Ch6. Multiple Regression: Estimation

1 The model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon_i, \quad i = 1, 2, \cdots, n, \tag{1}$$

The assumptions for ϵ_i and y_i are analogous to as those for simple linear regression, namely

1. $E(\epsilon_i) = 0$ for all $i = 1, 2, \cdots, n$, or, equivalently, $E(y_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}$.

2. $var(\epsilon_i) = \sigma^2$ for all $i = 1, 2, \cdots, n$, or, equivalently, $var(y_i) = \sigma^2$.

3. $cov(\epsilon_i, \epsilon_j) = 0$ for all $i \neq j$, or, equivalently, $cov(y_i, y_j) = 0$.

In matrix form, the model can be written as

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ beta_k \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdots \\ \epsilon_n \end{pmatrix}$$

or

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

The assumption on ϵ_i or y_i can be expressed as

1. $E(\boldsymbol{\epsilon}) = 0$ or $E(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{\beta}$.

2. $cov(\boldsymbol{\epsilon}) = \sigma^2 I$ or $cov(\boldsymbol{y}) = \sigma^2 \boldsymbol{I}$.

The matrix is $n \times (k+1)$ and is called the design matrix. In this chapter, we assume that $n > k+1$ and rank($\boldsymbol{X}$)=$k+1$.

2 Estimation of β and σ^2

2.1 Least squares estimator for β

The least squares approach is to seek the estimators of β which minimize the sums of squares if deviations of the n observed y's from their predicted values $\hat{y}$, i.e., minimize

$$\sum_{i=1}^{n} \hat{\epsilon}_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2.$$

Theorem 2.1 *If $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{X}$ is $n \times (k+1)$ of rank $k+1 < n$, then the least squares estimator of $\boldsymbol{\beta}$ is*

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

PROOF: Exercise.

2.2 Properties of the least squares estimator $\hat{\beta}$

Theorem 2.2 *If $E(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{\beta}$, then $\hat{\boldsymbol{\beta}}$ is an unbiased estimator for $\boldsymbol{\beta}$.*

PROOF:

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

Theorem 2.3 *If $cov(\boldsymbol{y}) = \sigma^2\boldsymbol{I}$, the covariance matrix for β is given by $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$.*

PROOF: Exercise.

Theorem 2.4 *(Gauss-Markov Theorem) If $E(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{\beta}$ and $cov(\boldsymbol{y}) = \sigma^2\boldsymbol{I}$, the least squares estimators $\hat{\beta}_j$, $j = 0, 1, \cdots, k$, have minimum variance among all linear unbiased estimators, i.e., the least squares estimators $\hat{\beta}_j$, $j = 0, 1, \cdots, k$ are best linear unbiased estimators (BLUE).*

PROOF: We consider a linear estimator $\boldsymbol{Ay}$ of $\boldsymbol{\beta}$ and seek the matrix $\boldsymbol{A}$ for which $\boldsymbol{Ay}$ is a minimum variance unbiased estimator of $\boldsymbol{\beta}$. Since $\boldsymbol{Ay}$ is to be unbiased for $\boldsymbol{\beta}$, we have

$$E(\boldsymbol{Ay}) = \boldsymbol{A}E(\boldsymbol{y}) = \boldsymbol{AX\beta} = \boldsymbol{\beta},$$

which gives the unbiasedness condition

$$\boldsymbol{AX} = \boldsymbol{I}$$

since the relationship $\boldsymbol{AX\beta} = \boldsymbol{\beta}$ must hold for any positive value of $\boldsymbol{\beta}$.

