Homework 6

Zhuodiao Kuang

10/03/2025

4.18. Let Y_{i1} denote a pre-treatment score and Y_{i2} a post-treatment score on the ith individual, $i = 1, \cdots, N$. For $i, l = 1, \cdots, N$ and $j = 1, 2$, let $\mathrm{E}(Y_{ij}) = \tau_j$, $\mathrm{Var}(Y_{ij}) = \sigma^2$ and $\mathrm{Corr}(Y_{i1}, Y_{l2}) = \rho$ if $i = l$, and 0 if $i \neq l$.

(a) Estimate the parameters τ_1 and τ_2 in this linear model.

(b) How will you obtain the estimated standard errors of the estimates in (a)?

(a) Estimate the parameters τ_1 and τ_2

We can express this problem using a general linear model, $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where the errors have a non-identity covariance structure, $E[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T] = \mathbf{V}$. The optimal estimator for $\boldsymbol{\beta}$ is the Generalized Least Squares (GLS) estimator:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}$$

First, let's define the matrices for this specific problem.

- **Response Vector Y:** We stack the observations for each individual, creating a $2N \times 1$ vector.

$$\mathbf{Y} = \begin{pmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \\ \vdots \\ Y_{N1} \\ Y_{N2} \end{pmatrix}$$

- **Parameter Vector $\boldsymbol{\beta}$:** The parameters to be estimated are τ_1 and τ_2.

$$\boldsymbol{\beta} = \begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix}$$

- **Design Matrix X:** The design matrix links the parameters to the expected values, $E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta}$. This results in a $2N \times 2$ matrix.

$$\mathbf{X} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{1}_N \otimes \mathbf{I}_2$$

where $\mathbf{1}_N$ is a column vector of N ones, $\mathbf{I}_2$ is the 2×2 identity matrix, and $\otimes$ denotes the Kronecker product.

- **Covariance Matrix V:** The problem states that measurements on different individuals are uncorrelated, while measurements on the same individual are correlated. This gives $\mathbf{V}$ a block-diagonal structure. The covariance for a single individual i is:

$$\mathbf{\Sigma} = \begin{pmatrix} \mathrm{Var}(Y_{i1}) & \mathrm{Cov}(Y_{i1}, Y_{i2}) \\ \mathrm{Cov}(Y_{i1}, Y_{i2}) & \mathrm{Var}(Y_{i2}) \end{pmatrix} = \begin{pmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{pmatrix} = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

The full $2N \times 2N$ covariance matrix is:

$$\mathbf{V} = \begin{pmatrix} \mathbf{\Sigma} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{\Sigma} \end{pmatrix} = \mathbf{I}_N \otimes \mathbf{\Sigma}$$

Now, we compute the components of the GLS estimator.

1. **Find $\mathbf{V}^{-1}$:** The inverse of a block-diagonal matrix is the block-diagonal matrix of the inverses. We first invert $\mathbf{\Sigma}$:

$$\mathbf{\Sigma}^{-1} = \frac{1}{(\sigma^2)^2(1-\rho^2)} \begin{pmatrix} \sigma^2 & -\rho\sigma^2 \\ -\rho\sigma^2 & \sigma^2 \end{pmatrix} = \frac{1}{\sigma^2(1-\rho^2)} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}$$

Therefore, $\mathbf{V}^{-1} = \mathbf{I}_N \otimes \mathbf{\Sigma}^{-1}$.

2. **Calculate $\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}$:**

$$\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X} = (\mathbf{1}_N^T \otimes \mathbf{I}_2)(\mathbf{I}_N \otimes \mathbf{\Sigma}^{-1})(\mathbf{1}_N \otimes \mathbf{I}_2)$$

Using the Kronecker product property $(A \otimes B)(C \otimes D) = (AC \otimes BD)$, this becomes:

$$\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X} = (\mathbf{1}_N^T\mathbf{I}_N\mathbf{1}_N) \otimes (\mathbf{I}_2\mathbf{\Sigma}^{-1}\mathbf{I}_2) = N\mathbf{\Sigma}^{-1}$$

3. **Calculate $(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}$:**

$$(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1} = (N\mathbf{\Sigma}^{-1})^{-1} = \frac{1}{N}\mathbf{\Sigma}$$

4. **Calculate $\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}$:** Let $\mathbf{Y}_i = [Y_{i1}, Y_{i2}]^T$. Then $\mathbf{Y} = [\mathbf{Y}_1^T, \dots, \mathbf{Y}_N^T]^T$.

