\sigma_e^2\right]^{-1}\left(\boldsymbol{K'}\widehat{\boldsymbol{\beta}}\right)/q}{\widehat{\sigma}_e^2} \\
&\sim \boldsymbol{F}_{q,n-r}
\end{aligned}$$

$$\text{Under } \Omega: \quad \boldsymbol{F} \sim \boldsymbol{F}_{q,n-r;\delta} \tag{2.57}$$

where $\delta = (\boldsymbol{K'}\boldsymbol{\beta})'\left[\boldsymbol{K'}\left(\boldsymbol{X'X}\right)^{-}\boldsymbol{K}\sigma_e^2\right]^{-1}(\boldsymbol{K'}\boldsymbol{\beta})$. These two forms of the $\boldsymbol{F}$ statistic are equivalent. The denominators in each is $\widehat{\sigma}_e^2$. So, all we need to show is that

$$\boldsymbol{Q}_\omega - \boldsymbol{Q}_\Omega = \left(\boldsymbol{K'}\widehat{\boldsymbol{\beta}}\right)'\left[\boldsymbol{K'}\left(\boldsymbol{X'X}\right)^{-}\boldsymbol{K}\sigma_e^2\right]^{-1}\left(\boldsymbol{K'}\widehat{\boldsymbol{\beta}}\right) \tag{2.58}$$

Proposition 2.7.1 *Let* $\boldsymbol{\psi}^*_{(q\times 1)} = \boldsymbol{D}_{(q\times q)}\boldsymbol{\psi}$. *Note that* $\boldsymbol{\psi} = \boldsymbol{K'\beta}$ *and* $\boldsymbol{\psi}^* = \boldsymbol{DK'\beta}$. *For*

$$\begin{aligned}
H_1: &\quad \boldsymbol{\psi} = \boldsymbol{0} \\
H_2: &\quad \boldsymbol{\psi}^* = \boldsymbol{0}
\end{aligned}$$

Then $SS_{H_1} = SS_{H_2}$ *if* $\boldsymbol{D}$ *is non-singular.*

Proof:

$$\begin{aligned}
SS_{H_2} &= \left(\boldsymbol{DK'}\widehat{\boldsymbol{\beta}}\right)'\left[\boldsymbol{DK'}\left(\boldsymbol{X'X}\right)^{-}\boldsymbol{D'K}\right]^{-1}\left(\boldsymbol{DK'}\widehat{\boldsymbol{\beta}}\right) \\
&= \widehat{\boldsymbol{\beta}}'\boldsymbol{KD'}\left(\boldsymbol{D'}\right)^{-1}\left[\boldsymbol{K'}\left(\boldsymbol{X'X}\right)^{-}\boldsymbol{K}\right]^{-1}\boldsymbol{D}^{-1}\boldsymbol{DK'}\widehat{\boldsymbol{\beta}} \\
&= \left(\boldsymbol{K'}\widehat{\boldsymbol{\beta}}\right)'\left[\boldsymbol{K'}\left(\boldsymbol{X'X}\right)^{-}\boldsymbol{K}\right]^{-1}\left(\boldsymbol{K'}\widehat{\boldsymbol{\beta}}\right) \\
&= SS_{H_1}
\end{aligned}$$

Now can show that

$$\boldsymbol{\psi}^* = \begin{bmatrix} \zeta_1 \\ \vdots \\ \zeta_q \end{bmatrix} = \boldsymbol{D\psi} \tag{2.59}$$

for non-singular $\boldsymbol{D}$. So $SS_{H_1} = SS_{H_2}$. Can also show that $SS_{H_2} = \boldsymbol{Q}_\omega - \boldsymbol{Q}_\Omega$. To show 2.59 let $\widehat{\boldsymbol{\psi}} = \boldsymbol{K'}\widehat{\beta} = \boldsymbol{A'y}$ and by the Gauss-Markoff Theorem

$$\mathrm{E}(\boldsymbol{A'y}) = \boldsymbol{A'X\beta} = \boldsymbol{A'\eta} = \boldsymbol{\psi}$$

Define $\boldsymbol{P'}_1 = \boldsymbol{AT}$, Gram-Schmidt process in matrix, $\boldsymbol{A} = \boldsymbol{P'}_1\boldsymbol{T}^{-1} \Rightarrow \boldsymbol{A'} = (\boldsymbol{T}^{-1})'\boldsymbol{P}_1$. $\boldsymbol{D}$ is a non-singular matrix. Now

$$\boldsymbol{A'\eta} = \boldsymbol{\psi} = \boldsymbol{A'}\left[\boldsymbol{P'}_1\boldsymbol{\zeta}_1 + \boldsymbol{P'}_2\boldsymbol{\zeta}_2\right] = (\boldsymbol{T}^{-1})'\boldsymbol{P}_1\boldsymbol{P'}_1\boldsymbol{\zeta}_1 + (\boldsymbol{T}^{-1})'\underbrace{\boldsymbol{P}_1\boldsymbol{P'}_2}_{\boldsymbol{0}}\boldsymbol{\zeta}_2$$

So, $SS_{H_1} = SS_{H_2}$ where

$$\begin{aligned} H_1 :&\quad \boldsymbol{\psi}^{q\times 1} = \boldsymbol{0} \\ H_2 :&\quad \boldsymbol{\zeta}_1^{q\times 1} = \boldsymbol{0} \end{aligned}$$

In order to calculate SS_{H_2} first need to get the $BLUE$ of $\boldsymbol{\zeta}_1$. Working in canonical scale we have

$$\begin{aligned} \mathrm{E}(\boldsymbol{z}) &= \boldsymbol{P'}_1\boldsymbol{\zeta}_1 + \boldsymbol{P'}_2\boldsymbol{\zeta}_2 \\ \mathrm{Var}(\boldsymbol{z}) &= \boldsymbol{I}\sigma_e^2 \end{aligned} \tag{2.60}$$

From the Gauss-Markoff Theorem, $BLUE$ of $\widehat{\boldsymbol{\psi}}^* = \widehat{\boldsymbol{\zeta}}_1$ can be written as

- $\boldsymbol{A'y}$ where the columns of $\boldsymbol{A} \in V_r \subset V_n$ spanned by $\boldsymbol{P'}_1$ and $\boldsymbol{P'}_2$.
- Further, $\mathrm{E}(\boldsymbol{A'y}) = \boldsymbol{\zeta}_1$

Both this conditions are satisfied for $\boldsymbol{A} = \boldsymbol{P'}_1$. From the Gauss-Markoff Theorem $\boldsymbol{A'}$ is unique. So

$$\begin{aligned} \widehat{\boldsymbol{\psi}}^* = \widehat{\boldsymbol{\zeta}}_1 = \boldsymbol{P}_1\boldsymbol{y} = \boldsymbol{P}_1\boldsymbol{P'}_1\boldsymbol{z}_1 = \boldsymbol{z}_1 \\ \mathrm{Var}(\widehat{\boldsymbol{\psi}}^*) = \mathrm{Var}(\widehat{\boldsymbol{\zeta}}_1) = \mathrm{Var}(\boldsymbol{z}_1) = \boldsymbol{I}\sigma_e^2 \end{aligned}$$

As a result

$$SS_{H_2} = \boldsymbol{z'}_1\left[\mathrm{Var}(\boldsymbol{z}_1)\right]^{-1}\boldsymbol{z}_1\frac{1}{\sigma_e^2} = \boldsymbol{z'}_1\boldsymbol{z}_1 = \boldsymbol{Q}_\omega - \boldsymbol{Q}_\Omega$$

So the test based on likelihood ratio

$$\begin{aligned} \boldsymbol{F} &= \frac{\left(\boldsymbol{Q}_\omega - \boldsymbol{Q}_\Omega\right)/q}{\widehat{\sigma}_e^2} \\ &= \frac{\left(\boldsymbol{K'}\widehat{\boldsymbol{\beta}}\right)'\left[\boldsymbol{K'}\left(\boldsymbol{X'X}\right)^-\boldsymbol{K}\right]^{-1}\left(\boldsymbol{K'}\widehat{\boldsymbol{\beta}}\right)/q}{\widehat{\sigma}_e^2} \end{aligned} \tag{2.61}$$

In summary

- $\Omega :\quad \boldsymbol{y} \sim \boldsymbol{N}\left(\boldsymbol{X\beta}, \boldsymbol{I}\sigma_e^2\right) \quad \boldsymbol{X\beta} = \boldsymbol{\eta} \in V_r \subset V_n$
- $BLUE\left(\boldsymbol{K'\beta}\right) = \boldsymbol{K'}\widehat{\beta}$ (also, $\widehat{\eta}$ is MLE of $\boldsymbol{\eta}$) if $\boldsymbol{K'\beta}$ is estimable, where $\widehat{\beta}$ is the solution to $\left(\boldsymbol{X'X}\right)\widehat{\boldsymbol{\beta}} = \boldsymbol{X'y} \Rightarrow \widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X'X}\right)^-\boldsymbol{X'y}$

- $\mathrm{Var}(\boldsymbol{K'}\widehat{\boldsymbol{\beta}}) = \boldsymbol{K'}\left(\boldsymbol{X'X}\right)^{-}\boldsymbol{K}\sigma_e^2$

Hypothesis testing:

$$H_0: \quad \boldsymbol{\psi} = \boldsymbol{K'}^{q\times p}\boldsymbol{\beta} = \boldsymbol{0}$$

is equivalent to

$$H: \quad \boldsymbol{\eta} \in V_{r-q} \subset V_r$$

$$\boldsymbol{F} = \frac{\left(\boldsymbol{K'}\widehat{\boldsymbol{\beta}}\right)'\left[\boldsymbol{K'}\left(\boldsymbol{X'X}\right)^{-}\boldsymbol{K}\right]^{-1}\left(\boldsymbol{K'}\widehat{\boldsymbol{\beta}}\right)/q}{\widehat{\sigma}_e^2} \tag{2.62}$$

Reject H_0 when $\boldsymbol{F} > \boldsymbol{F}_{\alpha;q,n-r}$. This is equivalent to likelihood ratio test.

2.8 Consequences of using wrong model

Suppose $H: \quad \boldsymbol{\psi}^{q\times 1} = \boldsymbol{0}$ is true and $\eta \in V_{r-q} \subset V_r \subset V_n$. Then, writing the model in the cannonical form

$$\boldsymbol{y} = \boldsymbol{P'}_2\boldsymbol{\zeta}_2 + \boldsymbol{e} \quad \mathrm{E}(\boldsymbol{y}) = \boldsymbol{\eta} = \boldsymbol{P'}_2\boldsymbol{\zeta}_2 \tag{2.63}$$

Then $BLUE$ of η is $\boldsymbol{P'}_2\widehat{\boldsymbol{\zeta}}_2$. Also

$$\begin{aligned}\widehat{\boldsymbol{\zeta}}_2 &= \left(\boldsymbol{P}_2\boldsymbol{P'}_2\right)^{-1}\boldsymbol{P}_2\boldsymbol{y} = \boldsymbol{P}_2\boldsymbol{y}\\ \widehat{\boldsymbol{\eta}}_\omega &= \boldsymbol{P'}_2\boldsymbol{P}_2\boldsymbol{y} = \boldsymbol{P'}_2\boldsymbol{z}_2\end{aligned}$$

Note that $\widehat{\boldsymbol{\eta}}$ is the projection of $\boldsymbol{y}$ onto V_{r-q}. We consider now the consequences of overfitting. Suppose H is true and $\boldsymbol{\eta} \in V_{r-q} \subset V_r \subset V_n$. But use model

$$\boldsymbol{y} = \boldsymbol{P'}_1\boldsymbol{\zeta}_1 + \boldsymbol{P'}_2\boldsymbol{\zeta}_2 + \boldsymbol{e} \tag{2.64}$$

Then $\boldsymbol{\eta}$ is estimated by

$$\begin{aligned}\widehat{\boldsymbol{\eta}}_\Omega &= \boldsymbol{P'}_1\widehat{\boldsymbol{\zeta}}_1 + \boldsymbol{P'}_2\widehat{\boldsymbol{\zeta}}_2\\ &= \boldsymbol{P'}_1\boldsymbol{z}_1 + \boldsymbol{P'}_2\boldsymbol{z}_2\end{aligned} \tag{2.65}$$

As a result

$$\begin{aligned}\mathrm{E}_\omega(\widehat{\boldsymbol{\eta}}_\Omega) &= \boldsymbol{P'}_1\mathrm{E}(\boldsymbol{z}_1) + \boldsymbol{P'}_2\mathrm{E}(\boldsymbol{z}_2)\\ &= \boldsymbol{0} + \boldsymbol{P'}_2\boldsymbol{\zeta}_2\\ &= \boldsymbol{\eta}\end{aligned} \tag{2.66}$$

and

$$\mathrm{Var}_\omega(\widehat{\boldsymbol{\eta}}_\Omega) = \boldsymbol{P'}_1\boldsymbol{P}_1\sigma_e^2 + \boldsymbol{P'}_2\boldsymbol{P}_2\sigma_e^2 \tag{2.67}$$

whereas

$$\mathrm{Var}_\omega(\widehat{\boldsymbol{\eta}}_\omega) = \boldsymbol{P'}_2\boldsymbol{P}_2\sigma_e^2 \tag{2.68}$$

Consider now the consequences of underfitting using the cannonical form. Suppose H is not true but the model

$$\boldsymbol{y} = \boldsymbol{P'}_2\boldsymbol{\zeta}_2 + \boldsymbol{e} \tag{2.69}$$

is used to estimate η. As a result

$$\mathrm{E}_\Omega(\widehat{\eta}_\omega) = \mathrm{E}(\boldsymbol{P'}_2\boldsymbol{z}_2) = \boldsymbol{P'}_2\boldsymbol{\zeta}_2 \tag{2.70}$$

whereas $\boldsymbol{\eta} = \boldsymbol{P'}_1\boldsymbol{\zeta}_1 + \boldsymbol{P'}_2\boldsymbol{\zeta}_2$. Also

$$\mathrm{Var}_\Omega(\widehat{\eta}_\omega) = \boldsymbol{P'}_2\boldsymbol{P}_2\sigma_e^2 \tag{2.71}$$

So, $\widehat{\eta}_\omega$ is biased, but has lower variance than

$$\mathrm{Var}_\Omega(\widehat{\eta}_\Omega) = \boldsymbol{P'}_1\boldsymbol{P}_1\sigma_e^2 + \boldsymbol{P'}_2\boldsymbol{P}_2\sigma_e^2 \tag{2.72}$$

Consider now a different approach to these problems. Suppose the true model is

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{e} \qquad \Omega \tag{2.73}$$

but the model

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{e} \qquad \omega \tag{2.74}$$

is fitted (underfitting), and the estimable function $\boldsymbol{K'\beta}_1$ is estimated as $\boldsymbol{K'}\widehat{\boldsymbol{\beta}}_\omega = \boldsymbol{A'y}$ where $\boldsymbol{A'} = \boldsymbol{K'}\left(\boldsymbol{X'}_1\boldsymbol{X}_1\right)^- \boldsymbol{X'}_1$. Now consider

$$\begin{aligned}\mathrm{E}_\Omega(\widehat{\eta}_\omega) &= \boldsymbol{A'}(\boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2)\\ &= \boldsymbol{K'\beta}_1 + \boldsymbol{A'X}_2\boldsymbol{\beta}_2\end{aligned}$$

because $\boldsymbol{A'X}_1 = \boldsymbol{K'}$. Note that in general, $\boldsymbol{A'X}_2 \neq \boldsymbol{K'}$ does not even have the same dimensions. Consider now the situation of overfitting. The model

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{e} \tag{2.75}$$

is used. Then

$$\begin{aligned}\mathrm{E}_\omega(\boldsymbol{K'}\widehat{\boldsymbol{\beta}}_\Omega) &= \boldsymbol{A'}\mathrm{E}_\omega(\boldsymbol{y})\\ &= \boldsymbol{A'X}_1\boldsymbol{\beta}_1\\ &= \boldsymbol{K'\beta}_1\end{aligned}$$

if $\boldsymbol{K'\beta}_1$ is estimable under Ω.

2.9 Variance of estimates by iteration

When the system of equations is large, solve by Gauss-Seidel iteration.

$$\mathrm{Var}(\boldsymbol{k'}\widehat{\boldsymbol{\beta}}) = \boldsymbol{k'}(\boldsymbol{X'X})^-\boldsymbol{k}\sigma_e^2 \tag{2.76}$$

The desired variance can be computed without computing $(\boldsymbol{X'X})^-$. Note the fact that $(\boldsymbol{X'X})^-\boldsymbol{k}$ is a solution to $(\boldsymbol{X'X})\boldsymbol{b} = \boldsymbol{k}$. So first solve this equation to obtain $\boldsymbol{b}$ by Gauss-Seidel iteration . Then

$$\boldsymbol{k'b}\sigma_e^2 = \mathrm{Var}(\boldsymbol{k'}\widehat{\boldsymbol{\beta}}) \tag{2.77}$$

Also if there is no solution to $(\boldsymbol{X'X})\boldsymbol{b} = \boldsymbol{k}$, then $\boldsymbol{k'\beta}$ is not estimable.

Chapter 3

Prediction of Random Effects

3.1 Best Prediction

Mean squared error of prediction: Let T be an unobservable random variable (genotypic value) and $\boldsymbol{y}$ (phenotypic values) a vector of observations that are related to T. In prediction, the goal is to define some function $\tilde{T}$ of $\boldsymbol{y}$ such that

$$\mathrm{E}(T - \tilde{T})^2, \tag{3.1}$$

the MSE of prediction, is minimum. Let $\widehat{T} = \mathrm{E}(T|\boldsymbol{y})$, and write

$$\begin{aligned} \mathrm{E}(T - \tilde{T})^2 &= \mathrm{E}(T - \widehat{T} + \widehat{T} - \tilde{T})^2 \\ &= \mathrm{E}\left[(T - \widehat{T})^2 + (\widehat{T} - \tilde{T})^2 + 2(T - \widehat{T})(\widehat{T} - \tilde{T})\right]. \end{aligned} \tag{3.2}$$

But,

$$\begin{aligned} \mathrm{E}\left[(T - \widehat{T})(\widehat{T} - \tilde{T})\right] &= \mathrm{E}_{\boldsymbol{y}}\left\{\mathrm{E}\left[(T - \widehat{T})(\widehat{T} - \tilde{T}) \mid \boldsymbol{y}\right]\right\} \\ &= \mathrm{E}_{\boldsymbol{y}}\left[(\widehat{T} - \widehat{T})(\widehat{T} - \tilde{T})\right] \\ &= 0. \end{aligned} \tag{3.3}$$

So,

$$\mathrm{E}(T - \tilde{T})^2 = \mathrm{E}\left[(T - \widehat{T})^2 + (\widehat{T} - \tilde{T})^2\right]. \tag{3.4}$$

The first term of (3.4) does not involve $\tilde{T}$, and the second term is minimum when $\tilde{T} = \widehat{T}$. So $\mathrm{E}(T - \tilde{T})^2$ is minimized by choosing $\tilde{T}$ to be $\widehat{T} = \mathrm{E}(T|\boldsymbol{y})$.

Correlation between predictor and predictand:

Proposition 3.1.1 *Can show that*

$$\rho(T, \tilde{T}) = \frac{Cov(T, \tilde{T})}{\sqrt{Var(T)\, Var(\tilde{T})}} \tag{3.5}$$

is maximized by choosing $\tilde{T} = \widehat{T}$.

Proof: Let $\mathrm{E}(\tilde{T}) = \theta$. Then,

$$\mathrm{Cov}(T, \tilde{T}) = \mathrm{E}\left[T(\tilde{T} - \theta)\right] = \mathrm{E}\left\{\left[(T - \widehat{T}) + \widehat{T}\right](\tilde{T} - \theta)\right\}, \tag{3.6}$$

but,

$$\mathrm{E}_{\boldsymbol{y}}\left\{\mathrm{E}\left[(T - \widehat{T})(\tilde{T} - \theta) \mid \boldsymbol{y}\right]\right\} = \mathrm{E}_{\boldsymbol{y}}\left[(\widehat{T} - \widehat{T})(\tilde{T} - \theta)\right] = 0. \tag{3.7}$$

So

$$\mathrm{Cov}(T, \tilde{T}) = \mathrm{E}\left[\widehat{T}(\tilde{T} - \theta)\right] = \mathrm{Cov}(\widehat{T}, \tilde{T}), \tag{3.8}$$

and $\mathrm{Cov}(T, \widehat{T}) = \mathrm{Cov}(\widehat{T}, \widehat{T}) = \mathrm{Var}(\widehat{T})$. Now,

$$\begin{aligned} \rho^2(T, \tilde{T}) &= \frac{\mathrm{Cov}^2(T, \tilde{T})}{\mathrm{Var}(T)\mathrm{Var}(\tilde{T})} \\ &= \frac{\mathrm{Cov}^2(\widehat{T}, \tilde{T})}{\mathrm{Var}(T)\mathrm{Var}(\tilde{T})} \\ &= \frac{\mathrm{Cov}^2(\widehat{T}, \tilde{T})}{\mathrm{Var}(\widehat{T})\mathrm{Var}(\tilde{T})} \frac{\mathrm{Var}(\widehat{T})}{\mathrm{Var}(T)} \\ &= \rho^2(\widehat{T}, \tilde{T}) \frac{\mathrm{Var}(\widehat{T})}{\mathrm{Var}(T)} \end{aligned} \tag{3.9}$$

This is maximum when $\tilde{T} = \widehat{T}$ and $\rho^2(\widehat{T}, \tilde{T}) = 1$. Note that

$$\frac{\mathrm{Var}(\widehat{T})}{\mathrm{Var}(T)} = \rho^2(T, \widehat{T})$$

Mean of selected candidates: Consider now the problem of maximizing the expected value of selected $T_i's$. Suppose there are n candidates and we want to choose k such that

$$\mathrm{E}\left[\frac{\sum_{i=1}^{k} T_{s_i}}{k}\right]$$

where $s_1, \ldots, s_k$ are the indices of the selected $T_i's$.

$$\mathrm{E}\left[\frac{\sum_{i=1}^{k} T_{s_i}}{k}\right] = \frac{1}{k}\mathrm{E}_{\boldsymbol{y}}\left[\mathrm{E}\left(\sum_{i=1}^{k} T_{s_i} \mid \boldsymbol{y}\right)\right] = \frac{1}{k}\mathrm{E}_{\boldsymbol{y}}\left[\sum_{i=1}^{k} \widehat{T}_{s_i}\right] \tag{3.10}$$

It is clear that selecting $s_1, \ldots, s_k$ to be the indices of highest ranking $\widehat{T}_i$ would maximize 3.10. This is a very general result that not depend on the joint distribution of $\boldsymbol{T}$ and $\boldsymbol{y}$. Here, the proportion selected $(\frac{k}{n})$ is a constant.

Truncation selection: Under truncation selection, the proportion selected is not a constant. Cochran (1951) showed that under truncation selection the mean of the selected candidates is maximized by selecting according to $\widehat{T}$ provided that the candidates are identically and independently distributed and the information used for prediction is also independently and identically distributed:

$$(\boldsymbol{y}_i, T_i)\text{iid}$$

Selecting based on conditional mean does not maximize probability of correct ordering.

- ***Example*** **:**

	$\Pr(\boldsymbol{T} \mid \boldsymbol{y}) = 0.1$	$\Pr(\boldsymbol{T} \mid \boldsymbol{y}) = 0.9$
T_1	1	2
T_2	20	1

$$\begin{aligned} \mathrm{E}(T_1|\boldsymbol{y}) &= 0.1 * 1 + 0.9 * 2 = 1.9 \\ \mathrm{E}(T_2|\boldsymbol{y}) &= 0.1 * 20 + 0.9 * 1 = 2.9 \end{aligned}$$

Given this value of $\boldsymbol{y}$, if you select T_2, you would be wrong in 90% of the time. Note $\mathrm{E}(T_2|\boldsymbol{y}) > \mathrm{E}(T_1|\boldsymbol{y})$.

Proposition 3.1.2 *Suppose $\boldsymbol{X}$ and $\boldsymbol{Y}$ are multivariate normal (MVN). In general*

$$\begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix} \sim \boldsymbol{N}\left(\begin{bmatrix} \boldsymbol{\mu_X} \\ \boldsymbol{\mu_Y} \end{bmatrix}, \begin{bmatrix} \sum_{\boldsymbol{X}} & \sum_{\boldsymbol{XY}} \\ \sum_{\boldsymbol{YX}} & \sum_{\boldsymbol{Y}} \end{bmatrix} \right) \tag{3.11}$$

Can show that

$$E(\boldsymbol{X}|\boldsymbol{Y}) = \boldsymbol{\mu_X} + \sum\nolimits_{\boldsymbol{XY}} \sum\nolimits_{\boldsymbol{Y}}^{-1} (\boldsymbol{Y} - \boldsymbol{\mu_Y}) = \widehat{\boldsymbol{X}}$$

Proof: Write $\boldsymbol{X} = \widehat{\boldsymbol{X}} + (\boldsymbol{X} - \widehat{\boldsymbol{X}})$. Observe that

- $\mathrm{E}(\boldsymbol{X} - \widehat{\boldsymbol{X}}) = \boldsymbol{0}$
- $\mathrm{Cov}\left[(\boldsymbol{X} - \widehat{\boldsymbol{X}}), \boldsymbol{Y'}\right] = \sum_{\boldsymbol{XY}} - \sum_{\boldsymbol{XY}} = \boldsymbol{0}$

Note that under multivariate normality a null correlation implies independence. So,

$$\begin{aligned} \mathrm{E}\left(\boldsymbol{X} \mid \boldsymbol{Y}\right) &= \mathrm{E}\left(\widehat{\boldsymbol{X}} \mid \boldsymbol{Y}\right) + \mathrm{E}\left[(\boldsymbol{X} - \widehat{\boldsymbol{X}}) \mid \boldsymbol{Y}\right] \\ &= \widehat{\boldsymbol{X}} + \mathrm{E}\left(\boldsymbol{X} \mid \boldsymbol{Y}\right) - \widehat{\boldsymbol{X}} \\ &= \widehat{\boldsymbol{X}} + \boldsymbol{0} \\ &= \widehat{\boldsymbol{X}} \end{aligned} \tag{3.12}$$

and as a result $\widehat{\boldsymbol{X}}$ is a linear function of $\boldsymbol{Y}$. Consider now

$$
\begin{aligned}
\operatorname{Var}(\boldsymbol{X} \mid \boldsymbol{Y}) &= \underbrace{\operatorname{Var}\left(\widehat{\boldsymbol{X}} \mid \boldsymbol{Y}\right)}_{\mathbf{0}} + \operatorname{Var}\left[(\boldsymbol{X} - \widehat{\boldsymbol{X}}) \mid \boldsymbol{Y}\right] \\
&= \operatorname{Var}\left(\boldsymbol{X} - \widehat{\boldsymbol{X}}\right)
\end{aligned}
\tag{3.13}
$$

because $(\boldsymbol{X} - \widehat{\boldsymbol{X}})$ and $\boldsymbol{Y}$ are independent. Now

$$
\begin{aligned}
\operatorname{Var}\left(\boldsymbol{X} - \widehat{\boldsymbol{X}}\right) &= \operatorname{Cov}\left[(\boldsymbol{X} - \widehat{\boldsymbol{X}}), (\boldsymbol{X} - \widehat{\boldsymbol{X}})'\right] \\
&= \Sigma_{\boldsymbol{X}} - \Sigma_{\boldsymbol{XY}} \Sigma_{\boldsymbol{Y}}^{-1} \Sigma_{\boldsymbol{YX}} - \Sigma_{\boldsymbol{XY}} \Sigma_{\boldsymbol{Y}}^{-1} \Sigma_{\boldsymbol{YX}} \\
&+ \Sigma_{\boldsymbol{XY}} \Sigma_{\boldsymbol{Y}}^{-1} \Sigma_{\boldsymbol{YX}} \\
&= \Sigma_{\boldsymbol{X}} - \Sigma_{\boldsymbol{XY}} \Sigma_{\boldsymbol{Y}}^{-1} \Sigma_{\boldsymbol{YX}}
\end{aligned}
\tag{3.14}
$$

Suppose

$$
\begin{bmatrix} T \\ \boldsymbol{y} \end{bmatrix} \sim \boldsymbol{N}\left(\begin{bmatrix} \mu_T \\ \boldsymbol{\eta} \end{bmatrix}, \begin{bmatrix} \sigma_T^2 & \boldsymbol{c'} \\ \boldsymbol{c} & \boldsymbol{V} \end{bmatrix} \right) \tag{3.15}
$$

Then

$$
\operatorname{E}(\boldsymbol{T} \mid \boldsymbol{y}) = \mu_T + \boldsymbol{c'}\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{\eta}) = \widehat{T} \tag{3.16}
$$

Note $\widehat{T} = a + \boldsymbol{b'y}$ (linear in $\boldsymbol{y}$) where $a = \mu_T - \boldsymbol{c'V}^{-1}\boldsymbol{\eta}$ and $\boldsymbol{b'} = \boldsymbol{c'V}^{-1}$. We have seen that in general, the best predictor does not maximize the probability of correct ranking. However, under MVN, can show that the BP maximizes the probability of correct pairwise ranking. Let

$$
\begin{aligned}
d &= T_1 - T_2 \\
\widehat{d} &= \operatorname{E}(T_1 \mid \boldsymbol{y}) - \operatorname{E}(T_2 \mid \boldsymbol{y}) = \operatorname{E}(d \mid \boldsymbol{y})
\end{aligned}
$$

Ranking is correct when $\widehat{d}$ has the same sign as d. Note:

$$
\operatorname{E}(\widehat{d} \mid d) = \operatorname{E}(\widehat{d}) + \frac{\operatorname{Cov}(d, \widehat{d})}{\operatorname{Var}(d)} \left[d - \operatorname{E}(d)\right]
$$

If $\operatorname{E}(d) = \operatorname{E}(\widehat{d}) = 0$

$$
\begin{aligned}
\operatorname{E}(\widehat{d} \mid d) &= \frac{\operatorname{Cov}(d, \widehat{d})}{\operatorname{Var}(d)}\, d \\
&= \rho(d, \widehat{d}) \sqrt{\frac{\operatorname{Var}(\widehat{d})}{\operatorname{Var}(d)}}\, d
\end{aligned}
\tag{3.17}
$$

Also

$$
\begin{aligned}
\operatorname{Var}(\widehat{d} \mid d) &= \operatorname{Var}(\widehat{d}) - \frac{\operatorname{Cov}^2(d, \widehat{d})}{\operatorname{Var}(d)} \\
&= \operatorname{Var}(\widehat{d}) - \rho^2(d, \widehat{d}) \operatorname{Var}(\widehat{d}) \\
&= \operatorname{Var}(\widehat{d}) \left(1 - \rho^2(d, \widehat{d})\right)
\end{aligned}
\tag{3.18}
$$

$$
\begin{aligned}
\Pr(\widehat{d} > 0 \mid d = k > 0) &= 1 - \Phi\left(\frac{-\rho(d,\widehat{d})\sqrt{\frac{\mathrm{Var}(\widehat{d})}{\mathrm{Var}(d)}}\ k}{\sqrt{\mathrm{Var}(\widehat{d})}[1-\rho^2(d,\widehat{d})]^{-1/2}}\right) \\
&= 1 - \Phi\left(\frac{-\rho(d,\widehat{d})\frac{1}{\sqrt{\mathrm{Var}(d)}}\ k}{[1-\rho^2(d,\widehat{d})]^{-1/2}}\right)
\end{aligned}
\tag{3.19}
$$

This is maximized by choosing $\widehat{d}$ to maximize $\rho(d, \widehat{d})$.

3.2 Best Linear Prediction

Calculating the BP requires knowing the joint distribution of $(T, \boldsymbol{y})$. Further, it requires computing $\mathrm{E}(T \mid \boldsymbol{y})$. This may be non-linear and difficult to compute. Consider a predictor of the form:

$$\tilde{T} = a^* + \boldsymbol{b'}^*\boldsymbol{y} \quad \text{(linear in } \boldsymbol{y})$$

where a and $\boldsymbol{b'}$ are chosen such that $\mathrm{E}(T - \widehat{T})^2$ is minimum. So,

$$
\begin{aligned}
\mathrm{E}(T-\widehat{T})^2 &= \mathrm{E}(T - a^* - \boldsymbol{b'}^*\boldsymbol{y})^2 \\
&= \sigma_T^2 + \mu_T^2 - 2\mu_T a^* - 2\boldsymbol{b'}^*\boldsymbol{c} - \mu_T\boldsymbol{b'}^*\boldsymbol{\eta} + {a^*}^2 - 2a\boldsymbol{b'}^*\boldsymbol{\eta} + \boldsymbol{b'}^*\boldsymbol{V}\boldsymbol{b}^* \\
&+ \boldsymbol{b'}^*\boldsymbol{\eta}\boldsymbol{\eta'}\boldsymbol{b}^*
\end{aligned}
\tag{3.20}
$$

This is the expression for $MSEP(\tilde{T})$ regardless of the exact form of the joint distribution of T and $\boldsymbol{y}$. Claim that $a^* = a = \mu_t - \boldsymbol{c'}\boldsymbol{V}^{-1}\boldsymbol{\eta}$ and $\boldsymbol{b'}^* = \boldsymbol{b} = \boldsymbol{c'}\boldsymbol{V}^{-1}$. Suppose $a^* \neq a$ and $\boldsymbol{b}^* \neq \boldsymbol{b}$ has lower $MSEP$ than $a + \boldsymbol{by}$. But under MVN, $a + \boldsymbol{by}$ gives the BP. So, a^* and $\boldsymbol{b}^*$ cannot give lower value for $MSEP$. To verify let $\widehat{T} = a + \boldsymbol{b'y}$ and let $\tilde{T} = a^* + \boldsymbol{b'}^*\boldsymbol{y}$. Then

$$MSEP(\tilde{T}) = \mathrm{E}(T-\tilde{T})^2 = \mathrm{E}\left[(T-\widehat{T}) + (\widehat{T}-\tilde{T})\right]^2 \tag{3.21}$$

Note $\mathrm{E}(T - \widehat{T}) = 0$ and

$$\mathrm{E}(T-\widehat{T})\boldsymbol{y'} = \mathrm{Cov}\left[(T-\widehat{T}), \boldsymbol{y'}\right] = \boldsymbol{c'} - \boldsymbol{c'} = \boldsymbol{0'} \tag{3.22}$$

So $\mathrm{E}(T - \widehat{T})(\widehat{T} - \tilde{T}) = 0$ and

$$MSEP(\tilde{T}) = \mathrm{E}(T-\widehat{T})^2 + \mathrm{E}(\widehat{T}-\tilde{T})^2 \tag{3.23}$$

The first term does not involve $\tilde{T}$ and the second is minimum by choosing $\tilde{T} = \widehat{T}$. Consider now the problem of maximizing the correlation between T and $\tilde{T}$. Let $\tilde{T} = a^* + \boldsymbol{b'}^*\boldsymbol{y}$ be an arbitrary linear predictor.

$$\rho(T, \tilde{T}) = \frac{\boldsymbol{b'}^*\boldsymbol{c}}{\sqrt{\boldsymbol{b'}^*\boldsymbol{V}\boldsymbol{b}^*\sigma_T^2}} \tag{3.24}$$

This is the expression for $\rho(T, \widehat{T})$ regardless of the exact form of the joint distribution between T and $\boldsymbol{y}$. We know that choosing $a^* = a$ and $b^* = b$ maximizes $\rho(T, \tilde{T})$ under MVN. So among all linear predictors, the BLP maximizes $\rho(T, \widehat{T})$.Consider now a different approach. Let $T = \widehat{T} + (T - \widehat{T})$ where $\widehat{T}$ is the BLP. Then,

$$\text{Cov}(T, \tilde{T}) = \text{Cov}\left[\widehat{T} + (T - \widehat{T}), \tilde{T}\right] \tag{3.25}$$

But, we have already seen that for any linear predictor $\tilde{T}$

$$\text{Cov}\left[(T - \widehat{T}), \tilde{T}\right] = 0 \tag{3.26}$$

So,

$$\text{Cov}(T, \tilde{T}) = \text{Cov}(\widehat{T}, \tilde{T}) \tag{3.27}$$

and

$$\begin{aligned} \rho^2(T, \tilde{T}) &= \frac{\text{Cov}^2(T, \tilde{T})}{\text{Var}(T)\text{Var}(\tilde{T})} \\ &= \frac{\text{Cov}^2(\widehat{T}, \tilde{T})}{\text{Var}(T)\text{Var}(\tilde{T})} \\ &= \frac{\text{Cov}^2(\widehat{T}, \tilde{T})}{\text{Var}(\widehat{T})\text{Var}\tilde{T}} \frac{\text{Var}(\widehat{T})}{\text{Var}(T)} \\ &= \rho^2(\widehat{T}, \tilde{T}) \frac{\text{Var}(\widehat{T})}{\text{Var}(T)} \end{aligned} \tag{3.28}$$

Note that $\frac{\text{Var}(\widehat{T})}{\text{Var}(T)}$ does not depend on the choice of $\tilde{T}$. So, $\rho^2(T, \tilde{T})$ is maximized by the maximum $\rho^2(\widehat{T}, \tilde{T})$. This is maximum when $\tilde{T} = \widehat{T}$. Note that

$$\frac{\text{Var}(\widehat{T})}{\text{Var}(T)} = \rho^2(T, \widehat{T})$$

3.3 Best Linear Unbiased Prediction

If the true values of $\mu_T, \boldsymbol{\mu}_y, \boldsymbol{c'}$ and $\boldsymbol{V}$ are known,

$$\widehat{T} = \mu_T + \boldsymbol{c'}\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{\eta}) \tag{3.29}$$

is the *BLP*. Suppose,

$$\mathrm{E}(\boldsymbol{y}) = \boldsymbol{\eta} = \boldsymbol{X\beta} \tag{3.30}$$

where $\boldsymbol{X}$ is known and β is unknown. Further, suppose $\mu_T = \boldsymbol{\lambda}'\boldsymbol{\beta}$ is estimable. That means, there exists an $\boldsymbol{b}'$ such that

$$\mathrm{E}(\boldsymbol{b}'\boldsymbol{y}) = \boldsymbol{b}'\boldsymbol{X\beta} = \boldsymbol{\lambda}'\boldsymbol{\beta} \quad \text{for all}\beta \tag{3.31}$$

thus, $\boldsymbol{b}'\boldsymbol{X} = \boldsymbol{\lambda}'$. Then can predict T with

$$\widehat{T} = \boldsymbol{\lambda}'\widehat{\boldsymbol{\beta}} + \boldsymbol{c}'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \tag{3.32}$$

where $\widehat{\boldsymbol{\beta}}$ is a solution to

$$(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})\widehat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y} \tag{3.33}$$

Note that $\widehat{\boldsymbol{T}}$ can be written as

$$\widehat{\boldsymbol{T}} = \boldsymbol{b}'\boldsymbol{y} \tag{3.34}$$

where

$$\boldsymbol{b}' = \boldsymbol{\lambda}'(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{V}^{-1} - \boldsymbol{c}'\boldsymbol{V}^{-1}\left[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{V}^{-1}\right] \tag{3.35}$$

This predictor is called the *BLUP* of T. It can be shown that among all linear unbiased predictors of the form $\tilde{T} = \boldsymbol{b}'^{*}\boldsymbol{y}$, $MSEP(\tilde{T}) = \mathrm{E}(T - \tilde{T})^2$ is minimum for $\tilde{T} = \widehat{T}$. A useful result to prove the above is:

Lemma 3.3.1 *Cov* $\left[(T - \widehat{T}), \boldsymbol{b}'^{*}\boldsymbol{y}\right]$ *is a constant with respect to* $\boldsymbol{b}'^{*}$.