The covariance matrix for $\boldsymbol{Ay}$ is

$$cov(\boldsymbol{Ay}) = \boldsymbol{A}(\sigma^2\boldsymbol{I})\boldsymbol{A}' = \sigma^2\boldsymbol{AA}'.$$

The variance of the β_j's are on the diagonal of $\sigma^2\boldsymbol{AA}'$, and therefore we need to choose $\boldsymbol{A}$ (subject to
$bA\boldsymbol{X} = \boldsymbol{I}$) so that the diagonal elements of $\boldsymbol{AA}'$ are minimized. Since

$$\begin{aligned}\boldsymbol{AA}' &= [\boldsymbol{A} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'][\boldsymbol{A} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']' \\ &= [\boldsymbol{A} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'][\boldsymbol{A} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']' + (\boldsymbol{X}'\boldsymbol{X})^{-1}.\end{aligned}$$

Note in the last equality, $\boldsymbol{AX} = \boldsymbol{I}$ is used. Since $[\boldsymbol{A} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'][\boldsymbol{A} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']'$ is positive semidefinite, the diagonal elements are great than or equal to zero. These diagonal elements can be made equal to zero by choosing $\boldsymbol{A} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. (This value of $\boldsymbol{A}$ also satisfies the unbiasedness condition $\boldsymbol{AX} = \boldsymbol{I}$). The resulting minimum variance estimator of $\boldsymbol{\beta}$ is

$$\boldsymbol{Ay} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y},$$

which is equal to the least square estimator $\hat{\boldsymbol{\beta}}$.

Remark: The remarkable feature of the Gauss-Markov theorem is its distributional generality. The result holds for any distribution of $\boldsymbol{y}$; normality is not required. The only assumptions used in the proof are $E(\boldsymbol{y}) = \boldsymbol{X\beta}$ and $cov(\boldsymbol{y}) = \sigma^2\boldsymbol{I}$. If these assumptions do not hold, $\hat{\boldsymbol{\beta}}$ may be biased or each $\hat{\beta}_j$ may have a larger variance than that of some other estimator.

Corollary 2.1 *If $E(\boldsymbol{y}) = \boldsymbol{X\beta}$ and $cov(\boldsymbol{y}) = \sigma^2\boldsymbol{I}$, the best linear unbiased estimator of $\boldsymbol{a}'\boldsymbol{\beta}$ is $\boldsymbol{a}'\hat{\boldsymbol{\beta}}$, where $\boldsymbol{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$.*

A fourth property of $\hat{\boldsymbol{\beta}}$ is the following: the predicted value $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k = \hat{\boldsymbol{\beta}}' \boldsymbol{x}$ is invariant to simple linear changes of scale on the x's, where $\boldsymbol{x} = (1, x_1, x_2, \cdots, x_k)'$.

Theorem 2.5 *If $\boldsymbol{x} = (1, x_1, \cdots, x_k)'$ and $\boldsymbol{z} = (1, c_1 x_1, \cdots, c_k x_k)$, then $\hat{y} = \hat{\boldsymbol{\beta}}' \boldsymbol{x} = \hat{\boldsymbol{\beta}}'_z \boldsymbol{z}$, where $\hat{\boldsymbol{\beta}}'_z$ is the least squares estimator from the regression of y on $\boldsymbol{z}$.*

PROOF: We can write $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{D}$, where $\boldsymbol{D} = diag(1, c_1, c_2, \cdots, c_k)$. Substituting $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{D}$ into $\hat{\boldsymbol{\beta}}_z$, we have

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_z &= [(\boldsymbol{X}\boldsymbol{D})'(\boldsymbol{X}\boldsymbol{D})]^{-1}(\boldsymbol{X}\boldsymbol{D})'\boldsymbol{y} \\
&= \boldsymbol{D}^{-1}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \\
&= \boldsymbol{D}^{-1}\hat{\boldsymbol{\beta}}.
\end{aligned}$$

Hence,

$$\hat{\boldsymbol{\beta}}'_z \boldsymbol{z} = (\boldsymbol{D}^{-1}\hat{\boldsymbol{\beta}})'\boldsymbol{D}\boldsymbol{x} = \hat{\boldsymbol{\beta}}'\boldsymbol{x}.$$

2.3 An estimator for σ^2

By assumption 1, $E(y_i) = \boldsymbol{x}_i'\boldsymbol{\beta}$, and by assumption 2, $\sigma^2 = E[y_i - E(y_i)]^2$, we have

$$\sigma^2 = E(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2.$$

Hence, σ^2 can be estimated by

$$\begin{aligned} s^2 &= \frac{1}{n-k-1}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2 \\ &= \frac{1}{n-k-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) \\ &= \frac{SSE}{n-k-1}. \end{aligned}$$

With the denominator $n - k - 1$, s^2 is an unbiased estimator of σ^2.