$$\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y} = (\mathbf{1}_N^T \otimes \mathbf{I}_2)(\mathbf{I}_N \otimes \mathbf{\Sigma}^{-1})\mathbf{Y} = (\mathbf{1}_N^T \otimes \mathbf{\Sigma}^{-1})\mathbf{Y}$$

This results in summing $\mathbf{\Sigma}^{-1}\mathbf{Y}_i$ over all individuals:

$$\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y} = \sum_{i=1}^{N} \mathbf{\Sigma}^{-1}\mathbf{Y}_i$$

5. **Assemble the Estimator $\hat{\boldsymbol{\beta}}$:**

$$\hat{\boldsymbol{\beta}} = (\frac{1}{N}\mathbf{\Sigma}) \left(\sum_{i=1}^{N} \mathbf{\Sigma}^{-1}\mathbf{Y}_i \right) = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{\Sigma}\mathbf{\Sigma}^{-1})\mathbf{Y}_i = \frac{1}{N}\sum_{i=1}^{N}\mathbf{Y}_i$$

Calculating the sum:

$$\sum_{i=1}^{N}\mathbf{Y}_i = \sum_{i=1}^{N} \begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} = \begin{pmatrix} \sum Y_{i1} \\ \sum Y_{i2} \end{pmatrix} = \begin{pmatrix} N\bar{Y}_1 \\ N\bar{Y}_2 \end{pmatrix}$$

Therefore, the estimator is:

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\tau}_1 \\ \hat{\tau}_2 \end{pmatrix} = \frac{1}{N} \begin{pmatrix} N\bar{Y}_1 \\ N\bar{Y}_2 \end{pmatrix} = \begin{pmatrix} \bar{Y}_1 \\ \bar{Y}_2 \end{pmatrix}$$

The estimated parameters are the sample means:

$$\hat{\tau}_1 = \bar{Y}_1 = \frac{1}{N}\sum_{i=1}^{N} Y_{i1} \quad \text{and} \quad \hat{\tau}_2 = \bar{Y}_2 = \frac{1}{N}\sum_{i=1}^{N} Y_{i2}$$

(b) How will you obtain the estimated standard errors?

The standard errors are the square roots of the diagonal elements of the estimated covariance matrix of $\hat{\boldsymbol{\beta}}$.

1. **Find the Covariance Matrix of $\hat{\boldsymbol{\beta}}$:** The theoretical covariance matrix for the GLS estimator is given by $(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}$. We already calculated this in part (a).

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1} = \frac{1}{N}\boldsymbol{\Sigma} = \frac{\sigma^2}{N}\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

   The variances of our estimators are the diagonal elements:

$$\mathrm{Var}(\hat{\tau}_1) = \frac{\sigma^2}{N} \quad \text{and} \quad \mathrm{Var}(\hat{\tau}_2) = \frac{\sigma^2}{N}$$

2. **Estimate the Unknown Variance σ^2:** The formula for the variances involves the unknown population parameter σ^2. To find the *estimated* standard errors, we must first find an estimate for σ^2 from the data. A reasonable and unbiased estimator for the common variance σ^2 is the pooled sample variance, calculated from the model residuals. The residuals are $\hat{\epsilon}_{i1} = Y_{i1} - \hat{\tau}_1 = Y_{i1} - \bar{Y}_1$ and $\hat{\epsilon}_{i2} = Y_{i2} - \hat{\tau}_2 = Y_{i2} - \bar{Y}_2$. The pooled variance estimator $\hat{\sigma}^2$ is:

$$\hat{\sigma}^2 = \frac{1}{2(N-1)}\left[\sum_{i=1}^{N}(Y_{i1} - \bar{Y}_1)^2 + \sum_{i=1}^{N}(Y_{i2} - \bar{Y}_2)^2\right]$$

3. **Calculate the Estimated Standard Errors:** Finally, we substitute $\hat{\sigma}^2$ into the variance formulas and take the square root. The estimated standard errors for $\hat{\tau}_1$ and $\hat{\tau}_2$ are identical:

$$\mathrm{SE}(\hat{\tau}_1) = \sqrt{\widehat{\mathrm{Var}}(\hat{\tau}_1)} = \sqrt{\frac{\hat{\sigma}^2}{N}}$$

$$\mathrm{SE}(\hat{\tau}_2) = \sqrt{\widehat{\mathrm{Var}}(\hat{\tau}_2)} = \sqrt{\frac{\hat{\sigma}^2}{N}}$$

4.19. Consider the model (4.5.1). For the full-rank case, show that $\widehat{\boldsymbol{\beta}}_{GLS}$ and $\widehat{\boldsymbol{\beta}}$ coincide when

(a) $\mathbf{y} \in \mathcal{C}(\mathbf{X})$, or

(b) $\mathbf{V} = (1-\rho)\mathbf{I} + \rho\mathbf{J}$ and $\mathbf{y}$ is orthogonal to $\mathbf{X}$, the first column of $\mathbf{X}$ being $\mathbf{1}_N$.