Proof: Note that when $\mathrm{E}(\boldsymbol{b}'^{*}\boldsymbol{y}) = \boldsymbol{b}'^{*}\boldsymbol{X\beta} = \boldsymbol{\lambda}'\boldsymbol{\beta}$ for all $\boldsymbol{\beta} \Rightarrow \boldsymbol{b}'^{*}\boldsymbol{X} = \boldsymbol{\lambda}'$

$$\begin{aligned}
&\mathrm{Cov}\left[(T - \widehat{T}), \boldsymbol{b}'^{*}\boldsymbol{y}\right] \\
&= \mathrm{Cov}[(T - \boldsymbol{\lambda}'(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y} - \boldsymbol{c}'\boldsymbol{V}^{-1}\boldsymbol{y} \\
&+ \boldsymbol{c}'\boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}), \boldsymbol{b}'^{*}\boldsymbol{y}] \\
&= \boldsymbol{c}'\boldsymbol{b}'^{*} - \boldsymbol{\lambda}'(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-}\underbrace{\boldsymbol{X}'\boldsymbol{b}'^{*}}_{\boldsymbol{\lambda}} - \boldsymbol{c}'\boldsymbol{b}'^{*} + \boldsymbol{c}'\boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-}\underbrace{\boldsymbol{X}'\boldsymbol{b}'^{*}}_{\boldsymbol{\lambda}} \\
&= -\boldsymbol{\lambda}'(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-}\boldsymbol{\lambda} + \boldsymbol{c}'\boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-}\boldsymbol{\lambda}
\end{aligned} \tag{3.36}$$

Let $\tilde{T} = \boldsymbol{b}'^{*}\boldsymbol{y}$ and $\widehat{T} = \boldsymbol{b}'\boldsymbol{y}$ where for both, $\boldsymbol{b}'^{*}\boldsymbol{X} = \boldsymbol{b}'\boldsymbol{X} = \boldsymbol{\lambda}'$. So,

$$\mathrm{Cov}\left[(T - \widehat{T}), (\widehat{T} - \tilde{T})\right] = \mathrm{Cov}\left[(T - \widehat{T}), \widehat{T}\right] - \mathrm{Cov}\left[(T - \widehat{T}), \tilde{T})\right] = \mathbf{0} \tag{3.37}$$

Now we can prove that $MSE(\tilde{T}) = \mathrm{E}(T - \tilde{T})^2$ is minimum when $\tilde{T} = \widehat{T}$, for $\tilde{T} = \boldsymbol{b}'^{*}\boldsymbol{y}$ and $\boldsymbol{b}'^{*}\boldsymbol{X} = \boldsymbol{\lambda}'$.

$$MSE(\tilde{T}) = \mathrm{E}(T - \tilde{T})^2 = \mathrm{E}\left[(T - \widehat{T}) + (\widehat{T} - \tilde{T})\right]^2 \tag{3.38}$$

Note that $\mathrm{E}(T-\widehat{T})=0$. So,

$$\mathrm{E}(T-\widehat{T})(\widehat{T}-\tilde{T})=\mathrm{Cov}\left[(T-\widehat{T}),(\widehat{T}-\tilde{T})\right]=0 \tag{3.39}$$

and

$$\mathit{MSEP}(\tilde{T})=\mathrm{E}(T-\widehat{T})^2+\mathrm{E}(\widehat{T}-\tilde{T})^2 \tag{3.40}$$

The first term is free of $\tilde{T}$ and the second term is minimum when $\tilde{T}=\widehat{T}$. So $\mathit{MSEP}(\tilde{T})$ is minimized by choosing $\tilde{T}=\widehat{T}$. Note that there is a linear unbiased predictor that has lower *MSEP* than *BLUP*, namely *BLP*. In animal breeding we model genotypic values as

$$T=\mathrm{E}(T)+T-\mathrm{E}(T)=\boldsymbol{\lambda}'\boldsymbol{\beta}+u \tag{3.41}$$

where $\mathrm{E}(u)=0$. Now *BLUP* of u is

$$\widehat{u}=\boldsymbol{c}'\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})=\boldsymbol{b}'\boldsymbol{y} \tag{3.42}$$

where $\boldsymbol{b}'=\boldsymbol{c}'\boldsymbol{V}^{-1}\left[\boldsymbol{I}-\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{V}^{-1}\right]$ and for $\mathrm{E}(\widehat{u})=\boldsymbol{b}'\boldsymbol{X}\boldsymbol{\beta}=\boldsymbol{0}$ for all $\boldsymbol{\beta}\Rightarrow\boldsymbol{b}'\boldsymbol{X}=\boldsymbol{0}$. Let $\tilde{u}=\boldsymbol{b}'^{*}\boldsymbol{y}$ be another predictor with $\mathrm{E}(\tilde{u})=\boldsymbol{b}'^{*}\boldsymbol{X}\boldsymbol{\beta}=\boldsymbol{0}$ for all β. So, $\boldsymbol{b}'^{*}\boldsymbol{X}=\boldsymbol{0}'$. Then,

$$\mathrm{Cov}\left[(u-\widehat{u}),\tilde{u}\right]=0 \tag{3.43}$$

$$\begin{aligned}
&\mathrm{Cov}\left[u-\boldsymbol{c}'\boldsymbol{V}^{-1}\boldsymbol{y}+\boldsymbol{c}'\boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y},\boldsymbol{b}'^{*}\boldsymbol{y}\right]\\
&=\boldsymbol{c}'\boldsymbol{b}^{*}-\boldsymbol{c}'\boldsymbol{b}^{*}+\boldsymbol{c}'\boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-}\underbrace{\boldsymbol{X}'\boldsymbol{b}^{*}}_{\boldsymbol{0}}\\
&=0
\end{aligned} \tag{3.44}$$

So,

$$\mathrm{Cov}(u,\tilde{u})=\mathrm{Cov}\left[\widehat{u}+(u-\widehat{u}),\tilde{u}\right]=\mathrm{Cov}(\widehat{u},\tilde{u}) \tag{3.45}$$

It follows that

$$\mathrm{Cov}(u,\widehat{u})=\mathrm{Cov}(\widehat{u},\widehat{u})=\mathrm{Var}(\widehat{u}) \tag{3.46}$$

and that

$$\mathrm{Var}(u-\widehat{u})=\mathrm{Var}(u)-2\mathrm{Cov}(u,\widehat{u})+\mathrm{Var}(\widehat{u})=\mathrm{Var}(u)-\mathrm{Var}(\widehat{u}) \tag{3.47}$$

Consider

$$\begin{aligned}
\rho^2(u,\tilde{u})&=\frac{\mathrm{Cov}^2(u,\tilde{u})}{\mathrm{Var}(u)\mathrm{Var}(\tilde{u})}\\
&=\frac{\mathrm{Cov}^2(\widehat{u},\tilde{u})}{\mathrm{Var}(\widehat{u}),\mathrm{Var}(\tilde{u})}\frac{\mathrm{Var}(\widehat{u})}{\mathrm{Var}(u)}\\
&=\rho^2(\widehat{u},\tilde{u})\frac{\mathrm{Var}(\widehat{u})}{\mathrm{Var}(u)}
\end{aligned} \tag{3.48}$$

So to maximize $\rho^2(u, \tilde{u})$ chose $\tilde{u} = \widehat{u}$. Those among all predictors of the form $\tilde{u} = \boldsymbol{b'}^*\boldsymbol{y}$ with $\boldsymbol{b'}^*\boldsymbol{X} = \boldsymbol{0'}$ $\quad \tilde{u} = \widehat{u} = \boldsymbol{b'y}$ has the highest $\rho(u, \widehat{u})$.

Consider a mixed linear model

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e} \tag{3.49}$$

with

$$\begin{aligned} &\mathrm{E}(\boldsymbol{y}) = \boldsymbol{X\beta}, \quad \mathrm{E}(\boldsymbol{u}) = \boldsymbol{0}, \quad \mathrm{E}(\boldsymbol{e}) = \boldsymbol{0} \\ &\mathrm{Var}(\boldsymbol{u}) = \boldsymbol{G}, \quad \mathrm{Var}(\boldsymbol{e}) = \boldsymbol{R}, \quad \mathrm{Cov}(\boldsymbol{u}, \boldsymbol{e}) = \boldsymbol{0}, \quad \mathrm{Var}(\boldsymbol{y}) = \boldsymbol{ZGZ'} + \boldsymbol{R} = \boldsymbol{V} \end{aligned} \tag{3.50}$$

So,

$$\mathrm{Cov}(\boldsymbol{u}, \boldsymbol{y'}) = \boldsymbol{C'} = \boldsymbol{GZ'} \quad BLUP(\boldsymbol{u}) = \widehat{\boldsymbol{u}} = \boldsymbol{GZ'V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \tag{3.51}$$

What if incorrect $\boldsymbol{G}$ and $\boldsymbol{V}$ are used. Still,

$$\mathrm{E}(\boldsymbol{X}\widehat{\boldsymbol{\beta}}) = \boldsymbol{X\beta} \tag{3.52}$$

where $\widehat{\boldsymbol{\beta}}$ was obtained using an incorrect V. For example, suppose $\widehat{\boldsymbol{\beta}}$ is obtained of *OLS* as $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X'X})^{-}\boldsymbol{X'y}$. Then,

$$\mathrm{E}(\boldsymbol{X}\widehat{\boldsymbol{\beta}}) = \underbrace{\boldsymbol{X}(\boldsymbol{X'X})^{-}\boldsymbol{X'X}}_{\boldsymbol{X}}\boldsymbol{\beta} = \boldsymbol{X\beta} \tag{3.53}$$

because $\boldsymbol{IX} = \boldsymbol{X}$ and $\boldsymbol{X\beta}$ is estimable. So,

$$\mathrm{E}(\widehat{\boldsymbol{u}}) = \boldsymbol{0} \tag{3.54}$$

even when the wrong values for $\boldsymbol{G}$ and $\boldsymbol{V}$ are used.

3.4 Henderson's Mixed Model Equations

Can obtain *BLUE* and *BLUP* efficiently by solving Henderson's Mixed Model Equations (H.M.M.E). Suppose,

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e} = \begin{bmatrix} \boldsymbol{X} & \boldsymbol{Z} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{u} \end{bmatrix} + \boldsymbol{e} \tag{3.55}$$

where both $\boldsymbol{\beta}$ and $\boldsymbol{u}$ are fixed effects, and $\mathrm{Var}(\boldsymbol{y}) = \mathrm{Var}(\boldsymbol{e}) = \boldsymbol{R}$. Then the *GLS* equations are

$$\boldsymbol{W'R}^{-1}\boldsymbol{W} \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}} \end{bmatrix} = \boldsymbol{W'R}^{-1}\boldsymbol{y} \tag{3.56}$$

where $\boldsymbol{W} = \begin{bmatrix} \boldsymbol{X} & \boldsymbol{Z} \end{bmatrix}$ so,

$$\begin{bmatrix} \boldsymbol{X'} \\ \boldsymbol{Z'} \end{bmatrix} \boldsymbol{R}^{-1} \begin{bmatrix} \boldsymbol{X} & \boldsymbol{Z} \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X'R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z'R}^{-1}\boldsymbol{y} \end{bmatrix} \tag{3.57}$$

$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z} \end{bmatrix} \begin{bmatrix} \widehat{\beta} \\ \widehat{\boldsymbol{u}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{y} \end{bmatrix} \tag{3.58}$$

When $\boldsymbol{u}$ is random with $\mathrm{E}(\boldsymbol{u}) = \boldsymbol{0}$ and $\mathrm{Var}(\boldsymbol{u}) = \boldsymbol{G}$ and $\mathrm{Cov}(\boldsymbol{u}, \boldsymbol{e}') = \boldsymbol{0}$, *BLUE* of estimable functions of $\boldsymbol{\beta}$ and *BLUP* of $\boldsymbol{u}$ can be obtained from H.M.M.E

$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1} \end{bmatrix} \begin{bmatrix} \widehat{\beta} \\ \widehat{\boldsymbol{u}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{y} \end{bmatrix} \tag{3.59}$$

BLUE of estimable $\boldsymbol{k}'\boldsymbol{\beta}$ is $\boldsymbol{k}'\widehat{\boldsymbol{\beta}}$ and *BLUP* of $\boldsymbol{u}$ is $\widehat{\boldsymbol{u}}$.

To see the above is true, we need to consider a very useful result in the inverse of partitioned matrices. This result is useful in deriving $\boldsymbol{A}^{-1}$ also. Consider the nonsingular matrix

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix} \tag{3.60}$$

Then,

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}^{11} & \boldsymbol{A}^{12} \\ \boldsymbol{A}^{21} & \boldsymbol{A}^{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \tag{3.61}$$

and consider now

$$\begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}^{12} \\ \boldsymbol{A}^{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{I} \end{bmatrix} \tag{3.62}$$

Next we premultiply the first row of the above matrix equation by $(\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1})$ and then subtract it from the second row. As a result we can write

$$\boldsymbol{0}\boldsymbol{A}^{12} + (\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})\boldsymbol{A}^{22} = \boldsymbol{I}$$

and

$$\boldsymbol{A}^{22} = (\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^{-1} \tag{3.63}$$

Now consider

$$\begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}^{11} \\ \boldsymbol{A}^{21} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{0} \end{bmatrix} \tag{3.64}$$

and premultiply now the second row of this matrix equation by $(\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1})$ and then subtract it from the first row. This results in

$$(\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})\boldsymbol{A}^{11} = \boldsymbol{I}$$

and

$$\boldsymbol{A}^{11} = (\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1} \tag{3.65}$$

In order to compute $\boldsymbol{A}^{12}$ make use of the fact that $\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$. Consider now the partinoned result

$$\begin{bmatrix} \boldsymbol{A}^{11} & \boldsymbol{A}^{12} \\ \boldsymbol{A}^{21} & \boldsymbol{A}^{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \tag{3.66}$$

After the multiplication of the two matrices in the left hand side we can write

$$\boldsymbol{A}^{11}\boldsymbol{A}_{12} + \boldsymbol{A}^{12}\boldsymbol{A}_{22} = \boldsymbol{0} \tag{3.67}$$

and consequently

$$\begin{aligned}\boldsymbol{A}^{12} &= -\boldsymbol{A}^{11}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} \\ &= -(\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\end{aligned} \tag{3.68}$$

Can also write $\boldsymbol{A}^{22}$ in the same way from $\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$,

$$\boldsymbol{A}_{21}\boldsymbol{A}^{12} + \boldsymbol{A}_{22}\boldsymbol{A}^{22} = \boldsymbol{I}$$

and consequently

$$\begin{aligned}\boldsymbol{A}^{22} &= \boldsymbol{A}_{22}^{-1}(\boldsymbol{I} - \boldsymbol{A}_{21}\boldsymbol{A}^{12}) \\ &= \boldsymbol{A}_{22}^{-1} + \boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}(\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\end{aligned} \tag{3.69}$$

So,

$$\begin{aligned}\boldsymbol{A}^{22} &= (\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^{-1} \\ &= \boldsymbol{A}_{22}^{-1} - \boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}(\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\end{aligned} \tag{3.70}$$

Now based on $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}$ can write

$$\boldsymbol{A}_{11}\boldsymbol{A}^{12} + \boldsymbol{A}_{12}\boldsymbol{A}^{22} = \boldsymbol{0}$$

and as a result

$$\boldsymbol{A}^{12} = -\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{A}^{22} \tag{3.71}$$

Using again $\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$ can write

$$\boldsymbol{A}^{21}\boldsymbol{A}_{11} + \boldsymbol{A}^{22}\boldsymbol{A}_{12} = \boldsymbol{0}$$

and consequently

$$\boldsymbol{A}^{21} = -\boldsymbol{A}^{22}\boldsymbol{A}_{12}\boldsymbol{A}_{11}^{-1} \tag{3.72}$$

Now from $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}$

$$\boldsymbol{A}_{11}\boldsymbol{A}^{11} + \boldsymbol{A}_{12}\boldsymbol{A}^{21} = \boldsymbol{0}$$

and consequently

$$\begin{aligned}\boldsymbol{A}^{11} &= \boldsymbol{A}_{11}^{-1}(\boldsymbol{I} - \boldsymbol{A}_{12}\boldsymbol{A}^{21}) \\ &= \boldsymbol{A}_{11}^{-1} + \boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{A}^{22}\boldsymbol{A}_{12}\boldsymbol{A}_{11}^{-1}\end{aligned} \tag{3.73}$$

Based on the results discussed above can write

$$\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{bmatrix}^{-1} = \begin{bmatrix}\boldsymbol{A}_{11}^{-1} + \boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{A}^{22}\boldsymbol{A}_{12}\boldsymbol{A}_{11}^{-1} & -\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{A}^{22} \\ -\boldsymbol{A}^{22}\boldsymbol{A}_{12}\boldsymbol{A}_{11}^{-1} & (\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^{-1}\end{bmatrix} \tag{3.74}$$

and conclude that

$$\begin{bmatrix} \boldsymbol{A}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} -\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12} \\ \boldsymbol{I} \end{bmatrix} \boldsymbol{A}^{22} \begin{bmatrix} -\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & \boldsymbol{I} \end{bmatrix} = \boldsymbol{A}^{-1} \tag{3.75}$$

with $\boldsymbol{A}^{22} = (\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^{-1}$. It is useful to observe that when $\boldsymbol{A}_{22}$ is 1×1, 3.75 becomes

$$\underbrace{\begin{bmatrix} \boldsymbol{A}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}}_{\boldsymbol{B}} + \underbrace{\begin{bmatrix} -\boldsymbol{A}_{11}^{-1}\boldsymbol{a}_{12} \\ \boldsymbol{I} \end{bmatrix}}_{r} \boldsymbol{a}_{22} \underbrace{\begin{bmatrix} -\boldsymbol{a}_{21}\boldsymbol{A}_{11}^{-1} & 1 \end{bmatrix}}_{s} = \left(b_{1j} + \frac{r_i s_j}{a^{22}} \right) \tag{3.76}$$

with $a^{22} = (a_{22} - a_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{a}_{12})^{-1}$. In calculating the inverse of the additive relationship matrix $\boldsymbol{A}_{11}^{-1}\boldsymbol{a}_{12}$ is easily determined and is very sparse.

Consider now the inverse of

$$\boldsymbol{V} = \boldsymbol{ZGZ}' + \boldsymbol{R} = \left(\boldsymbol{R} - \boldsymbol{Z}(-\boldsymbol{G})\boldsymbol{Z}'\right) \tag{3.77}$$

with $\boldsymbol{R} = \boldsymbol{A}_{22}, \boldsymbol{Z} = \boldsymbol{A}_{21}, -\boldsymbol{G} = \boldsymbol{A}_{11}^{-1}$ and $\boldsymbol{Z}' = \boldsymbol{A}_{12}$, then

$$\begin{aligned} \boldsymbol{V}^{-1} &= \left(\boldsymbol{R} - \boldsymbol{Z}(-\boldsymbol{G})\boldsymbol{Z}'\right)^{-1} \\ &= \boldsymbol{R}^{-1} + \boldsymbol{R}^{-1}\boldsymbol{Z}\left(-\boldsymbol{G}^{-1} - \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{R}^{-1} \\ &= \boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\boldsymbol{Z}\underbrace{\left(\boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1}\right)^{-1}}_{\boldsymbol{P}}\boldsymbol{Z}'\boldsymbol{R}^{-1} \end{aligned} \tag{3.78}$$

Next a proof of the fact that Henderson's mixed model equations (HMME) gives *BLUE* and *BLUP* is provided. Consider

$$\begin{aligned} &\boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X}\widehat{\boldsymbol{\beta}} + \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z}\widehat{\boldsymbol{u}} = \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{y} \\ &\boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X}\widehat{\boldsymbol{\beta}} + \left(\boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1}\right)\widehat{\boldsymbol{u}} = \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{y} \end{aligned}$$

from the second equation can obtain $\widehat{\boldsymbol{u}} = \boldsymbol{P}\left(\boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{y} - \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)$ and then substitute $\widehat{\boldsymbol{u}}$ in the first equation. Then

$$\begin{aligned} &\boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X}\widehat{\boldsymbol{\beta}} + \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{ZP}\left(\boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{y} - \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right) = \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{y} \\ &\boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{ZPZ}'\boldsymbol{R}^{-1}\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{y} - \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{ZPZ}'\boldsymbol{R}^{-1}\boldsymbol{y} \\ &\boldsymbol{X}'\left(\boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\boldsymbol{ZPZ}'\boldsymbol{R}^{-1}\right)\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{X}'\left(\boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\boldsymbol{ZPZ}'\boldsymbol{R}^{-1}\right)\boldsymbol{y} \\ &\left(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right)\widehat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y} \end{aligned}$$

the generalized least squares equations which give *BLUE.BLUP* of $\boldsymbol{u}$ is

$$\begin{aligned} \boldsymbol{GZ}'\boldsymbol{V}^{-1}\left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right) &= \boldsymbol{GZ}'\left(\boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\boldsymbol{ZPZ}'\boldsymbol{R}^{-1}\right)(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \\ &= \left(\boldsymbol{GZ}'\boldsymbol{R}^{-1} - \boldsymbol{GZ}'\boldsymbol{R}^{-1}\boldsymbol{ZPZ}'\boldsymbol{R}^{-1}\right)(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \\ &= (\boldsymbol{G} - \boldsymbol{GZ}'\boldsymbol{R}^{-1}\boldsymbol{ZP})\underbrace{\left(\boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{y} - \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)}_{\text{Rhs of MME}} \end{aligned} \tag{3.79}$$

Note that

$$(\boldsymbol{G}^{-1} + \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z})\boldsymbol{P} = \boldsymbol{I}$$

and as a result

$$\boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z}\boldsymbol{P} = \boldsymbol{I} - \boldsymbol{G}^{-1}\boldsymbol{P}$$

So,

$$\begin{aligned}
BLUP(\boldsymbol{u}) &= \left[\boldsymbol{G} - \boldsymbol{G}(\boldsymbol{I} - \boldsymbol{G}^{-1}\boldsymbol{P})\right]\left(\boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{y} - \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right) \\
&= [\boldsymbol{G} - \boldsymbol{G} + \boldsymbol{P}]\left(\boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{y} - \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right) \\
&= \boldsymbol{P}\left(\boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{y} - \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right) \\
&= \widehat{\boldsymbol{u}}
\end{aligned} \tag{3.80}$$

Consider now the variance of *HMME* estimates. $\boldsymbol{X}$ is assumed to have full column rank. If $\boldsymbol{X}$ is not full column in the following results one has to use a generalized inverse instead of the unique inverse. From *HMME* can write

$$\begin{bmatrix}\widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}\end{bmatrix} = \underbrace{\begin{bmatrix}\boldsymbol{C}^{11} & \boldsymbol{C}^{12} \\ \boldsymbol{C}^{21} & \boldsymbol{C}^{22}\end{bmatrix}}_{\boldsymbol{C}^{-1}}\begin{bmatrix}\boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{y}\end{bmatrix} \tag{3.81}$$

and the variance can be written as

$$\begin{aligned}
\mathrm{Var}\begin{bmatrix}\widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}\end{bmatrix} &= \boldsymbol{C}^{-1}\begin{bmatrix}\boldsymbol{X}' \\ \boldsymbol{Z}'\end{bmatrix}\boldsymbol{R}^{-1}(\boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}' + \boldsymbol{R})\boldsymbol{R}^{-1}\begin{bmatrix}\boldsymbol{X} & \boldsymbol{Z}\end{bmatrix}\boldsymbol{C}^{-1} \\
&= \boldsymbol{C}^{-1}\begin{bmatrix}\boldsymbol{X}' \\ \boldsymbol{Z}'\end{bmatrix}(\boldsymbol{R}^{-1}\boldsymbol{Z})\boldsymbol{G}(\boldsymbol{Z}'\boldsymbol{R}^{-1})\begin{bmatrix}\boldsymbol{X} & \boldsymbol{Z}\end{bmatrix}\boldsymbol{C}^{-1} \\
&+ \boldsymbol{C}^{-1}\begin{bmatrix}\boldsymbol{X}' \\ \boldsymbol{Z}'\end{bmatrix}\boldsymbol{R}^{-1}\begin{bmatrix}\boldsymbol{X} & \boldsymbol{Z}\end{bmatrix}\boldsymbol{C}^{-1} \\
&= \boldsymbol{C}^{-1}\begin{bmatrix}\boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z}\end{bmatrix}\boldsymbol{G}\begin{bmatrix}\boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z}\end{bmatrix}\boldsymbol{C}^{-1} \\
&+ \boldsymbol{C}^{-1}\begin{bmatrix}\boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z}\end{bmatrix}\boldsymbol{C}^{-1}
\end{aligned} \tag{3.82}$$

Note that $\boldsymbol{C}^{-1}\boldsymbol{C} = \boldsymbol{I}$ and,

$$\begin{bmatrix}\boldsymbol{C}^{11} & \boldsymbol{C}^{12} \\ \boldsymbol{C}^{21} & \boldsymbol{C}^{22}\end{bmatrix}\begin{bmatrix}\boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1}\end{bmatrix} = \begin{bmatrix}\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}\end{bmatrix} \tag{3.83}$$

So,

$$\begin{bmatrix}\boldsymbol{C}^{11} & \boldsymbol{C}^{12} \\ \boldsymbol{C}^{21} & \boldsymbol{C}^{22}\end{bmatrix}\begin{bmatrix}\boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z}\end{bmatrix} = \begin{bmatrix}\boldsymbol{I} & -\boldsymbol{C}^{12}\boldsymbol{G}^{-1} \\ \boldsymbol{0} & \boldsymbol{I} - \boldsymbol{C}^{22}\boldsymbol{G}^{-1}\end{bmatrix} \tag{3.84}$$

and

$$\text{Var}\begin{bmatrix}\widehat{\boldsymbol{\beta}}\\ \widehat{\boldsymbol{u}}\end{bmatrix} = \begin{bmatrix}-\boldsymbol{C}^{12}\boldsymbol{G}^{-1}\\ \boldsymbol{I}-\boldsymbol{C}^{22}\boldsymbol{G}^{-1}\end{bmatrix}\boldsymbol{G}\begin{bmatrix}\boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z}\end{bmatrix}\boldsymbol{C}^{-1} + \begin{bmatrix}\boldsymbol{I} & \boldsymbol{C}^{12}\boldsymbol{G}^{-1}\\ \boldsymbol{0} & \boldsymbol{I}-\boldsymbol{C}^{22}\boldsymbol{G}^{-1}\end{bmatrix}\boldsymbol{C}^{-1}$$

Also note that $\boldsymbol{C}\boldsymbol{C}^{-1} = \boldsymbol{I}$ and,

$$\begin{bmatrix}\boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z}\end{bmatrix}\boldsymbol{C}^{-1} = \begin{bmatrix}-\boldsymbol{G}^{-1}\boldsymbol{C}^{21} & \boldsymbol{I}-\boldsymbol{G}^{-1}\boldsymbol{C}^{22}\end{bmatrix} \tag{3.85}$$

and,

$$\begin{aligned}\text{Var}\begin{bmatrix}\widehat{\boldsymbol{\beta}}\\ \widehat{\boldsymbol{u}}\end{bmatrix} &= \begin{bmatrix}-\boldsymbol{C}^{12}\\ \boldsymbol{G}-\boldsymbol{C}^{22}\end{bmatrix}\begin{bmatrix}-\boldsymbol{G}^{-1}\boldsymbol{C}^{21} & \boldsymbol{I}-\boldsymbol{G}^{-1}\boldsymbol{C}^{22}\end{bmatrix}\\ &\quad+\begin{bmatrix}\boldsymbol{C}^{11}-\boldsymbol{C}^{12}\boldsymbol{G}^{-1}\boldsymbol{C}^{21} & \boldsymbol{C}^{12}-\boldsymbol{C}^{12}\boldsymbol{G}^{-1}\boldsymbol{C}^{22}\\ (\boldsymbol{I}-\boldsymbol{C}^{22}\boldsymbol{G}^{-1})\boldsymbol{C}^{21} & (\boldsymbol{I}-\boldsymbol{C}^{22}\boldsymbol{G}^{-1})\boldsymbol{C}^{22}\end{bmatrix}\\ &= \begin{bmatrix}\boldsymbol{C}^{11} & \boldsymbol{0}\\ \boldsymbol{0} & (\boldsymbol{G}-\boldsymbol{C}^{22})\end{bmatrix}\end{aligned} \tag{3.86}$$

So,

$$\text{Var}\begin{bmatrix}\widehat{\boldsymbol{\beta}}\\ \widehat{\boldsymbol{u}}\end{bmatrix} = \begin{bmatrix}\boldsymbol{C}^{11} & \boldsymbol{0}\\ \boldsymbol{0} & (\boldsymbol{G}-\boldsymbol{C}^{22})\end{bmatrix} \tag{3.87}$$

Also consider,

$$\begin{aligned}\text{Cov}\begin{bmatrix}\widehat{\boldsymbol{\beta}}, & (\widehat{\boldsymbol{u}}-\boldsymbol{u})'\end{bmatrix} &= -\text{Cov}(\widehat{\boldsymbol{\beta}}, \boldsymbol{u}')\\ &= -\begin{bmatrix}\boldsymbol{C}^{11} & \boldsymbol{C}^{12}\end{bmatrix}\begin{bmatrix}\boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z}\\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z}\end{bmatrix}\boldsymbol{G}\\ &= \boldsymbol{C}^{12}\boldsymbol{G}^{-1}\boldsymbol{G}\\ &= \boldsymbol{C}^{12}\end{aligned} \tag{3.88}$$

and,

$$\begin{aligned}\text{Var}(\widehat{\boldsymbol{u}}-\boldsymbol{u}) &= \text{Var}(\widehat{\boldsymbol{u}}) + \text{Var}(\boldsymbol{u}) - 2\text{Cov}(\boldsymbol{u}, \widehat{\boldsymbol{u}}')\\ &= \text{Var}(\widehat{\boldsymbol{u}}) + \text{Var}(\boldsymbol{u}) - 2\text{Var}(\widehat{\boldsymbol{u}})\\ &= \text{Var}(\boldsymbol{u}) - \text{Var}(\widehat{\boldsymbol{u}})\\ &= \boldsymbol{G} - (\boldsymbol{G}-\boldsymbol{C}^{22})\\ &= \boldsymbol{C}^{22}\end{aligned} \tag{3.89}$$

Finally,

$$\text{Var}\begin{bmatrix}\widehat{\boldsymbol{\beta}}\\ \widehat{\boldsymbol{u}}-\boldsymbol{u}\end{bmatrix} = \begin{bmatrix}\boldsymbol{C}^{11} & \boldsymbol{C}^{12}\\ \boldsymbol{C}^{21} & \boldsymbol{C}^{22}\end{bmatrix} \tag{3.90}$$

Assume now

$$\begin{bmatrix}\boldsymbol{u}\\ \boldsymbol{e}\end{bmatrix} \sim \boldsymbol{N}(\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix}, \begin{bmatrix}\boldsymbol{G} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{R}\end{bmatrix}) \tag{3.91}$$

The joint density of $\boldsymbol{u}$ and $\boldsymbol{y}$ is given by

$$
\begin{aligned}
&f(\boldsymbol{u}, \boldsymbol{y}; \boldsymbol{\beta}, \boldsymbol{G}, \boldsymbol{R}) = \\
&\left(\frac{1}{2\pi}\right)^{n/2} \mid \boldsymbol{R} \mid^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{Zu})'\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{Zu})\right\} \\
&\left(\frac{1}{2\pi}\right)^{q/2} \mid \boldsymbol{G} \mid^{-1/2} \exp\left\{\boldsymbol{u}'\boldsymbol{G}^{-1}\boldsymbol{u}\right\}
\end{aligned}
$$

Henderson obtained the MME by maximizing the above relation with respect to $\boldsymbol{\beta}$ and $\boldsymbol{u}$. Note that this is not really maximum likelihood. But the mixed model equations give,

$$
BLUP(T) = \widehat{T} = \boldsymbol{\lambda}'\widehat{\boldsymbol{\beta}} + \boldsymbol{C}'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \tag{3.92}
$$

where under MVN $\widehat{\boldsymbol{\beta}}$ is the MLE of $\boldsymbol{\beta}$ and

$$
\boldsymbol{\lambda}'\boldsymbol{\beta} + \boldsymbol{C}'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta}) = \mathrm{E}(T \mid \boldsymbol{y}) \tag{3.93}
$$

So,$BLUP$(T) is the MLE of $\mathrm{E}(T \mid \boldsymbol{y})$.