Theorem 2.6 *If $E(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{\beta}$ and $cov(\boldsymbol{y}) = \sigma^2\boldsymbol{I}$, then*

$$E(s^2) = \sigma^2.$$

PROOF: Exercise.

Corollary 2.2 *An unbiased estimator of $cov(\hat{\boldsymbol{\beta}})$ is given by*

$$\widehat{cov}(\hat{\boldsymbol{\beta}}) = s^2(\boldsymbol{X}'\boldsymbol{X})^{-1}.$$

Theorem 2.7 *If $E(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{\beta}$, $cov(\boldsymbol{y}) = \sigma^2\boldsymbol{I}$, and $E(\epsilon_i^4) = 3\sigma^4$ for the linear model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, then s^2 is the best (minimum variance) quadratic unbiased estimator of σ^2.*

PROOF: See Graybill (1954) or Wang and Chow (1994, pp.161-163).

3 The model in centered form

In matrix form, the centered model for the linear multiple regression becomes

$$\boldsymbol{y} = (\boldsymbol{j}, \boldsymbol{X}_c)\begin{pmatrix}\alpha \\ \boldsymbol{\beta}_1\end{pmatrix} + \boldsymbol{\epsilon},$$

where $\boldsymbol{j}$ is a vector of 1's, $\boldsymbol{\beta}_1 = (\beta_1, \beta_2, \cdots, \beta_k)'$,

$$\boldsymbol{X}_c = (\boldsymbol{I} - \frac{1}{n}\boldsymbol{J})\boldsymbol{X}_1 = \begin{pmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \cdots & x_{1k} - \bar{x}_k \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & \cdots & x_{2k} - \bar{x}_k \\ \vdots & \vdots & & \vdots \\ x_{n1} - \bar{x}_1 & x_{n2} - \bar{x}_2 & \cdots & x_{nk} - \bar{x}_k \end{pmatrix}.$$

The matrix $\boldsymbol{I} - \frac{1}{n}\boldsymbol{J}$ is sometimes called the centering matrix.

The corresponding least squares estimator becomes

$$\begin{aligned} \begin{pmatrix} \hat{\alpha} \\ \hat{\boldsymbol{\beta}}_1 \end{pmatrix} &= [(\boldsymbol{j}, \boldsymbol{X}_c)'(\boldsymbol{j}, \boldsymbol{X}_c)]^{-1}(\boldsymbol{j}, \boldsymbol{X}_c)'\boldsymbol{y} \\ &= \begin{pmatrix} n & 0 \\ 0 & \boldsymbol{X}_c'\boldsymbol{X}_c \end{pmatrix}^{-1} \begin{pmatrix} n\bar{y} & \boldsymbol{X}_c'\boldsymbol{y} \end{pmatrix} \\ &= \begin{pmatrix} \bar{y} \\ (\boldsymbol{X}_c'\boldsymbol{X}_c)^{-1}\boldsymbol{X}_c'\boldsymbol{y} \end{pmatrix}. \end{aligned}$$

4 Normal model

4.1 Assumptions

Normality assumption:

$$\boldsymbol{y} \; is \; N_n(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I}) \; or \; \boldsymbol{\epsilon} \; is \; N_n(0, \sigma^2 \boldsymbol{I}).$$

Under normality, $cov(\boldsymbol{y}) = cov(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{I}$ implies that the y's are independent as well as uncorrelated.