The **OLS estimator** ($\hat{\beta}$) and **GLS estimator** ($\hat{\beta}_{GLS}$) are defined as:

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

$$\hat{\beta}_{GLS} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

We'll show they're equal in each case.

(a) When $\mathbf{y} \in C(\mathbf{X})$

This condition means the response vector $\mathbf{y}$ lies perfectly in the **column space** of the design matrix $\mathbf{X}$. In other words, there's no random error; $\mathbf{y}$ is an exact linear combination of the columns of $\mathbf{X}$. This allows us to write:

$$\mathbf{y} = \mathbf{Xc}$$

for some vector $\mathbf{c}$.

1. **OLS Estimator**: Substitute $\mathbf{y} = \mathbf{Xc}$ into the OLS formula:
$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{Xc}) = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})\mathbf{c}$$
Since $\mathbf{X}$ is full rank, $(\mathbf{X}'\mathbf{X})$ is invertible, so $(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X}) = \mathbf{I}$.
$$\hat{\beta} = \mathbf{Ic} = \mathbf{c}$$

2. **GLS Estimator**: Do the same for the GLS formula:
$$\hat{\beta}_{GLS} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}(\mathbf{Xc}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})\mathbf{c}$$
Similarly, $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})$ is invertible.
$$\hat{\beta}_{GLS} = \mathbf{Ic} = \mathbf{c}$$

Since both estimators are equal to the vector $\mathbf{c}$, they coincide.

(b) When $\mathbf{y}$ is orthogonal to $\mathbf{X}$ with a specific covariance structure

Here we have three conditions: * $\mathbf{V} = (1-\rho)\mathbf{I} + \rho\mathbf{J}$ (exchangable structure) * $\mathbf{y}$ is orthogonal to $\mathbf{X}$, meaning $\mathbf{X}'\mathbf{y} = \mathbf{0}$ * The first column of $\mathbf{X}$ is $\mathbf{1}_N$ (the model has an intercept)

1. **OLS Estimator**: The **orthogonality** condition makes this simple.
$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{y}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{0} = \mathbf{0}$$
The OLS estimate is the zero vector. We now show the GLS estimate is also zero.

2. **GLS Estimator**: The GLS estimate $\hat{\beta}_{GLS}$ will be zero if the term $\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ is zero. The inverse of the covariance matrix $\mathbf{V}$ is:
$$\mathbf{V}^{-1} = \frac{1}{1-\rho}\left(\mathbf{I} - \frac{\rho}{1-\rho+N\rho}\mathbf{J}\right)$$
Let's compute $\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$:
$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} = \frac{1}{1-\rho}\left[\mathbf{X}'\mathbf{y} - \frac{\rho}{1-\rho+N\rho}\mathbf{X}'\mathbf{J}\mathbf{y}\right]$$
We know $\mathbf{X}'\mathbf{y} = \mathbf{0}$, which simplifies the expression:
$$= -\frac{\rho}{(1-\rho)(1-\rho+N\rho)}\mathbf{X}'\mathbf{J}\mathbf{y}$$
The term $\mathbf{Jy}$ is a vector where each element is the sum of all elements in $\mathbf{y}$, so $\mathbf{Jy} = (\sum y_i)\mathbf{1}_N$. Because the first column of $\mathbf{X}$ is $\mathbf{1}_N$, the orthogonality condition $\mathbf{X}'\mathbf{y} = \mathbf{0}$ implies that $\mathbf{1}_N'\mathbf{y} = 0$, which means $\sum y_i = 0$. Substituting this result gives:
$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} = -\frac{\rho(0)}{(1-\rho)(1-\rho+N\rho)}\mathbf{X}'\mathbf{1}_N = \mathbf{0}$$
Since $\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} = \mathbf{0}$, the GLS estimator is:
$$\hat{\beta}_{GLS} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}(\mathbf{0}) = \mathbf{0}$$

4.20. Show that $\mathbf{c}'\boldsymbol{\beta}$ is estimable in the model (4.5.1) if and only if it is estimable in the model (4.5.3).

