Ch6. Multiple Regression

6.1 The model

Multiple Linear Regression Model with p independent variables is defined as

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i \tag{6.1}$$

where $i = 1, ..., n$ (n is the sample size), or in matrix form

$$\boldsymbol{y}_{n\times 1} = \boldsymbol{X}_{n\times r}\boldsymbol{\beta}_{r\times 1} + \boldsymbol{\epsilon}_{n\times 1}$$

where $r = p + 1$.

- What are $\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}$ and $\boldsymbol{\epsilon}$?

- Main assumptions of the model are
  1. $E(\boldsymbol{\epsilon}) = \boldsymbol{0}$;
  2. $Cov(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{I}$, thus ϵ_i or y_i are uncorrelated for $i = 1, \ldots, n$;
  3. $\boldsymbol{X}$ is full column rank (for now).

6.2 Least Squares Estimation (LSE)

6.2.1 LSE

Minimize the sum of squares of deviations of observed and predicted values $(\boldsymbol{y} - \boldsymbol{X}\hat{\beta})'(\boldsymbol{y} - \boldsymbol{X}\hat{\beta})$, we have

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}.$$

Proof.

6.2.1 LSE – residuals and estimator of σ^2

- Vector of residuals $\hat{\epsilon} = \boldsymbol{y} - \boldsymbol{X}\hat{\beta}$:

$$\begin{aligned} \hat{\epsilon} &= \boldsymbol{y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \\ &= [\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y} \\ &= [\boldsymbol{I} - \boldsymbol{H}]\boldsymbol{y} \end{aligned}$$

  where $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$.
- Fitted value: $\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\beta} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{H}\boldsymbol{y}$.
- The unbiased estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{\hat{\epsilon}'\hat{\epsilon}}{n-r}.$$

6.2.1 LSE – residuals and estimator of σ^2

- $\boldsymbol{X}'\hat{\epsilon} = \boldsymbol{0}$ $\quad (\boldsymbol{X}'\boldsymbol{H} = \boldsymbol{X}'$, $\boldsymbol{H}\boldsymbol{X} = \boldsymbol{X}$ and $\boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{H}) = \boldsymbol{0}$, $(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{X} = \boldsymbol{0})$;
- $\boldsymbol{H}$ is symmetric idempotent;
- $\hat{\boldsymbol{y}}'\hat{\boldsymbol{\epsilon}} = 0$;
- $\boldsymbol{I} - \boldsymbol{H}$ is symmetric idempotent;
- $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ $\quad$ (unbiased estimator);
- $\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\sigma^2$;
- $tr(\boldsymbol{I} - \boldsymbol{H}) = n - r$;
- $\hat{\epsilon}'\hat{\epsilon} = tr(\boldsymbol{y}\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{H}))$;
- $E(\boldsymbol{y}\boldsymbol{y}') = \sigma^2\boldsymbol{I} + \boldsymbol{X}\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{X}'$;
- $E\left(\frac{\hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}}}{n-r}\right) = \sigma^2$.

6.2.2 Generalized Least Squares Estimation

Assume that $\mathrm{Cov}(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{V}$, $\boldsymbol{V}$ known

- Estimation of $\boldsymbol{\beta}$
  - Ordinary Least Squares Estimator:
    minimize $(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \Rightarrow \quad \hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$
  - Generalized Least Squares Estimator: minimize
    $(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{a})'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{a})$
    Let

$$
\begin{aligned}
S &= (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{a})'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{a}) \\
&= \boldsymbol{y}'\boldsymbol{V}^{-1}\boldsymbol{y} - 2\boldsymbol{y}'\boldsymbol{V}^{-1}\boldsymbol{X}\boldsymbol{a} + \boldsymbol{a}'\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\boldsymbol{a} \\
\frac{\partial S}{\partial \boldsymbol{a}} &= -2\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\boldsymbol{a} \;=\; 0 \\
\Rightarrow \quad \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y} &= \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\boldsymbol{a} \\
\Rightarrow \quad \tilde{\boldsymbol{a}} &= (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}
\end{aligned}
$$

    Note: If $\boldsymbol{V} = \sigma^2 \boldsymbol{I}$, OLS = GLS
  - Weighted Least Squares Estimator.

6.2.3 Properties of LSE

Gauss-Markov Theorem: The best linear unbiased estimator (b.l.u.e.): Let $\boldsymbol{t}$ be a vector and we need to construct the b.l.u.e. of $\boldsymbol{t}'\beta$.