Under MVN can show that

$$
BLUP(\boldsymbol{u}) = \widehat{\boldsymbol{u}} = \mathrm{E}(\boldsymbol{u} \mid \boldsymbol{w}) \tag{3.94}
$$

where

$$
\boldsymbol{w} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} = \overbrace{\left[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{V}^{-1}\right]}^{\boldsymbol{M}}\boldsymbol{y} \tag{3.95}
$$

and $\mathrm{E}(\boldsymbol{w}) = \boldsymbol{0}$. Let

$$
\boldsymbol{P} = (\boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{V}^{-1}) \tag{3.96}
$$

and note that $\boldsymbol{PVP} = \boldsymbol{P}$ and $\boldsymbol{M} = \boldsymbol{VP}$. Then,

$$
\mathrm{Var}(\boldsymbol{w}) = \boldsymbol{VPVPV} = \boldsymbol{VPV} \tag{3.97}
$$

Note also that

$$
\boldsymbol{VPVV}^{-1}\boldsymbol{VPV} = \boldsymbol{VPVPV} = \boldsymbol{VPV} \tag{3.98}
$$

So, $\boldsymbol{V}^{-1}$ is a generalized inverse of $\mathrm{Var}(\boldsymbol{w}) = \boldsymbol{VPV}$. Now we can show that

$$
\begin{aligned}
&BLUP(\boldsymbol{u})\\
&= \mathrm{Cov}(\boldsymbol{u}, \boldsymbol{w}')\mathrm{Var}(\boldsymbol{w})\boldsymbol{w}\\
&= (\boldsymbol{GZ}'\boldsymbol{PV})\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})\\
&= \boldsymbol{GZ}'\boldsymbol{P}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})\\
&= \boldsymbol{GZ}'(\boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{V}^{-1})(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})\\
&= \boldsymbol{GZ}'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) - \boldsymbol{GZ}'\boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{V}^{-1})\boldsymbol{y}\\
&+ \boldsymbol{GZ}'\boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{V}^{-1})\boldsymbol{X}\widehat{\boldsymbol{\beta}})\\
&= \boldsymbol{C}'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) - \boldsymbol{GZ}'\boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-}\underbrace{(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y} - (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})\widehat{\boldsymbol{\beta}})}_{0}\\
&= \boldsymbol{C}'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})
\end{aligned}
$$

Can show that under MVN $\boldsymbol{C}^{*}\boldsymbol{W}^{-}\boldsymbol{w} = \mathrm{E}(\boldsymbol{u} \mid \boldsymbol{w})$ where $\mathrm{Cov}(\boldsymbol{u}, \boldsymbol{w}') = \boldsymbol{C}^{*}$ and $\mathrm{Var}(\boldsymbol{w}) = \boldsymbol{W}$. Write $\boldsymbol{u} = \widehat{\boldsymbol{u}} + (\boldsymbol{u} - \widehat{\boldsymbol{u}})$ and consequently

$$
\begin{aligned}
\mathrm{E}(\boldsymbol{u} \mid \boldsymbol{w}) &= \mathrm{E}(\widehat{\boldsymbol{u}} \mid \boldsymbol{w}) + \mathrm{E}\left[(\boldsymbol{u} - \widehat{\boldsymbol{u}}) \mid \boldsymbol{w}\right]\\
&= \widehat{\boldsymbol{u}} + \mathrm{E}\left[(\boldsymbol{u} - \widehat{\boldsymbol{u}}) \mid \boldsymbol{w}\right]
\end{aligned}
\tag{3.99}
$$

So, $\mathrm{E}(\boldsymbol{u} \mid \boldsymbol{w}) = \widehat{\boldsymbol{u}}$ if $\mathrm{E}\left[(\boldsymbol{u} - \widehat{\boldsymbol{u}}) \mid \boldsymbol{w}\right] = \boldsymbol{0}$. But, $\mathrm{E}(\boldsymbol{u} - \widehat{\boldsymbol{u}}) = \boldsymbol{0}$ and as a result $\mathrm{Cov}\left[(\boldsymbol{u} - \widehat{\boldsymbol{u}}), \boldsymbol{w}'\right] = \boldsymbol{0}$. Note that we had shown earlier that any linear function that had $\mathrm{E}(.) = 0$ has $\mathrm{Cov}() = 0$ with $(u - \widehat{u})$. Note also that the joint distribution of $\boldsymbol{w}$ and $\boldsymbol{u}$ is MVN so, $(\boldsymbol{u} - \widehat{\boldsymbol{u}})$ and $\boldsymbol{w}$ are independent and

$$
\mathrm{E}\left[(\boldsymbol{u} - \widehat{\boldsymbol{u}}) \mid \boldsymbol{w}\right] = \mathrm{E}(\boldsymbol{u} - \widehat{\boldsymbol{u}}) = \boldsymbol{0} \tag{3.100}
$$

Both $\boldsymbol{\beta}$ and $\boldsymbol{u}$ MVN (fixed versus random)

$$
\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}\\ \boldsymbol{e}\end{bmatrix} \sim \boldsymbol{N}(\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix}, \begin{bmatrix}\boldsymbol{D} & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{G} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{R}\end{bmatrix}) \tag{3.101}
$$

where

$$
\boldsymbol{\theta} = \begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}\end{bmatrix} \quad \text{and} \quad \sum = \begin{bmatrix}\boldsymbol{D} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{G}\end{bmatrix} \tag{3.102}
$$

and consider the model

$$
\begin{aligned}
\boldsymbol{y} &= \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e}\\
&= \boldsymbol{w\theta} + \boldsymbol{e}
\end{aligned}
\tag{3.103}
$$

with $\boldsymbol{w} = \begin{bmatrix} \boldsymbol{X} & \boldsymbol{Z} \end{bmatrix}$. Can write

$$\begin{aligned} f(\boldsymbol{\theta}, \boldsymbol{y}, \sum, \boldsymbol{R}) &= f(\boldsymbol{y} \mid \boldsymbol{\theta}, \sum, \boldsymbol{R}) f(\boldsymbol{\theta}, \sum, \boldsymbol{R}) \\ &\simeq \mid \boldsymbol{R} \mid^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{w}\boldsymbol{\theta})'\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{w}\boldsymbol{\theta})\right\} \\ &\quad \mid \sum \mid^{-1/2} \exp\left\{-\frac{1}{2}\boldsymbol{\theta}'\sum{}^{-1}\boldsymbol{\theta}\right\} \qquad (3.104) \\ &= \frac{\exp\{-\frac{1}{2}\overbrace{\left[(\boldsymbol{y} - \boldsymbol{w}\boldsymbol{\theta})'\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{w}\boldsymbol{\theta}) + \boldsymbol{\theta}'\sum^{-1}\boldsymbol{\theta}\right]}^{\boldsymbol{Q}}\}}{\mid \sum \mid^{1/2} \mid \boldsymbol{R} \mid^{1/2}} \end{aligned}$$

Consider now

$$\begin{aligned} \boldsymbol{Q} &= \boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{y} - 2\boldsymbol{\theta}'\boldsymbol{w}'\boldsymbol{R}^{-1}\boldsymbol{y} + \boldsymbol{\theta}'\boldsymbol{w}'\boldsymbol{R}^{-1}\boldsymbol{w}\boldsymbol{\theta} + \boldsymbol{\theta}'\sum{}^{-1}\boldsymbol{\theta} \\ &= \boldsymbol{y}\boldsymbol{R}^{-1}\boldsymbol{y} + \boldsymbol{\theta}'\left(\boldsymbol{w}'\boldsymbol{R}^{-1}\boldsymbol{w} + \sum{}^{-1}\right)\boldsymbol{\theta} - 2\boldsymbol{\theta}'\boldsymbol{w}'\boldsymbol{R}^{-1}\boldsymbol{y} \\ &= \boldsymbol{y}\boldsymbol{R}^{-1}\boldsymbol{y} + (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})'\left(\boldsymbol{w}'\boldsymbol{R}^{-1}\boldsymbol{w} + \sum{}^{-1}\right)(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\theta}}'\left(\boldsymbol{w}'\boldsymbol{R}^{-1}\boldsymbol{w} + \sum{}^{-1}\right)\widehat{\boldsymbol{\theta}} \end{aligned} \tag{3.105}$$

where $\left(\boldsymbol{w}'\boldsymbol{R}^{-1}\boldsymbol{w} + \sum^{-1}\right)\widehat{\boldsymbol{\theta}} = \boldsymbol{w}'\boldsymbol{R}^{-1}\boldsymbol{y}$. So,

$$\begin{aligned} f(\boldsymbol{\theta}, \boldsymbol{y}, \sum, \boldsymbol{R}) &\simeq \exp\left\{-\frac{1}{2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})'(\boldsymbol{w}'\boldsymbol{R}^{-1}\boldsymbol{w} + \sum{}^{-1})(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})\right\} \\ &\quad \exp\left\{-\frac{1}{2}\left[\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{y} - \widehat{\boldsymbol{\theta}}(\boldsymbol{w}'\boldsymbol{R}^{-1}\boldsymbol{w} + \sum{}^{-1})\widehat{\boldsymbol{\theta}}\right]\right\} \mid \sum \mid^{-1/2} \mid \boldsymbol{R} \mid^{-1/2} \end{aligned} \tag{3.106}$$

Also,

$$f(\boldsymbol{\theta} \mid \boldsymbol{y}, \sum, \boldsymbol{R}) = \frac{f(\boldsymbol{\theta}, \boldsymbol{y}, \sum, \boldsymbol{R})}{f(\boldsymbol{y}, \sum, \boldsymbol{R})} \tag{3.107}$$

where $f(\boldsymbol{y}, \sum, \boldsymbol{R}) = \int f(\boldsymbol{\theta}, \boldsymbol{y}, \sum, \boldsymbol{R}) d\boldsymbol{\theta}$;note that $\boldsymbol{y}$ is a constant with respect to $\boldsymbol{\theta}$. So,

$$f(\boldsymbol{\theta} \mid \boldsymbol{y}, \sum, \boldsymbol{R}) \simeq \frac{\exp\left\{-\frac{1}{2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})'(\boldsymbol{w}'\boldsymbol{R}^{-1}\boldsymbol{w} + \sum^{-1})(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})\right\}}{\mid \sum \mid^{-1/2} \mid \boldsymbol{R} \mid^{-1/2}} \tag{3.108}$$

and consequently

$$\boldsymbol{\theta} \mid \boldsymbol{y}, \sum, \boldsymbol{R} \sim \boldsymbol{N}\left[\widehat{\boldsymbol{\theta}}, (\boldsymbol{w}'\boldsymbol{R}^{-1}\boldsymbol{w} + \sum{}^{-1})^{-1}\right] \tag{3.109}$$

with $\mathrm{E}(\boldsymbol{\theta} \mid \boldsymbol{y}) = \widehat{\boldsymbol{\theta}}$ and $\mathrm{Var}(\boldsymbol{\theta} \mid \boldsymbol{y}) = (\boldsymbol{w}'\boldsymbol{R}^{-1}\boldsymbol{w} + \sum^{-1})^{-1}$ where,

$$\begin{aligned} \boldsymbol{w}'\boldsymbol{R}^{-1}\boldsymbol{w} + \sum{}^{-1} &= \begin{bmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X} + \boldsymbol{D}^{-1} & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1} \end{bmatrix} \\ \boldsymbol{w}'\boldsymbol{R}^{-1}\boldsymbol{y} &= \begin{bmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{y} \end{bmatrix} \end{aligned}$$

Note that if the variance of β is considered to be very large, $\boldsymbol{D}^{-1} \rightarrow \mathbf{0}$ and $\widehat{\boldsymbol{\theta}}$ becomes the *BLUP* solution.

Uncertainty about parameters is expressed by a density function. So,

$$f(\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{y}, \sum, \boldsymbol{R}) = f(\boldsymbol{y} \mid \boldsymbol{\beta}, \boldsymbol{u}, \sum, \boldsymbol{R}) f(\boldsymbol{\beta} \mid \sum) f(\boldsymbol{u} \mid \sum) f(\sum) f(\boldsymbol{R}) \tag{3.110}$$

where $f(\boldsymbol{\beta} \mid \sum)$ represents the "belief" density (normal) and $f(\sum)$ and $f(\boldsymbol{R})$ can be assumed 1 for particular values and 0 for all other values. In this way can obtain *BLUP*. However need to keep in mind that when you pretend $\boldsymbol{\beta} \sim \boldsymbol{N}(\mathbf{0}, \boldsymbol{D})$, or when you use Bayesian inference, can calculate $\mathrm{E}(T \mid \boldsymbol{y})$ but using this will not give maximum expected genetic progress. That is because we are not using the using the correct joint distribution of T and $\boldsymbol{y}$. Note that here expected genetic progress is a frequency (or sampling) definition. Also when we talk about correct joint distribution that is from a sampling point of view.

3.5 Genetic evaluation in populations undergoing selection

Selection results in:

- complex distribution of data and usual assumptions are violated.
- genetic parameters are changed, i.e. $\mathrm{E}(\boldsymbol{u}) \neq \mathbf{0}$

Example: Suppose y_1, y_2, y_3, y_4 are phenotypic records from four full sibs. Under additive inheritance,

$$\mathrm{Cov}(u_i, u_j) = \frac{1}{2}\sigma_a^2$$

Let y_{s_1}, and y_{s_2} be the phenotypic values of the highest ranking animals. Then,

$$\mathrm{Cov}(u_{s_1}, u_{s_2}) > \frac{1}{2}\sigma_a^2$$

Can show that if data used for selection are a subset of $\boldsymbol{y}$, $\mathrm{E}(\boldsymbol{u} \mid \boldsymbol{y})$ can be computed ignoring selection.

Two approaches have been used to model selection. A simple cow culling problem is used below to describe these two models of selection.

Model I:

	year1	year2
cow1	y_{11}	y_{12}
cow2	y_{21}	$-$

Also $y_{11} > y_{21}$ so, a second observation is obtained on cow 1. Under MVN can calculate distribution of

$$\begin{bmatrix} y_{11} \\ y_{21} \\ y_{12} \end{bmatrix} \quad \text{given } y_{11} > y_{21}$$

Henderson showed that given one-cycle of selection of this "type" $BLUP$ can be computed from usual MME. Note that with this model the distribution of

$$\begin{bmatrix} y_{11} \\ y_{21} \end{bmatrix}$$

is altered by selection.

Model II:

	A			B	
	year1	year2		year1	year2
cow1	y_{11}	y_{12}	cow1	y_{11}	$-$
cow2	y_{21}	$-$	cow2	y_{21}	y_{22}

Let $\boldsymbol{y}'_A = (y_{11}, y_{21}, y_{12})$ and $\boldsymbol{y}'_B = (y_{11}, y_{21}, y_{22})$. Suppose $\boldsymbol{y}_A$ is realized when $y_{11} > y_{21}$ and $\boldsymbol{y}_B$ is realized when $y_{11} \leq y_{21}$. In this model the distribution of

$$\begin{bmatrix} y_{11} \\ y_{21} \end{bmatrix}$$

is not changed by selection. Note that in both models the distribution of

$$\begin{bmatrix} y_{11} \\ y_{21} \\ y_{s2} \end{bmatrix}$$

is changed by selection.

3.5.1 Genetic evaluation

Consider genetic evaluation of N individuals from multiple, possibly overlapping, generations. The pedigree for these N individuals can take a large but finite number of possibilities: $P_1, P_2, \ldots, P_k$.

Under random mating, the joint density of phenotypic ($\boldsymbol{y}$) and genotypic ($\boldsymbol{u}$) values, conditional of pedigree $P = P_i$ is denoted by $f(\boldsymbol{u}, \boldsymbol{y}|P = P_i)$. Then, genetic evaluations are based on

$$f(\boldsymbol{u}|\boldsymbol{y}, P = P_i) = \frac{f(\boldsymbol{u}, \boldsymbol{y}|P = P_i)}{f(\boldsymbol{y}|P = P_i)}, \tag{3.111}$$

where $f(\boldsymbol{y}|P = P_i)$ is the marginal density of the phenotypic values:

$$f(\boldsymbol{y}|P = P_i) = \int f(\boldsymbol{u}, \boldsymbol{y}|P = P_i) d\boldsymbol{u}.$$

Under selection and non-random mating, the joint density of $\boldsymbol{y}$, $\boldsymbol{u}$, and P is different from the joint density of these random variables under random mating. Thus, (3.111) cannot be used for genetic evaluation.

As described below, selection based on some variable $\boldsymbol{z}$ is modeled by using a random variable s with sample space $\{1, 2, \ldots, k\}$ and distribution $D_s(\boldsymbol{\theta}(\boldsymbol{z}))$. Note that the distribution of s depends on $\boldsymbol{z}$ through the parameter vector $\boldsymbol{\theta}(\boldsymbol{z})$.

Selection is now modeled by specifying that data from pedigree P_i is realized only when $s = i$. Thus, the joint density of $\boldsymbol{y}$, $\boldsymbol{u}$, and P conditional on selection can be written in terms of their random-mating density as

$$\begin{aligned} g(\boldsymbol{u}, \boldsymbol{y}, P = P_i) &= \frac{f(\boldsymbol{u}, \boldsymbol{y}, P = P_i) \Pr(s = i|\boldsymbol{u}, \boldsymbol{y}, P = P_i)}{\Pr(s = i)} \\ &\propto \Pr(P = P_i) f(\boldsymbol{u}, \boldsymbol{y}|P = P_i) \Pr(s = i|\boldsymbol{u}, \boldsymbol{y}, P = P_i) \end{aligned} \tag{3.112}$$

where $\Pr(P = P_i)$ is the marginal probability of pedigree P_i under random mating. Now, genetic evaluations are based on

$$\begin{aligned} g(\boldsymbol{u}|\boldsymbol{y}, P = P_i, s = i) &= \frac{\Pr(P = P_i) f(\boldsymbol{u}, \boldsymbol{y}|P = P_i) \Pr(s = i|\boldsymbol{u}, \boldsymbol{y}, P = P_i)}{\int \Pr(P = P_i) f(\boldsymbol{u}, \boldsymbol{y}|P = P_i) \Pr(s = i|\boldsymbol{u}, \boldsymbol{y}, P = P_i) d\boldsymbol{u}} \\ &= \frac{\Pr(P = P_i) f(\boldsymbol{u}, \boldsymbol{y}|P = P_i) \Pr(s = i|\boldsymbol{u}, \boldsymbol{y}, P = P_i)}{\Pr(P = P_i) f(\boldsymbol{y}|P = P_i) \Pr(s = i|\boldsymbol{y}, P = P_i)} \\ &= \frac{f(\boldsymbol{u}, \boldsymbol{y}|P = P_i)}{f(\boldsymbol{y}|P = P_i)} \frac{\Pr(s = i|\boldsymbol{u}, \boldsymbol{y}, P = P_i)}{\Pr(s = i|\boldsymbol{y}, P = P_i)}. \end{aligned} \tag{3.113}$$

In the above equation, when the ratio

$$\frac{\Pr(s = i|\boldsymbol{u}, \boldsymbol{y}, P = P_i)}{\Pr(s = i|\boldsymbol{y}, P = P_i)} \tag{3.114}$$

is unity, (3.113) reduces to (3.111) and selection can be ignored. The numerator and denominator of (3.114) are identical under the following conditions:

1. s is independent of $\boldsymbol{y}$, $\boldsymbol{u}$, and P:

$$\Pr(s = i|\boldsymbol{u}, \boldsymbol{y}, P = P_i) = \Pr(s = i)$$

   and

$$\Pr(s = i|\boldsymbol{y}, P = P_i) = \Pr(s = i)$$

2. conditional on $\boldsymbol{y}$ and P, s is independent of $\boldsymbol{u}$:

$$\Pr(s = i|\boldsymbol{u}, \boldsymbol{y}, P = P_i) = \Pr(s = i|\boldsymbol{y}, P = P_i)$$

The first condition is true when selection is based on some criterion unrelated to $\boldsymbol{y}$, $\boldsymbol{u}$, and P. The second condition is true for selection based on $\boldsymbol{y}$ and any other criterion that is conditionally independent of $\boldsymbol{u}$, given $\boldsymbol{y}$. This includes data based selection. However, if selection partly depends on $\boldsymbol{y}_{mis}$, not contained in $\boldsymbol{y}$, and

$$f(\boldsymbol{y}_{mis}, \boldsymbol{u}|\boldsymbol{y}) \neq f(\boldsymbol{y}_{mis}|\boldsymbol{y}) f(\boldsymbol{u}|\boldsymbol{y}),$$

neither of the conditions for ignoring selection is true.

3.5.2 *BLUP* with selection

Under *MVN*

$$BLUP(\boldsymbol{u}) = \widehat{\boldsymbol{u}} = \mathrm{E}(\boldsymbol{u} \mid \boldsymbol{w}) \tag{3.115}$$

So, as long as selection was based on $\boldsymbol{w} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}$, $BLUP(\boldsymbol{u})$ can be obtained from *HMME*. Note that

$$f_s(\boldsymbol{u}_i \mid \boldsymbol{w}_i) = \frac{f(\boldsymbol{u}_i, \boldsymbol{w}_i)}{f(\boldsymbol{w}_i)} \frac{\Pr(s(\boldsymbol{z}) = i \mid \boldsymbol{u}_i, \boldsymbol{w}_i)}{\Pr(s(\boldsymbol{z}) = i \mid \boldsymbol{w}_i)} \tag{3.116}$$

and that $\boldsymbol{z} = L'\boldsymbol{y}$ and because

$$\mathrm{E}(\boldsymbol{z}) = L'\boldsymbol{X}\boldsymbol{\beta} = 0 \quad \text{for any} \quad \boldsymbol{\beta} \quad \text{only if} \quad L'\boldsymbol{X} = 0 \tag{3.117}$$

if $L'\boldsymbol{X} \neq 0 \Rightarrow$ did not select on $\boldsymbol{w}$ and as a result cannot ignore the selection process. So if selection is across fixed effects cannot ignore selection. However for $\boldsymbol{T} = \boldsymbol{k}'\boldsymbol{\beta} + \boldsymbol{u}_i$

$$BLUP(\boldsymbol{T}) = \boldsymbol{k}'\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{u}_i} \quad \text{and is equal to the } MLE \text{ of } \mathrm{E}(\boldsymbol{T} \mid \boldsymbol{y})$$

Note that if $\boldsymbol{\beta}$ were known, under *MVN*, we can compute $\mathrm{E}(\boldsymbol{u} \mid \boldsymbol{y})$ or $\mathrm{E}(\boldsymbol{T} \mid \boldsymbol{y})$. This is not affected by selection across fixed effects! But, if we have good estimates "lots of data" to compute $\boldsymbol{\beta}$, what we get is almost $\mathrm{E}(\boldsymbol{u} \mid \boldsymbol{y})$ and $\mathrm{E}(\boldsymbol{T} \mid \boldsymbol{y})$. Remember that *MLE* are consistent! Note also that for the *BLUP* property to hold true, we don't need any assumptions, only need to know the second moments. Normality is required just for *BLUP* to be equal to the conditional mean.

3.6 Genetic evaluation under additive inheritance

In order to setup the *MME* need the inverse of $\mathrm{Var}(\boldsymbol{u}) = \boldsymbol{G}$. Under additive inheritance,

$$\mathrm{Var}(\boldsymbol{u}) = \boldsymbol{G} = \boldsymbol{A}\sigma_a^2 \tag{3.118}$$

where $\boldsymbol{A}$ is the additive relationship matrix with elements:

$$a_{ij} = 2\Pr(\text{a random allele from } i \text{ is IBD to a random allele in } j)$$

If i is not a descendant of j then,

$$\begin{aligned} a_{ij} &= \frac{1}{2}(a_{is_j} + a_{id_j}) \\ a_{ii} &= 1 + \frac{a_{s_i d_i}}{2} \end{aligned}$$

This leads to the tabular method to compute $\boldsymbol{A}$. Following steps provide $\boldsymbol{A}$:

1. Number individuals such that parents precede offspring.

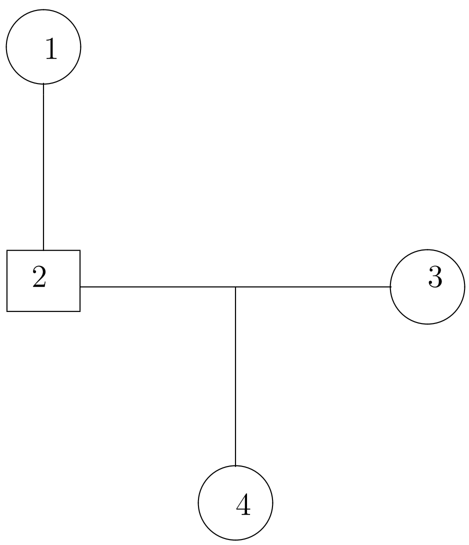
2. For founders (individuals without parents) enter 1 on diagonal and 0 on off-diagonals.

3. For non-founder i calculate row elements 1 to $i-1$ as the average of the parental row elements.

4. Set the diagonal element i to $1+\frac{a_{s_i d_i}}{2}$.

5. Fill columns by symmetry.

Consider the following pedigree as an example:



For this pedigree the $\boldsymbol{A}$ matrix is given by:

$$\begin{bmatrix} 1 & 0.5 & 0 & 0.25 \\ 0.5 & 1 & 0 & 0.5 \\ 0 & 0 & 1 & 0.5 \\ 0.25 & 0.5 & 0.5 & 1 \end{bmatrix}$$

In matrix notation the tabular method becomes

$$\boldsymbol{A}_i = \begin{bmatrix} \boldsymbol{A}_{i-1} & \boldsymbol{A}_{i-1}\boldsymbol{q}_i \\ \boldsymbol{q}_i'\boldsymbol{A}_{i-1} & 1+\frac{a_{s_i d_i}}{2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{a}_{12} \\ \boldsymbol{a}_{21} & a_{22} \end{bmatrix}$$

where $\boldsymbol{A}_i$ is the relationship matrix expanded up to individual i, $\boldsymbol{q}_i$ has only at most 2 non-zero elements ($=\frac{1}{2}$) corresponding to the parents of i. For our example

$$
\begin{aligned}
&\boldsymbol{A}_1 = 1\\
&\boldsymbol{A}_2 = \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_1\boldsymbol{q}_2 \\ \boldsymbol{q}_2'\boldsymbol{A}_1 & 1 \end{bmatrix} \quad \boldsymbol{q}_2' = \frac{1}{2} \Rightarrow \boldsymbol{A}_2 = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1 \end{bmatrix}\\
&\boldsymbol{A}_3 = \begin{bmatrix} \boldsymbol{A}_2 & \boldsymbol{A}_2\boldsymbol{q}_3 \\ \boldsymbol{q}_3'\boldsymbol{A}_2 & 1 \end{bmatrix} \quad \boldsymbol{q}_3' = \begin{bmatrix} 0 & 0 \end{bmatrix} \Rightarrow \boldsymbol{A}_3 = \begin{bmatrix} 1 & 1/2 & 0 \\ 1/2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\\
&\boldsymbol{A}_4 = \begin{bmatrix} \boldsymbol{A}_3 & \boldsymbol{A}_3\boldsymbol{q}_4 \\ \boldsymbol{q}_4'\boldsymbol{A}_3 & 1 \end{bmatrix} \quad \boldsymbol{q}_4' = \begin{bmatrix} 0 & 1/2 & 1/2 \end{bmatrix} \Rightarrow \boldsymbol{A}_4 = \begin{bmatrix} 1 & 1/2 & 0 & 1/4 \\ 1/2 & 1 & 0 & 1/2 \\ 0 & 0 & 1 & 1/2 \\ 1/4 & 1/2 & 1/2 & 1 \end{bmatrix}
\end{aligned}
$$

Recall the inverse of a partitioned matrix

$$
\begin{aligned}
\boldsymbol{A} &= \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{a}_{12} \\ \boldsymbol{a}_{21} & a_{22} \end{bmatrix}\\
\boldsymbol{A}^{-1} &= \begin{bmatrix} \boldsymbol{A}_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -\boldsymbol{A}_{11}^{-1}\boldsymbol{a}_{12} \\ 1 \end{bmatrix} a^{22} \begin{bmatrix} -\boldsymbol{a}_{21}\boldsymbol{A}_{11}^{-1} & 1 \end{bmatrix}
\end{aligned}
$$

where $a^{22} = (a_{22} - \boldsymbol{a}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{a}_{12})^{-1}$. So,

$$
\begin{aligned}
\boldsymbol{A}_{i-1}^{-1} &= \begin{bmatrix} \boldsymbol{A}_{i-1}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -\boldsymbol{A}_{i-1}^{-1}\boldsymbol{A}_{i-1}\boldsymbol{q}_i \\ 1 \end{bmatrix} a^{22} \begin{bmatrix} -\boldsymbol{q}_i'\boldsymbol{A}_{i-1}\boldsymbol{A}_{i-1}^{-1} & 1 \end{bmatrix}\\
&= \begin{bmatrix} \boldsymbol{A}_{i-1}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -\boldsymbol{q}_i \\ 1 \end{bmatrix} a^{22} \begin{bmatrix} -\boldsymbol{q}_i' & 1 \end{bmatrix}
\end{aligned}
$$

where $a^{22} = (a_{22} - \boldsymbol{a}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{a}_{12})^{-1}$ and in general

$$
a^{ii} = (a_{ii} - \boldsymbol{q}_i'\boldsymbol{A}_{i-1}\boldsymbol{A}_{i-1}^{-1}\boldsymbol{A}_{i-1}\boldsymbol{q}_i)^{-1} = (a_{ii} - \boldsymbol{q}_i'\boldsymbol{A}_{i-1}\boldsymbol{q}_i)^{-1} \tag{3.119}
$$

where $a_{ii} = 1 + \frac{a_{s_i d_i}}{2}$. Consider now the example previously discussed

$$\boldsymbol{A}_1^{-1} = 1$$

$$\boldsymbol{q}_2 = 1/2 \Rightarrow (a_{22} - \boldsymbol{q}_2'\boldsymbol{A}_1\boldsymbol{q}_2)^{-1} = (1 - 0.25)^{-1} = \frac{4}{3}$$

$$\boldsymbol{A}_2^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -1/2 \\ 1 \end{bmatrix} 4/3 \begin{bmatrix} -1/2 & 1 \end{bmatrix} = \begin{bmatrix} 1+1/3 & -2/3 \\ -2/3 & 4/3 \end{bmatrix}$$

$$\boldsymbol{q}_3' = \begin{bmatrix} 0 & 0 \end{bmatrix} \Rightarrow (a_{33} - \boldsymbol{q}_3'\boldsymbol{A}_2\boldsymbol{q}_3)^{-1} = (1 - 0)^{-1} = 1$$

$$\boldsymbol{A}_3^{-1} = \begin{bmatrix} 1+1/3 & -2/3 & 0 \\ -2/3 & 4/3 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} 1 \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1+1/3 & -2/3 & 0 \\ -2/3 & 4/3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\boldsymbol{q}_4' = \begin{bmatrix} 0 & 1/2 & 1/2 \end{bmatrix} \Rightarrow (a_{44} - \boldsymbol{q}_4'\boldsymbol{A}_3\boldsymbol{q}_4)^{-1} = (1 - 1/2)^{-1} = 2$$

$$\boldsymbol{A}_4^{-1} = \begin{bmatrix} 1+1/3 & -2/3 & 0 & 0 \\ -2/3 & 4/3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 \\ -1/2 \\ -1/2 \\ 1 \end{bmatrix} 2 \begin{bmatrix} 0 & -1/2 & -1/2 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1+1/3 & -2/3 & 0 & 0 \\ -2/3 & 4/3+1/2 & 1/2 & -1 \\ 0 & 1/2 & 1+1/2 & -1 \\ 0 & -1 & -1 & 2 \end{bmatrix}$$

Note that the following rule can be used to compute a^{ii} for a pedigree without inbreeding:

$$a^{ii} = \frac{4}{m+2} \tag{3.120}$$

where m is the number of unknown parents. The following algorithm can be used to obtain $\boldsymbol{A}^{-1}$:

1. Set $\boldsymbol{A}^{-1} = \boldsymbol{0}$.
2. Compute a^{ii} for all animals first. Note that these values can be computed without computing the whole $\boldsymbol{A}$ matrix.

3. For each animal add the following to $\boldsymbol{A}^{-1}$

$$
\begin{aligned}
&a^{ii} \text{ to } (i,i)\\
&-\tfrac{1}{2}a^{ii} \text{ to } (i,s_i),(s_i,i),(i,d_i),(d_i,i)\\
&-\tfrac{1}{4}a^{ii} \text{ to } (s_i,s_i),(s_i,d_i),(d_i,s_i),(d_i,d_i)
\end{aligned}
$$

4. Omit entries for missing parents.

Chapter 4

Estimation of Variance Components

4.1 Maximum Likelihood

Consider $\boldsymbol{y}$, sampled from a distribution with density $f(\boldsymbol{y};\boldsymbol{\theta})$, for $\theta \in \Omega$. In this density the argument is $\boldsymbol{y}$. The likelihood is defined to be a function of $\boldsymbol{\theta}$:

$$L(\boldsymbol{\theta};\boldsymbol{y}) \propto f(\boldsymbol{y};\boldsymbol{\theta}) \tag{4.1}$$

Then $MLE(\boldsymbol{\theta}) = \widehat{\boldsymbol{\theta}}$ where

$$L(\widehat{\boldsymbol{\theta}};\boldsymbol{y}) = \mathrm{Max}_{\theta\in\Omega} L(\boldsymbol{\theta};\boldsymbol{y}) \tag{4.2}$$

This idea was made popular by Fisher, but may have been used earlier. Consider now the following intuitive explanation. Suppose that have data of type:

$$y_1 \dots y_n \quad \text{i.i.d} \quad \boldsymbol{N}(\mu_T, \sigma_T^2)$$

Can draw histogram from data and find the expected distribution for $\widehat{\boldsymbol{\theta}} = \begin{bmatrix} \widehat{\mu} & \widehat{\sigma^2} \end{bmatrix}$. Can "superimpose" expected distribution with empirical distribution by choosing $\mu = \widehat{\mu}, \sigma^2 = \widehat{\sigma^2}$. Can say that the maximum likelihood helps choose which particular distribution fits "best" the data.

4.1.1 Cauchy - Schwartz Inequality

Lemma 4.1.1

$$Var(T) \geq \frac{Cov^2(T,Y)}{Var(Y)} \tag{4.3}$$

Proof: Let,

$$\widehat{T} = \mathrm{E}(T) + \frac{\mathrm{Cov}(T,Y)}{\mathrm{Var}(Y)}[Y - \mathrm{E}(Y)]$$

and write

$$T = \widehat{T} + \underbrace{(T - \widehat{T})}_{Z}$$

recall also that $\mathrm{E}(Z) = 0$ and $\mathrm{Cov}(Y, Z) = 0 \Rightarrow \mathrm{Cov}(\widehat{T}, Z) = 0$ So,

$$\begin{aligned}\mathrm{Var}(T) &= \mathrm{Var}(\widehat{T}) + \mathrm{Var}(Z) \\ &= \frac{\mathrm{Cov}^2(T, Y)}{\mathrm{Var}(Y)} + \mathrm{Var}(Z)\end{aligned}$$

So,

$$\mathrm{Var}(T) \geq \frac{\mathrm{Cov}^2(T, Y)}{\mathrm{Var}(Y)} \tag{4.4}$$

unless, T is a linear function of Y. Then, $Z = 0$ and $\mathrm{Var}(T) = \frac{\mathrm{Cov}^2(T,Y)}{\mathrm{Var}(Y)}$.