Consider the linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{V}, \tag{4.5.1}$$

where $\mathrm{r}(\mathbf{X}) = r \leq p$ and $\mathbf{V}$ is a known $N \times N$ symmetric, p.d. matrix. The other assumptions of the general linear model still hold. We show that the least squares procedure produces optimal estimates (in the sense of the Gauss–Markov theorem). Since $\mathbf{V}$ is p.d., from Result 2.4.5, we have that $\mathbf{K} = \mathbf{V}^{1/2}$ is a symmetric and n.n.d. matrix. Let $\mathbf{L} = \mathbf{K}^{-1}$. Then $\mathbf{V}^{-1} = \mathbf{L}'\mathbf{L}$. Let

$$\mathbf{z} = \mathbf{L}\mathbf{y}, \quad \mathbf{B} = \mathbf{L}\mathbf{X}, \quad \text{and} \quad \boldsymbol{\eta} = \mathbf{L}\boldsymbol{\varepsilon}. \tag{4.5.2}$$

Since $\mathrm{r}(\mathbf{X}) = r$ and $\mathbf{L}$ is nonsingular, it follows from property 5 of Result 1.3.11 that $\mathrm{r}(\mathbf{B}) = r$. It may also be verified that

$$\mathrm{E}(\boldsymbol{\eta}) = \mathbf{0}, \quad \text{and} \quad \mathrm{Var}(\boldsymbol{\eta}) = \mathbf{L}(\sigma^2\mathbf{V})\mathbf{L}' = \sigma^2\mathbf{L}\mathbf{K}\mathbf{K}'\mathbf{L}' = \sigma^2\mathbf{I}_N.$$

Consider the "transformed" linear model

$$\mathbf{z} = \mathbf{B}\boldsymbol{\beta} + \boldsymbol{\eta}, \quad \mathrm{Var}(\boldsymbol{\eta}) = \sigma^2\mathbf{I}_N, \tag{4.5.3}$$

Solution to Problem 4.20

A function $\mathbf{c}'\beta$ is estimable in one model if and only if it's estimable in the other because the transformation between them uses an invertible matrix, which preserves the fundamental properties of the design matrix's column space.

1. **Condition for Estimability**: A linear function $\mathbf{c}'\beta$ is **estimable** if and only if the vector $\mathbf{c}$ lies in the column space of the design matrix's transpose.
   - For model (4.5.1), $\mathbf{y} = \mathbf{X}\beta + \epsilon$, the condition is $\mathbf{c} \in C(\mathbf{X}')$.
   - For the transformed model (4.5.3), $\mathbf{z} = \mathbf{B}\beta + \eta$, the condition is $\mathbf{c} \in C(\mathbf{B}')$.
2. **The Proof**: We just need to show that the two column spaces, $C(\mathbf{X}')$ and $C(\mathbf{B}')$, are identical. From the problem definition, the transformed design matrix is $\mathbf{B} = \mathbf{L}\mathbf{X}$, where $\mathbf{L}$ is a **non-singular** (invertible) matrix. The transpose is $\mathbf{B}' = (\mathbf{L}\mathbf{X})' = \mathbf{X}'\mathbf{L}'$.

(The proof relies on a key result from linear algebra: multiplying a matrix by a non-singular matrix does not change its column space. Since $\mathbf{L}'$ is non-singular, we have:

$$C(\mathbf{X}') = C(\mathbf{X}'\mathbf{L}')$$

Because $C(\mathbf{X}') = C(\mathbf{B}')$, the condition for estimability is exactly the same for both models. Therefore, a function $\mathbf{c}'\beta$ is estimable in the general model if and only if it is estimable in the transformed model.)

4.21. In the model (4.5.1), show that

(a) $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})\boldsymbol{\beta}$ is estimable.

(b) $\mathbf{c}'\boldsymbol{\beta}$ is estimable if $\mathbf{c}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^-(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}) = \mathbf{c}'$.

Here is the solution to problem 4.21.

(a) Show that $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})\beta$ is estimable.

A function $\mathbf{L}\beta$ is **estimable** if and only if there's a matrix $\mathbf{A}$ such that $\mathbf{L} = \mathbf{A}\mathbf{X}$.

In this case, we have:

$$\mathbf{L} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$$

We need to find a matrix $\mathbf{A}$ that satisfies the condition. Let's choose:

$$\mathbf{A} = \mathbf{X}'\mathbf{V}^{-1}$$

Now, we can check if this choice of $\mathbf{A}$ works:

$$\mathbf{A}\mathbf{X} = (\mathbf{X}'\mathbf{V}^{-1})\mathbf{X} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$$

This is exactly equal to our $\mathbf{L}$. Since we found a matrix $\mathbf{A}$ that satisfies the condition, the function $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})\beta$ is estimable.