4.2 Maximum likelihood estimators for β and σ^2

Theorem 4.1 *If $\boldsymbol{y}$ is $N_n(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I})$, where $\boldsymbol{X}$ is $n \times (k+1)$ of rank $k+1 < n$, the maximum likelihood estimators of $\boldsymbol{\beta}$ and σ^2 are*

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

$$\hat{\sigma}^2 = \frac{1}{n}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}).$$

PROOF: Exercise.

The maximum likelihood estimator $\hat{\boldsymbol{\beta}}$ is the same as the least squares estimator $\hat{\boldsymbol{\beta}}$. The estimator $\hat{\sigma}^2$ is biased since the denominator is n rather $n-k-1$. We often use the unbiased estimator s^2 to estimate σ^2.

4.3 Properties of $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$

Theorem 4.2 *Suppose $\boldsymbol{y}$ is $N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I})$, where $\boldsymbol{X}$ is $n \times (k+1)$ of rank $k+1 < n$ and $\boldsymbol{\beta} = (\beta_0, \cdots, \beta_k)'$. Then the maximum likelihood estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ have the following distributional properties:*

(i) $\hat{\boldsymbol{\beta}}$ *is* $N_{k+1}(\hat{\boldsymbol{\beta}}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1})$.

(ii) $n\hat{\sigma}^2/\sigma^2$ *is* $\chi^2(n-k-1)$*, or equivalently,* $(n-k-1)s^2/\sigma^2$ *is* $\chi^2(n-k-1)$.

(iii) $\hat{\boldsymbol{\beta}}$ *and* $\hat{\sigma}^2$ *(or* s^2*) are independent.*

PROOF: (i) Since $\boldsymbol{y}$ is normal, $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ is a linear function of $\boldsymbol{y}$, $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ and $cov(\hat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$, $\hat{\boldsymbol{\beta}} \sim N_{k+1}(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1})$.

(ii)

$$n\hat{\sigma}^2/\sigma^2 = \frac{\boldsymbol{y}'}{\sigma}(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})\boldsymbol{X}')\frac{\boldsymbol{y}}{\sigma},$$

and $\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})\boldsymbol{X}'$ is idempotent, hence $n\hat{\sigma}^2/\sigma^2$ is $\chi^2(n-k-1)$.

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

$$n\hat{\sigma}^2 = \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})\boldsymbol{X}')\boldsymbol{y},$$

and that $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})\boldsymbol{X}') = \boldsymbol{O}$, we have $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ (or s^2) are independent.

Theorem 4.3 *If $\boldsymbol{y}$ is $N_n(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$, then $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are jointly sufficient for $\boldsymbol{\beta}$ and σ^2.*

PROOF: Using the Neyman factorization theorem. For details, see Rencher and Schaalje (2008, pp.159-160).

Since $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are jointly sufficient for $\boldsymbol{\beta}$ and σ^2, no other estimators can improve on the information they extract from the sample to estimate $\boldsymbol{\beta}$ and σ^2. Thus, it is not surprising that $\hat{\boldsymbol{\beta}}$ and s^2 are minimum variance unbiased estimators.

Theorem 4.4 *If $\boldsymbol{y}$ is $N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I})$, then $\hat{\boldsymbol{\beta}}$ and s^2 have minimum variance among all unbiased estimators.*

PROOF: See Graybill (1976, P.176) or Christensen (1996, pp.25-27).

5 R^2 in fixed-x regression

The proportion of the total sum of squares due to regression is measured by

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST},$$

where $SST = \sum_{i=1}^n (y_i - \bar{y})^2$, $SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \hat{\boldsymbol{\beta}}' \boldsymbol{X}' \boldsymbol{y} - n\bar{y}^2$, and

$$SST = SSR + SSE,$$

where $SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$.

The R^2 is called the *coefficient of determination* or the *squared multiple correlation*. The positive square root R is called the *multiple correlation coefficient*. If the x's were random, R would estimate a population multiple correlation.

We list some properties of R^2 and R.