- Let $\boldsymbol{\lambda}'\boldsymbol{y}$ be a linear function of the observations and an estimator of $\boldsymbol{t}'\beta$.
- If $\boldsymbol{\lambda}'\boldsymbol{y}$ is an unbiased estimator of $\boldsymbol{t}'\beta$, $E(\boldsymbol{\lambda}'\boldsymbol{y}) = \boldsymbol{t}'\beta$.

  But $E(\boldsymbol{\lambda}'\boldsymbol{y}) = \boldsymbol{\lambda}'E(\boldsymbol{y}) = \boldsymbol{\lambda}'\boldsymbol{X}\beta$
  Hence, $\boldsymbol{\lambda}'\boldsymbol{X}\beta = \boldsymbol{t}'\beta$ which is true for all β
  $\Rightarrow \quad \boldsymbol{\lambda}'\boldsymbol{X} = \boldsymbol{t}'$
- Find the linear unbiased estimator of $\boldsymbol{t}'\beta$ which has minimum variance.

$$\mathrm{Var}(\boldsymbol{\lambda}'\boldsymbol{y}) = \sigma^2\boldsymbol{\lambda}'\boldsymbol{V}\boldsymbol{\lambda}$$

6.2.3 Properties of LSE

Gauss-Markov Theorem: $W = \boldsymbol{\lambda}'\boldsymbol{V}\boldsymbol{\lambda}$ is minimum if

$$\boldsymbol{\lambda}' = \boldsymbol{t}'(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}$$

subject to the constraint that $\boldsymbol{X}'\boldsymbol{\lambda} = \boldsymbol{t}$; i.e. GLS is the b.l.u.e.

6.3 Maximum likelihood estimation

A normal model is defined by (6.1) with an additional assumption

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon}_{n\times 1} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}),$$

where σ^2 is unknown.

- The likelihood function is

$$L(\boldsymbol{\beta}, \sigma^2) = (\frac{1}{2\pi\sigma^2})^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})}$$

Take log, we have

$$\begin{aligned} & logL(\boldsymbol{\beta}, \sigma^2) \\ = & -\frac{n}{2}log2\pi - \frac{n}{2}log\sigma^2 - \frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}) \\ = & -\frac{n}{2}log2\pi - \frac{n}{2}log\sigma^2 - \frac{1}{2\sigma^2}(\boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{y}'\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}) \end{aligned}$$

6.3 Maximum likelihood estimation

- MLE for $\boldsymbol{\beta}$ and σ^2

$$
\begin{aligned}
\frac{\partial logL}{\partial \boldsymbol{\beta}} &= \frac{1}{\sigma^2}(\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}) \\
\frac{\partial logL}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})
\end{aligned}
$$

- Put the above 2 equations to zero and we obtain the MLEs:

$$
\begin{aligned}
\text{MLE}(\boldsymbol{\beta}) &= \hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \\
\text{MLE}(\sigma^2) &= \hat{\sigma}^2 = \frac{1}{n}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) \\
&= \frac{1}{n}(\boldsymbol{y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y})'(\boldsymbol{y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}) \\
&= \frac{1}{n}\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{y} \\
&= \frac{1}{n}\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y} = \frac{1}{n}[\boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y}]
\end{aligned}
$$

6.3 Maximum likelihood estimation: properties

- Distribution of $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \sim N(\boldsymbol{\beta}, (\boldsymbol{X}'\boldsymbol{X})^{-1}\sigma^2)$ Since

$$\begin{aligned} \mathrm{E}(\hat{\boldsymbol{\beta}}) &= \mathrm{E}[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}] = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathrm{E}(\boldsymbol{y}) \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{\beta} \\ \mathrm{Cov}(\hat{\boldsymbol{\beta}}) &= \mathrm{Cov}((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\sigma^2. \end{aligned}$$

- Let $SSE = \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y}$:
  - $\hat{\sigma}^2 = \frac{SSE}{n}$ is the MLE
  - $\tilde{\sigma}^2 = \frac{SSE}{n-r(\boldsymbol{X})}$ is an unbiased estimator of σ^2 with $r(\boldsymbol{X}) = \text{rank } of\ \boldsymbol{X}$.
- $\hat{\boldsymbol{\beta}}$ and SSE are independent.
- Distribution of $\hat{\sigma}^2$

$$\frac{[n - r(\boldsymbol{X})]\tilde{\sigma}^2}{\sigma^2} \sim \chi^2_{(n-r(\boldsymbol{X}))}.$$