(b) Show that $\mathbf{c}'\beta$ is estimable if $\mathbf{c}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}) = \mathbf{c}'$.

A function $\mathbf{c}'\beta$ is **estimable** if and only if the vector $\mathbf{c}'$ is in the row space of the design matrix $\mathbf{X}$, which we write as $\mathbf{c}' \in R(\mathbf{X})$.

A key theorem in linear models states that the row space of $\mathbf{X}$ is identical to the row space of $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$, because $\mathbf{V}^{-1}$ is a positive definite matrix.

$$R(\mathbf{X}) = R(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})$$

Let's simplify the notation by setting $\mathbf{G} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$. The condition for estimability is therefore equivalent to showing that $\mathbf{c}' \in R(\mathbf{G})$.

The condition given in the problem is:

$$\mathbf{c}'\mathbf{G}^{-}\mathbf{G} = \mathbf{c}'$$

This equation is a fundamental property from linear algebra. It holds if and only if $\mathbf{c}'$ can be written as a linear combination of the rows of $\mathbf{G}$—in other words, if $\mathbf{c}'$ **is in the row space of** $\mathbf{G}$.

Since the given condition implies that $\mathbf{c}' \in R(\mathbf{G})$, and we know that $R(\mathbf{G}) = R(\mathbf{X})$, the condition therefore implies that $\mathbf{c}'\beta$ is estimable.

4.22. Verify Result 4.5.2.

Result 4.5.2. If $\mathbf{c}'\boldsymbol{\beta}$ is an estimable function of $\boldsymbol{\beta}$, then the statistic $\mathbf{c}'\boldsymbol{\beta}^0_{GLS}$ is the unique b.l.u.e. of $\mathbf{c}'\boldsymbol{\beta}$. For two estimable functions $\mathbf{c}_1'\boldsymbol{\beta}$ and $\mathbf{c}_2'\boldsymbol{\beta}$,

$$\text{Cov}(\mathbf{c}_1'\boldsymbol{\beta}^0_{GLS}, \mathbf{c}_2'\boldsymbol{\beta}^0_{GLS}) = \sigma^2\mathbf{c}_1'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{c}_2.$$

The proof mimics the proof of Result 4.4.1 for the transformed model (4.5.3). Its detail is left as Exercise 4.22.

4.5 Generalized least squares

Consider the linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{V}, \tag{4.5.1}$$

where $\mathrm{r}(\mathbf{X}) = r \le p$ and $\mathbf{V}$ is a known $N \times N$ symmetric, p.d. matrix. The other assumptions of the general linear model still hold. We show that the least squares procedure produces optimal estimates (in the sense of the Gauss–Markov theorem). Since $\mathbf{V}$ is p.d., from Result 2.4.5, we have that $\mathbf{K} = \mathbf{V}^{1/2}$ is a symmetric and n.n.d. matrix. Let $\mathbf{L} = \mathbf{K}^{-1}$. Then $\mathbf{V}^{-1} = \mathbf{L}'\mathbf{L}$. Let

$$\mathbf{z} = \mathbf{L}\mathbf{y}, \quad \mathbf{B} = \mathbf{L}\mathbf{X}, \quad \text{and} \quad \boldsymbol{\eta} = \mathbf{L}\boldsymbol{\varepsilon}. \tag{4.5.2}$$

Since $\mathrm{r}(\mathbf{X}) = r$ and $\mathbf{L}$ is nonsingular, it follows from property 5 of Result 1.3.11 that $\mathrm{r}(\mathbf{B}) = r$. It may also be verified that

$$\mathrm{E}(\boldsymbol{\eta}) = \mathbf{0}, \quad \text{and} \quad \mathrm{Var}(\boldsymbol{\eta}) = \mathbf{L}(\sigma^2\mathbf{V})\mathbf{L}' = \sigma^2\mathbf{L}\mathbf{K}\mathbf{K}'\mathbf{L}' = \sigma^2\mathbf{I}_N.$$

Consider the "transformed" linear model

$$\mathbf{z} = \mathbf{B}\boldsymbol{\beta} + \boldsymbol{\eta}, \quad \mathrm{Var}(\boldsymbol{\eta}) = \sigma^2\mathbf{I}_N, \tag{4.5.3}$$

To verify Result 4.5.2, we follow the textbook's hint and use the **transformed model (4.5.3)**. This strategy works because the transformed model is a standard OLS model, allowing us to apply the familiar Gauss-Markov theorem and then translate the results back into the terms of the original GLS model.