More generally, let $\boldsymbol{y}$ be a vector and $\widehat{T}$ the *BLP* of $\boldsymbol{T}$. Then consider $T = \widehat{T} + Z$ where

$$\widehat{T} = \mathrm{E}(T) + \boldsymbol{c}'V^{-1}(\boldsymbol{y} - \boldsymbol{X\beta})$$

and $\mathrm{E}(Z) = 0$ and $\mathrm{Cov}(\boldsymbol{y}, Z) = 0 \Rightarrow \mathrm{Cov}(\widehat{T}, Z) = 0$ So,

$$\begin{aligned}\mathrm{Var}(T) &= \mathrm{Var}(\widehat{T}) + \mathrm{Var}(Z) \\ &= \boldsymbol{c}'V^{-1}\boldsymbol{c} + \mathrm{Var}(Z)\end{aligned}$$

and consequently

$$\mathrm{Var}(T) \geq \boldsymbol{c}'V^{-1}\boldsymbol{c} \tag{4.5}$$

4.1.2 Jensen's Inequality

Lemma 4.1.2 *If $g(x)$ is a convex function and $E(x) = \mu$,*

$$E[g(x)] \geq g(\mu) \tag{4.6}$$

with equality only when x is the constant μ.

Proof: Convex functions have the property:

$$g(x) \geq g(x_0) + g'(x_0)(x - x_0)$$

Now let $x_0 = \mu$. Then,

$$g(x) \geq g(\mu) + g'(\mu)(x - \mu)$$

and

$$\mathrm{E}[g(x)] \geq g(\mu)$$

because $\mathrm{E}(x) = \mu$.

4.1.3 Kullback-Leibler Inequality

Lemma 4.1.3 *Suppose $f(\boldsymbol{y};\boldsymbol{\theta})$ is the density function for random variable $\boldsymbol{y}$ with $\boldsymbol{\theta} \in \Omega$.*

$$E_{\theta^*} \log \frac{f(\boldsymbol{y};\boldsymbol{\theta^*})}{f(\boldsymbol{y};\boldsymbol{\theta})} \geq 0 \quad \textit{for} \quad \boldsymbol{\theta} \in \Omega \tag{4.7}$$

Proof:

$$\mathrm{E}_{\theta^*} \log \frac{f(\boldsymbol{y};\boldsymbol{\theta^*})}{f(\boldsymbol{y};\boldsymbol{\theta})} = \mathrm{E}_{\theta^*} \left[\underbrace{- \log \frac{f(\boldsymbol{y};\boldsymbol{\theta})}{f(\boldsymbol{y};\boldsymbol{\theta^*})}}_{\text{convex function}} \right] \tag{4.8}$$

Now using Jensen's Inequality,

$$\mathrm{E}_{\theta^*} \left[- \log \frac{f(\boldsymbol{y};\boldsymbol{\theta})}{f(\boldsymbol{y};\boldsymbol{\theta^*})} \right] \geq - \log \mathrm{E}_{\theta^*} \left[\frac{f(\boldsymbol{y};\boldsymbol{\theta})}{f(\boldsymbol{y};\boldsymbol{\theta^*})} \right] = - \log 1 = 0$$

because

$$\mathrm{E}_{\theta^*} \left[\frac{f(\boldsymbol{y};\boldsymbol{\theta})}{f(\boldsymbol{y};\boldsymbol{\theta^*})} \right] = \int \frac{f(\boldsymbol{y};\boldsymbol{\theta})}{f(\boldsymbol{y};\boldsymbol{\theta^*})} f(\boldsymbol{y};\boldsymbol{\theta^*}) d\boldsymbol{y} = \int f(\boldsymbol{y};\boldsymbol{\theta}) d\boldsymbol{y} = 1$$

4.1.4 Consistency of Maximum Likelihood Estimates

Suppose the data $\boldsymbol{y}$ can be partitioned as $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n$ *iid* vectors. Then

$$\log f(\boldsymbol{y};\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(\boldsymbol{y}_i;\boldsymbol{\theta}) \propto \log L(\boldsymbol{\theta};\boldsymbol{y}) \tag{4.9}$$

From the Kullback-Leibler Inequality,

$$\mathrm{E}_{\boldsymbol{\theta}^*} \left[\log f(\boldsymbol{y};\boldsymbol{\theta}^*) - \log f(\boldsymbol{y};\boldsymbol{\theta}^* \pm \boldsymbol{\delta}) \right] \geq 0 \tag{4.10}$$

But as $n \to \infty$

$$\frac{1}{n} \sum_{i=1}^{n} \log f(\boldsymbol{y}_i;\boldsymbol{\theta}) \to \mathrm{E} \log f(\boldsymbol{y}_i;\boldsymbol{\theta}) \tag{4.11}$$

So as $n \to \infty$,

$$\frac{1}{n} \left[\log L(\boldsymbol{\theta}^*;\boldsymbol{y}) - \log L(\boldsymbol{\theta}^* \pm \boldsymbol{\delta};\boldsymbol{y}) \right] \geq 0 \tag{4.12}$$

Also as $n \to \infty$ the maximum of $L(\boldsymbol{\theta};\boldsymbol{y})$ is at $\boldsymbol{\theta}^*$. In words the previous results are described as follows:

- $\frac{1}{n} \log L(\boldsymbol{\theta};\boldsymbol{y})$ converges to its expected value.
- The $\mathrm{E} \log L(\boldsymbol{\theta};\boldsymbol{y})$ is maximum at $\boldsymbol{\theta}^*$.
- In "large" samples $\frac{1}{n} \log L(\boldsymbol{\theta};\boldsymbol{y})$ as well as $L(\boldsymbol{\theta};\boldsymbol{y})$ are maximized at $\boldsymbol{\theta}^*$.

4.1.5 Cramèr-Rao Lower Bond

Lemma 4.1.4 *Let* $\boldsymbol{s}(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \log f(\boldsymbol{y}; \boldsymbol{\theta})$ *Then*

$$E_{\theta^*}\left[|\boldsymbol{s}(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta^*}} \right] = \mathbf{0} \tag{4.13}$$

Proof:

$$\begin{aligned}
\mathrm{E}_{\theta^*}\left[|\boldsymbol{s}(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta^*}} \right] &= \int \left| \frac{\partial}{\partial \boldsymbol{\theta}} \log f(\boldsymbol{y}; \boldsymbol{\theta}) \right|_{\boldsymbol{\theta}=\boldsymbol{\theta^*}} f(\boldsymbol{y}; \boldsymbol{\theta}^*)\, d\boldsymbol{y} \\
&= \int \frac{1}{f(\boldsymbol{y}; \boldsymbol{\theta}^*)} \left| \frac{\partial}{\partial \boldsymbol{\theta}} f(\boldsymbol{y}; \boldsymbol{\theta}) \right|_{\boldsymbol{\theta}=\boldsymbol{\theta^*}} f(\boldsymbol{y}; \boldsymbol{\theta}^*)\, d\boldsymbol{y} \\
&= \left| \frac{\partial}{\partial \boldsymbol{\theta}} \underbrace{\int f(\boldsymbol{y}; \boldsymbol{\theta})\, d\boldsymbol{y}}_{1} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta^*}} \\
&= 0
\end{aligned}$$

So, the expected log likelihood has a root at the true value. However this result is not as powerful as earlier result.

Lemma 4.1.5 *Suppose* $\widehat{\theta}_i$ *is an unbiased estimator of* θ_i^*. *Then,*

$$Var(\widehat{\theta}_i) \geq b^{ii} \tag{4.14}$$

where $\boldsymbol{B} = Var[s(\theta^*)]$.

Proof: $\widehat{\theta}_i$ is an unbiased estimator of θ_i^* so,

$$\mathrm{E}_{\theta^*}(\widehat{\theta}_i) = \theta_i^* = \int \widehat{\theta}_i f(\boldsymbol{y}; \boldsymbol{\theta}^*)\, d\boldsymbol{y}$$

Note that

$$\left[\frac{\partial}{\partial \theta_i} \mathrm{E}(\widehat{\theta}_i) \right]_{\theta_i = \theta_i^*} = 1 \tag{4.15}$$

but also

$$\begin{aligned}
\frac{\partial}{\partial \theta_i} \mathrm{E}_{\theta^*}(\widehat{\theta}_i) &= \int \widehat{\theta}_i \underbrace{\frac{1}{f(\boldsymbol{y}; \boldsymbol{\theta}^*)} \left[\frac{\partial f(\boldsymbol{y}; \boldsymbol{\theta})}{\partial \theta_i} \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}}_{\frac{\partial}{\partial \theta_i} \log f(\boldsymbol{y}; \boldsymbol{\theta})} f(\boldsymbol{y}; \boldsymbol{\theta}^*)\, d\boldsymbol{y} \\
&= \int \widehat{\theta}_i \underbrace{\left[\frac{\partial}{\partial \theta_i} \log f(\boldsymbol{y}; \boldsymbol{\theta}) \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}}_{s(\boldsymbol{\theta}^*)} f(\boldsymbol{y}; \boldsymbol{\theta}^*)\, d\boldsymbol{y} \\
&= \int \widehat{\theta}_i s_i f(\boldsymbol{y}; \boldsymbol{\theta}^*)\, d\boldsymbol{y} \\
&= \mathrm{E}_{\theta^*}\left(\widehat{\theta}_i, s_i \right) \\
&= \mathrm{Cov}\left(\widehat{\theta}_i, s_i \right) \\
&= 1
\end{aligned}$$

because $\mathrm{E}_{\theta^*}\left[s(\boldsymbol{\theta}^*)\right] = 0 \Rightarrow \mathrm{E}_{\theta^*}(s_i) = 0$

$$\frac{\partial}{\partial \theta_j}\mathrm{E}_{\theta^*}(\widehat{\theta}_i) = \frac{\partial}{\partial \theta_j}\theta_i = 0 = \mathrm{Cov}\left(\widehat{\theta}_i, s_j\right) \tag{4.16}$$

So,

$$\mathrm{Cov}\left(\widehat{\theta}_i, \boldsymbol{s}'\right) = \begin{bmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{bmatrix} = \boldsymbol{c}_i' \tag{4.17}$$

From the generalized Cauchy-Schwarz inequality and nature of $\boldsymbol{c}_i'$,

$$\begin{aligned} &\mathrm{Var}(\widehat{\theta}_i) \geq \boldsymbol{c}_i'\boldsymbol{B}^{-1}\boldsymbol{c}_i \\ &\boldsymbol{c}_i' = \begin{bmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{bmatrix} \end{aligned}$$

and conclude that

$$\mathrm{Var}(\widehat{\theta}_i) \geq b^{ii} \tag{4.18}$$

with equality achieved when $\widehat{\theta}_i$ is a linear function of $\boldsymbol{s}$.

4.1.6 Newton Raphson Algorithm

Newton Raphson Algorithm is used for finding the root of a function. The use of this algorithm allows the examination of properties of *MLE*. Let $g(x)$ be some function of x. Want to find the value of x such that for $x = \widehat{x} \rightarrow g(\widehat{x}) = 0$. The reasoning is based on the use of Taylor series to approximate $g(x)$ as:

$$g(x) \approx g(x_0) + g'(x_0)(x - x_0) \tag{4.19}$$

Now set (4.19) to zero and solve for $\widehat{x}$ as

$$\begin{aligned} &g(x_0) + g'(x_0)(\widehat{x} - x_0) = 0 \\ &g'(x_0)(\widehat{x} - x_0) = -g(x_0) \\ &\widehat{x} = x_0 - \frac{g(x_0)}{g'(x_0)} \end{aligned}$$

Consider now a vector of functions

$$\begin{aligned} \boldsymbol{g}(\boldsymbol{x}) &\approx \boldsymbol{g}(\boldsymbol{x_0}) + \underbrace{\frac{\partial \boldsymbol{g}(\boldsymbol{x_0})}{\partial \boldsymbol{x}'}}_{\boldsymbol{B}}(\boldsymbol{x} - \boldsymbol{x_0}) \\ &\approx \boldsymbol{g}(\boldsymbol{x_0}) + \boldsymbol{B}(\boldsymbol{x} - \boldsymbol{x_0}) \end{aligned} \tag{4.20}$$

The last result represents the Taylor series approximation of $\boldsymbol{g}(\boldsymbol{x})$. The next step is to set (4.20) equal to zero and solve for $\widehat{\boldsymbol{x}}$.

$$\begin{aligned} &\boldsymbol{B}(x - x_0) = -\boldsymbol{g}(\boldsymbol{x_0}) \\ &\widehat{\boldsymbol{x}} = \boldsymbol{x_0} - \boldsymbol{B}^{-1}\boldsymbol{g}(\boldsymbol{x_0}) \end{aligned} \tag{4.21}$$

4.1.7 Application of Newton Raphson to MLE

Let $U(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}; \boldsymbol{y})$. Want to find $\widehat{\boldsymbol{\theta}}$ such that

$$\frac{\partial}{\partial \boldsymbol{\theta}} \left[U(\boldsymbol{\theta}) \right]_{\theta=\widehat{\theta}} = \boldsymbol{s}(\widehat{\boldsymbol{\theta}}) = \mathbf{0} \tag{4.22}$$

Applying Newton Raphson algorithm gives a way to obtain MLE,

$$\boldsymbol{\theta}^{[i]} = \boldsymbol{\theta}^{[i-1]} - \boldsymbol{B}^{-1}\boldsymbol{s} \tag{4.23}$$

where

$$\boldsymbol{B}(\boldsymbol{\theta}_{i-1}) = \left[\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_{i-1}} = \left[\frac{\partial^2 U(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_{i-1}} \tag{4.24}$$

4.1.8 Asymptotic distribution of MLE

Approx.distribution that gets "better" as $n \to \infty$.

Theorem 4.1.1 *For*

$$\boldsymbol{s}(\boldsymbol{\theta}^*) = \left| \frac{\partial}{\partial \boldsymbol{\theta}} \log f(\boldsymbol{y}; \boldsymbol{\theta}) \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \tag{4.25}$$

and $U(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}; \boldsymbol{y})$, *can show that*

$$\begin{aligned} Var[\boldsymbol{s}(\boldsymbol{\theta}^*)] &= \left\{ -E_{\boldsymbol{\theta}^*} \left[\left| \frac{\partial^2 U(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \right] \right\} \\ &= \left\{ \int \frac{1}{f(\boldsymbol{y}; \boldsymbol{\theta}^*)} \left| \frac{\partial}{\partial \theta_j} f(\boldsymbol{y}; \boldsymbol{\theta}) \frac{\partial}{\partial \theta_k} f(\boldsymbol{y}; \boldsymbol{\theta}) \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} d\boldsymbol{y} \right\} \\ &= \left\{ E_{\boldsymbol{\theta}^*} \left[\left| \frac{\partial U(\boldsymbol{\theta})}{\partial \theta_j} \frac{\partial U(\boldsymbol{\theta})}{\partial \theta_k} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \right] \right\} \end{aligned} \tag{4.26}$$

Proof:

$$\begin{aligned} \left| \frac{\partial^2 U(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} &= \frac{\partial}{\partial \theta_k} \left[\frac{1}{f(\boldsymbol{y}; \boldsymbol{\theta}^*)} \left| \frac{\partial f(\boldsymbol{y}; \boldsymbol{\theta})}{\partial \theta_j} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \right] \\ &= \frac{1}{f(\boldsymbol{y}; \boldsymbol{\theta}^*)} \left| \frac{\partial^2 f(\boldsymbol{y}; \boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \\ &- \frac{1}{[f(\boldsymbol{y}; \boldsymbol{\theta}^*)]^2} \left| \frac{\partial f(\boldsymbol{y}; \boldsymbol{\theta})}{\partial \theta_k} \frac{\partial f(\boldsymbol{y}; \boldsymbol{\theta})}{\partial \theta_j} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \end{aligned} \tag{4.27}$$

Now,

$$
\begin{aligned}
\mathrm{E}_{\boldsymbol{\theta}^*}\left[\left|\frac{\partial^2 U(\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right] &= \int f(\boldsymbol{y};\boldsymbol{\theta}^*)\left|\frac{\partial^2 U(\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} d\boldsymbol{y} \\
&= \int \left|\frac{\partial^2 f(\boldsymbol{y};\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} d\boldsymbol{y} \\
&\quad - \int \frac{1}{f(\boldsymbol{y};\boldsymbol{\theta}^*)}\left|\frac{\partial f(\boldsymbol{y};\boldsymbol{\theta})}{\partial\theta_k}\frac{\partial f(\boldsymbol{y};\boldsymbol{\theta})}{\partial\theta_j}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} d\boldsymbol{y} \\
&= -\int \frac{1}{f(\boldsymbol{y};\boldsymbol{\theta}^*)}\left|\frac{\partial f(\boldsymbol{y};\boldsymbol{\theta})}{\partial\theta_k}\frac{\partial f(\boldsymbol{y};\boldsymbol{\theta})}{\partial\theta_j}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} d\boldsymbol{y}
\end{aligned}
\tag{4.28}
$$

because the first term in the last equation is

$$
\int \left|\frac{\partial^2 f(\boldsymbol{y};\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} d\boldsymbol{y} = \left|\frac{\partial^2}{\partial\theta_j\partial\theta_k}\underbrace{\int f(\boldsymbol{y};\boldsymbol{\theta})d\boldsymbol{y}}_{1}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} = 0 \tag{4.29}
$$

Similarly,

$$
\left|\frac{\partial U(\boldsymbol{\theta})}{\partial\theta_j}\frac{\partial U(\boldsymbol{\theta})}{\partial\theta_k}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} = \frac{1}{[f(\boldsymbol{y};\boldsymbol{\theta}^*)]^2}\left|\frac{\partial f(\boldsymbol{y};\boldsymbol{\theta})}{\partial\theta_k}\frac{\partial f(\boldsymbol{y};\boldsymbol{\theta})}{\partial\theta_j}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \tag{4.30}
$$

And as a result

$$
\begin{aligned}
\mathrm{E}_{\boldsymbol{\theta}^*}\left[\left|\frac{\partial U(\boldsymbol{\theta})}{\partial\theta_j}\frac{\partial U(\boldsymbol{\theta})}{\partial\theta_k}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right] &= \int \frac{1}{f(\boldsymbol{y};\boldsymbol{\theta}^*)}\left|\frac{\partial f(\boldsymbol{y};\boldsymbol{\theta})}{\partial\theta_k}\frac{\partial f(\boldsymbol{y};\boldsymbol{\theta})}{\partial\theta_j}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} d\boldsymbol{y} \\
&= -\mathrm{E}_{\boldsymbol{\theta}^*}\left[\left|\frac{\partial^2 U(\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right] \\
&= \mathrm{E}_{\boldsymbol{\theta}^*}\left[s_j(\boldsymbol{\theta}^*)s_k(\boldsymbol{\theta}^*)\right]
\end{aligned}
\tag{4.31}
$$

Recall that $\mathrm{E}\left[\boldsymbol{s}(\boldsymbol{\theta}^*)\right] = \mathbf{0}$. So,

$$
\mathrm{Var}[\boldsymbol{s}(\boldsymbol{\theta}^*)] = \mathrm{E}_{\boldsymbol{\theta}^*}\left[\boldsymbol{s}(\boldsymbol{\theta}^*)\boldsymbol{s}'(\boldsymbol{\theta}^*)\right] = -\mathrm{E}_{\boldsymbol{\theta}^*}\left[\boldsymbol{B}(\boldsymbol{\theta}^*)\right] = -\boldsymbol{B}^* \tag{4.32}
$$

where

$$
\boldsymbol{B}(\boldsymbol{\theta}^*) = \left|\frac{\partial^2 U(\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}
$$

Proposition 4.1.1 *For a data set consisting of iid vectors of observations,* $\boldsymbol{y}_1, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$

$$
\boldsymbol{B}(\boldsymbol{\theta}^*) = \sum \boldsymbol{B}_i(\boldsymbol{\theta}^*) = n\bar{\boldsymbol{B}}(\boldsymbol{\theta}^*) \tag{4.33}
$$

where

$$
\boldsymbol{B}_i(\boldsymbol{\theta}^*) = \left[\frac{\partial^2 \log f(\boldsymbol{y}_i;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \tag{4.34}
$$

and as $n \to \infty$, $\bar{\boldsymbol{B}}(\boldsymbol{\theta}^*) \to E[\boldsymbol{B}_i(\boldsymbol{\theta}^*)] = \boldsymbol{B}_i^*$.

This holds because "arithmetic mean $\bar{x}$ converges to $\mathrm{E}(x_i)$". So for large n

$$\boldsymbol{B}(\boldsymbol{\theta}^*) \approx n\mathrm{E}\left[\boldsymbol{B}_i(\boldsymbol{\theta}^*)\right] = n\boldsymbol{B}_i^* = \boldsymbol{B}^*$$

or in words for large n the observed information matrix is approximately equal to the expected information matrix.

Recall that MLE are consistent. So, for large n, $\widehat{\boldsymbol{\theta}}$ is close to $\boldsymbol{\theta}^*$ and

$$\widehat{\boldsymbol{\theta}} \simeq \boldsymbol{\theta}^* - (\boldsymbol{B}^*)^{-1}\boldsymbol{s}(\boldsymbol{\theta}^*) \tag{4.35}$$

The result in 4.35 can be obtained by the use of Taylor series approximation:

$$\begin{aligned}
&\boldsymbol{s}(\boldsymbol{\theta}) = \boldsymbol{s}(\boldsymbol{\theta}^*) + \underbrace{\frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{s}(\boldsymbol{\theta}^*)}_{\boldsymbol{B}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) = \mathbf{0} \\
&\boldsymbol{s}(\boldsymbol{\theta}^*) - \boldsymbol{B}\boldsymbol{\theta}^* = -\boldsymbol{B}\widehat{\boldsymbol{\theta}} \\
&\boldsymbol{B}^{-1} \mid \boldsymbol{B}\boldsymbol{\theta}^* - \boldsymbol{s}(\boldsymbol{\theta}^*) = \boldsymbol{B}\widehat{\boldsymbol{\theta}} \\
&\boldsymbol{\theta}^* - \boldsymbol{B}^{-1}\boldsymbol{s}(\boldsymbol{\theta}^*) = \widehat{\boldsymbol{\theta}}
\end{aligned} \tag{4.36}$$

Note also that the approximation in 4.35 gets better as n gets larger. Based on 4.35 it can be seen that,

$$\mathrm{E}(\widehat{\boldsymbol{\theta}}) \simeq \boldsymbol{\theta}^* - (\boldsymbol{B}^*)^{-1}\underbrace{\mathrm{E}[\boldsymbol{s}(\boldsymbol{\theta}^*)]}_{0} = \boldsymbol{\theta}^* \tag{4.37}$$

$\widehat{\boldsymbol{\theta}}$ is asymptotically unbiased. Also,

$$\begin{aligned}
\mathrm{Var}(\widehat{\boldsymbol{\theta}}) &\simeq \mathrm{Var}\left[\boldsymbol{\theta}^* - (\boldsymbol{B}^*)^{-1}\boldsymbol{s}(\boldsymbol{\theta}^*)\right] \\
&= (\boldsymbol{B}^*)^{-1}\mathrm{Var}[\boldsymbol{s}(\boldsymbol{\theta}^*)](\boldsymbol{B}^*)^{-1} \\
&= (\boldsymbol{B}^*)^{-1}(-\boldsymbol{B}^*)(\boldsymbol{B}^*)^{-1} \\
&= -(\boldsymbol{B}^*)^{-1} \\
&= \left[\mathrm{Var}[\boldsymbol{s}(\boldsymbol{\theta}^*)]\right]^{-1}
\end{aligned} \tag{4.38}$$

so we conlude that $\mathrm{Var}(\widehat{\boldsymbol{\theta}}) = \left[\mathrm{Var}[\boldsymbol{s}(\boldsymbol{\theta}^*)]\right]^{-1}$ which is the inverse of the Cramer - Rao lower bound for unbiased estimators. Further, if $\boldsymbol{y}_i$ iid $\boldsymbol{s}(\boldsymbol{\theta}^*)$ is the sum of iid $\boldsymbol{s}_i(\boldsymbol{\theta}^*)$. So for large n, $\boldsymbol{s}(\boldsymbol{\theta}^*) \sim \boldsymbol{\aleph}(\mathbf{0}, -\boldsymbol{B}^*)$ and because of 4.35 $\widehat{\boldsymbol{\theta}} \sim \boldsymbol{\aleph}(\boldsymbol{\theta}^*, -(\boldsymbol{B}^*)^{-1})$.

4.1.9 Asymptotic distribution of the likelihood ratio

$$\lambda = \frac{\max_{\omega} L(\boldsymbol{\theta}; \boldsymbol{y})}{\max_{\Omega} L(\boldsymbol{\theta}; \boldsymbol{y})} \qquad H_0 : \boldsymbol{\theta} \in \boldsymbol{\omega} \tag{4.39}$$

where,

Ω is a p dimensional space

ω is a $p-q$ dimensional supspace of Ω

Lemma 4.1.6 *Can show that as $n \to \infty$ and if H_0 is true,*

$$-2\log\lambda \sim \boldsymbol{\chi}_q^2 \tag{4.40}$$

Proof: Consider the following notation,

$$\max_{\theta\in\Omega}\log L(\boldsymbol{\theta};\boldsymbol{y}) = u(\widehat{\boldsymbol{\theta}}_\Omega)$$

and after using a second order Taylor series approximation can write

$$u(\widehat{\boldsymbol{\theta}}_\Omega) \approx u(\boldsymbol{\theta}^*) + (\widehat{\boldsymbol{\theta}}_\Omega - \boldsymbol{\theta}^*)'\boldsymbol{s}(\boldsymbol{\theta}^*) + \frac{1}{2}(\widehat{\boldsymbol{\theta}}_\Omega - \boldsymbol{\theta}^*)'\boldsymbol{B}^*(\widehat{\boldsymbol{\theta}}_\Omega - \boldsymbol{\theta}^*) \tag{4.41}$$

Recall the fact that for large n,

$$(\widehat{\boldsymbol{\theta}}_\Omega - \boldsymbol{\theta}^*) \simeq -(\boldsymbol{B}^*)^{-1}\boldsymbol{s}(\boldsymbol{\theta}^*) \tag{4.42}$$

Now using 4.42 in 4.41 can write

$$\begin{aligned} u(\widehat{\boldsymbol{\theta}}_\Omega) &\simeq u(\boldsymbol{\theta}^*) - \boldsymbol{s}'(\boldsymbol{\theta}^*)(\boldsymbol{B}^*)^{-1}\boldsymbol{s}(\boldsymbol{\theta}^*) + \frac{1}{2}\boldsymbol{s}'(\boldsymbol{\theta}^*)\underbrace{(\boldsymbol{B}^*)^{-1}\boldsymbol{B}^*}_{\boldsymbol{I}}(\boldsymbol{B}^*)^{-1}\boldsymbol{s}(\boldsymbol{\theta}^*) \\ &= u(\boldsymbol{\theta}^*) - \frac{1}{2}\boldsymbol{s}'(\boldsymbol{\theta}^*)(\boldsymbol{B}^*)^{-1}\boldsymbol{s}(\boldsymbol{\theta}^*) \end{aligned} \tag{4.43}$$

Note that under ω there are $p-q$ free parameters and q fixed parameters and consequently can write

$$(\widehat{\boldsymbol{\theta}}_\omega - \boldsymbol{\theta}^*) = \begin{bmatrix} \widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^* \\ \vdots \\ \widehat{\boldsymbol{\theta}}_{p-q} - \boldsymbol{\theta}_{p-q}^* \\ \boldsymbol{\theta}_{p-q+1}^* - \boldsymbol{\theta}_{p-q+1}^* \\ \vdots \\ \boldsymbol{\theta}_p^* - \boldsymbol{\theta}_p^* \end{bmatrix} = \begin{bmatrix} \boldsymbol{d}_1 \\ \boldsymbol{0} \end{bmatrix} \tag{4.44}$$

Then,

$$\max_{\theta\in\omega}\log L(\boldsymbol{\theta};\boldsymbol{y}) = u(\widehat{\boldsymbol{\theta}}_\omega) = u(\boldsymbol{\theta}^*) + \boldsymbol{d}_1'\boldsymbol{s}_1(\boldsymbol{\theta}^*) + \frac{1}{2}\boldsymbol{d}_1'\boldsymbol{B}_{11}^*\boldsymbol{d}_1 \tag{4.45}$$

but we know that $\boldsymbol{d}_1 \simeq -(\boldsymbol{B}_{11}^*)^{-1}\boldsymbol{s}_1(\boldsymbol{\theta}^*)$ so under ω,

$$\begin{aligned} u(\widehat{\boldsymbol{\theta}}_\omega) &\simeq u(\boldsymbol{\theta}^*) - \boldsymbol{s}_1'(\boldsymbol{\theta}^*)(\boldsymbol{B}_{11}^*)^{-1}\boldsymbol{s}_1(\boldsymbol{\theta}^*) + \frac{1}{2}\boldsymbol{s}_1'(\boldsymbol{\theta}^*)\underbrace{(\boldsymbol{B}_{11}^*)^{-1}\boldsymbol{B}_{11}^*}_{\boldsymbol{I}}(\boldsymbol{B}_{11}^*)^{-1}\boldsymbol{s}_1(\boldsymbol{\theta}^*) \\ &= u(\boldsymbol{\theta}^*) - \frac{1}{2}\boldsymbol{s}_1'(\boldsymbol{\theta}^*)(\boldsymbol{B}_{11}^*)^{-1}\boldsymbol{s}_1(\boldsymbol{\theta}^*) \end{aligned} \tag{4.46}$$

Thus,

$$-2\log\lambda = \boldsymbol{s}'(\boldsymbol{\theta}^*)(\boldsymbol{B}^*)^{-1}\boldsymbol{s}(\boldsymbol{\theta}^*) - \boldsymbol{s}_1'(\boldsymbol{\theta}^*)(\boldsymbol{B}_{11}^*)^{-1}\boldsymbol{s}_1(\boldsymbol{\theta}^*) \tag{4.47}$$

But for large n,

$$\boldsymbol{s}^{p\times 1}(\boldsymbol{\theta}^*) \sim \aleph(\mathbf{0}, \boldsymbol{B}^*) \quad \text{and} \quad \boldsymbol{s}_1^{(p-q)\times 1}(\boldsymbol{\theta}^*) \sim \aleph(\mathbf{0}, \boldsymbol{B}_{11}^*) \tag{4.48}$$

and consequently,

$$\begin{aligned} -2\log\lambda &= \boldsymbol{s}'(\boldsymbol{\theta}^*)(\boldsymbol{B}^*)^{-1}\boldsymbol{s}(\boldsymbol{\theta}^*) - \boldsymbol{s}_1'(\boldsymbol{\theta}^*)(\boldsymbol{B}_{11}^*)^{-1}\boldsymbol{s}_1(\boldsymbol{\theta}^*) \\ &= \boldsymbol{\chi}_q^2 \end{aligned} \tag{4.49}$$

The last result is obtained using the result 6.5.2 from appendix.