1. The range of R^2 is $0 \leq R^2 \leq 1$. If all the $\hat{\beta}_j$'s were zero, except for $\hat{\beta}_0$, R^2 would be zero. (This event has probability zero for continuous data.) If all the y-values fell on the fitted surface, that is, if $y_i = \hat{y}_i$, $i = 1, 2, \cdots, n$, then R^2 would be 1.

2. $R = r_{y\hat{y}}$; that is, the multiple correlation is equal to the simple correlation between the observed y_i's and the fitted $\hat{y}_i$'s.

3. Adding a variable x to the model increases (can not decrease) the value of R^2.

4. If $\beta_1 = \beta_2 = \cdots = \beta_k = 0$, then

$$E(R^2) = \frac{k}{n-1}.$$

   Note that the $\hat{\beta}_j$'s will not be zero when the β_j's are zero.

5. R^2 cannot be partitioned into k components, each of which is uniquely attributable to an x_j, unless the x's are mutually orthogonal, that is, unless $\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)(x_{im} - \bar{x}_m) = 0$ for $j \neq m$.

6. R^2 is invariant to full-rank linear transformations on the x's and to a scale change on y (but not invariant to a joint linear transformation including y and the x's).

Adjusted R^2 (AdjR^2)

$$
\begin{aligned}
R_a^2 = AdjR^2 &= \frac{(R^2 - k/(n-1))(n-1)}{n-k-1} \\
&= \frac{(n-1)R^2 - k}{n-k-1} \\
&= 1 - \frac{SSE/(n-k-1)}{SST/(n-1)}
\end{aligned}
$$

R^2 can also be expressed in terms of sample variances and covariances:

$$\begin{aligned} R^2 &= \frac{\hat{\boldsymbol{\beta}}_1' \boldsymbol{X}_c' \boldsymbol{X}_c \hat{\boldsymbol{\beta}}_1}{\sum_{i=1}^n (y_i - \bar{y})^2} \\ &= \frac{\boldsymbol{s}_{yx}' \boldsymbol{S}_{xx}^{-1} (n-1) \boldsymbol{S}_{xx} \boldsymbol{S}_{xx}^{-1} \boldsymbol{s}_{yx}}{\sum_{i=1}^n (y_i - \bar{y})^2} \\ &= \frac{\boldsymbol{s}_{yx}' \boldsymbol{S}_{xx}^{-1} \boldsymbol{s}_{yx}}{s_y^2} \end{aligned}$$

Note that $\hat{\boldsymbol{\beta}}_1 = (n-1)(\boldsymbol{X}_c' \boldsymbol{X}_c)^{-1} \frac{\boldsymbol{X}_c' \boldsymbol{y}}{n-1} = (\frac{\boldsymbol{X}_c' \boldsymbol{X}_c}{n-1})^{-1} \frac{\boldsymbol{X}_c' \boldsymbol{y}}{n-1} = \boldsymbol{S}_{xx}^{-1} \boldsymbol{s}_{yx}$. This form of R^2 will facilitate a comparison with R^2 for the random-x case.

Geometrically, R is the cosine of the angle θ between $\boldsymbol{y}$ and $\hat{\boldsymbol{y}}$ corrected for their means. The mean of $\hat{\boldsymbol{y}}$ is $\bar{y}$, the same as the mean of $\boldsymbol{y}$. Thus, the centered form of $\boldsymbol{y}$ and $\hat{\boldsymbol{y}}$ are $\boldsymbol{y} - \bar{y}\boldsymbol{j}$ and $\hat{\boldsymbol{y}} - \bar{y}\boldsymbol{j}$.

$$\cos\theta = \frac{(\boldsymbol{y} - \bar{y}\boldsymbol{j})'(\hat{\boldsymbol{y}} - \bar{y}\boldsymbol{j})}{\sqrt{[(\boldsymbol{y} - \bar{y}\boldsymbol{j})'(\boldsymbol{y} - \bar{y}\boldsymbol{j})][(\hat{\boldsymbol{y}} - \bar{y}\boldsymbol{j})'(\hat{\boldsymbol{y}} - \bar{y}\boldsymbol{j})]}}.$$