The transformed model is:

$$\mathbf{z} = \mathbf{B}\beta + \eta, \quad \text{with } \mathrm{Var}(\eta) = \sigma^2\mathbf{I}_N$$

Verifying the b.l.u.e. Property

1. **Start with the OLS Result**: In the transformed model, the standard Gauss-Markov theorem states that if $\mathbf{c}'\beta$ is an estimable function, its **Best Linear Unbiased Estimator (b.l.u.e.)** is $\mathbf{c}'\beta^0_{OLS}$, where β^0_{OLS} is a solution to the OLS normal equations:

$$\beta^0_{OLS} = (\mathbf{B}'\mathbf{B})^-\mathbf{B}'\mathbf{z}$$

2. **Translate the Terms**: Now, we substitute the definitions of $\mathbf{B}$ and $\mathbf{z}$ back to the terms of the original model (4.5.1), using the fact that $\mathbf{V}^{-1} = \mathbf{L}'\mathbf{L}$:

   - $\mathbf{B}'\mathbf{B} = (\mathbf{L}\mathbf{X})'(\mathbf{L}\mathbf{X}) = \mathbf{X}'\mathbf{L}'\mathbf{L}\mathbf{X} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$

- $\mathbf{B}'\mathbf{z} = (\mathbf{LX})'(\mathbf{Ly}) = \mathbf{X}'\mathbf{L}'\mathbf{Ly} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$

3. **Show Equivalence**: Plugging these translated terms back into the formula for β^0_{OLS} gives:

$$\beta^0_{OLS} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{y})$$

   This expression is precisely the definition of the **GLS estimator**, β^0_{GLS}. Since the b.l.u.e. in the transformed model is algebraically identical to the GLS estimator in the original model, the first part of the result is verified.

Verifying the Covariance Formula

1. **Start with the OLS Result**: For the standard OLS model (4.5.3), the covariance between two estimable functions, $\mathbf{c}_1'\beta$ and $\mathbf{c}_2'\beta$, is given by:

$$\mathrm{Cov}(\mathbf{c}_1'\beta^0_{OLS}, \mathbf{c}_2'\beta^0_{OLS}) = \sigma^2\mathbf{c}_1'(\mathbf{B}'\mathbf{B})^{-}\mathbf{c}_2$$

2. **Translate the Terms**: As we established in the previous section:

$$\mathbf{B}'\mathbf{B} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$$

3. **Show Equivalence**: Substituting this into the OLS covariance formula directly yields the GLS covariance formula:

$$\mathrm{Cov}(\mathbf{c}_1'\beta^0_{GLS}, \mathbf{c}_2'\beta^0_{GLS}) = \sigma^2\mathbf{c}_1'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{c}_2$$

This verifies the second part of Result 4.5.2. The proof demonstrates that GLS is simply OLS applied to a transformed version of the data.

Appendix

4.18

That sign, $\otimes$, represents the **Kronecker product**, an operation performed on any two matrices of arbitrary size. It's a way of creating a larger block matrix from smaller ones.

How It Works

Simply put, you multiply the **entire second matrix** by **each individual element** of the first matrix. The result is a larger matrix composed of these smaller matrix "blocks."

Let's say you have two matrices, $\mathbf{A}$ (size $m \times n$) and $\mathbf{B}$ (size $p \times q$). The Kronecker product $\mathbf{A} \otimes \mathbf{B}$ will be a larger matrix of size $(mp) \times (nq)$.

The formula looks like this:

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2n}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{pmatrix}$$

You take the first element of $\mathbf{A}$ (a_{11}) and multiply it by the entire $\mathbf{B}$ matrix. This becomes the top-left block of your new matrix. Then, you do the same for the second element of $\mathbf{A}$'s first row (a_{12}), and that becomes the next block to the right, and so on.

A Concrete Example

Let's use two simple 2x2 matrices to see it in action.

Let $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 0 & 5 \\ 6 & 7 \end{pmatrix}$.