4.2 MLE of variance components

Consider the mixed model (animal model)

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^* + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e} \tag{4.50}$$

where

- $\boldsymbol{u}$ animal effect
- $\boldsymbol{\beta}^*$ unknown true value

Assume also that

$$\boldsymbol{y} \sim \aleph(\boldsymbol{X}\boldsymbol{\beta}^*, \boldsymbol{V}^*) \tag{4.51}$$

where $\boldsymbol{V}^* = \boldsymbol{Z}\boldsymbol{A}\boldsymbol{Z}'\sigma_a^{2*} + \boldsymbol{I}\sigma_e^{2*}$. Then we can write the likelihood as

$$L(\boldsymbol{\beta}, \sigma_a^2, \sigma_e^2; \boldsymbol{y}) = \frac{\exp\left\{\frac{1}{2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})\right\}}{(2\pi)^{n/2} \mid \boldsymbol{V} \mid^{1/2}} \tag{4.52}$$

and the loglikelihood as

$$\begin{aligned} U = \log L = K - \frac{1}{2}\log \mid \boldsymbol{V} \mid + \frac{1}{2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}) \\ \approx -\frac{1}{2}\log \mid \boldsymbol{V} \mid + \frac{1}{2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}) \end{aligned} \tag{4.53}$$

In order to be able to use the Newton-Raphson algorithm we need the first and second derivatives of the loglikelihood. These expressions can be computed using the results regarding derivatives of matrices and determinants described in Appendix. Consequently,

$$\begin{aligned} \frac{\partial U}{\partial \boldsymbol{\beta}} &= \frac{1}{2}\frac{\partial}{\partial \boldsymbol{\beta}}\left(2\boldsymbol{y}'\boldsymbol{V}^{-1}\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\boldsymbol{\beta}\right) \\ &= \frac{1}{2}\left[2\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y} - 2\left(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right)\boldsymbol{\beta}\right] \end{aligned} \tag{4.54}$$

By setting the last equation to zero and solving for $\boldsymbol{\beta}$ we obtain

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-} \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y} \tag{4.55}$$

Similarly the second order derivative can be obtained as

$$\frac{\partial^2 U}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} = -\left(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right) \tag{4.56}$$

Now consider

$$\begin{aligned}
\frac{\partial U}{\partial\sigma_a^2} &= -\frac{1}{2}\frac{\partial \log \mid \boldsymbol{V} \mid}{\partial\sigma_a^2} - \frac{1}{2}\frac{\partial}{\partial\sigma_a^2}\left[(\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta})\right] \\
&= -\frac{1}{2}\mathrm{tr}\left[\boldsymbol{V}^{-1}\left(\frac{\partial\boldsymbol{V}}{\partial\sigma_a^2}\right)\right] + \frac{1}{2}(\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{V}^{-1}\left(\frac{\partial\boldsymbol{V}}{\partial\sigma_a^2}\right)\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta}) \\
&= -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_a\right) + \frac{1}{2}(\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{V}^{-1}\boldsymbol{V}_a\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta})
\end{aligned} \tag{4.57}$$

where

$$\boldsymbol{V}_a = \frac{\partial\boldsymbol{V}}{\partial\sigma_a^2} = \frac{\partial}{\partial\sigma_a^2}\left(\boldsymbol{ZAZ}'\sigma_a^2 + \boldsymbol{I}\sigma_e^2\right) = \boldsymbol{ZAZ}' \tag{4.58}$$

Also

$$\begin{aligned}
\frac{\partial^2 U}{\partial\beta\partial\sigma_a^2} &= \frac{\partial}{\partial\sigma_a^2}\left[\boldsymbol{X}'\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta})\right] \\
&= \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{V}_a\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta})
\end{aligned} \tag{4.59}$$

and

$$\begin{aligned}
\frac{\partial^2 U}{\partial\sigma_a^2\partial\sigma_a^2} &= -\frac{1}{2}\mathrm{tr}\left[\frac{\partial(\boldsymbol{V}^{-1}\boldsymbol{V}_a)}{\partial\sigma_a^2}\right] + \frac{1}{2}\frac{\partial}{\partial\sigma_a^2}\left[(\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{V}^{-1}\boldsymbol{V}_a\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta})\right] \\
&= \frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_a\boldsymbol{V}^{-1}\boldsymbol{V}_a\right) - \frac{1}{2}(\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{V}^{-1}\boldsymbol{V}_a\boldsymbol{V}^{-1}\boldsymbol{V}_a\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta}) \\
&- \frac{1}{2}(\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{V}^{-1}\boldsymbol{V}_a\boldsymbol{V}^{-1}\boldsymbol{V}_a\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta}) \\
&= \frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_a\boldsymbol{V}^{-1}\boldsymbol{V}_a\right) - (\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{V}^{-1}\boldsymbol{V}_a\boldsymbol{V}^{-1}\boldsymbol{V}_a\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta})
\end{aligned} \tag{4.60}$$

Consider now the derivatives of the loglikelihood with respect to σ_e^2

$$\begin{aligned}
\frac{\partial U}{\partial\sigma_e^2} &= -\frac{1}{2}\frac{\partial \log \mid \boldsymbol{V} \mid}{\partial\sigma_e^2} - \frac{1}{2}\frac{\partial}{\partial\sigma_e^2}\left[(\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta})\right] \\
&= -\frac{1}{2}\mathrm{tr}\left[\boldsymbol{V}^{-1}\left(\frac{\partial\boldsymbol{V}}{\partial\sigma_e^2}\right)\right] + \frac{1}{2}(\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{V}^{-1}\left(\frac{\partial\boldsymbol{V}}{\partial\sigma_e^2}\right)\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta}) \\
&= -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_e\right) + \frac{1}{2}(\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{V}^{-1}\boldsymbol{V}_e\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta})
\end{aligned} \tag{4.61}$$

where

$$\boldsymbol{V}_e = \frac{\partial \boldsymbol{V}}{\partial \sigma_e^2} = \frac{\partial}{\partial \sigma_a^2}\left(\boldsymbol{ZAZ}'\sigma_a^2 + \boldsymbol{I}\sigma_e^2\right) = \boldsymbol{I} \tag{4.62}$$

So,

$$\frac{\partial U}{\partial \sigma_e^2} = -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}^{-1}\right) + \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{V}^{-1}\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta}) \tag{4.63}$$

Also,

$$\begin{aligned} \frac{\partial^2 U}{\partial \beta \partial \sigma_e^2} &= \frac{\partial}{\partial \sigma_e^2}\left[\boldsymbol{X}'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta})\right] \\ &= \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta}) \end{aligned} \tag{4.64}$$

$$\begin{aligned} \frac{\partial^2 U}{\partial \sigma_a^2 \partial \sigma_e^2} &= -\frac{1}{2}\mathrm{tr}\left[\frac{\partial(\boldsymbol{V}^{-1}\boldsymbol{V}_a)}{\partial \sigma_e^2}\right] + \frac{1}{2}\frac{\partial}{\partial \sigma_e^2}\left[(\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{V}^{-1}\boldsymbol{V}_a\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta})\right] \\ &= \frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}^{-1}\boldsymbol{V}_a\right) - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{V}^{-1}\boldsymbol{V}^{-1}\boldsymbol{V}_a\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta}) \\ &- \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{V}^{-1}\boldsymbol{V}^{-1}\boldsymbol{V}_a\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta}) \\ &= \frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}^{-1}\boldsymbol{V}_a\right) - (\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{V}^{-1}\boldsymbol{V}^{-1}\boldsymbol{V}_a\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta}) \end{aligned} \tag{4.65}$$

and

$$\begin{aligned} \frac{\partial^2 U}{\partial \sigma_e^2 \partial \sigma_e^2} &= \frac{\partial}{\partial \sigma_e^2}\left[-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}^{-1}\right) + \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{V}^{-1}\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta})\right] \\ &= \frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}^{-1}\right) - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{V}^{-1}\boldsymbol{V}^{-1}\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta}) \\ &- \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{V}^{-1}\boldsymbol{V}^{-1}\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta}) \\ &= \frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}^{-1}\right) - (\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{V}^{-1}\boldsymbol{V}^{-1}\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta}) \end{aligned} \tag{4.66}$$

Now we can apply the Newton-Raphson algorithm

$$\widehat{\boldsymbol{\theta}}^{[i]} = \widehat{\boldsymbol{\theta}}^{[i-1]} - \boldsymbol{Bs} \tag{4.67}$$

where

$$
\begin{aligned}
\boldsymbol{B} &= \left[\frac{\partial U(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}^{[i-1]}} \\
&= \begin{bmatrix}
-\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{V}_a\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta}) & \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta}) \\
 & \begin{array}{c}\frac{1}{2}\text{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_a\boldsymbol{V}^{-1}\right) - \\ (\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{V}^{-1}\boldsymbol{V}_a\boldsymbol{V}^{-1}\boldsymbol{V}_a\boldsymbol{V}^{-1} \\ (\boldsymbol{y}-\boldsymbol{X\beta})\end{array} & \begin{array}{c}\frac{1}{2}\text{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}^{-1}\boldsymbol{V}_a\right) - \\ (\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{V}^{-1}\boldsymbol{V}^{-1}\boldsymbol{V}_a\boldsymbol{V}^{-1} \\ (\boldsymbol{y}-\boldsymbol{X\beta})\end{array} \\
 & & \begin{array}{c}\frac{1}{2}\text{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}^{-1}\right) - \\ (\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{V}^{-1}\boldsymbol{V}^{-1}\boldsymbol{V}^{-1} \\ (\boldsymbol{y}-\boldsymbol{X\beta})\end{array}
\end{bmatrix}
\end{aligned} \tag{4.68}
$$

and

$$
\boldsymbol{s} = \begin{bmatrix}
\boldsymbol{X}'\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta}) \\
-\frac{1}{2}\text{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_a\right) + \frac{1}{2}(\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{V}^{-1}\boldsymbol{V}_a\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta}) \\
-\frac{1}{2}\text{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_e\right) + \frac{1}{2}(\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{V}^{-1}\boldsymbol{V}_e\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta})
\end{bmatrix} \tag{4.69}
$$

note that here $\boldsymbol{V}_e = \boldsymbol{I}$.

Consider a more general model where $\boldsymbol{\beta}^{p\times 1}$ and $\boldsymbol{V}$ is a function of k variance and covariance components. Then the first derivatives are

$$
\begin{aligned}
\frac{\partial U}{\partial\boldsymbol{\beta}} &= \boldsymbol{X}'\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta}) \\
&\vdots \\
\frac{\partial U}{\partial\theta_{p+1}} &= -\frac{1}{2}\text{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_1\right) + \frac{1}{2}(\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{V}^{-1}\boldsymbol{V}_1\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta}) \\
&\vdots \\
\frac{\partial U}{\partial\theta_{p+i}} &= -\frac{1}{2}\text{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_i\right) + \frac{1}{2}(\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{V}^{-1}\boldsymbol{V}_i\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta}) \\
&\vdots \\
\frac{\partial U}{\partial\theta_{p+k}} &= -\frac{1}{2}\text{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_k\right) + \frac{1}{2}(\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{V}^{-1}\boldsymbol{V}_k\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta})
\end{aligned} \tag{4.70}
$$

where $\boldsymbol{V}_i = \frac{\partial \boldsymbol{V}}{\partial \theta_{p+i}}$ and the second derivatives are

$$\frac{\partial^2 U}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = -\left(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right)$$

$$\begin{aligned}\frac{\partial^2 U}{\partial \boldsymbol{\beta} \partial \theta_{p+i}} &= -\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{V}_i\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta}) \\ \frac{\partial^2 U}{\partial \theta_{p+i} \partial \theta_{p+j}} &= \frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_i\boldsymbol{V}^{-1}\boldsymbol{V}_j\right) - (\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{V}^{-1}\boldsymbol{V}_i\boldsymbol{V}^{-1}\boldsymbol{V}_j\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta})\end{aligned} \tag{4.71}$$

Consider a two trait problem where

$$\begin{aligned}\boldsymbol{y}_1 &= \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{Z}_1\boldsymbol{u}_1 + \boldsymbol{e}_1 \\ \boldsymbol{y}_2 &= \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{Z}_2\boldsymbol{u}_2 + \boldsymbol{e}_2\end{aligned}$$

The previous equations can be written as

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{e} \tag{4.72}$$

where

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{X}_2 \end{bmatrix} \quad , \quad \boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Z}_2 \end{bmatrix} \tag{4.73}$$

and

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} \quad , \quad \boldsymbol{u} = \begin{bmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \end{bmatrix} \quad , \quad \boldsymbol{e} = \begin{bmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \end{bmatrix} \tag{4.74}$$

with $\boldsymbol{y} \sim \aleph\left(\boldsymbol{X\beta}, \boldsymbol{V}\right)$, with $\boldsymbol{V} = \boldsymbol{Z}\mathrm{Var}(\boldsymbol{u})\boldsymbol{Z}' + \mathrm{Var}(\boldsymbol{e})$. Also

$$\begin{aligned}\mathrm{Var}(\boldsymbol{u}) &= \begin{bmatrix} \boldsymbol{A}\sigma^2_{a_1} & \boldsymbol{A}\sigma_{a_{12}} \\ \boldsymbol{A}\sigma_{a_{12}} & \boldsymbol{A}\sigma^2_{a_2} \end{bmatrix} \\ \mathrm{Var}(\boldsymbol{e}) &= \begin{bmatrix} \boldsymbol{I}\sigma^2_{e_1} & \boldsymbol{R}^*_{12}\sigma_{e_{12}} \\ \boldsymbol{R}^*_{21}\sigma_{e_{12}} & \boldsymbol{I}\sigma^2_{e_2} \end{bmatrix}\end{aligned} \tag{4.75}$$

Consequently,

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{Z}_1\boldsymbol{A}\boldsymbol{Z}'_1\sigma^2_{a_1} & \boldsymbol{Z}_1\boldsymbol{A}\boldsymbol{Z}'_2\sigma_{a_{12}} \\ \boldsymbol{Z}_2\boldsymbol{A}\boldsymbol{Z}'_1\sigma_{a_{12}} & \boldsymbol{Z}_2\boldsymbol{A}\boldsymbol{Z}'_2\sigma^2_{a_2} \end{bmatrix} + \begin{bmatrix} \boldsymbol{I}\sigma^2_{e_1} & \boldsymbol{R}^*_{12}\sigma_{e_{12}} \\ \boldsymbol{R}^*_{21}\sigma_{e_{12}} & \boldsymbol{I}\sigma^2_{e_2} \end{bmatrix} \tag{4.76}$$

and the first derivatives of $\boldsymbol{V}$ with respect to the variance components are

$$\begin{aligned}\frac{\partial \boldsymbol{V}}{\partial \sigma^2_{a_1}} &= \begin{bmatrix} \boldsymbol{Z}_1\boldsymbol{A}\boldsymbol{Z}'_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \\ &\vdots \\ \frac{\partial \boldsymbol{V}}{\partial \sigma_{e_{12}}} &= \begin{bmatrix} \boldsymbol{0} & \boldsymbol{R}^*_{12} \\ \boldsymbol{R}^*_{21} & \boldsymbol{0} \end{bmatrix}\end{aligned} \tag{4.77}$$

Now in order to describe $\boldsymbol{R}^*_{12}$ (note that $(\boldsymbol{R}^*_{12})' = \boldsymbol{R}^*_{21}$) consider the following data

	Trait1	Trait2
1	x	x
2	x	x
3	x	$-$
4	$-$	x
5	x	$-$

where x means that the observation is present and $-$ that the observation is missing. For this data

$$\text{Var}(\boldsymbol{e}) = \begin{bmatrix} \sigma^2_{e_1} & 0 & 0 & 0 & \sigma_{e_{12}} & 0 & 0 \\ 0 & \sigma^2_{e_1} & 0 & 0 & 0 & \sigma_{e_{12}} & 0 \\ 0 & 0 & \sigma^2_{e_1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_{e_1} & 0 & 0 & 0 \\ \sigma_{e_{12}} & 0 & 0 & 0 & \sigma^2_{e_2} & 0 & 0 \\ 0 & \sigma_{e_{12}} & 0 & 0 & 0 & \sigma^2_{e_2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{e_2} \end{bmatrix} \tag{4.78}$$

So

$$\boldsymbol{R}^*_{12} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{4.79}$$

Asymptotic Variance Matrix

$$\text{Var}\left(\widehat{\boldsymbol{\theta}}\right) \approx -\left(\boldsymbol{B}^*\right)^{-1} = -\text{E}\left(\boldsymbol{B}(\boldsymbol{\theta^*})\right)^{-1} \tag{4.80}$$

The expected values of second derivatives (expectations taken with respect to $\boldsymbol{\theta}$ not $\boldsymbol{\theta}^*$) are

$$\begin{aligned} -\text{E}\left[\frac{\partial U}{\partial\beta\partial\beta'}\right] &= -\text{E}\left[\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right] = \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X} \\ -\text{E}\left[\frac{\partial U}{\partial\beta\partial\theta_{p+i}}\right] &= -\text{E}\left[\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{V}_i\boldsymbol{V}^{-1}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right)\right] = \boldsymbol{0} \end{aligned} \tag{4.81}$$

and

$$
\begin{aligned}
&- \mathrm{E}\left[\frac{\partial U}{\partial \theta_{p+i} \partial \theta_{p+j}}\right] \\
&= -\mathrm{E}\left[\frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_i\boldsymbol{V}^{-1}\boldsymbol{V}_j\right) - (\boldsymbol{y}-\boldsymbol{X\beta})'\,\boldsymbol{V}^{-1}\boldsymbol{V}_i\boldsymbol{V}^{-1}\boldsymbol{V}_j\boldsymbol{V}^{-1}\,(\boldsymbol{y}-\boldsymbol{X\beta})\right] \\
&= -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_i\boldsymbol{V}^{-1}\boldsymbol{V}_j\right) + \mathrm{E}\left[\mathrm{tr}\boldsymbol{V}^{-1}\boldsymbol{V}_i\boldsymbol{V}^{-1}\boldsymbol{V}_j\boldsymbol{V}^{-1}\,(\boldsymbol{y}-\boldsymbol{X\beta})\,(\boldsymbol{y}-\boldsymbol{X\beta})'\right] \\
&= -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_i\boldsymbol{V}^{-1}\boldsymbol{V}_j\right) + \mathrm{tr}\boldsymbol{V}^{-1}\boldsymbol{V}_i\boldsymbol{V}^{-1}\boldsymbol{V}_j\boldsymbol{V}^{-1}\,\underbrace{\mathrm{E}\left[(\boldsymbol{y}-\boldsymbol{X\beta})\,(\boldsymbol{y}-\boldsymbol{X\beta})'\right]}_{\boldsymbol{V}} \\
&= -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_i\boldsymbol{V}^{-1}\boldsymbol{V}_j\right) + \mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_i\boldsymbol{V}^{-1}\boldsymbol{V}_j\boldsymbol{V}^{-1}\right) \\
&= \frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_i\boldsymbol{V}^{-1}\boldsymbol{V}_j\right)
\end{aligned}
\tag{4.82}
$$

Then the asymptotic variance matrix becomes

$$
\mathrm{Var}\left(\widehat{\boldsymbol{\theta}}\right) \approx \begin{bmatrix} \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X} & \mathbf{0} \\ \mathbf{0} & \left\{\frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_i\boldsymbol{V}^{-1}\boldsymbol{V}_j\right)\right\} \end{bmatrix}^{-1} = -\left(\boldsymbol{B}^*\right)^{-1} \tag{4.83}
$$

Fisher method of scoring

$$
\widehat{\boldsymbol{\theta}}_i = \widehat{\boldsymbol{\theta}_{i-1}} - \boldsymbol{B}^*(\widehat{\boldsymbol{\theta}}_i)\boldsymbol{s}(\widehat{\boldsymbol{\theta}}_i) \tag{4.84}
$$

Note the difference between Newton-Raphson where

$$
\frac{\partial^2 U}{\partial \theta_i \partial \theta_j} = -\frac{1}{2}\mathrm{tr}\boldsymbol{V}^{-1}\boldsymbol{V}_i\boldsymbol{V}^{-1}\boldsymbol{V}_j + (\boldsymbol{y}-\boldsymbol{X\beta})'\,\boldsymbol{V}^{-1}\boldsymbol{V}_i\boldsymbol{V}^{-1}\boldsymbol{V}_j\boldsymbol{V}^{-1}\,(\boldsymbol{y}-\boldsymbol{X\beta}) \tag{4.85}
$$

is used and Fisher method of scoring where

$$
\mathrm{E}\left[\frac{\partial^2 U}{\partial \theta_i \partial \theta_j}\right] = \frac{1}{2}\mathrm{tr}\boldsymbol{V}^{-1}\boldsymbol{V}_i\boldsymbol{V}^{-1}\boldsymbol{V}_j \tag{4.86}
$$

is used.

Functional iteration

Consider the function $g(x)$, want x such that $g(x) = 0$. Then rewrite $g(x)$ as $h(x) - x = 0$. Then the functional iteration process is given by

$$
x^{[i]} = h\left(x^{[i-1]}\right)
$$

We look now at how can functional iteration be used in the case of the general problem disscused previously. Remember that the first derivatives of the loglikelihood were

$$
\begin{aligned}
\frac{\partial U}{\partial \boldsymbol{\beta}} &= \boldsymbol{X}'\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta}) \\
&\vdots \\
\frac{\partial U}{\partial \theta_{p+1}} &= -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_1\right) + \frac{1}{2}(\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{V}^{-1}\boldsymbol{V}_1\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta}) \\
&\vdots \\
\frac{\partial U}{\partial \theta_{p+i}} &= -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_i\right) + \frac{1}{2}(\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{V}^{-1}\boldsymbol{V}_i\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta}) \\
&\vdots \\
\frac{\partial U}{\partial \theta_{p+k}} &= -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_k\right) + \frac{1}{2}(\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{V}^{-1}\boldsymbol{V}_k\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta})
\end{aligned}
\tag{4.87}
$$

Now using the principle of functional iteration we can write

$$
\begin{aligned}
&\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y} - \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\widehat{\boldsymbol{\beta}}) = \boldsymbol{0} \\
&-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_1\right) + \frac{1}{2}(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})'\boldsymbol{V}^{-1}\boldsymbol{V}_1\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}) = \boldsymbol{0} \\
&\vdots \\
&-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_k\right) + \frac{1}{2}(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})'\boldsymbol{V}^{-1}\boldsymbol{V}_k\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})
\end{aligned}
\tag{4.88}
$$

Note that

$$
\mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_i\right) = \mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_i\boldsymbol{V}^{-1}\boldsymbol{V}\right) = \mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_i\boldsymbol{V}^{-1}\underbrace{(\sum_j \boldsymbol{V}_j\theta_j)}_{\boldsymbol{V}}\right) \tag{4.89}
$$

because any variance covariance matrix function of several parameters can be written as $\sum_{i=1}^{k}\boldsymbol{V}_i\theta_i$. As a result

$$
\mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_i\right) = \mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_i\boldsymbol{V}^{-1}\underbrace{(\sum_j \boldsymbol{V}_j\theta_j)}_{\boldsymbol{V}}\right) = \sum_j \mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_i\boldsymbol{V}^{-1}\boldsymbol{V}_j\right)\theta_j \tag{4.90}
$$

and now

$$\begin{aligned}\frac{\partial U}{\partial\theta_{p+i}} &= -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_i\right)+\frac{1}{2}(\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{V}^{-1}\boldsymbol{V}_i\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta}) \\ -\frac{1}{2}\sum_j \mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_i\boldsymbol{V}^{-1}\boldsymbol{V}_j\right)\theta_j &+ \frac{1}{2}(\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{V}^{-1}\boldsymbol{V}_i\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta})\end{aligned} \tag{4.91}$$

Set 4.91 to zero and then use for functional iteration

$$\left\{\mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_i\boldsymbol{V}^{-1}\boldsymbol{V}_j\right)\right\}\begin{bmatrix}\theta_{p+1}\\ \vdots \\ \theta_{p+k} = \left(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)'\boldsymbol{V}^{-1}\boldsymbol{V}_i\boldsymbol{V}^{-1}\left(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)\end{bmatrix} \tag{4.92}$$

Note that $\boldsymbol{V}$ in the matrix of traces containes $\theta's$ from round $n-1$ while $\theta_{p+1},\ldots,\theta_{p+k}$ are from round n. Consider the case of three variance components for ilustration

$$\begin{bmatrix}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ & \mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_1\boldsymbol{V}^{-1}\boldsymbol{V}_1\right) & \mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_1\boldsymbol{V}^{-1}\boldsymbol{V}_2\right) & \mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_1\boldsymbol{V}^{-1}\boldsymbol{V}_3\right)\\ & & \mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_2\boldsymbol{V}^{-1}\boldsymbol{V}_2\right) & \mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_2\boldsymbol{V}^{-1}\boldsymbol{V}_3\right)\\ & & & \mathrm{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{V}_3\boldsymbol{V}^{-1}\boldsymbol{V}_3\right)\end{bmatrix}$$
$$\times\begin{bmatrix}\boldsymbol{\beta}\\ \theta_{p+1}\\ \theta_{p+2}\\ \theta_{p+3}\end{bmatrix} = \begin{bmatrix}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}\\ \left(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)'\boldsymbol{V}^{-1}\boldsymbol{V}_1\boldsymbol{V}^{-1}\left(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)\\ \left(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)'\boldsymbol{V}^{-1}\boldsymbol{V}_2\boldsymbol{V}^{-1}\left(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)\\ \left(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)'\boldsymbol{V}^{-1}\boldsymbol{V}_3\boldsymbol{V}^{-1}\left(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)\end{bmatrix} \tag{4.93}$$

Values for the desired parameters can be obtained through an iterative process from the previous equation.

4.2.1 EM algorithm

Suppose that $L(\boldsymbol{\theta};\boldsymbol{y})$ is hard to compute, but $L(\boldsymbol{\theta};\boldsymbol{y},\boldsymbol{m})$ is easy to compute where $\boldsymbol{m}$ is an additional variable. Further suppose

$$Q(\boldsymbol{\theta};\boldsymbol{\theta}^{[i-1]}) = \mathrm{E}\left[U(\boldsymbol{\theta};\boldsymbol{y},\boldsymbol{m}) \mid \boldsymbol{y};\boldsymbol{\theta}^{[i-1]}\right] \tag{4.94}$$

is easy to compute. Note that in 4.94 $\boldsymbol{\theta}$ is the argument of this likelihood and $\boldsymbol{\theta}^{[i-1]}$ is the value of the parameter used in computing the expected value. Then, can maximize $L(\boldsymbol{\theta};\boldsymbol{y})$ with respect to $\boldsymbol{\theta}$ by the EM algorithm as follows

1. E step: Compute $Q(\boldsymbol{\theta};\boldsymbol{\theta}^{[i-1]})$

2. M step: Maximize $Q(\boldsymbol{\theta};\boldsymbol{\theta}^{[i-1]})$,

$$\mathrm{Max}_{\theta\in\omega}Q(\boldsymbol{\theta};\boldsymbol{\theta}^{[i-1]}) = Q(\boldsymbol{\theta}^{i};\boldsymbol{\theta}^{[i-1]})$$

Can show that every step of this algorithm will result in a higher value for $L(\boldsymbol{\theta}; \boldsymbol{y})$. Note that

$\boldsymbol{y}$ - observed data or incomplete data;

$\boldsymbol{m}$ - missing data;

form the complete data.

Proposition 4.2.1 *Can show that*

$$L\left(\boldsymbol{\theta}^{i}; \boldsymbol{y}\right) \geq L\left(\boldsymbol{\theta}^{[i-1]}; \boldsymbol{y}\right) \tag{4.95}$$

Proof: Consider the density of complete data

$$f(\boldsymbol{y}, \boldsymbol{m}; \boldsymbol{\theta}) = f(\boldsymbol{y}; \boldsymbol{\theta}) f(\boldsymbol{m} \mid \boldsymbol{y}; \boldsymbol{\theta}) \tag{4.96}$$

Take the logarithm of the previous expression

$$\log f(\boldsymbol{y}, \boldsymbol{m}; \boldsymbol{\theta}) = \log f(\boldsymbol{y}; \boldsymbol{\theta}) + \log f(\boldsymbol{m} \mid \boldsymbol{y}; \boldsymbol{\theta}) \tag{4.97}$$

Let

$$Q\left(\boldsymbol{\theta}; \boldsymbol{\theta}^{[i-1]}\right) = \mathrm{E}\left[\log f(\boldsymbol{y}, \boldsymbol{m}; \boldsymbol{\theta}) \mid \boldsymbol{y}; \boldsymbol{\theta}^{[i-1]}\right] \tag{4.98}$$

$$U\left(\boldsymbol{\theta}; \boldsymbol{y}\right) = \log f(\boldsymbol{y}; \boldsymbol{\theta}) \tag{4.99}$$

$$H\left(\boldsymbol{\theta}; \boldsymbol{\theta}^{[i-1]}\right) = \mathrm{E}\left[\log f(\boldsymbol{m} \mid \boldsymbol{y}; \boldsymbol{\theta}) \mid \boldsymbol{y}; \boldsymbol{\theta}^{[i-1]}\right] \tag{4.100}$$

Now using 4.97, 4.99, 4.100 in 4.98 can write

$$Q\left(\boldsymbol{\theta}; \boldsymbol{\theta}^{[i-1]}\right) = U\left(\boldsymbol{\theta}; \boldsymbol{y}\right) + H\left(\boldsymbol{\theta}; \boldsymbol{\theta}^{[i-1]}\right) \tag{4.101}$$

Note that

- $Q\left(\boldsymbol{\theta}; \boldsymbol{\theta}^{[i-1]}\right)$ is the conditional expectation of complete data loglikelihood, given the incomplete data.
- $U\left(\boldsymbol{\theta}; \boldsymbol{y}\right)$ is the loglikelihood for the incomplete data.
- $H\left(\boldsymbol{\theta}; \boldsymbol{\theta}^{[i-1]}\right)$ is the conditional expectation of missing data loglikelihood given the observed data ($\boldsymbol{y}$).

Note also that

$$Q\left(\boldsymbol{\theta}^{[i]}; \boldsymbol{\theta}^{[i-1]}\right) \geq Q\left(\boldsymbol{\theta}^{[i-1]}; \boldsymbol{\theta}^{[i-1]}\right) \tag{4.102}$$

because $\boldsymbol{\theta}^{[i]}$ is obtained by maximizing $Q\left(\boldsymbol{\theta}; \boldsymbol{\theta}^{[i-1]}\right)$. Then using 4.101 in 4.102 can write

$$U\left(\boldsymbol{\theta}^{[i]}; \boldsymbol{y}\right) + H\left(\boldsymbol{\theta}^{[i]}; \boldsymbol{\theta}^{[i-1]}\right) \geq U\left(\boldsymbol{\theta}^{[i-1]}; \boldsymbol{y}\right) + H\left(\boldsymbol{\theta}^{[i-1]}; \boldsymbol{\theta}^{[i-1]}\right)$$

Then

$$U\left(\boldsymbol{\theta}^{[i]};\boldsymbol{y}\right)-U\left(\boldsymbol{\theta}^{[i-1]};\boldsymbol{y}\right)\geq H\left(\boldsymbol{\theta}^{[i-1]};\boldsymbol{\theta}^{[i-1]}\right)-H\left(\boldsymbol{\theta}^{[i]};\boldsymbol{\theta}^{[i-1]}\right) \qquad (4.103)$$

and using 4.100

$$\begin{aligned}
&U\left(\boldsymbol{\theta}^{[i]};\boldsymbol{y}\right)-U\left(\boldsymbol{\theta}^{[i-1]};\boldsymbol{y}\right)\\
&\geq \mathrm{E}\left[\log f(\boldsymbol{m}\mid\boldsymbol{y};\boldsymbol{\theta}^{[i-1]})\mid\boldsymbol{y};\boldsymbol{\theta}^{[i-1]}\right]-\mathrm{E}\left[\log f(\boldsymbol{m}\mid\boldsymbol{y};\boldsymbol{\theta}^{[i]})\mid\boldsymbol{y};\boldsymbol{\theta}^{[i-1]}\right]\\
&\geq \mathrm{E}\left[\log f(\boldsymbol{m}\mid\boldsymbol{y};\boldsymbol{\theta}^{[i-1]})-\log f(\boldsymbol{m}\mid\boldsymbol{y};\boldsymbol{\theta}^{[i]})\mid\boldsymbol{y};\boldsymbol{\theta}^{[i-1]}\right] \qquad (4.104)\\
&\geq \mathrm{E}\left[\log\frac{f(\boldsymbol{m}\mid\boldsymbol{y};\boldsymbol{\theta}^{[i-1]})}{f(\boldsymbol{m}\mid\boldsymbol{y};\boldsymbol{\theta}^{[i]})}\mid\boldsymbol{y};\boldsymbol{\theta}^{[i-1]}\right]\\
&\geq 0
\end{aligned}$$

based on the Kullback-Leibler Inequality. So can conclude that

$$U\left(\boldsymbol{\theta}^{[i]};\boldsymbol{y}\right)\geq U\left(\boldsymbol{\theta}^{[i-1]};\boldsymbol{y}\right) \qquad (4.105)$$

and consequently

$$L\left(\boldsymbol{\theta}^{i};\boldsymbol{y}\right)\geq L\left(\boldsymbol{\theta}^{[i-1]};\boldsymbol{y}\right) \qquad (4.106)$$

Proposition 4.2.2 *Can also show that at convergence of the EM algorithm i.e. when* $\boldsymbol{\theta}^{[i-1]}=\boldsymbol{\theta}^{[i]}$

$$\frac{\partial}{\partial\boldsymbol{\theta}}U(\boldsymbol{\theta};\boldsymbol{y})\mid_{\boldsymbol{\theta}^{[i]}}=\mathbf{0} \qquad (4.107)$$

i.e. we have reached a root of $U(\boldsymbol{\theta};\boldsymbol{y})$*.*

Proof: At step $i-1$, $\boldsymbol{\theta}^{[i]}$ is obtained by

$$\mathrm{Max}_{\theta\in\omega}Q(\boldsymbol{\theta};\boldsymbol{\theta}^{[i-1]})$$

which implies that

$$\frac{\partial}{\partial\boldsymbol{\theta}}Q(\boldsymbol{\theta};\boldsymbol{\theta}^{[i-1]})\mid_{\boldsymbol{\theta}^{[i]}}=\mathbf{0} \qquad (4.108)$$

Note also that from 4.101

$$U\left(\boldsymbol{\theta};\boldsymbol{y}\right)=Q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{[i-1]}\right)-H\left(\boldsymbol{\theta};\boldsymbol{\theta}^{[i-1]}\right) \qquad (4.109)$$

Thus,

$$\begin{aligned}
\frac{\partial}{\partial\boldsymbol{\theta}}U(\boldsymbol{\theta};\boldsymbol{y})\mid_{\boldsymbol{\theta}^{[i]}} &= -\frac{\partial}{\partial\boldsymbol{\theta}}H(\boldsymbol{\theta};\boldsymbol{\theta}^{[i-1]})\mid_{\boldsymbol{\theta}^{[i]}}\\
&= -\frac{\partial}{\partial\boldsymbol{\theta}}\mathrm{E}\left[\log f(\boldsymbol{m}\mid\boldsymbol{y};\boldsymbol{\theta});\boldsymbol{\theta}^{[i-1]}\right]\mid_{\boldsymbol{\theta}^{[i]}} \qquad (4.110)\\
&= -\mathrm{E}\left[\frac{\partial}{\partial\boldsymbol{\theta}}\left\{\log f(\boldsymbol{m}\mid\boldsymbol{y};\boldsymbol{\theta})\right\}\mid_{\boldsymbol{\theta}^{[i]}}\mid\boldsymbol{y};\boldsymbol{\theta}^{[i-1]}\right]
\end{aligned}$$

But as convergence is reached, $\boldsymbol{\theta}^{[i-1]} = \boldsymbol{\theta}^{[i]}$ and

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\theta}} U(\boldsymbol{\theta}; \boldsymbol{y}) \mid_{\boldsymbol{\theta}^{[i]}} &= -\mathrm{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}} \{\log f(\boldsymbol{m} \mid \boldsymbol{y}; \boldsymbol{\theta})\} \mid_{\boldsymbol{\theta}^{[i]}} \mid \boldsymbol{y}; \boldsymbol{\theta}^{[i]} \right] \\ &= \mathbf{0} \end{aligned} \tag{4.111}$$

4.2.2 Use of EM algorithm to estimate variance components

Consider the following model

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e} \tag{4.112}$$

with

$$\begin{aligned} &\mathrm{E}(\boldsymbol{y}) = \boldsymbol{X\beta}, \quad \mathrm{E}(\boldsymbol{u}) = \mathbf{0}, \quad \mathrm{E}(\boldsymbol{e}) = \mathbf{0} \\ &\mathrm{Var}(\boldsymbol{u}) = \boldsymbol{Z} \underbrace{\boldsymbol{A}\sigma_u^2}_{\boldsymbol{G}} \boldsymbol{Z}', \quad \mathrm{Var}(\boldsymbol{e}) = \underbrace{\boldsymbol{I}\sigma_e^2}_{\boldsymbol{R}}, \quad \mathrm{Cov}(\boldsymbol{u}, \boldsymbol{e}') = \mathbf{0} \end{aligned} \tag{4.113}$$

To obtain estimates of $\boldsymbol{\beta}, \sigma_u^2$ and σ_e^2 by MLE using the EM algorithm let

$\boldsymbol{y}$ be the incomplete data

$\boldsymbol{u}$ be the missing data

and $\boldsymbol{y}$ and $\boldsymbol{u}$ form the complete data. Then,

$$\begin{aligned} L(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{u}) &\approx f(\boldsymbol{y} \mid \boldsymbol{u}; \boldsymbol{\theta}) f(\boldsymbol{u}; \boldsymbol{\theta}) \\ &\approx \frac{\exp\left\{-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{Zu})' \boldsymbol{R}^{-1} (\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{Zu})\right\}}{\left|\boldsymbol{R}\right|^{1/2}} \\ &\times \frac{\exp\left\{-\frac{1}{2}\boldsymbol{u}'\boldsymbol{G}^{-1}\boldsymbol{u}\right\}}{\left|\boldsymbol{G}\right|^{1/2}} \end{aligned} \tag{4.114}$$

Let $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{Zu}$, then

$$U(\theta; \boldsymbol{y}, \boldsymbol{u}) \approx -\frac{1}{2}\log\left|\boldsymbol{R}\right| - \frac{1}{2}\log\left|\boldsymbol{G}\right| - \frac{1}{2}\boldsymbol{e}'\boldsymbol{R}^{-1}\boldsymbol{e} - \boldsymbol{u}'\boldsymbol{G}^{-1}\boldsymbol{u} \tag{4.115}$$

and

$$\begin{aligned} Q\left(\boldsymbol{\theta}; \boldsymbol{\theta}^{[i-1]}\right) = &-\frac{1}{2}\log\left|\boldsymbol{R}\right| - \frac{1}{2}\log\left|\boldsymbol{G}\right| - \frac{1}{2}\left[\mathrm{tr}\left(\boldsymbol{R}^{-1}\mathrm{Var}\left(\boldsymbol{e} \mid \boldsymbol{y}; \boldsymbol{\theta}^{[i-1]}\right)\right) + \widehat{\boldsymbol{e}}'\boldsymbol{R}^{-1}\widehat{\boldsymbol{e}}\right] \\ &- \frac{1}{2}\left[\mathrm{tr}\left(\boldsymbol{G}^{-1}\mathrm{Var}\left(\boldsymbol{u} \mid \boldsymbol{y}; \boldsymbol{\theta}^{[i-1]}\right)\right) + \widehat{\boldsymbol{u}}'\boldsymbol{G}^{-1}\widehat{\boldsymbol{u}}\right] \end{aligned} \tag{4.116}$$

where

$$\widehat{\boldsymbol{e}} = \mathrm{E}\left(\boldsymbol{e} \mid \boldsymbol{y}; \boldsymbol{\theta}^{[i-1]}\right)$$

$$\widehat{\boldsymbol{u}} = \mathrm{E}\left(\boldsymbol{u} \mid \boldsymbol{y}; \boldsymbol{\theta}^{[i-1]}\right)$$