6.3 Maximum likelihood estimation: Example 6.1

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad i = 1, 2, \ldots, n; \quad \varepsilon_i \overset{ind}{\sim} N(0,\ \sigma^2).$$

In matrix notation, $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{bmatrix}, \quad \boldsymbol{X}'\boldsymbol{y} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{bmatrix}$$

$$\begin{aligned} (\boldsymbol{X}'\boldsymbol{X})^{-1} &= \frac{1}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \begin{bmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{bmatrix} \\ &= \frac{1}{n\sum_{i=1}^n (x_i - \bar{x})^2} \begin{bmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{bmatrix} \end{aligned}$$

6.3 Maximum likelihood estimation: Example 6.1

$$
\begin{aligned}
\hat{\beta} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \\
&= \frac{1}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}\begin{bmatrix} \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i \\ \\ n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i \end{bmatrix} \\
&= \begin{bmatrix} \bar{y} - \hat{\beta}_1 \bar{x} \\ \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \end{bmatrix} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}
\end{aligned}
$$

6.3 Maximum likelihood estimation: Example 6.1

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-2}(\boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y}) \\
&= \frac{1}{n-2}[\sum_{i=1}^{n} y_i^2 - \hat{\beta}_0 \sum_{i=1}^{n} y_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i y_i] \\
&= \frac{1}{n-2}[\sum_{i=1}^{n} y_i^2 - \bar{y} \sum_{i=1}^{n} y_i + \hat{\beta}_1 \bar{x} \sum_{i=1}^{n} y_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i y_i] \\
&= \frac{1}{n-2}[\sum_{i=1}^{n} (y_i - \bar{y})^2 - \hat{\beta}_1 (\sum_{i=1}^{n} x_i y_i - n \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i)] \\
&= \frac{1}{n-2}[\sum_{i=1}^{n} (y_i - \bar{y})^2 - \frac{[\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})]^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}]
\end{aligned}
$$

6.4. The model in centered form

- The centered form of the multiple linear regression model is

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i \\ &= \alpha + b_1(x_{i1} - \bar{x}_1) + b_2(x_{i2} - \bar{x}_2) + \cdots + b_p(x_{ip} - \bar{x}_p) + \epsilon_i \end{aligned}$$

  where $\alpha = \beta_0 + \beta_1\bar{x}_1 + \beta_2\bar{x}_2 + \ldots + \beta_k\bar{x}_p$, and $\bar{x}_j$ is the average of $\{x_{ij}, i = 1, \ldots, n\}$ for $j = 1, 2, \ldots, p$.
- The matrix form can be expressed as

$$\boldsymbol{y}_{n\times 1} = \begin{pmatrix} \mathbf{1} & \boldsymbol{Z} \end{pmatrix} \begin{pmatrix} \alpha \\ \mathbf{b} \end{pmatrix} + \boldsymbol{\epsilon}_{n\times 1},$$

  where $(\boldsymbol{Z} = \boldsymbol{X}_1 - \mathbf{1}\bar{\boldsymbol{X}}')$, $\bar{\boldsymbol{X}}' = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p)$ and $\mathbf{b}' = (b_1, \ldots, b_p)$.
- We can prove that

$$\hat{\alpha} = \hat{\beta}_0 + \hat{\beta}_1\bar{x}_1 + \hat{\beta}_2\bar{x}_2 + \ldots + \hat{\beta}_p\bar{x}_p$$

  and $\hat{\mathbf{b}} = \hat{\boldsymbol{\beta}}_1$.

6.4. The model in centered form

$$\begin{aligned}
\mathrm{Var}(\hat{\boldsymbol{\beta}}) &= \mathrm{Var}\begin{pmatrix} \hat{\beta}_0 \\ \hat{\mathbf{b}} \end{pmatrix} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\sigma^2 \\
&= \begin{bmatrix} \frac{1}{n} + \bar{\boldsymbol{x}}'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\bar{\boldsymbol{x}} & -\bar{\boldsymbol{x}}'(\boldsymbol{Z}'\boldsymbol{Z})^{-1} \\ -(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\bar{\boldsymbol{x}} & (\boldsymbol{Z}'\boldsymbol{Z})^{-1} \end{bmatrix} \sigma^2
\end{aligned}$$