To find $\mathbf{A} \otimes \mathbf{B}$, we follow the pattern:

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} 1 \cdot \mathbf{B} & 2 \cdot \mathbf{B} \\ 3 \cdot \mathbf{B} & 4 \cdot \mathbf{B} \end{pmatrix}$$

Now, we just substitute matrix $\mathbf{B}$ into that structure:

1. **Top-Left Block:** $1 \cdot \mathbf{B} = 1 \cdot \begin{pmatrix} 0 & 5 \\ 6 & 7 \end{pmatrix} = \begin{pmatrix} 0 & 5 \\ 6 & 7 \end{pmatrix}$
2. **Top-Right Block:** $2 \cdot \mathbf{B} = 2 \cdot \begin{pmatrix} 0 & 5 \\ 6 & 7 \end{pmatrix} = \begin{pmatrix} 0 & 10 \\ 12 & 14 \end{pmatrix}$
3. **Bottom-Left Block:** $3 \cdot \mathbf{B} = 3 \cdot \begin{pmatrix} 0 & 5 \\ 6 & 7 \end{pmatrix} = \begin{pmatrix} 0 & 15 \\ 18 & 21 \end{pmatrix}$
4. **Bottom-Right Block:** $4 \cdot \mathbf{B} = 4 \cdot \begin{pmatrix} 0 & 5 \\ 6 & 7 \end{pmatrix} = \begin{pmatrix} 0 & 20 \\ 24 & 28 \end{pmatrix}$

Finally, we assemble these blocks into one large 4x4 matrix:

$$\mathbf{A} \otimes \mathbf{B} = \left(\begin{array}{cc|cc} 0 & 5 & 0 & 10 \\ 6 & 7 & 12 & 14 \\ \hline 0 & 15 & 0 & 20 \\ 18 & 21 & 24 & 28 \end{array}\right)$$

Key Properties

- **Not Commutative:** Order matters! In general, $\mathbf{A} \otimes \mathbf{B} \neq \mathbf{B} \otimes \mathbf{A}$.
- **Associative:** $(\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C})$.
- **Used in various fields:** The Kronecker product is especially useful in quantum mechanics (for describing systems of multiple particles) and in solving systems of linear matrix equations, like in the statistics problem you referenced.

Of course. That step relies on a key rule of the Kronecker product and then on the definition of block matrix multiplication. Let's break it down.

Part 1: Simplifying the Kronecker Product

The simplification from `(1_N^T $\otimes$ I_2)(I_N $\otimes$ $\Sigma^{-1}$)` to `(1_N^T $\otimes$ $\Sigma^{-1}$)` uses the **mixed-product property** of Kronecker products.

The property states that if the matrix products $\mathbf{AC}$ and $\mathbf{BD}$ are defined, then:

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC} \otimes \mathbf{BD})$$

Let's apply this to our expression: * $\mathbf{A} = \mathbf{1}_N^T$ (a $1 \times N$ row vector) * $\mathbf{B} = \mathbf{I}_2$ (the 2×2 identity matrix) * $\mathbf{C} = \mathbf{I}_N$ (the $N \times N$ identity matrix) * $\mathbf{D} = \mathbf{\Sigma}^{-1}$ (a 2×2 matrix)

Now, we compute the products $\mathbf{AC}$ and $\mathbf{BD}$: * $\mathbf{AC} = \mathbf{1}_N^T \mathbf{I}_N = \mathbf{1}_N^T$ (Multiplying a vector by the identity matrix doesn't change it). * $\mathbf{BD} = \mathbf{I}_2 \mathbf{\Sigma}^{-1} = \mathbf{\Sigma}^{-1}$ (Multiplying a matrix by the identity matrix doesn't change it).

Substituting these back into the property's result, $(\mathbf{AC} \otimes \mathbf{BD})$, gives us:

$$(\mathbf{1}_N^T \otimes \mathbf{\Sigma}^{-1})$$

This confirms the first part of the simplification.

Part 2: From Block Matrix to Summation

Now we need to understand how multiplying this result by $\mathbf{Y}$ leads to a sum.

1. **Expand the Kronecker Product:** Let's see what the matrix $(\mathbf{1}_N^T \otimes \mathbf{\Sigma}^{-1})$ actually looks like.

   - $\mathbf{1}_N^T$ is the row vector $[1, 1, \ldots, 1]$.
   - We multiply each element of $\mathbf{1}_N^T$ by the matrix $\mathbf{\Sigma}^{-1}$:

   $$[1 \cdot \mathbf{\Sigma}^{-1}, 1 \cdot \mathbf{\Sigma}^{-1}, \ldots, 1 \cdot \mathbf{\Sigma}^{-1}] = [\mathbf{\Sigma}^{-1}, \mathbf{\Sigma}^{-1}, \ldots, \mathbf{\Sigma}^{-1}]$$

   This is a large $2 \times 2N$ "row" of matrices.