Now consider

$$\begin{aligned}
\mathrm{E}\left(\boldsymbol{u} \mid \boldsymbol{y}; \boldsymbol{\theta}^{[i-1]}\right) &= \widehat{\boldsymbol{u}} \\
&= \mathrm{Cov}\left(\boldsymbol{u}, \boldsymbol{y}'\right) \mathrm{Var}^{-1}(\boldsymbol{y}) \left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^{[i-1]}\right) \\
&= \boldsymbol{G}^{[i-1]}\boldsymbol{Z}' \left(\boldsymbol{Z}\boldsymbol{G}^{[i-1]}\boldsymbol{Z}' + \boldsymbol{R}^{[i-1]}\right)^{-1} \left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^{[i-1]}\right) \\
&= \left(\boldsymbol{Z}' \left(\boldsymbol{R}^{-1}\right)^{[i-1]} \boldsymbol{Z} + \left(\boldsymbol{G}^{-1}\right)^{[i-1]}\right)^{-1} \boldsymbol{Z}'\boldsymbol{R}^{-1} \left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^{[i-1]}\right)
\end{aligned} \tag{4.117}$$

and

$$\begin{aligned}
\mathrm{Var}\left(\boldsymbol{u} \mid \boldsymbol{y}; \boldsymbol{\theta}^{[i-1]}\right) &= \mathrm{Var}(\boldsymbol{u}) - \mathrm{Cov}\left(\boldsymbol{u}, \boldsymbol{y}'\right) \mathrm{Var}^{-1}(\boldsymbol{y}) \mathrm{Cov}\left(\boldsymbol{y}, \boldsymbol{u}'\right) \\
&= \boldsymbol{G}^{[i-1]} - \boldsymbol{G}^{[i-1]}\boldsymbol{Z}' \left(\boldsymbol{Z}\boldsymbol{G}^{[i-1]}\boldsymbol{Z}' + \boldsymbol{R}^{[i-1]}\right)^{-1} \boldsymbol{Z}'\boldsymbol{G}^{[i-1]} \\
&= \left(\boldsymbol{Z}\boldsymbol{G}^{[i-1]}\boldsymbol{Z}' + \boldsymbol{R}^{[i-1]}\right)^{-1} \\
&= \boldsymbol{C}_{uu}^{-1}
\end{aligned} \tag{4.118}$$

here we have used the following result from partition matrices

$$\begin{aligned}
\boldsymbol{A}^{22} &= \left(\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\right)^{-1} \\
&= \boldsymbol{A}_{22}^{-1} + \boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\left(\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\right)^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}
\end{aligned} \tag{4.119}$$

Now consider

$$\begin{aligned}
\mathrm{E}\left(\boldsymbol{e} \mid \boldsymbol{y}; \boldsymbol{\theta}^{[i-1]}\right) &= \widehat{\boldsymbol{e}} \\
&= \mathrm{E}\left[\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{u}\right) \mid \boldsymbol{y}; \boldsymbol{\theta}^{[i-1]}\right] \\
&= \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\widehat{\boldsymbol{u}}
\end{aligned} \tag{4.120}$$

and

$$\begin{aligned}
\mathrm{Var}\left(\boldsymbol{e} \mid \boldsymbol{y}; \boldsymbol{\theta}^{[i-1]}\right) &= \boldsymbol{Z}\mathrm{Var}\left(\boldsymbol{u} \mid \boldsymbol{y}; \boldsymbol{\theta}^{[i-1]}\right)\boldsymbol{Z}' \\
&= \boldsymbol{Z}\boldsymbol{C}_{uu}^{-1}\boldsymbol{Z}'
\end{aligned} \tag{4.121}$$

Now we proced with the E-step in equation 4.116

$$\begin{aligned}
Q\left(\boldsymbol{\theta}; \boldsymbol{\theta}^{[i-1]}\right) = &-\frac{1}{2}\log\left|\boldsymbol{R}\right| - \frac{1}{2}\log\left|\boldsymbol{G}\right| - \frac{1}{2}\left[\mathrm{tr}\left(\boldsymbol{R}^{-1}\boldsymbol{Z}\boldsymbol{C}_{uu}^{-1}\boldsymbol{Z}'\right) + \widehat{\boldsymbol{e}}'\boldsymbol{R}^{-1}\widehat{\boldsymbol{e}}\right] \\
&- \frac{1}{2}\left[\mathrm{tr}\left(\boldsymbol{G}^{-1}\boldsymbol{C}_{uu}^{-1}\right) + \widehat{\boldsymbol{u}}'\boldsymbol{G}^{-1}\widehat{\boldsymbol{u}}\right]
\end{aligned} \tag{4.122}$$

And for the case considered $\boldsymbol{G} = \boldsymbol{A}\sigma_u^2$ and $\boldsymbol{R} = \boldsymbol{I}\sigma_e^2$ so

$$\begin{aligned} Q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{[i-1]}\right) &= -\frac{1}{2}n\log\left(\sigma_e^2\right)^{[i]} - \frac{1}{2}\left|\boldsymbol{A}\left(\sigma_u^2\right)^{[i]}\right| \\ &- \frac{1}{2}\left[\mathrm{tr}\left(\boldsymbol{I}\left(\frac{1}{\sigma_e^2}\right)^{[i]}\boldsymbol{Z}\boldsymbol{C}_{uu}^{-1}\boldsymbol{Z}'\right) + \widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}\left(\frac{1}{\sigma_e^2}\right)^{[i]}\right] \\ &- \frac{1}{2}\left[\mathrm{tr}\left(\boldsymbol{A}^{-1}\boldsymbol{C}_{uu}^{-1}\left(\frac{1}{\sigma_u^2}\right)^{[i]}\right) + \left(\frac{1}{\sigma_u^2}\right)^{[i]}\widehat{\boldsymbol{u}}'\boldsymbol{A}^{-1}\widehat{\boldsymbol{u}}\right] \end{aligned} \tag{4.123}$$

For the M-step take the first derivatives with respect to $\boldsymbol{\beta}, \sigma_u^2$ and σ_e^2 and then set to zero and solve for the parameters.So

$$\begin{aligned} \frac{\partial Q}{\partial \boldsymbol{\beta}} &= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\beta}}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\widehat{\boldsymbol{u}}\right)'\boldsymbol{R}^{-1}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\widehat{\boldsymbol{u}}\right) \\ &= \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{y} - \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z}\widehat{\boldsymbol{u}} - \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X}\boldsymbol{\beta} \\ &= 0 \end{aligned} \tag{4.124}$$

Now we can solve for $\boldsymbol{\beta}$

$$(\boldsymbol{\beta})^{[i]} = \left(\boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X}\right)^{-}\boldsymbol{X}'\boldsymbol{R}^{-1}\left(\boldsymbol{y} - \boldsymbol{Z}\widehat{\boldsymbol{u}}\right) \tag{4.125}$$

Next consider

$$\begin{aligned} \frac{\partial Q}{\partial \sigma_u^2} &= -\frac{1}{2}\frac{q}{\sigma_u^2} + \frac{1}{2}\frac{1}{\left(\sigma_u^2\right)^2}\mathrm{tr}\left(\boldsymbol{A}^{-1}\boldsymbol{C}_{uu}^{-1}\right) + \frac{1}{2}\frac{1}{\left(\sigma_u^2\right)^2}\widehat{\boldsymbol{u}}'\boldsymbol{A}^{-1}\widehat{\boldsymbol{u}} \\ &= 0 \end{aligned} \tag{4.126}$$

and solve for σ_u^2

$$\left(\sigma_u^2\right)^{[i]} = \frac{1}{q}\left[\mathrm{tr}\left(\boldsymbol{A}^{-1}\boldsymbol{C}_{uu}^{-1}\right) + \widehat{\boldsymbol{u}}'\boldsymbol{A}^{-1}\widehat{\boldsymbol{u}}\right] \tag{4.127}$$

And finally consider

$$\frac{\partial Q}{\partial \sigma_e^2} = -\frac{1}{2}\frac{n}{\sigma_e^2} + \frac{1}{2}\frac{\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}}{\left(\sigma_e^2\right)^2} + \frac{1}{2}\frac{1}{\left(\sigma_e^2\right)^2}\mathrm{tr}\left(\boldsymbol{Z}\boldsymbol{C}_{uu}^{-1}\boldsymbol{Z}'\right) \tag{4.128}$$

and solve for σ_e^2

$$\left(\sigma_e^2\right)^{[i]} = \frac{1}{n}\left[\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}} + \mathrm{tr}\left(\boldsymbol{Z}\boldsymbol{C}_{uu}^{-1}\boldsymbol{Z}'\right)\right] \tag{4.129}$$

Simplification for Exponential family of distributions

Let $\boldsymbol{x} = \begin{bmatrix}\boldsymbol{y} \\ \boldsymbol{m}\end{bmatrix}$ be the complete data. Suppose

$$f(\boldsymbol{x};\boldsymbol{\theta}) = \frac{b(\boldsymbol{x})}{a(\boldsymbol{\theta})}\exp\left\{\boldsymbol{t}'\boldsymbol{\theta}\right\} \tag{4.130}$$

where $\boldsymbol{t}$ is a sufficient statistic, and

$$a(\boldsymbol{\theta}) = \int b(\boldsymbol{x})\exp\left\{\boldsymbol{t}'\boldsymbol{\theta}\right\} d\boldsymbol{x} \tag{4.131}$$

Then,

$$\frac{\partial \log a(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{a(\boldsymbol{\theta})}\frac{\partial a(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \int \frac{\boldsymbol{t}b(\boldsymbol{x})\exp\left\{\boldsymbol{t}'\boldsymbol{\theta}\right\}}{a(\boldsymbol{\theta})} d\boldsymbol{x} = \int \boldsymbol{t}f(\boldsymbol{x};\boldsymbol{\theta})d\boldsymbol{x} = \mathrm{E}\left(\boldsymbol{t};\boldsymbol{\theta}\right) \tag{4.132}$$

This result will be used in the M-step as follows

$$Q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{[i-1]}\right) = \boldsymbol{c} - \log a(\boldsymbol{\theta}) + \mathrm{E}\left(\boldsymbol{t}' \mid \boldsymbol{y};\boldsymbol{\theta}^{[i-1]}\right)\boldsymbol{\theta} \tag{4.133}$$

where $\boldsymbol{c} = \mathrm{E}\left(\log b(\boldsymbol{x}) \mid \boldsymbol{y};\boldsymbol{\theta}^{[i-1]}\right)$ is not a function of $\boldsymbol{\theta}$. So in the M-step

$$\begin{aligned}\frac{\partial Q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{[i-1]}\right)}{\partial \boldsymbol{\theta}} &= -\frac{\partial \log a(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \mathrm{E}\left(\boldsymbol{t}' \mid \boldsymbol{y};\boldsymbol{\theta}^{[i-1]}\right)\\ &= -\mathrm{E}\left(\boldsymbol{t};\boldsymbol{\theta}\right) + \mathrm{E}\left(\boldsymbol{t}' \mid \boldsymbol{y};\boldsymbol{\theta}^{[i-1]}\right)\\ &= 0\end{aligned} \tag{4.134}$$

So Q is maximized by solving through iteration

$$\mathrm{E}\left(\boldsymbol{t};\boldsymbol{\theta}\right) = \mathrm{E}\left(\boldsymbol{t}' \mid \boldsymbol{y};\boldsymbol{\theta}^{[i-1]}\right) \tag{4.135}$$

in the case when we have incomplete data.

Note that when complete data is available, that is when $\boldsymbol{x}$ is observed,

$$\log f(\boldsymbol{x};\boldsymbol{\theta}) \approx \boldsymbol{t}'\boldsymbol{\theta} - \log a(\boldsymbol{\theta}) \tag{4.136}$$

and as a result

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\theta}}\log f(\boldsymbol{x};\boldsymbol{\theta}) &= \boldsymbol{t} - \mathrm{E}\left(\boldsymbol{t};\boldsymbol{\theta}\right)\\ &= \boldsymbol{t} - \boldsymbol{P}\boldsymbol{\theta}\end{aligned} \tag{4.137}$$

By setting the first derivative equal to zero and solving for $\boldsymbol{\theta}$ we have

$$\widehat{\boldsymbol{\theta}} = \boldsymbol{P}^{-1}\boldsymbol{t} \tag{4.138}$$

4.2.3 REML - residual maximum likelihood

Consider the case when we do not write the likelihood given $\boldsymbol{y}$ but given $\boldsymbol{K}'\boldsymbol{y}$ where

$$\begin{aligned}&\mathrm{E}\left(\boldsymbol{K}'\boldsymbol{y}\right) = \boldsymbol{K}'\boldsymbol{X}\boldsymbol{\beta} = 0 \quad \text{for} \quad \text{all} \quad \boldsymbol{\beta}\\ &\Rightarrow \boldsymbol{K}'\boldsymbol{X} = 0\end{aligned} \tag{4.139}$$

with $\boldsymbol{K}$ a $(n-r)\times n$ matrix and rank$(\boldsymbol{K}') = n - r$. Note that $\boldsymbol{w} = \boldsymbol{K}'\boldsymbol{y}$ are $n-r$ linearly independent error contrasts because $\mathrm{E}\left(\boldsymbol{K}'\boldsymbol{y}\right) = 0$. Remember that we have seen that for *BLUP* that

$\mathrm{E}(\boldsymbol{u}) = \mathbf{0}$

$BLUP(\boldsymbol{u}) = \mathrm{E}(\boldsymbol{u} \mid \boldsymbol{w})$

In order to do EM for REML estimates need to look at

$\mathrm{E}(\boldsymbol{u} \mid \boldsymbol{w})$, $\mathrm{Var}(\boldsymbol{u} \mid \boldsymbol{w})$

and $\mathrm{E}(\boldsymbol{e} \mid \boldsymbol{w})$, $\mathrm{Var}(\boldsymbol{e} \mid \boldsymbol{w})$

Recall that $\boldsymbol{y} \in V_n$ can be written as

$$\boldsymbol{y} = \widehat{\boldsymbol{y}} + \widehat{\boldsymbol{e}} \tag{4.140}$$

where $\widehat{\boldsymbol{y}} \in V_r$, the subspace spanned by the columns of $\boldsymbol{X}$, and $\widehat{\boldsymbol{e}} \perp V_r$. Consequently $\widehat{\boldsymbol{y}}$ can be written as $\boldsymbol{Ay}$ and $\widehat{\boldsymbol{e}}$ as $(\boldsymbol{I} - \boldsymbol{A})\,\boldsymbol{y}$ where $\boldsymbol{A} = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-}\boldsymbol{X}'$. Further we have shown that $\widehat{\boldsymbol{y}}$ and $\widehat{\boldsymbol{e}}$ are unique. It can be shown that this fact implies that $\boldsymbol{A}$ and $(\boldsymbol{I} - \boldsymbol{A})$ are unique too. To see this fact suppose $\boldsymbol{A}$ and $\boldsymbol{B}$ are both projections of $\boldsymbol{y}$ onto V_r. Then

$$(\boldsymbol{A} - \boldsymbol{B})\,\boldsymbol{y} = \mathbf{0} \quad \text{for} \quad \text{all} \quad \boldsymbol{y} \tag{4.141}$$

due to the fact that $\widehat{\boldsymbol{y}}$ is unique, so $\boldsymbol{Ay} = \boldsymbol{By} = \widehat{\boldsymbol{y}}$. Note that this holds for all $\boldsymbol{y}$. Consider now

$$\begin{aligned}
\boldsymbol{y} &= \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \Rightarrow \boldsymbol{A} \quad \text{and} \quad \boldsymbol{B} \quad \text{have the same first column} \\
&\vdots \\
\boldsymbol{y} &= \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \quad \Rightarrow \boldsymbol{A} \quad \text{and} \quad \boldsymbol{B} \quad \text{have the same last column}
\end{aligned} \tag{4.142}$$

So we can conclude that $\boldsymbol{A} = \boldsymbol{B}$.

Theorem 4.2.1 *For*

$$\boldsymbol{y} \sim \aleph\left(\boldsymbol{X\beta}, \boldsymbol{V}\right), \quad rank\left(\boldsymbol{X}^{n\times p}\right) = r; \quad rank\left(\boldsymbol{K}'^{(n-r)\times n}\right) = (n-r) \tag{4.143}$$

and $\boldsymbol{K}'\boldsymbol{X} = \mathbf{0}; \boldsymbol{w} = \boldsymbol{K}'\boldsymbol{y}$,*and for* $\boldsymbol{u}$ *and* $\boldsymbol{y}$ *MVN, can show that*

$$\begin{aligned}
E(\boldsymbol{u} \mid \boldsymbol{w}) &= Cov(\boldsymbol{u}, \boldsymbol{w}')\ Var^{-1}(\boldsymbol{w})\boldsymbol{w} \\
&= Cov(\boldsymbol{u}, \boldsymbol{y}')\,\boldsymbol{K}\left(\boldsymbol{K}'\boldsymbol{VK}\right)^{-1}\boldsymbol{K}'\boldsymbol{y}
\end{aligned} \tag{4.144}$$

is

$$\mathrm{BLUP}(\boldsymbol{u}) = \widehat{\boldsymbol{u}} = Cov(\boldsymbol{u}, \boldsymbol{y}')\,\boldsymbol{V}^{-1}\left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right) \tag{4.145}$$

where $\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}$

Proof: Note that by replacing $\widehat{\boldsymbol{\beta}}$ into $BLUP(\boldsymbol{u})$ can write

$$\widehat{\boldsymbol{u}} = \operatorname{Cov}(\boldsymbol{u}, \boldsymbol{y}') \left[\boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-}\boldsymbol{X}'\boldsymbol{V}^{-1}\right]\boldsymbol{y} \tag{4.146}$$

First we will prove the following lemma:

Lemma 4.2.1 *For* $\boldsymbol{y} \sim \aleph(\boldsymbol{X\beta}, \boldsymbol{V})$, *rank* $\left(\boldsymbol{X}^{n\times p}\right) = r$; *rank* $\left(\boldsymbol{K}'^{(n-r)\times n}\right) = (n-r)$ *and with the property that* $\boldsymbol{K}'\boldsymbol{X} = \mathbf{0}$ *can show that*

$$\boldsymbol{K}\left(\boldsymbol{K}'\boldsymbol{V}\boldsymbol{K}\right)^{-1}\boldsymbol{K}' = \boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-}\boldsymbol{X}'\boldsymbol{V}^{-1} \tag{4.147}$$

Proof: Note that

$$\boldsymbol{K}'\boldsymbol{X} = \underbrace{\boldsymbol{K}'\boldsymbol{V}^{1/2}}_{\boldsymbol{K}^{*'}}\underbrace{\boldsymbol{V}^{-1/2}\boldsymbol{X}}_{\boldsymbol{X}^{*}} = \boldsymbol{K}^{*'}\boldsymbol{X}^{*} = \mathbf{0} \tag{4.148}$$

Let $\boldsymbol{V}^{*}$ be the space spanned by the columns of $\boldsymbol{X}^{*}$. From 4.148 it follows that the columns of $\boldsymbol{K}^{*}$ span the orthocomplement of $\boldsymbol{V}^{*}$, denoted by $\boldsymbol{V}^{*\perp}$. So,

$$\boldsymbol{K}^{*}\left(\boldsymbol{K}^{*'}\boldsymbol{K}^{*}\right)^{-1}\boldsymbol{K}^{*'}$$

is the projection matrix onto $\boldsymbol{V}^{*\perp}$. But,

$$\left[\boldsymbol{I} - \boldsymbol{X}^{*}\left(\boldsymbol{X}^{*'}\boldsymbol{X}^{*}\right)^{-1}\boldsymbol{X}^{*'}\right]$$

is also a projection matrix onto $\boldsymbol{V}^{*\perp}$. Thus,

$$\boldsymbol{K}^{*}\left(\boldsymbol{K}^{*'}\boldsymbol{K}^{*}\right)^{-1}\boldsymbol{K}^{*'} = \left[\boldsymbol{I} - \boldsymbol{X}^{*}\left(\boldsymbol{X}^{*'}\boldsymbol{X}^{*}\right)^{-}\boldsymbol{X}^{*'}\right]$$

and by pre and postmultiplying by $\boldsymbol{V}^{-1/2}$,

$$\begin{aligned}\boldsymbol{V}^{-1/2}\boldsymbol{K}^{*}\left(\boldsymbol{K}^{*'}\boldsymbol{K}^{*}\right)^{-1}\boldsymbol{K}^{*'}\boldsymbol{V}^{-1/2} &= \boldsymbol{V}^{-1/2}\boldsymbol{V}^{-1/2} \\ &\quad - \boldsymbol{V}^{-1/2}\boldsymbol{X}^{*}\left(\boldsymbol{X}^{*'}\boldsymbol{X}^{*}\right)^{-}\boldsymbol{X}^{*'}\boldsymbol{V}^{-1/2}\end{aligned} \tag{4.149}$$

Now using the fact that $\boldsymbol{K}^{*'} = \boldsymbol{K}'\boldsymbol{V}^{1/2}$ and $\boldsymbol{X}^{*} = \boldsymbol{V}^{-1/2}\boldsymbol{X}$ in 4.149 results in

$$\begin{aligned}&\underbrace{\boldsymbol{V}^{-1/2}\boldsymbol{V}^{1/2}}_{\boldsymbol{I}}\boldsymbol{K}\left(\boldsymbol{K}'\boldsymbol{V}^{1/2}\boldsymbol{V}^{1/2}\boldsymbol{K}\right)^{-1}\boldsymbol{K}'\underbrace{\boldsymbol{V}^{1/2}\boldsymbol{V}^{-1/2}}_{\boldsymbol{I}} \\ &= \boldsymbol{V}^{-1} - \underbrace{\boldsymbol{V}^{-1/2}\boldsymbol{V}^{-1/2}}_{\boldsymbol{V}^{-1}}\boldsymbol{X}\left(\boldsymbol{X}'\underbrace{\boldsymbol{V}^{-1/2}\boldsymbol{V}^{-1/2}}_{\boldsymbol{V}^{-1}}\boldsymbol{X}\right)^{-}\boldsymbol{X}'\underbrace{\boldsymbol{V}^{-1/2}\boldsymbol{V}^{-1/2}}_{\boldsymbol{V}^{-1}}\end{aligned} \tag{4.150}$$

So,

$$\boldsymbol{K}\left(\boldsymbol{K}'\boldsymbol{V}\boldsymbol{K}\right)^{-1}\boldsymbol{K}' = \boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-}\boldsymbol{X}'\boldsymbol{V}^{-1} \tag{4.151}$$

Now we return to the proof of the theorem. Note that using the result of the lemma,

$$\begin{aligned} \mathrm{E}\left(\boldsymbol{u} \mid \boldsymbol{w}\right) &= \mathrm{Cov}\left(\boldsymbol{u}, \boldsymbol{y}'\right)\boldsymbol{K}\left(\boldsymbol{K}'\boldsymbol{V}\boldsymbol{K}\right)^{-1}\boldsymbol{K}'\boldsymbol{y} \\ &= \mathrm{Cov}\left(\boldsymbol{u}, \boldsymbol{y}'\right)\boldsymbol{V}^{-1}\left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right) \\ &= BLUP(\boldsymbol{u}) \end{aligned} \tag{4.152}$$

where $\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-}\boldsymbol{X}'\boldsymbol{V}^{-1}$ and consequently does not depend on $\boldsymbol{K}$.

Consider now

$$\begin{aligned} BLUP(\boldsymbol{e}) = \widehat{\boldsymbol{e}} &= \mathrm{Cov}\left(\boldsymbol{e}, \boldsymbol{y}'\right)\boldsymbol{V}^{-1}\left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right) \\ &= \boldsymbol{R}\boldsymbol{V}^{-1}\left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right) \end{aligned} \tag{4.153}$$

want to show that

$$BLUP(\boldsymbol{e}) = \widehat{\boldsymbol{e}} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{Z}\widehat{u} \tag{4.154}$$

where $\widehat{u} = BLUP(\boldsymbol{u})$. Can write

$$\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{Z}\widehat{u} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}'\boldsymbol{V}^{-1}\left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right) \tag{4.155}$$

Note that

$$\begin{aligned} &\mathrm{Var}(\boldsymbol{y}) = \left(\boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}' + \boldsymbol{R}\right) = \boldsymbol{V} \\ &\boldsymbol{V}\boldsymbol{V}^{-1} = \boldsymbol{I} \end{aligned}$$

Then by postmultiplying by $\boldsymbol{V}^{-1}$ the variance of $\boldsymbol{y}$ can write

$$\begin{aligned} &\left(\boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}' + \boldsymbol{R}\right)\boldsymbol{V}^{-1} = \boldsymbol{I} \Rightarrow \\ &\boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}'\boldsymbol{V}^{-1} + \boldsymbol{R}\boldsymbol{V}^{-1} = \boldsymbol{I} \Rightarrow \\ &\boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}'\boldsymbol{V}^{-1} = \boldsymbol{I} - \boldsymbol{R}\boldsymbol{V}^{-1} \end{aligned} \tag{4.156}$$

Now using 4.156 in 4.155 results in

$$\begin{aligned} \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{Z}\widehat{u} &= \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} - \left(\boldsymbol{I} - \boldsymbol{R}\boldsymbol{V}^{-1}\right)\left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right) \\ &= \boldsymbol{R}\boldsymbol{V}^{-1}\left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right) \end{aligned} \tag{4.157}$$

Now can conclude that

$$\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{Z}\widehat{u} = \boldsymbol{R}\boldsymbol{V}^{-1}\left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right) = \widehat{\boldsymbol{e}} = BLUP(\boldsymbol{e}) \tag{4.158}$$

Note that the ML estimate of σ_e^2 is equal to $\frac{\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}}{n}$ for a fixed linear model, but the REML estimate of σ_e^2 will be equal to $\frac{\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}}{n-r}$ where $r = \text{rank}(\boldsymbol{X})$. In REML,

$$
\begin{aligned}
&\boldsymbol{w} = \boldsymbol{K}'\boldsymbol{y} \sim \aleph\left(\boldsymbol{0}, \boldsymbol{K}'\boldsymbol{V}\boldsymbol{K}\right) \\
&f(\boldsymbol{w}) \approx \frac{1}{\left|\boldsymbol{K}'\boldsymbol{V}\boldsymbol{K}\right|^{1/2}} \exp\left\{-\frac{1}{2}\left[\boldsymbol{w}'\left(\boldsymbol{K}'\boldsymbol{V}\boldsymbol{K}\right)^{-1}\boldsymbol{w}\right]\right\} \\
&U(\boldsymbol{w}) = \log f(\boldsymbol{w}) \approx -\frac{1}{2}\log\left|\boldsymbol{K}'\boldsymbol{V}\boldsymbol{K}\right| - \frac{1}{2}\boldsymbol{y}'\boldsymbol{K}\left(\boldsymbol{K}'\boldsymbol{V}\boldsymbol{K}\right)^{-1}\boldsymbol{K}'\boldsymbol{y}
\end{aligned}
$$

Now consider the first derivative of the loglikelihood

$$
\begin{aligned}
&\frac{\partial U}{\partial \theta_i} = \\
&-\frac{1}{2}\text{tr}\left(\boldsymbol{K}'\boldsymbol{V}\boldsymbol{K}\right)^{-1}\frac{\partial}{\partial\theta_i}\left(\boldsymbol{K}'\boldsymbol{V}\boldsymbol{K}\right) \\
&+\frac{1}{2}\boldsymbol{y}'\boldsymbol{K}\left(\boldsymbol{K}'\boldsymbol{V}\boldsymbol{K}\right)^{-1}\frac{\partial}{\partial\theta_i}\left(\boldsymbol{K}'\boldsymbol{V}\boldsymbol{K}\right)\left(\boldsymbol{K}'\boldsymbol{V}\boldsymbol{K}\right)^{-1}\boldsymbol{K}'\boldsymbol{y} \\
&= -\frac{1}{2}\text{tr}\left(\boldsymbol{K}'\boldsymbol{V}\boldsymbol{K}\right)^{-1}\boldsymbol{K}'\boldsymbol{V}_i\boldsymbol{K} \\
&+\frac{1}{2}\boldsymbol{y}'\underbrace{\boldsymbol{K}\left(\boldsymbol{K}'\boldsymbol{V}\boldsymbol{K}\right)^{-1}\boldsymbol{K}'}_{\boldsymbol{P}}\boldsymbol{V}_i\underbrace{\boldsymbol{K}\left(\boldsymbol{K}'\boldsymbol{V}\boldsymbol{K}\right)^{-1}\boldsymbol{K}'}_{\boldsymbol{P}}\boldsymbol{y}
\end{aligned}
\tag{4.159}
$$

where $\boldsymbol{V}_i = \frac{\partial \boldsymbol{V}}{\partial \theta_i}$. As a result can write

$$
\frac{\partial U}{\partial \theta_i} = -\frac{1}{2}\text{tr}\boldsymbol{P}\boldsymbol{V}_i + \frac{1}{2}\boldsymbol{y}'\boldsymbol{P}\boldsymbol{V}_i\boldsymbol{P}\boldsymbol{y} \tag{4.160}
$$

In order to look at the second derivatives of the loglikelihood first look at the derivative of $\boldsymbol{P}$.

$$
\begin{aligned}
\frac{\partial \boldsymbol{P}}{\partial \theta_i} &= \frac{\partial}{\partial \theta_i}\boldsymbol{K}\left(\boldsymbol{K}'\boldsymbol{V}\boldsymbol{K}\right)^{-1}\boldsymbol{K}' \\
&= -\boldsymbol{K}\left(\boldsymbol{K}'\boldsymbol{V}\boldsymbol{K}\right)^{-1}\boldsymbol{K}'\boldsymbol{V}_i\boldsymbol{K}\left(\boldsymbol{K}'\boldsymbol{V}\boldsymbol{K}\right)^{-1}\boldsymbol{K}' \\
&= -\boldsymbol{P}\boldsymbol{V}_i\boldsymbol{P}
\end{aligned}
\tag{4.161}
$$

Now can write the second derivative of the loglikelihood

$$
\begin{aligned}
\frac{\partial U}{\partial \theta_i \partial \theta_j} &= \frac{1}{2}\text{tr}\boldsymbol{P}\boldsymbol{V}_j\boldsymbol{P}\boldsymbol{V}_i - \frac{1}{2}\boldsymbol{y}'\boldsymbol{P}\boldsymbol{V}_j\boldsymbol{P}\boldsymbol{V}_i\boldsymbol{P}\boldsymbol{y} - \frac{1}{2}\boldsymbol{y}'\boldsymbol{P}\boldsymbol{V}_i\boldsymbol{P}\boldsymbol{V}_j\boldsymbol{P}\boldsymbol{y} \\
&= \frac{1}{2}\text{tr}\boldsymbol{P}\boldsymbol{V}_j\boldsymbol{P}\boldsymbol{V}_i - \boldsymbol{y}'\boldsymbol{P}\boldsymbol{V}_j\boldsymbol{P}\boldsymbol{V}_i\boldsymbol{P}\boldsymbol{y}
\end{aligned}
\tag{4.162}
$$

After computing the first and second derivatives the parameters desired can be computed using one of the following procedures:

1. **Newton-Raphson**

$$\boldsymbol{\theta}^{[i]} = \boldsymbol{\theta}^{[i-1]} - \boldsymbol{B}^{-1}\boldsymbol{s}$$

where

$$\boldsymbol{B} = \left[\frac{\partial U}{\partial\theta_i\partial\theta_j}\right]_{\theta=\theta^{[i-1]}} \quad \text{and} \quad \boldsymbol{s} = \frac{\partial U}{\partial\theta_i}$$

2. **Functional Iteration**

Note that

$$\frac{\partial U}{\partial\theta_i} = -\frac{1}{2}\left\{\text{tr}\boldsymbol{P}\boldsymbol{V}_i - \boldsymbol{y}'\boldsymbol{P}\boldsymbol{V}_i\boldsymbol{P}\boldsymbol{y}\right\}$$

and that $\boldsymbol{P}\boldsymbol{V}\boldsymbol{P} = \boldsymbol{P}$. In order to do functional iteration set to zero the first derivative and then rearange the terms of the term that involves the trace

$$\begin{aligned}
\{\text{tr}\boldsymbol{P}\boldsymbol{V}_i\} &= \{\boldsymbol{y}'\boldsymbol{P}\boldsymbol{V}_i\boldsymbol{P}\boldsymbol{y}\} \\
\{\text{tr}\boldsymbol{P}\boldsymbol{V}\boldsymbol{P}\boldsymbol{V}_i\} &= \{\boldsymbol{y}'\boldsymbol{P}\boldsymbol{V}_i\boldsymbol{P}\boldsymbol{y}\} \\
\{\text{tr}\boldsymbol{P}\boldsymbol{V}_i\boldsymbol{P}\boldsymbol{V}\} &= \{\boldsymbol{y}'\boldsymbol{P}\boldsymbol{V}_i\boldsymbol{P}\boldsymbol{y}\} \\
\left\{\text{tr}\boldsymbol{P}\boldsymbol{V}_i\boldsymbol{P}\underbrace{\sum \boldsymbol{V}_j\theta_j}_{\boldsymbol{V}}\right\} &= \{\boldsymbol{y}'\boldsymbol{P}\boldsymbol{V}_i\boldsymbol{P}\boldsymbol{y}\} \\
\left\{\sum \text{tr}\boldsymbol{P}\boldsymbol{V}_i\boldsymbol{P}\boldsymbol{V}_j\theta_j\right\} &= \{\boldsymbol{y}'\boldsymbol{P}\boldsymbol{V}_i\boldsymbol{P}\boldsymbol{y}\}
\end{aligned} \tag{4.163}$$

Then the set of equations used for functional iteration will be

$$\underbrace{\{\text{tr}\boldsymbol{P}\boldsymbol{V}_i\boldsymbol{P}\boldsymbol{V}_j\}}_{\text{matrix}}\boldsymbol{\theta} = \underbrace{\{\boldsymbol{y}'\boldsymbol{P}\boldsymbol{V}_i\boldsymbol{P}\boldsymbol{y}\}}_{\text{vector}} \tag{4.164}$$

3. **REML - EM**

Consider the following model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e} \tag{4.165}$$

with $\boldsymbol{u} \sim \aleph\left(\boldsymbol{0}, \boldsymbol{A}\sigma_u^2\right)$ and $\boldsymbol{e} \sim \aleph\left(\boldsymbol{0}, \boldsymbol{I}\sigma_e^2\right)$. Want to maximize

$$L\left(\boldsymbol{\theta}; \boldsymbol{K}'\boldsymbol{y}\right) \quad \text{where} \quad \text{E}\left(\boldsymbol{K}'\boldsymbol{y}\right) = \boldsymbol{0} \quad \text{and} \quad \boldsymbol{\theta}' = \begin{bmatrix}\sigma_u^2 & \sigma_e^2\end{bmatrix} \tag{4.166}$$

Let

$$\boldsymbol{w} = \boldsymbol{K}'\boldsymbol{y} = \underbrace{\boldsymbol{K}'\boldsymbol{X}}_{0}\boldsymbol{\beta} + \boldsymbol{K}'\boldsymbol{Z}\boldsymbol{u} + \boldsymbol{K}'\boldsymbol{e} = \boldsymbol{K}'\boldsymbol{Z}\boldsymbol{u} + \boldsymbol{K}'\boldsymbol{e} \tag{4.167}$$

be the incomplete data, and let $\boldsymbol{u}$ and $\boldsymbol{e}$ form the complete data. Then the complete data likelihood is

$$L(\boldsymbol{\theta}; \boldsymbol{u}, \boldsymbol{e}) \approx \frac{\exp\left\{\frac{\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{u}}{2\sigma_u^2}\right\}}{\left|\boldsymbol{A}\sigma_u^2\right|^{1/2}} \times \frac{\exp\left\{\frac{\boldsymbol{e}'\boldsymbol{e}}{2\sigma_e^2}\right\}}{\left|\boldsymbol{I}\sigma_e^2\right|} \tag{4.168}$$