Then

$$\begin{aligned}
\mathrm{Var}(\hat{\boldsymbol{b}}) &= (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\sigma^2 \\
\mathrm{Var}(\hat{\beta}_0) &= \frac{\sigma^2}{n} + \bar{\boldsymbol{x}}'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\bar{\boldsymbol{x}}\sigma^2 \\
&= \frac{\sigma^2}{n} + \bar{\boldsymbol{x}}'\,\mathrm{Var}(\hat{\mathbf{b}})\,\bar{\boldsymbol{x}} \\
\mathrm{Cov}(\hat{\beta}_0, \hat{\mathbf{b}}') &= -\bar{\boldsymbol{x}}'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\sigma^2 = -\bar{\boldsymbol{x}}'\mathrm{Var}(\hat{\mathbf{b}})
\end{aligned}$$

6.5. Partitioning Total Sum of Squares

- SST (Total sum of squares corrected for the mean),

$$\begin{aligned} SST &= \boldsymbol{y}'\boldsymbol{y} - \frac{1}{n}\boldsymbol{y}'\boldsymbol{1}\boldsymbol{1}'\boldsymbol{y} = \boldsymbol{y}'(\boldsymbol{I} - \frac{1}{n}\boldsymbol{1}\boldsymbol{1}')\boldsymbol{y} \\ \frac{SST}{\sigma^2} &\sim \chi^2_{(n-1,\ \frac{\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} - \frac{1}{n}(\boldsymbol{1}'\boldsymbol{X}\boldsymbol{\beta})^2}{2\sigma^2})} \end{aligned}$$

- SSR (Sum of squares of regression) = SST - SSE,

$$\begin{aligned} SSR &= \boldsymbol{y}'(\boldsymbol{I} - \frac{1}{n}\boldsymbol{1}\boldsymbol{1}')\boldsymbol{y} - \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y} \\ &= \boldsymbol{y}'(\boldsymbol{H} - \frac{1}{n}\boldsymbol{1}\boldsymbol{1}')\boldsymbol{y} \\ \frac{SSR}{\sigma^2} &\sim \chi^2_{(k,\ \frac{\mathbf{b}'(\boldsymbol{Z}'\boldsymbol{Z})\mathbf{b}}{2\sigma^2})} \end{aligned}$$

- $R^2 = \frac{SSR}{SST}$.

6.5. Partitioning Total Sum of Squares: Anova Table

Source	df	SS	MS	F-statistics
Regression	$r(\boldsymbol{X})-1$	$\hat{\mathbf{b}}'\boldsymbol{Z}'\boldsymbol{y}$	$\frac{SSR}{r(\boldsymbol{X})-1}$	$F=\frac{MSR}{MSE}$
Error	$n-r(\boldsymbol{X})$	$\boldsymbol{y}'\boldsymbol{y}-\hat{\beta}'\boldsymbol{X}'\boldsymbol{y}$	$\frac{SSE}{n-r(\boldsymbol{X})}$	
Total	$n-1$	$\boldsymbol{y}'(\boldsymbol{I}-\frac{1}{n}\mathbf{1}\mathbf{1}')\boldsymbol{y}$		

- SSR is independent of SSE
- $F \sim F_{(r(\boldsymbol{X})-1,\ n-r(\boldsymbol{X}),\ \frac{\mathbf{b}'(\boldsymbol{Z}'\boldsymbol{Z})\mathbf{b}}{2\sigma^2})}$
- Under $H_0:\ \mathbf{b}=\mathbf{0}$,

$$F \sim F_{(r(\boldsymbol{X})-1,\ n-r(\boldsymbol{X}),\ 0)}.$$

6.6. Model Misspecification – Misspecification of the error structure

- Suppose the true model is

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \mathrm{Cov}(\boldsymbol{y}) = \sigma^2 \boldsymbol{V},$$

- but the working model is

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \mathrm{Cov}(\boldsymbol{y}) = \sigma^2 \boldsymbol{I}.$$

- This will still have an unbiased estimate of $\boldsymbol{\beta}$, but it is not the BLUE.

6.6. Model Misspecification –Misspecification of the mean

We consider the following two models

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}, \quad \mathrm{Cov}(\boldsymbol{y}) = \sigma^2 \boldsymbol{I}, \tag{6.2}$$

and

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}, \quad \mathrm{Cov}(\boldsymbol{y}) = \sigma^2 \boldsymbol{I}. \tag{6.3}$$

- Under-fitting: if the true model is (6.2), but use the model (6.3);
- Over-fitting: if the true model is (6.3), but use the model (6.2).