2. **Perform the Block Matrix Multiplication:** Now we multiply this block matrix by the vector $\mathbf{Y}$, which is a "column" of smaller vectors, $\mathbf{Y}_i$.

   $$[\mathbf{\Sigma}^{-1}, \mathbf{\Sigma}^{-1}, \ldots, \mathbf{\Sigma}^{-1}] \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_N \end{pmatrix}$$

   Just like regular matrix multiplication, this is the first element of the row times the first element of the column, plus the second times the second, and so on.

   $$= (\mathbf{\Sigma}^{-1})(\mathbf{Y}_1) + (\mathbf{\Sigma}^{-1})(\mathbf{Y}_2) + \cdots + (\mathbf{\Sigma}^{-1})(\mathbf{Y}_N)$$

3. **Write as a Summation:** This resulting expression is precisely the definition of a sum:

   $$\sum_{i=1}^{N} \mathbf{\Sigma}^{-1} \mathbf{Y}_i$$

So, the operation effectively applies the transformation $\mathbf{\Sigma}^{-1}$ to each individual's data vector $\mathbf{Y}_i$ and then sums up the results.

4.19 How to calculate V inverse

You can calculate the inverse of the matrix $\mathbf{V} = (1-\rho)\mathbf{I} + \rho\mathbf{J}$ by assuming its inverse has a similar structure and then solving for the unknown coefficients.

The inverse is:

$$\mathbf{V}^{-1} = \frac{1}{1-\rho}\mathbf{I} - \frac{\rho}{(1-\rho)(1+(n-1)\rho)}\mathbf{J}$$

This formula holds provided that $\rho \neq 1$ and $\rho \neq -1/(n-1)$.

Understanding the Matrix Structure First, let's understand what the matrix $\mathbf{V}$ looks like. Here, $\mathbf{I}$ is the $n \times n$ identity matrix and $\mathbf{J}$ is the $n \times n$ matrix of all ones.

$$\mathbf{V} = (1-\rho)\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} + \rho\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}$$

Combining these gives a matrix with **1s on the diagonal** and ρ**s everywhere else**. This is known as a **compound symmetric** or **exchangeable** matrix structure.

$$\mathbf{V} = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}$$

Calculation Method

Because the original matrix $\mathbf{V}$ has a simple structure of the form $a\mathbf{I} + b\mathbf{J}$, it's logical to propose that its inverse, $\mathbf{V}^{-1}$, also has this structure.

Let's assume $\mathbf{V}^{-1} = c\mathbf{I} + d\mathbf{J}$ and find the coefficients c and d.

1. **Set up the identity equation.** By definition, a matrix multiplied by its inverse is the identity matrix: $\mathbf{V}\mathbf{V}^{-1} = \mathbf{I}$.
$$\big((1-\rho)\mathbf{I} + \rho\mathbf{J}\big)(c\mathbf{I} + d\mathbf{J}) = \mathbf{I}$$

2. **Expand the product.** We distribute the terms:
$$c(1-\rho)\mathbf{I} + d(1-\rho)\mathbf{J} + c\rho\mathbf{J} + d\rho\mathbf{J}^2 = \mathbf{I}$$
A key property of the matrix of all ones is that $\mathbf{J}^2 = n\mathbf{J}$, where n is the dimension of the matrix. Substituting this in gives:
$$c(1-\rho)\mathbf{I} + \big(d(1-\rho) + c\rho + dn\rho\big)\mathbf{J} = \mathbf{I}$$

3. **Solve for the coefficients.** The equation must hold for both the $\mathbf{I}$ and $\mathbf{J}$ parts. We can think of the right side as $1 \cdot \mathbf{I} + 0 \cdot \mathbf{J}$. By comparing the coefficients on both sides, we get a system of two equations:
   - **For I:** $c(1-\rho) = 1 \implies c = \frac{1}{1-\rho}$
   - **For J:** $d(1-\rho) + c\rho + dn\rho = 0$

   Now, substitute the value of c we just found into the second equation:
$$d(1-\rho) + \frac{\rho}{1-\rho} + dn\rho = 0$$
   Factor out d:
$$d\big((1-\rho) + n\rho\big) = -\frac{\rho}{1-\rho}$$
$$d\big(1 + (n-1)\rho\big) = -\frac{\rho}{1-\rho}$$
   Finally, solve for d:
$$d = -\frac{\rho}{(1-\rho)(1+(n-1)\rho)}$$

4. **Construct the inverse matrix.** Now that we have c and d, we can write the final form of the inverse:
$$\mathbf{V}^{-1} = c\mathbf{I} + d\mathbf{J} = \frac{1}{1-\rho}\mathbf{I} - \frac{\rho}{(1-\rho)(1+(n-1)\rho)}\mathbf{J}$$
   This method is often simpler than using more general formulas like the Sherman-Morrison formula, which gives the same result.