Note that

$$\begin{aligned} \left|\boldsymbol{A}\sigma_u^2\right| &= \left|\boldsymbol{A}\right|\left(\sigma_u^2\right)^q \\ \left|\boldsymbol{I}\sigma_e^2\right| &= \left|\boldsymbol{I}\right|\left(\sigma_e^2\right)^n \end{aligned} \tag{4.169}$$

and then the loglikelihood becomes

$$U(\boldsymbol{\theta};\boldsymbol{u},\boldsymbol{e}) \approx -\frac{q}{2}\log\left(\sigma_u^2\right) - \frac{\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{u}}{2\sigma_u^2} - \frac{n}{2}\log\left(\sigma_e^2\right) - \frac{\boldsymbol{e}'\boldsymbol{e}}{2\sigma_e^2} \tag{4.170}$$

In the E-step of the EM algorithm we obtain

$$\begin{aligned} \mathrm{E}(U) &= Q\left(\boldsymbol{\theta}^{[i]},\boldsymbol{\theta}^{[i-1]}\right) \\ &= \frac{q}{2}\log\left(\sigma_u^2\right) - \frac{1}{2\sigma_u^2}\left[\mathrm{tr}\boldsymbol{A}^{-1}\mathrm{Var}\left(\boldsymbol{u}\mid\boldsymbol{w}\right) + \widehat{\boldsymbol{u}}'\boldsymbol{A}^{-1}\widehat{\boldsymbol{u}}\right] \\ &- \frac{n}{2}\log\left(\sigma_e^2\right) - \frac{1}{2\sigma_e^2}\left[\mathrm{tr}\mathrm{Var}\left(\boldsymbol{e}\mid\boldsymbol{w}\right) + \widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}\right] \end{aligned} \tag{4.171}$$

where $\widehat{\boldsymbol{u}} = \mathrm{E}\left(\boldsymbol{u}\mid\boldsymbol{w}\right) = BLUP(\boldsymbol{u})$ and $\mathrm{Var}\left(\boldsymbol{u}\mid\boldsymbol{w}\right) = \mathrm{Var}\left(\widehat{\boldsymbol{u}}-\boldsymbol{u}\right) = \boldsymbol{C}^{22}$. Note that

$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z} + \frac{\boldsymbol{A}^{-1}}{\sigma_u^2} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{C}^{11} & \boldsymbol{C}^{12} \\ \boldsymbol{C}^{21} & \boldsymbol{C}^{22} \end{bmatrix} \quad \text{with} \quad \boldsymbol{R} = \boldsymbol{I}\sigma_e^2 \tag{4.172}$$

Also, $\widehat{\boldsymbol{e}} = \mathrm{E}\left(\boldsymbol{e}\mid\boldsymbol{w}\right) = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{Z}\widehat{\boldsymbol{u}}$ and $\mathrm{Var}\left(\boldsymbol{e}\mid\boldsymbol{w}\right) = \mathrm{Var}\left(\widehat{\boldsymbol{e}}-\boldsymbol{e}\right)$ Now

$$\begin{aligned} \widehat{\boldsymbol{e}} - \boldsymbol{e} &= \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{Z}\widehat{\boldsymbol{u}} - \boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} \\ &= -\boldsymbol{X}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right) - \boldsymbol{Z}\left(\widehat{\boldsymbol{u}}-\boldsymbol{u}\right) \\ &= -\begin{bmatrix} \boldsymbol{X} & \boldsymbol{Z} \end{bmatrix}\begin{bmatrix} \widehat{\boldsymbol{\beta}}-\boldsymbol{\beta} \\ \widehat{\boldsymbol{u}}-\boldsymbol{u} \end{bmatrix} \end{aligned} \tag{4.173}$$

and as a result can write

$$\begin{aligned} \mathrm{Var}\left(\widehat{\boldsymbol{e}}-\boldsymbol{e}\right) &= \begin{bmatrix} \boldsymbol{X} & \boldsymbol{Z} \end{bmatrix}\mathrm{Var}\begin{bmatrix} \widehat{\boldsymbol{\beta}}-\boldsymbol{\beta} \\ \widehat{\boldsymbol{u}}-\boldsymbol{u} \end{bmatrix}\begin{bmatrix} \boldsymbol{X}' \\ \boldsymbol{Z}' \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{X} & \boldsymbol{Z} \end{bmatrix}\underbrace{\begin{bmatrix} \boldsymbol{C}^{11} & \boldsymbol{C}^{12} \\ \boldsymbol{C}^{21} & \boldsymbol{C}^{22} \end{bmatrix}}_{\boldsymbol{C}^{-1}}\begin{bmatrix} \boldsymbol{X}' \\ \boldsymbol{Z}' \end{bmatrix} \end{aligned} \tag{4.174}$$

Using these results can write

$$\begin{aligned} Q\left(\boldsymbol{\theta}^{[i]},\boldsymbol{\theta}^{[i-1]}\right) &= \frac{q}{2}\log\left(\sigma_u^2\right) - \frac{1}{2\sigma_u^2}\left[\mathrm{tr}\boldsymbol{A}^{-1}\boldsymbol{C}^{22} + \widehat{\boldsymbol{u}}'\boldsymbol{A}^{-1}\widehat{\boldsymbol{u}}\right] \\ &- \frac{n}{2}\log\left(\sigma_e^2\right) - \frac{1}{2\sigma_e^2}\left[\mathrm{tr}\begin{bmatrix} \boldsymbol{X}'\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{Z} \end{bmatrix}\boldsymbol{C}^{-1} + \widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}\right] \end{aligned} \tag{4.175}$$

Now in the M-step we take the first derivatives

$$\begin{aligned}\frac{\partial Q}{\partial \sigma_u^2} &= -\frac{q}{2\sigma_u^2} + \frac{1}{2\left(\sigma_u^2\right)^2}\left[\mathrm{tr}\boldsymbol{A}^{-1}\boldsymbol{C}^{22} + \widehat{\boldsymbol{u}}'\boldsymbol{A}^{-1}\widehat{\boldsymbol{u}}\right] \\ \frac{\partial Q}{\partial \sigma_e^2} &= -\frac{n}{2\sigma_e^2} + \frac{1}{2\left(\sigma_e^2\right)^2}\left[\mathrm{tr}\begin{bmatrix}\boldsymbol{X}'\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{Z}\end{bmatrix}\boldsymbol{C}^{-1} + \widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}\right]\end{aligned} \tag{4.176}$$

and then in order to maximize set them to zero and solve for σ_u^2 and σ_e^2

$$\begin{aligned}\left(\sigma_u^2\right)^{[i]} &= \frac{\mathrm{tr}\boldsymbol{A}^{-1}\boldsymbol{C}^{22} + \widehat{\boldsymbol{u}}'\boldsymbol{A}^{-1}\widehat{\boldsymbol{u}}}{q} \\ \left(\sigma_e^2\right)^{[i]} &= \frac{\mathrm{tr}\begin{bmatrix}\boldsymbol{X}'\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{Z}\end{bmatrix}\boldsymbol{C}^{-1} + \widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}}{n}\end{aligned} \tag{4.177}$$

4.2.4 REML in terms of mixed model equations

In the previous section the desired parameters are estimated with respect to $\boldsymbol{V}^{-1}$. It might be convenient to express them in terms of the mixed model equations. For $\boldsymbol{u} \sim \aleph\left(\boldsymbol{0}, \boldsymbol{G} = \boldsymbol{A}\sigma_u^2\right)$ and $\boldsymbol{K}'\boldsymbol{y} \mid \boldsymbol{u} \sim \aleph\left(\boldsymbol{0}, \boldsymbol{K}'\boldsymbol{R}\boldsymbol{K}\right)$ can write

$$\begin{aligned}&f\left(\boldsymbol{u}, \boldsymbol{K}'\boldsymbol{y}\right) = f\left(\boldsymbol{K}'\boldsymbol{y} \mid \boldsymbol{u}\right) f\left(\boldsymbol{u}\right) \\ &\approx \frac{\exp\left\{-\frac{1}{2}\left(\boldsymbol{y}-\boldsymbol{Z}\boldsymbol{u}\right)'\boldsymbol{K}\left(\boldsymbol{K}'\boldsymbol{R}\boldsymbol{K}\right)^{-1}\boldsymbol{K}'\left(\boldsymbol{y}-\boldsymbol{Z}\boldsymbol{u}\right)\right\}}{\left|\boldsymbol{K}'\boldsymbol{R}\boldsymbol{K}\right|^{1/2}} \\ &\times \frac{\exp\left\{-\frac{1}{2}\boldsymbol{u}'\boldsymbol{G}^{-1}\boldsymbol{u}\right\}}{\left|\boldsymbol{G}\right|^{1/2}}\end{aligned} \tag{4.178}$$

We use the following notation

$$\boldsymbol{M} = \boldsymbol{K}\left(\boldsymbol{K}'\boldsymbol{R}\boldsymbol{K}\right)^{-1}\boldsymbol{K}' = \boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{R}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{R}^{-1}$$

As a result we can write

$$f\left(\boldsymbol{u}, \boldsymbol{K}'\boldsymbol{y}\right) \approx \frac{\exp\left\{-\frac{1}{2}\left[\boldsymbol{y}'\boldsymbol{M}\boldsymbol{y} - 2\boldsymbol{y}'\boldsymbol{M}\boldsymbol{Z}\boldsymbol{u} + \boldsymbol{u}'\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z}\boldsymbol{u} + \boldsymbol{u}'\boldsymbol{G}^{-1}\boldsymbol{u}\right]\right\}}{\left|\boldsymbol{K}'\boldsymbol{R}\boldsymbol{K}\right|^{1/2}\left|\boldsymbol{G}\right|^{1/2}} \tag{4.179}$$

By completing squares can write

$$\begin{aligned}&f\left(\boldsymbol{u}, \boldsymbol{K}'\boldsymbol{y}\right) \approx \\ &\frac{\exp\left\{-\frac{1}{2}\left[\boldsymbol{y}'\boldsymbol{M}\boldsymbol{y} + \left(\boldsymbol{u}-\widehat{\boldsymbol{u}}\right)'\left[\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z} + \boldsymbol{G}^{-1}\right]\left(\boldsymbol{u}-\widehat{\boldsymbol{u}}\right) - \widehat{\boldsymbol{u}}'\left[\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z} + \boldsymbol{G}^{-1}\right]\widehat{\boldsymbol{u}}\right]\right\}}{\left|\boldsymbol{K}'\boldsymbol{R}\boldsymbol{K}\right|^{1/2}\left|\boldsymbol{G}\right|^{1/2}}\end{aligned} \tag{4.180}$$

where

$$\widehat{\boldsymbol{u}} = \left[\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z} + \boldsymbol{G}^{-1}\right]^{-1}\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{y}$$

Now we can compute the marginal distribution

$$
\begin{aligned}
f\left(\boldsymbol{K}'\boldsymbol{y}\right) &= \int f\left(\boldsymbol{u}, \boldsymbol{K}'\boldsymbol{y}\right) d\boldsymbol{u} \\
&\approx \left|\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z} + \boldsymbol{G}^{-1}\right|^{1/2} \left|\boldsymbol{K}'\boldsymbol{R}\boldsymbol{K}\right|^{1/2} \left|\boldsymbol{G}\right|^{1/2} \\
&\times \exp\left\{\frac{1}{2}\left[\boldsymbol{y}'\boldsymbol{M}\boldsymbol{y} - \widehat{\boldsymbol{u}}'\left[\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z} + \boldsymbol{G}^{-1}\right]\widehat{\boldsymbol{u}}\right]\right\}
\end{aligned}
\tag{4.181}
$$

Now consider the following result

Proposition 4.2.3 *For $\boldsymbol{K}^{'(n-r)\times n}$ and $X^{n\times p}$ can show that*

$$
\frac{\left|\boldsymbol{K}'\boldsymbol{R}\boldsymbol{K}\right|}{\left|\boldsymbol{K}'\boldsymbol{K}\right|} = \frac{\left|\boldsymbol{R}\right|\left|\boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X}\right|}{\left|\boldsymbol{X}'\boldsymbol{X}\right|}
\tag{4.182}
$$

Proof: Consider the following equality

$$
\begin{bmatrix}\boldsymbol{K}' \\ \boldsymbol{X}'\end{bmatrix} \boldsymbol{R}\begin{bmatrix}\boldsymbol{K} & \boldsymbol{X}\end{bmatrix} = \begin{bmatrix}\boldsymbol{K}'\boldsymbol{R}\boldsymbol{K} & \boldsymbol{K}'\boldsymbol{R}\boldsymbol{X} \\ \boldsymbol{X}'\boldsymbol{R}\boldsymbol{K} & \boldsymbol{X}'\boldsymbol{R}\boldsymbol{X}\end{bmatrix}
\tag{4.183}
$$

Now can write in terms of the determinants of both sides

$$
\left|\boldsymbol{R}\right|\begin{vmatrix}\boldsymbol{K}'\boldsymbol{K} & \boldsymbol{K}'\boldsymbol{X} \\ \boldsymbol{X}'\boldsymbol{K} & \boldsymbol{X}'\boldsymbol{X}\end{vmatrix} = \left|\boldsymbol{K}'\boldsymbol{R}\boldsymbol{K}\right|\left|\boldsymbol{X}'\boldsymbol{R}\boldsymbol{X} - \boldsymbol{X}'\boldsymbol{R}\boldsymbol{K}\left(\boldsymbol{K}'\boldsymbol{R}\boldsymbol{K}\right)^{-1}\boldsymbol{K}'\boldsymbol{R}\boldsymbol{X}\right|
\tag{4.184}
$$

But $\boldsymbol{K}'\boldsymbol{X} = \boldsymbol{0}$ and so,

$$
\left|\boldsymbol{R}\right|\begin{vmatrix}\boldsymbol{K}'\boldsymbol{K} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{X}'\boldsymbol{X}\end{vmatrix} = \left|\boldsymbol{K}'\boldsymbol{R}\boldsymbol{K}\right|\left|\boldsymbol{X}'\boldsymbol{R}\left(\boldsymbol{R}^{-1} - \boldsymbol{K}\left(\boldsymbol{K}'\boldsymbol{R}\boldsymbol{K}\right)^{-1}\boldsymbol{K}'\right)\boldsymbol{R}\boldsymbol{X}\right|
\tag{4.185}
$$

Now can write

$$
\begin{aligned}
\left|\boldsymbol{R}\right|\left|\boldsymbol{K}'\boldsymbol{K}\right|\left|\boldsymbol{X}'\boldsymbol{X}\right| &= \left|\boldsymbol{K}'\boldsymbol{R}\boldsymbol{K}\right|\left|\boldsymbol{X}'\boldsymbol{R}\left(\boldsymbol{R}^{-1} - \boldsymbol{M}\right)\boldsymbol{R}\boldsymbol{X}\right| \\
&= \left|\boldsymbol{K}'\boldsymbol{R}\boldsymbol{K}\right|\left|\boldsymbol{X}'\boldsymbol{R}\left(\boldsymbol{R}^{-1} - \boldsymbol{R}^{-1} + \boldsymbol{R}^{-1}\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{R}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{R}^{-1}\right)\boldsymbol{R}\boldsymbol{X}\right| \\
&= \left|\boldsymbol{K}'\boldsymbol{R}\boldsymbol{K}\right|\left|\boldsymbol{X}'\underbrace{\boldsymbol{R}\boldsymbol{R}^{-1}}_{\boldsymbol{I}}\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{R}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\underbrace{\boldsymbol{R}^{-1}\boldsymbol{R}}_{\boldsymbol{I}}\boldsymbol{X}\right| \\
&= \left|\boldsymbol{K}'\boldsymbol{R}\boldsymbol{K}\right|\left|\boldsymbol{X}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{R}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{X}\right| \\
&= \left|\boldsymbol{K}'\boldsymbol{R}\boldsymbol{K}\right|\frac{\left|\boldsymbol{X}'\boldsymbol{X}\right|^{2}}{\left|\boldsymbol{X}'\boldsymbol{R}\boldsymbol{X}\right|}
\end{aligned}
\tag{4.186}
$$

Finally can conclude that

$$
\frac{\left|\boldsymbol{K}'\boldsymbol{R}\boldsymbol{K}\right|}{\left|\boldsymbol{K}'\boldsymbol{K}\right|} = \frac{\left|\boldsymbol{R}\right|\left|\boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X}\right|}{\left|\boldsymbol{X}'\boldsymbol{X}\right|}
\tag{4.187}
$$

Note that $\left|\boldsymbol{K}'\boldsymbol{K}\right|$ and $\left|\boldsymbol{X}'\boldsymbol{X}\right|$ are constants and as a result we can write

$$\left|\boldsymbol{K}'\boldsymbol{R}\boldsymbol{K}\right| \approx \left|\boldsymbol{R}\right| \left|\boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X}\right| \tag{4.188}$$

Now multiplying both sides by $\left|\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z}+\boldsymbol{G}^{-1}\right|$ can write

$$\left|\boldsymbol{K}'\boldsymbol{R}\boldsymbol{K}\right| \left|\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z}+\boldsymbol{G}^{-1}\right| \approx \left|\boldsymbol{R}\right| \left|\boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X}\right| \left|\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z}+\boldsymbol{G}^{-1}\right| \tag{4.189}$$

But the determinant of $\boldsymbol{C}$, the mixed model equation matrix, can be written as

$$\left|\boldsymbol{C}\right| = \left|\boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X}\right| \left|\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z}+\boldsymbol{G}^{-1}\right| \tag{4.190}$$

As a result can conclude that

$$\left|\boldsymbol{K}'\boldsymbol{R}\boldsymbol{K}\right| \left|\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z}+\boldsymbol{G}^{-1}\right| \approx \left|\boldsymbol{R}\right| \left|\boldsymbol{C}\right| \tag{4.191}$$

Using the result of 4.191 in 4.181 can write

$$\begin{aligned} f\left(\boldsymbol{K}'\boldsymbol{y}\right) \approx & \left|\boldsymbol{R}\right|^{1/2} \left|\boldsymbol{C}\right|^{1/2} \left|\boldsymbol{G}\right|^{1/2} \\ & \times \exp\left\{\frac{1}{2}\left[\boldsymbol{y}'\boldsymbol{M}\boldsymbol{y} - \widehat{\boldsymbol{u}}'\left[\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z}+\boldsymbol{G}^{-1}\right]\widehat{\boldsymbol{u}}\right]\right\} \end{aligned} \tag{4.192}$$

It can be also shown by elimination of rows in the following matrix

$$\begin{bmatrix} \boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{y} & \boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{y} & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{y} & \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z} + \frac{\boldsymbol{A}^{-1}}{\sigma_u^2} \end{bmatrix}$$

that

$$\boldsymbol{y}'\boldsymbol{M}\boldsymbol{y} - \widehat{\boldsymbol{u}}'\left[\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z}+\boldsymbol{G}^{-1}\right]\widehat{\boldsymbol{u}} = \boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{y} - \widehat{\boldsymbol{\theta}}\boldsymbol{W}'\boldsymbol{R}^{-1}\boldsymbol{y} \tag{4.193}$$

where

$$\boldsymbol{W} = \begin{bmatrix}\boldsymbol{X} & \boldsymbol{Z}\end{bmatrix} \quad \text{and} \quad \widehat{\boldsymbol{\theta}} = \begin{bmatrix}\widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}\end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z}+\boldsymbol{G}^{-1} \end{bmatrix}^{-1} \boldsymbol{W}'\boldsymbol{y} \tag{4.194}$$

As a result can conclude that

$$\begin{aligned} f\left(\boldsymbol{K}'\boldsymbol{y}\right) \approx & \left|\boldsymbol{R}\right|^{1/2} \left|\boldsymbol{C}\right|^{1/2} \left|\boldsymbol{G}\right|^{1/2} \\ & \times \exp\left\{\frac{1}{2}\left[\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{y} - \widehat{\boldsymbol{\theta}}\boldsymbol{W}'\boldsymbol{R}^{-1}\boldsymbol{y}\right]\right\} \end{aligned} \tag{4.195}$$

and from here obtain the loglikelihood and proceed with REML.

4.2.5 Minimum variance unbiased estimators - MINVAR

Consider

$$\boldsymbol{s}(\boldsymbol{\theta}) = \frac{\partial U}{\partial \boldsymbol{\theta}} = -\frac{1}{2}\underbrace{\{\mathrm{tr}\boldsymbol{P}\boldsymbol{V}_i\}}_{\boldsymbol{s}_1} + \frac{1}{2}\underbrace{\{\boldsymbol{y}'\boldsymbol{P}\boldsymbol{V}_i\boldsymbol{P}\boldsymbol{y}\}}_{\boldsymbol{s}_2} \tag{4.196}$$

where $\boldsymbol{P} = \boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-}\boldsymbol{X}'\boldsymbol{V}^{-1}$ and $\boldsymbol{V}_i = \frac{\partial \boldsymbol{V}}{\partial \boldsymbol{\theta}}$. Note also that $\mathrm{E}(\boldsymbol{y}) = \boldsymbol{0}$. Now consider

$$\begin{aligned}
\mathrm{E}(\boldsymbol{s}_2) &= \{\mathrm{E}\left(\mathrm{tr}\boldsymbol{y}'\boldsymbol{P}\boldsymbol{V}_i\boldsymbol{P}\boldsymbol{y}\right)\} \\
&= \{\mathrm{E}\left(\mathrm{tr}\boldsymbol{P}\boldsymbol{V}_i\boldsymbol{P}\boldsymbol{y}\boldsymbol{y}'\right)\} \quad \text{note} \quad \mathrm{E}\left(\boldsymbol{y}\boldsymbol{y}'\right) = \boldsymbol{V}^* - \mathrm{E}(\boldsymbol{y}')\mathrm{E}(\boldsymbol{y}) \\
&= \{\mathrm{tr}\boldsymbol{P}\boldsymbol{V}_i\boldsymbol{P}\boldsymbol{V}^*\} \quad \text{where} \quad \boldsymbol{V}^* \quad \text{is the true Var-Cov matrix of } \boldsymbol{y} \\
&= \left\{\mathrm{tr}\boldsymbol{P}\boldsymbol{V}_i\boldsymbol{P}\left(\sum_j \boldsymbol{V}_j\boldsymbol{\theta}_j^*\right)\right\} \\
&= \left\{\sum_j \mathrm{tr}\boldsymbol{P}\boldsymbol{V}_i\boldsymbol{P}\boldsymbol{V}_j\boldsymbol{\theta}_j^*\right\} \\
&= \{\mathrm{tr}\boldsymbol{P}\boldsymbol{V}_i\boldsymbol{P}\boldsymbol{V}_j\}\,\boldsymbol{\theta}^* \\
&= \boldsymbol{F}\boldsymbol{\theta}^*
\end{aligned} \tag{4.197}$$

where $\boldsymbol{F}$ is the same matrix as the one used in Functional Iteration. Now can get an unbiased estimate of $\boldsymbol{\theta}^*$ as

$$\begin{aligned}
\widehat{\boldsymbol{\theta}^*} &= \boldsymbol{F}^{-1}\boldsymbol{s}_2(\boldsymbol{\theta}) \\
&= -\boldsymbol{F}^{-1}\boldsymbol{s}_1(\boldsymbol{\theta}) + \boldsymbol{F}^{-1}\left[\boldsymbol{s}_1(\boldsymbol{\theta}) + \boldsymbol{s}_2(\boldsymbol{\theta})\right] \\
&= -\boldsymbol{F}^{-1}\boldsymbol{s}_1(\boldsymbol{\theta}) + \boldsymbol{F}^{-1}\boldsymbol{s}(\boldsymbol{\theta})
\end{aligned} \tag{4.198}$$

Note that by Cramer-Rao lower bound if $\boldsymbol{\theta} = \boldsymbol{\theta}^*$ then $\widehat{\boldsymbol{\theta}^*}$ is the minimum variance unbiased estimator of $\boldsymbol{\theta}^*$.

Chapter 5

Properties of Two-stage Estimators and Predictors

Chapter 6

Appendix

6.1 Some useful results in vector algebra

Definition 6.1.1 *V_n will be used to denote the set of all $n \times 1$ vectors.*

Definition 6.1.2 *Let $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_s\}$ be a set of vectors in V_n. The vector space V spanned by $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_s\}$ is the set of all vectors that are linear combinations of $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_s$ and $\mathbf{0}$.*

Note that $V \subseteq V_n$.

Proposition 6.1.1 *If $\boldsymbol{x}$ and $\boldsymbol{y}$ are in V so is $a\boldsymbol{x} + b\boldsymbol{y}$ for any scalars a and b.*

Proof: Let $\boldsymbol{A} = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_s]$. Then, from definition (6.1.2), x and y can be written as $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{c}_1$ and $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{c}_2$ for some vectors $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$. Then,

$$a\boldsymbol{x} + b\boldsymbol{y} = \boldsymbol{A}(a\boldsymbol{c}_1 + b\boldsymbol{c}_2) = \boldsymbol{A}\boldsymbol{c}_3 \in V \tag{6.1}$$

Definition 6.1.3 *Vectors $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_r\}$ in V_n are linearly dependent if there exists a set of scalars $\{a_1, a_2, \ldots, a_r\}$ not all zero such that $a_1\boldsymbol{\alpha}_1 + a_2\boldsymbol{\alpha}_2 + \ldots + a_r\boldsymbol{\alpha}_r = \mathbf{0}$. Otherwise $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_r\}$ are said to be linearly independent.*

Definition 6.1.4 *A basis for a vector space V is a set of linearly independent vectors that span V.*

Lemma 6.1.1 *Every vector space V has a basis.*

Definition 6.1.5 *The dimension of a vector space V is equal to the number of columns in any basis of V.*

Notation: The notation $V_r \subset V_n$ is used to denote that V_r is a r dimensional vector space in V_n.

Lemma 6.1.2 *Any set of r linearly independent vectors in $V_r \subset V_n$, is a basis for V_r.*

Definition 6.1.6 *The length of a vector $\boldsymbol{X}$ is defined as $\| \boldsymbol{X} \| = (\boldsymbol{X'X})^{1/2}$.*

Note that $\| \boldsymbol{X}+\boldsymbol{Y} \|^2 = (\boldsymbol{X'X}+\boldsymbol{X'Y}+\boldsymbol{Y'X}+\boldsymbol{Y'Y})$.

Definition 6.1.7 *Two vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ are said to be orthogonal, $(\boldsymbol{X} \perp \boldsymbol{Y})$, if and only if $\boldsymbol{X'Y} = 0$.*

Note that if $\boldsymbol{X'Y} = 0$ then $\| \boldsymbol{X}+\boldsymbol{Y} \|^2 = \| \boldsymbol{X} \|^2 + \| \boldsymbol{Y} \|^2$

Lemma 6.1.3 *Orthogonal vectors are linearly independent.*

Proof: Let $\boldsymbol{A}$ be a orthonormal basis. Need to show that if $\boldsymbol{Ax}=\boldsymbol{0}$ then $\boldsymbol{x}=\boldsymbol{0}$. Let

$$\boldsymbol{Ax}=\boldsymbol{0} \tag{6.2}$$

by multipling both sides with the transpose of $\boldsymbol{A}$ we have

$$\boldsymbol{A'Ax}=\boldsymbol{A'0} \tag{6.3}$$

now because the vectors included in A are orthogonal $\boldsymbol{A'A}=\boldsymbol{I}$ and as a result

$$\boldsymbol{Ix}=\boldsymbol{0} \Rightarrow \boldsymbol{x}=\boldsymbol{0} \tag{6.4}$$

Definition 6.1.8 *Any set of r orthogonal nonzero vectors in $V_r \subset V_n$ is a basis for $V_r \subset V_n$.*

Definition 6.1.9 *A basis $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_r\}$ for $V_r \subseteq V_n$ is called orthonormal if the r vectors $\boldsymbol{\alpha}_i$ are pairwise orthogonal and have unit norm.*

Lemma 6.1.4 *Given an arbitrary basis $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_r\}$ for $V_r \subseteq V_n$, there exists an orthogonal basis $\{\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_r\}$ for $V_r \subseteq V_n$, where each $\boldsymbol{\gamma}_i$ is a linear combination of $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_r\}$.*

Such an orthogonal basis can be constructed using the Gram-Schmidt process. Let

$$\begin{aligned} \boldsymbol{\beta_1} &= \boldsymbol{\alpha_1} \\ \boldsymbol{\beta_2} &= \boldsymbol{\alpha_2} - \boldsymbol{c_{21}\beta_1} \end{aligned} \tag{6.5}$$

where we want to chose $\boldsymbol{c_{21}}$ such that

$$\begin{aligned} &\boldsymbol{\beta_1\beta_2} = \boldsymbol{0} \\ &\boldsymbol{\beta_1'}(\boldsymbol{\alpha_2} - \boldsymbol{c_{21}\beta_1}) = \boldsymbol{0} \\ &\boldsymbol{\beta_1'\alpha_2} = \boldsymbol{c_{21}\beta_1'\beta_1} \Rightarrow \boldsymbol{c_{21}} = \frac{\boldsymbol{\beta_1'\alpha_2}}{\boldsymbol{\beta_1'\beta_1}} \end{aligned} \tag{6.6}$$

Then

$$\gamma_{\boldsymbol{i}} = \frac{\boldsymbol{\beta_i}}{\sqrt{\boldsymbol{\beta_i'\beta_i}}} \tag{6.7}$$

Note that $\boldsymbol{\beta_2}$ cannot be $\boldsymbol{0}$. Otherwise $\boldsymbol{\alpha_1}$ and $\boldsymbol{\alpha_2}$ are not linearly independent.

Theorem 6.1.1 *For $V_r \subset V_n$, any vector in V_n can be uniquely written as*

$$\boldsymbol{y} = \widehat{\boldsymbol{y}} + \widehat{\boldsymbol{e}} \tag{6.8}$$

where $\widehat{\boldsymbol{y}} \in V_r$ and $\widehat{\boldsymbol{e}} \perp V_r$.

Proof: Let $\boldsymbol{U}$ be a matrix with $\boldsymbol{r}$ orthogonal column vectors in $V_r \subset V_n$. Let also $\widehat{\boldsymbol{y}} = \boldsymbol{U}\boldsymbol{c} \in V_r$ where $\boldsymbol{c} = \boldsymbol{U'}\boldsymbol{y}$. We can write

$$\widehat{\boldsymbol{e}} = \boldsymbol{y} - \widehat{\boldsymbol{y}} \tag{6.9}$$

and by introducing these values in (6.9) and then multiplying both sides of the resulting equation with $\boldsymbol{U'}$ we obtain

$$\begin{aligned} \boldsymbol{U'}\widehat{\boldsymbol{e}} &= \boldsymbol{U'}[\boldsymbol{y} - \boldsymbol{U}\boldsymbol{c}] \\ &= \boldsymbol{U'}\boldsymbol{y} - \boldsymbol{U'}\boldsymbol{U}\boldsymbol{c} \\ &= \boldsymbol{U'}\boldsymbol{y} - \boldsymbol{c} \\ &= \boldsymbol{0} \end{aligned} \tag{6.10}$$

$\boldsymbol{U'}\widehat{\boldsymbol{e}} = 0 \Rightarrow \widehat{\boldsymbol{e}} \perp V_r \Rightarrow \widehat{\boldsymbol{e}}$ is in the orthocomplement of V_r. We will show now that this representation is unique. Suppose

$$\boldsymbol{y} = \widehat{\boldsymbol{y}^*} + \widehat{\boldsymbol{e}^*} \tag{6.11}$$

where $\widehat{\boldsymbol{y}^*} \in V_r$ and $\widehat{\boldsymbol{e}^*} \perp V_r$. Using (6.11) in (6.8) results in

$$(\widehat{\boldsymbol{y}} - \widehat{\boldsymbol{y}^*}) + (\widehat{\boldsymbol{e}} - \widehat{\boldsymbol{e}^*}) = \boldsymbol{0} \tag{6.12}$$

where $(\widehat{\boldsymbol{y}} - \widehat{\boldsymbol{y}^*}) \in V_r$ and $(\widehat{\boldsymbol{e}} - \widehat{\boldsymbol{e}^*}) \perp V_r$. But (6.12) implies that

$$(\widehat{\boldsymbol{y}} - \widehat{\boldsymbol{y}^*}) = -(\widehat{\boldsymbol{e}} - \widehat{\boldsymbol{e}^*}) \perp V_r \tag{6.13}$$

As a result $(\widehat{\boldsymbol{y}} - \widehat{\boldsymbol{y}^*}) \perp V_r$ but because $(\widehat{\boldsymbol{y}} - \widehat{\boldsymbol{y}^*})$ also $\in V_r$, $(\widehat{\boldsymbol{y}} - \widehat{\boldsymbol{y}^*}) = \boldsymbol{0}$ and consequently $\widehat{\boldsymbol{y}}$ is unique. We conclude that there is a unique decomposition of $\boldsymbol{y}$.

Theorem 6.1.2 *Given V_r and $\boldsymbol{y} \in V_n$, $\| \boldsymbol{y} - \tilde{\boldsymbol{y}} \|^2$ is minimum for $\tilde{\boldsymbol{y}} \in V_r$ at $\tilde{\boldsymbol{y}} = \hat{\boldsymbol{y}}$ the projection of $\boldsymbol{y}$ on V_r.*

Proof: Note that we can write

$$\boldsymbol{y} - \tilde{\boldsymbol{y}} = (\boldsymbol{y} - \widehat{\boldsymbol{y}}) + (\widehat{\boldsymbol{y}} - \tilde{\boldsymbol{y}}) \tag{6.14}$$

as a result

$$\begin{aligned} \| \boldsymbol{y} - \tilde{\boldsymbol{y}} \|^2 &= (\boldsymbol{y} - \widehat{\boldsymbol{y}})'(\boldsymbol{y} - \widehat{\boldsymbol{y}}) + (\boldsymbol{y} - \widehat{\boldsymbol{y}})'(\widehat{\boldsymbol{y}} - \tilde{\boldsymbol{y}}) \\ &+ (\widehat{\boldsymbol{y}} - \tilde{\boldsymbol{y}})'(\boldsymbol{y} - \widehat{\boldsymbol{y}}) + (\widehat{\boldsymbol{y}} - \tilde{\boldsymbol{y}})'(\widehat{\boldsymbol{y}} - \tilde{\boldsymbol{y}}) \end{aligned} \tag{6.15}$$

Note the fact that

$$\begin{aligned} \widehat{\boldsymbol{y}} - \tilde{\boldsymbol{y}} &\in V_r \\ \boldsymbol{y} - \widehat{\boldsymbol{y}} &\perp V_r \end{aligned} \tag{6.16}$$

and as a result $(\boldsymbol{y}-\widehat{\boldsymbol{y}})'(\widehat{\boldsymbol{y}}-\tilde{\boldsymbol{y}}) = 0$ and $(\widehat{\boldsymbol{y}}-\tilde{\boldsymbol{y}})'(\boldsymbol{y}-\widehat{\boldsymbol{y}}) = 0$. In order to minimize (6.15) with respect to $\tilde{\boldsymbol{y}}$ we need to set $\tilde{\boldsymbol{y}} = \widehat{\boldsymbol{y}}$, because the choice of $\tilde{\boldsymbol{y}}$ has no effect on the first term of the relation and as a result it depends only on the last term.