Note that

$$\begin{aligned}(\boldsymbol{y}-\bar{y}\boldsymbol{j})'(\hat{\boldsymbol{y}}-\bar{y}\boldsymbol{j}) &= [(\hat{\boldsymbol{y}}-\bar{y}\boldsymbol{j})+(\boldsymbol{y}-\hat{\boldsymbol{y}})]'(\hat{\boldsymbol{y}}-\bar{y}\boldsymbol{j}) \\ &= (\hat{\boldsymbol{y}}-\bar{y}\boldsymbol{j})'(\hat{\boldsymbol{y}}-\bar{y}\boldsymbol{j})+(\boldsymbol{y}-\hat{\boldsymbol{y}})'(\hat{\boldsymbol{y}}-\bar{y}\boldsymbol{j}) \\ &= (\hat{\boldsymbol{y}}-\bar{y}\boldsymbol{j})'(\hat{\boldsymbol{y}}-\bar{y}\boldsymbol{j})+0.\end{aligned}$$

Hence,

$$\cos\theta = \frac{\sqrt{(\hat{\boldsymbol{y}}-\bar{y}\boldsymbol{j})'(\hat{\boldsymbol{y}}-\bar{y}\boldsymbol{j})}}{(\boldsymbol{y}-\bar{y}\boldsymbol{j})'(\boldsymbol{y}-\bar{y}\boldsymbol{j})} = \sqrt{\frac{SSR}{SST}} = R.$$

6 Generalized least squares: $cov(\boldsymbol{y}) = \sigma^2\boldsymbol{V}$

The model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\epsilon}, \quad E(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{\beta}, \quad cov(\boldsymbol{y}) = \Sigma = \sigma^2\boldsymbol{V},$$

Theorem 6.1 *Let $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\epsilon}$, $E(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{\beta}$, and $cov(\boldsymbol{y}) = cov(\boldsymbol{\epsilon}) = \sigma^2 V$, where $\boldsymbol{X}$ is a full-rank matrix and $\boldsymbol{V}$ is a known positive definite matrix. For this model, we obtain the following results:*

(i) The best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$ is

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

(ii) The covariance matrix for $\hat{\boldsymbol{\beta}}$ is

$$cov(\boldsymbol{\beta}) = \sigma^2(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}.$$

(iii) An unbiased estimator of σ^2 is

$$s^2 = \frac{(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})}{n-k-1}.$$

PROOF: (i) Since $\boldsymbol{V}$ is positive definite, there exists an $n \times n$ nonsingular matrix $\boldsymbol{P}$ such that $\boldsymbol{V} = \boldsymbol{P}\boldsymbol{P}'$. Multiplying $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ by $\boldsymbol{P}^{-1}$, we obtain

$$\boldsymbol{P}^{-1}\boldsymbol{y} = \boldsymbol{P}^{-1}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{P}^{-1}\boldsymbol{\epsilon},$$

Applying the least square approach to this transformed model, we get

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= [\boldsymbol{X}'(\boldsymbol{P}^{-1})'\boldsymbol{P}^{-1}\boldsymbol{X}]^{-1}\boldsymbol{X}'(\boldsymbol{P}^{-1})'\boldsymbol{P}^{-1}\boldsymbol{y} \\ &= (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}.\end{aligned}$$

Note that since $\boldsymbol{X}$ is full rank, $\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}$ is positive definite. The estimator $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}$ is usually called the **generalized least squares** estimator.

(ii) and (iii) are left as exercises.

Theorem 6.2 *If $\boldsymbol{y}$ is $N_n(\boldsymbol{X\beta}, \sigma^2\boldsymbol{V})$, where $\boldsymbol{X}$ is $n \times (k+1)$ of rank $k+1$ and $\boldsymbol{V}$ is a known positive definite matrix, then the maximum likelihood estimators for $\boldsymbol{\beta}$ and σ^2 are*

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y},$$
$$\hat{\sigma}^2 = \frac{1}{n}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}).$$

PROOF: Exercise.