6.2 Some useful results for solving systems of equations

Definition 6.2.1 *The column (row) rank of a matrix is equal to the number of linearly independent columns (rows).*

Proposition 6.2.1 *Let $\boldsymbol{A}$ be a $n \times p$ matrix with rank r. Then the system of linear equations*

$$\boldsymbol{Ax} = \boldsymbol{0} \tag{6.17}$$

has

1. *the unique solution $\boldsymbol{x} = \boldsymbol{0}$ if $r = p$*

2. *if $r < p$, the solutions are spanned by any set of $t = (p - r)$ linearly independent solutions.*

Proof: Let $\{x_1, x_2, \ldots, x_t\}$ be a set of linearly independent solutions of (6.17). Then each $\boldsymbol{x_i}$ is orthogonal to the rows of $\boldsymbol{A}$. So, any linear combination of $\{x_1, x_2, \ldots, x_t\}$ will also be orthogonal to the rows of $\boldsymbol{A}$. Note that t cannot be larger than $t = (p - r)$ otherwise the matrix

$$\begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{x}_1' \\ \vdots \\ \boldsymbol{x}_t' \end{bmatrix}$$

will have row rank $> p$ wich is impossible.

Proposition 6.2.2 *Let $\boldsymbol{A}$ be a $n \times p$ matrix with rank r and $\boldsymbol{b}$ an $n \times 1$ vector. Then the system of linear equations*

$$\boldsymbol{Ax} = \boldsymbol{b} \tag{6.18}$$

may not have a solution if $n > r$.

Proof: For $n > r$, we can write

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A_1} \\ \boldsymbol{L'A_1} \end{bmatrix} \tag{6.19}$$

As a result (6.18) will have a solution only if

$$\boldsymbol{b} = \begin{bmatrix} \boldsymbol{b_1} \\ \boldsymbol{L'b_1} \end{bmatrix} \tag{6.20}$$

Example:

$$\begin{cases} 3\boldsymbol{b_1} + 0\boldsymbol{b_2} + 3\boldsymbol{b_3} & = 6 \\ 0\boldsymbol{b_1} + 2\boldsymbol{b_2} + 2\boldsymbol{b_3} & = 10 \\ 6\boldsymbol{b_1} + 2\boldsymbol{b_2} + 8\boldsymbol{b_3} & = 22 \end{cases} \tag{6.21}$$

Here the third equation can be written as twice the first equation added to the second equation. Otherwise the system is not consistent.

Proposition 6.2.3 *For $\boldsymbol{Ax} = \boldsymbol{b}$ to be consistent $\boldsymbol{A}$ and $[\boldsymbol{A}\ \boldsymbol{b}]$ must have the same rank.*

Note that if we augment $\boldsymbol{A}$ with a linearly independent row $\boldsymbol{w}$, and $\boldsymbol{b}$ with element k the matrix

$$\begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{w} \end{bmatrix}$$

has rank $r + 1$ and the matrix

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{b} \\ \boldsymbol{w} & k \end{bmatrix}$$

also has rank $r + 1$ regardless of the value of k. So, if $\boldsymbol{Ax} = \boldsymbol{b}$ has a solution, so does

$$\begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{w} \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} \boldsymbol{b} \\ k \end{bmatrix} \tag{6.22}$$

for a linearly independent vector $\boldsymbol{w}$ and an arbitrary scalar k.

Proposition 6.2.4 *The set of all solutions of the consistent system $\boldsymbol{Ax} = \boldsymbol{b}$ can be written as $\boldsymbol{x} = \boldsymbol{u} + \boldsymbol{x_0}$ where $\boldsymbol{u}$ is the set of all solutions to $\boldsymbol{Au} = \boldsymbol{0}$ and $\boldsymbol{x_0}$ is a particular solution to $\boldsymbol{Ax} = \boldsymbol{b}$.*

Proof: Let $\boldsymbol{x_i}$ be an arbitrary solution to $\boldsymbol{Ax} = \boldsymbol{b}$ and $\boldsymbol{x_0}$ a particular solution. Then, $\boldsymbol{A(x_i - x_0)} = \boldsymbol{0}$ and $\boldsymbol{(x_i - x_0)}$ is a solution to $\boldsymbol{Au} = \boldsymbol{0}$. Letting $\boldsymbol{u} = \boldsymbol{(x_i - x_0)}$ gives $\boldsymbol{x_i} = \boldsymbol{u} + \boldsymbol{x_0}$. So any solution to $\boldsymbol{Ax} = \boldsymbol{b}$ can be written this way. Note that any $\boldsymbol{x_i} = \boldsymbol{u} + \boldsymbol{x_0}$ is a solution to $\boldsymbol{Ax} = \boldsymbol{b}$

$$\boldsymbol{Ax_i} = \boldsymbol{A(u + x_0)} = \boldsymbol{Au} + \boldsymbol{Ax_0} = \boldsymbol{0} + \boldsymbol{b} = \boldsymbol{b}$$

Definition 6.2.2 *For any* $\boldsymbol{b}$ *that can be written as* $\boldsymbol{Ax} = \boldsymbol{b}$ *a generalized inverse* $\boldsymbol{A}^{-}$ *is such that*

$$\boldsymbol{A}^{-}\boldsymbol{b} = \boldsymbol{x_0}$$

where $\boldsymbol{x_0}$ *is a solution to* $\boldsymbol{Ax} = \boldsymbol{b}$.

Note the fact that $\boldsymbol{Ax_0} = \boldsymbol{b} \Rightarrow \boldsymbol{AA}^{-}\boldsymbol{b} = \boldsymbol{b}$. Because every column of $\boldsymbol{A}$ is spanned by $\boldsymbol{A}, \Rightarrow \boldsymbol{AA}^{-}\boldsymbol{A} = \boldsymbol{A}$. This is a necessary but not sufficient condition for a generalized inverse. In order for $\boldsymbol{AA}^{-}\boldsymbol{A} = \boldsymbol{A}$ to be sufficient we need to show that it implies $\boldsymbol{A}^{-}\boldsymbol{b} = \boldsymbol{x_0}$. In order to show this fact multiply both sides of $\boldsymbol{AA}^{-}\boldsymbol{A} = \boldsymbol{A}$ by $\boldsymbol{x}$. Then we can write

$$\boldsymbol{AA}^{-}\boldsymbol{Ax} = \boldsymbol{Ax}$$

and because $\boldsymbol{Ax} = \boldsymbol{b}$ we obtain

$$\boldsymbol{AA}^{-}\boldsymbol{b} = \boldsymbol{b} \quad \forall\, \boldsymbol{\beta} \in V_r \Rightarrow \boldsymbol{A}^{-}\boldsymbol{b} = \boldsymbol{x_0} \quad \text{is a solution to} \quad \boldsymbol{Ax} = \boldsymbol{b}$$

and as a result we can conclude that $\boldsymbol{AA}^{-}\boldsymbol{A} = \boldsymbol{A}$ is a necessary and sufficient condition to define a generalized inverse. A complete solution to $\boldsymbol{Ax} = \boldsymbol{b}$ is given by

$$\boldsymbol{A}^{-}\boldsymbol{b} + (\boldsymbol{A}^{-}\boldsymbol{A} - \boldsymbol{I})\boldsymbol{c} \quad \text{where} \quad \boldsymbol{A}^{-}\boldsymbol{b} = \boldsymbol{x_0} \quad \text{and} \quad (\boldsymbol{A}^{-}\boldsymbol{A} - \boldsymbol{I})\boldsymbol{c} = \boldsymbol{u}$$

where $\boldsymbol{u}$ is a solution to $\boldsymbol{Au} = \boldsymbol{0}$. Note that

$$\boldsymbol{Au} = \boldsymbol{A}(\boldsymbol{A}^{-}\boldsymbol{A} - \boldsymbol{I})\boldsymbol{c} = (\boldsymbol{AA}^{-}\boldsymbol{A} - \boldsymbol{A})\boldsymbol{c} = (\boldsymbol{A} - \boldsymbol{A})\boldsymbol{c} = \boldsymbol{0}$$

Lemma 6.2.1 *If a* $(n \times r)$ *matrix* $\boldsymbol{X}$ *has rank* r*, then* $(\boldsymbol{X'X})$ *has also rank* r.

Proof: Suppose there is a $\boldsymbol{c'}$ such that $\boldsymbol{c'X'X} = \boldsymbol{u'X} = \boldsymbol{0}$ where $\boldsymbol{u} = \boldsymbol{Xc}$ and consequently $\boldsymbol{u} \in V_r$. But $\boldsymbol{u'X} = \boldsymbol{0}$ implies that $\boldsymbol{u} \perp V_r$ and as a result $\boldsymbol{u} = \boldsymbol{0}$. But because $\boldsymbol{X}$ is a full column rank matrix and $\boldsymbol{u} = \boldsymbol{Xc} = \boldsymbol{0}$ we conclude that $\boldsymbol{c} = \boldsymbol{0}$ and this implies that $\boldsymbol{X'X}$ is also full column rank.

Lemma 6.2.2 *If a* $(n \times p)$ *matrix* $\boldsymbol{X}$ *has rank* $r < p$*, then* $(\boldsymbol{X'X})$ *has also rank* r.

Proof: Because $\boldsymbol{X}$ is not a full rank matrix we can write

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_1\boldsymbol{L} \end{bmatrix} \quad \text{respectively} \quad \boldsymbol{X'} = \begin{bmatrix} \boldsymbol{X'}_1 \\ \boldsymbol{L'X'}_1 \end{bmatrix}$$

where rank($\boldsymbol{X}_1$)$= r$ and $\boldsymbol{X}_1\boldsymbol{L}$ is a linear combination of $\boldsymbol{X}_1$. Then

$$\boldsymbol{X'X} = \begin{bmatrix} \boldsymbol{X'}_1\boldsymbol{X}_1 & \boldsymbol{X'}_1\boldsymbol{X}_1\boldsymbol{L} \\ \boldsymbol{L'X'}_1\boldsymbol{X}_1 & \boldsymbol{X'}_1\boldsymbol{X}_1\boldsymbol{L'L} \end{bmatrix}$$

and due to the fact that $\boldsymbol{X'}_1\boldsymbol{X}_1\boldsymbol{L}$, $\boldsymbol{L'X'}_1\boldsymbol{X}_1$ and $\boldsymbol{X'}_1\boldsymbol{X}_1\boldsymbol{L'L}$ are linear combinations of $\boldsymbol{X'}_1\boldsymbol{X}_1$ we can conclude that the rank of $\boldsymbol{X'X}$ is equal to r.

6.2.1 Iterative Methods for solving equations

Consider the following system of linear equations that are consistent

$$\boldsymbol{Ab} = \boldsymbol{r}$$

the following iterative methods can be used to solve this system.

Iterative Method 1 (Gauss-Seidel)

$$b_i^{[n+1]} = \frac{r_i - \sum_{j=1}^{i-1} a_{ij} b_j^{[n+1]} - \sum_{j=i+1}^{n} a_{ij} b_j^{[n]}}{a_{ii}} \tag{6.23}$$

By using the following equation the convergence of the process is improved.

$$b_i^{[n+1]*} = b_i^{[n+1]} + \alpha(b_i^{[n+1]} - b_i^{[n]}) \tag{6.24}$$

Iterative Method 2 (Jacobi)

$$\boldsymbol{b}^{[n+1]} = \boldsymbol{D}^{-1}(\boldsymbol{r} - \boldsymbol{Ab}^{[n]}) + \boldsymbol{b}^{[n]} \tag{6.25}$$

where $\boldsymbol{D}$ is a diagonal matrix with the diagonal elements of $\boldsymbol{A}$. Again the convergence of the process can be improved by using the following relation.

$$\boldsymbol{b}^{[n+1]} = \boldsymbol{b}^{[n]} + \alpha(\boldsymbol{b}^{[n]} - \boldsymbol{b}^{[n-1]}) + \boldsymbol{D}^{-1}(\boldsymbol{r} - \boldsymbol{Ab}^{[n]}) \tag{6.26}$$

6.3 Derivatives of matrix expressions

6.3.1 Derivative of y'b

Suppose the vector $\boldsymbol{y}$ is not a function of $\boldsymbol{b}$. Note that $\boldsymbol{y'b} = \sum_j^n y_j b_j$. As a result

$$\begin{cases} \frac{\partial \boldsymbol{y'b}}{\partial b_1} & = y_1 \\ \frac{\partial \boldsymbol{y'b}}{\partial b_2} & = y_2 \\ & \vdots \\ \frac{\partial \boldsymbol{y'b}}{\partial b_n} & = y_n \end{cases}$$

In matrix notation, we have

$$\frac{\partial \boldsymbol{y'b}}{\partial \boldsymbol{b}} = \boldsymbol{y'} \tag{6.27}$$

6.3.2 Derivative of Xb

Suppose $\boldsymbol{X}$ is not a function of $\boldsymbol{b}$. Note that $\boldsymbol{Xb} = \sum_j \boldsymbol{x}_j b_j$, where $\boldsymbol{x}_j$ is the j-th column of $\boldsymbol{X}$. As a result

$$\begin{cases} \frac{\partial \boldsymbol{Xb}}{\partial b_1} & = \boldsymbol{x}_1 \\ \frac{\partial \boldsymbol{Xb}}{\partial b_2} & = \boldsymbol{x}_2 \\ & \vdots \\ \frac{\partial \boldsymbol{Xb}}{\partial b_n} & = \boldsymbol{x}_n \end{cases}$$

In matrix notation, we have

$$\frac{\partial \boldsymbol{Xb}}{\partial \boldsymbol{b}} = \boldsymbol{X} \tag{6.28}$$

6.3.3 Derivative of b'Ab

Now consider the derivative of $\boldsymbol{b'Ab} = \sum_i^n \sum_j^n \boldsymbol{a}_{ij} b_i b_j$ where $\boldsymbol{A}$ is not a function of $\boldsymbol{b}$.

$$\begin{cases} \frac{\partial \boldsymbol{X'AX}}{\partial \boldsymbol{x}_1} & = \boldsymbol{a}_{11} 2\boldsymbol{x}_1 + \sum_{i \neq 1}^n \boldsymbol{a}_{i1}\boldsymbol{x}_i + \sum_{j \neq 1}^n \boldsymbol{a}_{1j}\boldsymbol{x}_j = c_1'\boldsymbol{X} + r_1\boldsymbol{X} \\ \frac{\partial \boldsymbol{X'AX}}{\partial \boldsymbol{x}_2} & = \boldsymbol{a}_{22} 2\boldsymbol{x}_2 + \sum_{i \neq 2}^n \boldsymbol{a}_{i2}\boldsymbol{x}_i + \sum_{j \neq 2}^n \boldsymbol{a}_{2j}\boldsymbol{x}_j = c_2'\boldsymbol{X} + r_2\boldsymbol{X} \\ & \vdots \\ \frac{\partial \boldsymbol{X'AX}}{\partial \boldsymbol{x}_n} & = \boldsymbol{a}_{nn} 2\boldsymbol{x}_n + \sum_{i \neq n}^n \boldsymbol{a}_{in}\boldsymbol{x}_i + \sum_{j \neq n}^n \boldsymbol{a}_{nj}\boldsymbol{x}_j = c_n'\boldsymbol{X} + r_n\boldsymbol{X} \end{cases}$$

where c_n is the n-th column of A and r_n is the n-th row of A. As a result if we use matrix notation we can write

$$\frac{\partial \boldsymbol{X'AX}}{\partial \boldsymbol{X}} = (\boldsymbol{A'} + \boldsymbol{A})\boldsymbol{X} \tag{6.29}$$

Note that for

$$\boldsymbol{A'} = \boldsymbol{A} \quad \Rightarrow \quad \frac{\partial \boldsymbol{X'AX}}{\partial \boldsymbol{X}} = 2\boldsymbol{AX} \tag{6.30}$$

6.3.4 Product rule for matrices

Consider the case when matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are functions of x. Consider now

$$
\begin{aligned}
\frac{\partial}{\partial x}\boldsymbol{AB} &= \frac{\partial}{\partial x}\left\{\sum_k a_{ik}b_{kj}\right\} \\
&= \left\{\sum_k \frac{\partial}{\partial x}(a_{ik}b_{kj})\right\} \\
&= \left\{\sum_k \left[\left(\frac{\partial a_{ik}}{\partial x}\right)b_{kj} + a_{ik}\left(\frac{\partial b_{kj}}{\partial x}\right)\right]\right\} \\
&= \left\{\left(\frac{\partial \boldsymbol{a}_i'}{\partial x}\right)\boldsymbol{b}_j + \boldsymbol{a}_i'\left(\frac{\partial \boldsymbol{b}_j}{\partial x}\right)\right\} \\
&= \left(\frac{\partial \boldsymbol{A}}{\partial x}\right)\boldsymbol{B} + \boldsymbol{A}\left(\frac{\partial \boldsymbol{B}}{\partial x}\right)
\end{aligned} \tag{6.31}
$$

Note that when $\boldsymbol{B}$ is not a function of x then

$$
\begin{aligned}
\frac{\partial}{\partial x}\boldsymbol{AB} &= \left\{\frac{\partial}{\partial x}\boldsymbol{a}_i'\boldsymbol{b}_j\right\} \\
&= \left(\frac{\partial \boldsymbol{a}_i'}{\partial x}\right)\boldsymbol{b}_j \\
&= \left(\frac{\partial \boldsymbol{A}}{\partial x}\right)\boldsymbol{B}
\end{aligned} \tag{6.32}
$$

6.3.5 Derivative of a inverse

Consider a non-singular matrix $\boldsymbol{A}$ then $\boldsymbol{AA}^{-1} = \boldsymbol{I}$. Want $\frac{\partial \boldsymbol{A}^{-1}}{\partial x}$ and this can be obtained using the following result

$$
\frac{\partial}{\partial x}\left(\boldsymbol{AA}^{-1}\right) = \frac{\partial}{\partial x}(\boldsymbol{I}) = \boldsymbol{0} \tag{6.33}
$$

Now using the product rule described above can obtain

$$
\begin{aligned}
\frac{\partial}{\partial x}\left(\boldsymbol{AA}^{-1}\right) &= \left(\frac{\partial \boldsymbol{A}}{\partial x}\right)\boldsymbol{A}^{-1} + \boldsymbol{A}\left(\frac{\partial \boldsymbol{A}^{-1}}{\partial x}\right) = \boldsymbol{0} \\
\boldsymbol{A}\left(\frac{\partial \boldsymbol{A}^{-1}}{\partial x}\right) &= -\left(\frac{\partial \boldsymbol{A}}{\partial x}\right)\boldsymbol{A}^{-1} \\
\frac{\partial \boldsymbol{A}^{-1}}{\partial x} &= -\boldsymbol{A}^{-1}\left(\frac{\partial \boldsymbol{A}}{\partial x}\right)\boldsymbol{A}^{-1}
\end{aligned}
$$

6.3.6 Other usefull results regarding derivatives of matrices

Consider the case when we are interested in $\frac{\partial}{\partial x}\left(\boldsymbol{A}'\boldsymbol{B}\boldsymbol{A}\right)$ where $\boldsymbol{B}$ is not a function of x. Then

$$\begin{aligned}
\frac{\partial}{\partial x}\left(\boldsymbol{A}'\boldsymbol{B}\boldsymbol{A}\right) &= \frac{\partial}{\partial x}\left\{\boldsymbol{a}_i'\boldsymbol{B}\boldsymbol{a}_j\right\} \\
&= \frac{\partial}{\partial x}\left\{\mathrm{tr}\boldsymbol{B}\boldsymbol{a}_j\boldsymbol{a}_i'\right\} \\
&= \left\{\mathrm{tr}\boldsymbol{B}\frac{\partial}{\partial x}(\boldsymbol{a}_j\boldsymbol{a}_i')\right\} \\
&= \left\{\mathrm{tr}\boldsymbol{B}\left[(\frac{\partial \boldsymbol{a}_j}{\partial x})\boldsymbol{a}_i' + \boldsymbol{a}_j(\frac{\partial \boldsymbol{a}_i'}{\partial x})\right]\right\} \\
&= \mathrm{tr}\boldsymbol{B}\left(\frac{\partial \boldsymbol{A}}{\partial x}\boldsymbol{A} + \boldsymbol{A}\frac{\partial \boldsymbol{A}}{\partial x}\right) \\
&= \frac{\partial \boldsymbol{A}}{\partial x}\boldsymbol{B}\boldsymbol{A} + \boldsymbol{A}\boldsymbol{B}\frac{\partial \boldsymbol{A}}{\partial x}
\end{aligned} \tag{6.34}$$

Look now at $\frac{\partial}{\partial x}\left(\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\right)$ where $\boldsymbol{B}$ is not a function of x. Then using the previous result can write

$$\begin{aligned}
\frac{\partial}{\partial x}\left(\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\right) &= \left(\frac{\partial \boldsymbol{A}^{-1}}{\partial x}\right)\boldsymbol{B}\boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}\left(\frac{\partial \boldsymbol{A}^{-1}}{\partial x}\right) \\
&= -\left(\boldsymbol{A}^{-1}\left(\frac{\partial \boldsymbol{A}}{\partial x}\right)\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\left(\frac{\partial \boldsymbol{A}}{\partial x}\right)\boldsymbol{A}^{-1}\right)
\end{aligned} \tag{6.35}$$

And finally consider the case when we are interested in $\frac{\partial}{\partial x}\left(\boldsymbol{y}'\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{y}\right)$ where $\boldsymbol{y}$ and $\boldsymbol{B}$ not functions of x. Then

$$\begin{aligned}
&\frac{\partial}{\partial x}\left(\boldsymbol{y}'\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{y}\right) \\
&= \frac{\partial}{\partial x}\mathrm{tr}\left[\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{y}'\boldsymbol{y}\right] \\
&= -\mathrm{tr}\left[\left(\boldsymbol{A}^{-1}\left(\frac{\partial \boldsymbol{A}}{\partial x}\right)\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\left(\frac{\partial \boldsymbol{A}}{\partial x}\right)\boldsymbol{A}^{-1}\right)\boldsymbol{y}\boldsymbol{y}'\right] \\
&= -\boldsymbol{y}'\left(\boldsymbol{A}^{-1}\left(\frac{\partial \boldsymbol{A}}{\partial x}\right)\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\left(\frac{\partial \boldsymbol{A}}{\partial x}\right)\boldsymbol{A}^{-1}\right)\boldsymbol{y}
\end{aligned} \tag{6.36}$$

6.3.7 Derivatives involving determinants

Recall that the determinant of a matrix $\boldsymbol{A}$ is given by

$$|\boldsymbol{A}| = \sum_j a_{ij}A_{ij} \tag{6.37}$$

where A_{ij} is the cofactor of a_{ij}. Recall also that the inverse of a matrix $\boldsymbol{A}$ can be computed as

$$\boldsymbol{A}^{-1} = \frac{1}{|\boldsymbol{A}|}\begin{bmatrix} A_{11} & A_{21} & \dots \\ A_{12} & A_{22} & \dots \\ \vdots & \vdots & \vdots \\ A_{1n} & A_{2n} & \dots \end{bmatrix} \tag{6.38}$$

Now,

$$\begin{aligned} \frac{\partial |\boldsymbol{A}|}{\partial a_{ij}} &= A_{ij} \\ \frac{\partial |\boldsymbol{A}|}{\partial \boldsymbol{A}} &= \begin{bmatrix} A_{11} & A_{12} & \dots \\ A_{21} & A_{22} & \dots \\ \vdots & \vdots & \vdots \\ A_{n1} & A_{n2} & \dots \end{bmatrix} = \left(\boldsymbol{A}^{-1}\right)' |\boldsymbol{A}| \end{aligned} \tag{6.39}$$

and by using the chain rule can write

$$\begin{aligned} \frac{\partial |\boldsymbol{A}|}{\partial x} &= \sum_{ij} \frac{\partial |\boldsymbol{A}|}{\partial a_{ij}} \frac{\partial a_{ij}}{\partial x} \\ &= \text{tr} \frac{\partial |\boldsymbol{A}|}{\partial \boldsymbol{A}} \frac{\partial \boldsymbol{A}'}{\partial x} \\ &= \text{tr}\left[\left(\boldsymbol{A}^{-1}\right)' |\boldsymbol{A}| \frac{\partial \boldsymbol{A}'}{\partial x}\right] \\ &= \text{tr}\left[\left(\boldsymbol{A}^{-1}\right)' \frac{\partial \boldsymbol{A}'}{\partial x} |\boldsymbol{A}|\right] \end{aligned} \tag{6.40}$$

6.3.8 Derivatives of log. determinants

$$\begin{aligned} \frac{\partial \log |\boldsymbol{A}|}{\partial x} &= \frac{1}{|\boldsymbol{A}|} \frac{\partial |\boldsymbol{A}|}{\partial x} \\ &= \frac{1}{|\boldsymbol{A}|} \text{tr}\left[\left(\boldsymbol{A}^{-1}\right)' \frac{\partial \boldsymbol{A}'}{\partial x}\right] |\boldsymbol{A}| \\ &= \text{tr}\left[\left(\boldsymbol{A}^{-1}\right)' \frac{\partial \boldsymbol{A}'}{\partial x}\right] \end{aligned} \tag{6.41}$$

6.4 Expectations and variances of vectors and matrices

Proposition 6.4.1 *Under normality can show that*

$$Var(\boldsymbol{u} \mid \boldsymbol{y}) = Var(\boldsymbol{u} - \widehat{\boldsymbol{u}}) \tag{6.42}$$

where $\widehat{\boldsymbol{u}} = E(\boldsymbol{u} \mid \boldsymbol{y})$

Proof: We know that

$$\mathrm{E}\left(\boldsymbol{u} \mid \boldsymbol{y}\right) = \mathrm{E}(\boldsymbol{u}) + \mathrm{Cov}(\boldsymbol{u}, \boldsymbol{y}')\mathrm{Var}^{-1}(\boldsymbol{y})\left(\boldsymbol{y} - \mathrm{E}(\boldsymbol{y})\right) \tag{6.43}$$

and

$$\mathrm{Var}\left(\boldsymbol{u} \mid \boldsymbol{y}\right) = \mathrm{Var}\left(\boldsymbol{u}\right) - \mathrm{Cov}(\boldsymbol{u}, \boldsymbol{y}')\mathrm{Var}^{-1}(\boldsymbol{y})\mathrm{Cov}(\boldsymbol{y}, \boldsymbol{u}) \tag{6.44}$$

Thus

$$\begin{aligned} \mathrm{Var}\left(\boldsymbol{u} - \widehat{\boldsymbol{u}}\right) &= \mathrm{E}\left[\mathrm{Var}\left(\boldsymbol{u} - \widehat{\boldsymbol{u}}\right) \mid \boldsymbol{y}\right] + \mathrm{Var}\left[\mathrm{E}\left(\boldsymbol{u} - \widehat{\boldsymbol{u}}\right) \mid \boldsymbol{y}\right] \\ &= \mathrm{E}\left[\mathrm{Var}\left(\boldsymbol{u} \mid \boldsymbol{y}\right)\right] + \mathbf{0} \\ &= \mathrm{Var}\left(\boldsymbol{u} \mid \boldsymbol{y}\right) \end{aligned} \tag{6.45}$$

Proposition 6.4.2

$$E\left(\boldsymbol{y}'\boldsymbol{Q}\boldsymbol{y}\right) = tr\boldsymbol{Q}\boldsymbol{V} + \boldsymbol{\mu}'\boldsymbol{Q}\boldsymbol{\mu} \tag{6.46}$$

Proof:

$$\begin{aligned} \mathrm{E}\left(\boldsymbol{y}'\boldsymbol{Q}\boldsymbol{y}\right) &= \mathrm{E}\left[\mathrm{tr}\left(\boldsymbol{y}'\boldsymbol{Q}\boldsymbol{y}\right)\right] \\ &= \mathrm{E}\left[\mathrm{tr}\left(\boldsymbol{Q}\boldsymbol{y}\boldsymbol{y}'\right)\right] \\ &= \mathrm{tr}\left[\mathrm{E}\left(\boldsymbol{Q}\boldsymbol{y}\boldsymbol{y}'\right)\right] \\ &= \mathrm{tr}(\boldsymbol{Q})\mathrm{E}\left(\boldsymbol{y}\boldsymbol{y}'\right) \\ &= \mathrm{tr}(\boldsymbol{Q})\left(\boldsymbol{V} + \boldsymbol{\mu}\boldsymbol{\mu}'\right) \\ &= \mathrm{tr}\boldsymbol{Q}\boldsymbol{V} + \boldsymbol{\mu}'\boldsymbol{Q}\boldsymbol{\mu} \end{aligned} \tag{6.47}$$

6.5 Some usefull results with respect to the χ^2 distribution

Proposition 6.5.1 *The $\boldsymbol{\chi}^2$ distribution can be written as the squared length of a $\boldsymbol{N}(0,1)$ vector.*

In order to show this fact consider a vector $\boldsymbol{u}^{n\times 1} \sim \boldsymbol{N}(\mathbf{0}, \boldsymbol{\Sigma})$. We need to show that $\boldsymbol{u}'\boldsymbol{\Sigma}^{-1}\boldsymbol{u} \sim \boldsymbol{\chi}^2_n$. We can write

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2} \quad \text{and} \quad \boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{-1/2} \quad \text{where} \quad \boldsymbol{\Sigma}^{-1/2} = (\boldsymbol{\Sigma}^{1/2})^{-1}$$

where $\boldsymbol{\Sigma}^{-1/2}$ is a symmetric matrix(?). Now consider $\boldsymbol{z} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{u}$. Then we can write

$$\begin{aligned} &\mathrm{E}(\boldsymbol{z}) = \boldsymbol{\Sigma}^{-1/2}\mathrm{E}(\boldsymbol{u}) = \mathbf{0} \\ &\mathrm{Var}(\boldsymbol{z}) = \boldsymbol{\Sigma}^{-1/2}\mathrm{Var}(\boldsymbol{u})\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{I} \end{aligned}$$

So

$$\begin{aligned} &\boldsymbol{z} \sim \boldsymbol{N}(\mathbf{0}, \boldsymbol{I}) \quad \Rightarrow \quad \boldsymbol{z}'\boldsymbol{z} \sim \boldsymbol{\chi}^2_n \\ &\boldsymbol{z}'\boldsymbol{z} = \boldsymbol{u}'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{u} = \boldsymbol{u}'\boldsymbol{\Sigma}^{-1}\boldsymbol{u} \sim \boldsymbol{\chi}^2_n \end{aligned}$$

Note also that if $\boldsymbol{u} \sim \boldsymbol{N}(\boldsymbol{\eta}, \boldsymbol{\Sigma})$ then $\boldsymbol{u}'\boldsymbol{\Sigma}^{-1}\boldsymbol{u} \sim \boldsymbol{\chi}^2_{n;\boldsymbol{\delta}}$ where $\boldsymbol{\delta} = \boldsymbol{\eta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta}$.

Proposition 6.5.2 *If*

$$\boldsymbol{y}^{n\times 1} = \begin{bmatrix} \boldsymbol{y}_1^{p\times 1} \\ \\ \boldsymbol{y}_2^{(n-p)\times 1} \end{bmatrix} \sim \aleph\left(\boldsymbol{0}, \begin{bmatrix} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{21} & \boldsymbol{V}_{22} \end{bmatrix}\right) \tag{6.48}$$

then,

$$\begin{aligned} \boldsymbol{y}'\boldsymbol{V}^{-1}\boldsymbol{y} &\sim \chi_n^2 \\ \\ \boldsymbol{y}_1'\boldsymbol{V}_{11}^{-1}\boldsymbol{y}_1 &\sim \chi_p^2 \end{aligned} \tag{6.49}$$

and,

$$(\boldsymbol{y}'\boldsymbol{V}^{-1}\boldsymbol{y} - \boldsymbol{y}_1'\boldsymbol{V}_{11}^{-1}\boldsymbol{y}_1) \sim \chi^2_{(n-p)} \tag{6.50}$$

Proof: Let $\boldsymbol{P}^{n\times n}$ be a non-singular matrix. Can write $\boldsymbol{z} = \boldsymbol{P}\boldsymbol{y}$ then, $\mathrm{Var}(\boldsymbol{z}) = \boldsymbol{P}\boldsymbol{V}\boldsymbol{P}'$. Consequently,

$$\begin{aligned} \boldsymbol{z}'(\mathrm{Var}(\boldsymbol{z}))^{-1}\boldsymbol{z} &= \boldsymbol{z}'(\boldsymbol{P}\boldsymbol{V}\boldsymbol{P}')^{-1}\boldsymbol{z} \\ &= \boldsymbol{z}'(\boldsymbol{P}')^{-1}\boldsymbol{V}^{-1} \underbrace{\boldsymbol{P}^{-1}\boldsymbol{z}}_{\boldsymbol{P}^{-1}\boldsymbol{P}\boldsymbol{y}=\boldsymbol{y}} \\ &= \boldsymbol{y}'\boldsymbol{V}^{-1}\boldsymbol{y} \end{aligned} \tag{6.51}$$

So $\boldsymbol{z}'(\mathrm{Var}(\boldsymbol{z}))^{-1}\boldsymbol{z} = \boldsymbol{y}'\boldsymbol{V}^{-1}\boldsymbol{y}$. Now choose $\boldsymbol{P}$ such that

$$\begin{bmatrix} \boldsymbol{z}_1 \\ \boldsymbol{z}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 - \boldsymbol{V}_{21}\boldsymbol{V}_{11}^{-1}\boldsymbol{y}_1 \end{bmatrix} \tag{6.52}$$

To do that let

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ -\boldsymbol{V}_{21}\boldsymbol{V}_{11}^{-1} & \boldsymbol{I} \end{bmatrix} \tag{6.53}$$

and consider

$$\boldsymbol{P}\boldsymbol{y} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ -\boldsymbol{V}_{21}\boldsymbol{V}_{11}^{-1} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 - \boldsymbol{V}_{21}\boldsymbol{V}_{11}^{-1}\boldsymbol{y}_1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{z}_1 \\ \boldsymbol{z}_2 \end{bmatrix} \tag{6.54}$$

Note that $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ are independent. To see that consider

$$\begin{aligned} \mathrm{Cov}(\boldsymbol{z}_1, \boldsymbol{z}_2) &= \mathrm{Cov}(\boldsymbol{y}_1, \boldsymbol{y}_2 - \boldsymbol{V}_{21}\boldsymbol{V}_{11}^{-1}\boldsymbol{y}_1) \\ &= \boldsymbol{V}_{21} - \boldsymbol{V}_{21}\boldsymbol{V}_{11}^{-1}\boldsymbol{V}_{11} \\ &= \boldsymbol{V}_{21} - \boldsymbol{V}_{21} \\ &= \boldsymbol{0} \end{aligned} \tag{6.55}$$

Note also that

$$\mathrm{Var}(\boldsymbol{z}) = \begin{bmatrix} \boldsymbol{V}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \mathrm{Var}(\boldsymbol{z}_2) \end{bmatrix} \tag{6.56}$$

Because $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ are independent can write

$$\begin{aligned} \boldsymbol{z}'(\mathrm{Var}(\boldsymbol{z}))^{-1}\boldsymbol{z} &= \boldsymbol{y}_1'\boldsymbol{V}_{11}^{-1}\boldsymbol{y}_1 + \boldsymbol{z}_2'(\mathrm{Var}(\boldsymbol{z}_2))^{-1}\boldsymbol{z}_2 \\ &= \boldsymbol{y}'\boldsymbol{V}^{-1}\boldsymbol{y} \end{aligned} \tag{6.57}$$

So now

$$\begin{aligned} \boldsymbol{y}'\boldsymbol{V}^{-1}\boldsymbol{y} - \boldsymbol{y}_1'\boldsymbol{V}_{11}^{-1}\boldsymbol{y}_1 &= \boldsymbol{z}_2'(\mathrm{Var}(\boldsymbol{z}_2))^{-1}\boldsymbol{z}_2 \\ &\sim \boldsymbol{\chi}^2_{(n-p)} \end{aligned} \tag{6.58}$$