Modern Multivariate Statistical Techniques

Jinhong Du,15338039

April 14, 2018

Content

1. Suppose the error component $\mathbf{e}$ of the linear regression model has mean $\mathbf{0}$, but now has $Var(\mathbf{e}) = \sigma^2\mathbf{V}$, where $\mathbf{V}$ is a known $(n \times n)$ positive-definite symmetric matrix and $\sigma^2 > 0$ may not be necessarily known. Let $\hat{\boldsymbol{\beta}}_{gls}$ denote the generalized least-squares (GLS) estimator:

$$\hat{\boldsymbol{\beta}}_{gls} = \arg\min_{\beta} (\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta})^\top \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta})$$

Show that

$$\hat{\boldsymbol{\beta}}_{gls} = (\mathbf{Z}^\top \mathbf{V}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^\top \mathbf{V}^{-1}\mathbf{Y}$$

has expectation $\boldsymbol{\beta}$ and covariance matrix $Var(\hat{\boldsymbol{\beta}}_{gls}) = \sigma^2(\mathbf{Z}^\top \mathbf{V}^{-1}\mathbf{Z})^{-1}$.

Proof.

Method 1

Let

$$\begin{aligned} f(\boldsymbol{\beta}) &= (\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta})^\top \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta}) \\ &= \mathbf{Y}^\top \mathbf{V}^{-1}\mathbf{Y} - 2\boldsymbol{\beta}^\top \mathbf{Z}^\top \mathbf{V}^{-1}\mathbf{Y} + \boldsymbol{\beta}^\top \mathbf{Z}^\top \mathbf{V}^{-1}\mathbf{Z}\boldsymbol{\beta} \end{aligned}$$

From Matrix Cookbook, for vecter $\mathbf{x} \in \mathbb{R}^n$, matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$,

$$\begin{aligned} \frac{\mathrm{d}(\mathbf{x}^\top \mathbf{A})}{\mathrm{d}\mathbf{x}} &= \frac{\mathrm{d}(\mathbf{A}^\top \mathbf{x})}{\mathrm{d}\mathbf{x}} \\ &= \mathbf{A} \\ \frac{\mathrm{d}(\mathbf{x}^\top \mathbf{A}\mathbf{x})}{\mathrm{d}\mathbf{x}} &= (\mathbf{A} + \mathbf{A}^\top)\mathbf{x} \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{\beta}} &= -2\mathbf{Z}^\top \mathbf{V}^{-1}\mathbf{Y} + 2\mathbf{Z}^\top \mathbf{V}^{-1}\mathbf{Z}\boldsymbol{\beta} \\ &= 2\mathbf{Z}^\top \mathbf{V}^{-1}(\mathbf{Z}\boldsymbol{\beta} - \mathbf{Y}) \\ \frac{\mathrm{d}^2 f}{\mathrm{d}\boldsymbol{\beta}\mathrm{d}\boldsymbol{\beta}^\top} &= 2\mathbf{Z}^\top \mathbf{V}^{-1}\mathbf{Z} \end{aligned}$$

Since $\mathbf{Z}^\top \mathbf{V}^{-1}\mathbf{Z}$ is positive-definite, by setting $\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{\beta}} = \mathbf{0}$, we get the global minima

$$\hat{\boldsymbol{\beta}}_{gls} = (\mathbf{Z}^\top \mathbf{V}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^\top \mathbf{V}^{-1}\mathbf{Y}$$

Therefore,

$$\begin{aligned} \mathbb{E}\hat{\boldsymbol{\beta}}_{gls} &= \mathbb{E}[(\mathbf{Z}^\top \mathbf{V}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^\top \mathbf{V}^{-1}\mathbf{Y}] \\ &= (\mathbf{Z}^\top \mathbf{V}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^\top \mathbf{V}^{-1}\mathbb{E}(\mathbf{Y}) \end{aligned}$$

$$
\begin{aligned}
&= (\mathbf{Z}^\top\mathbf{V}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{V}^{-1}\mathbb{E}(\mathbf{Z}\boldsymbol{\beta}+\mathbf{e}) \\
&= (\mathbf{Z}^\top\mathbf{V}^{-1}\mathbf{Z})^{-1}(\mathbf{Z}^\top\mathbf{V}^{-1}\mathbf{Z})\boldsymbol{\beta} \\
&= \boldsymbol{\beta} \\
Var\hat{\boldsymbol{\beta}}_{gls} &= Var[(\mathbf{Z}^\top\mathbf{V}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{V}^{-1}\mathbf{Y}] \\
&= (\mathbf{Z}^\top\mathbf{V}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{V}^{-1}Var(\mathbf{Y})\mathbf{V}^{-1}\mathbf{Z}(\mathbf{Z}^\top\mathbf{V}^{-1}\mathbf{Z})^{-1} \\
&= (\mathbf{Z}^\top\mathbf{V}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{V}^{-1}Var(\mathbf{Z}\boldsymbol{\beta}+\mathbf{e})\mathbf{V}^{-1}\mathbf{Z}(\mathbf{Z}^\top\mathbf{V}^{-1}\mathbf{Z})^{-1} \\
&= \sigma^2(\mathbf{Z}^\top\mathbf{V}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{V}^{-1}\mathbf{V}\mathbf{V}^{-1}\mathbf{Z}(\mathbf{Z}^\top\mathbf{V}^{-1}\mathbf{Z})^{-1} \\
&= \sigma^2(\mathbf{Z}^\top\mathbf{V}^{-1}\mathbf{Z})^{-1}(\mathbf{Z}^\top\mathbf{V}^{-1}\mathbf{Z})(\mathbf{Z}^\top\mathbf{V}^{-1}\mathbf{Z})^{-1} \\
&= \sigma^2(\mathbf{Z}^\top\mathbf{V}^{-1}\mathbf{Z})^{-1}
\end{aligned}
$$

Method 2

Since $\mathbf{V}$ is positive-definite, by Cholesky decomposition,

$$\mathbf{V} = \mathbf{C}\mathbf{C}^\top$$

where $\mathbf{C}$ is a lower triangular matrix with real and positive diagonal entries.

Let

$$
\begin{aligned}
\mathbf{Y}^* &= \mathbf{C}^{-1}\mathbf{Y} \\
\mathbf{Z}^* &= \mathbf{C}^{-1}\mathbf{Z} \\
\mathbf{e}^* &= \mathbf{C}^{-1}\mathbf{e}
\end{aligned}
$$

Then

$$
\begin{aligned}
\mathbf{Y}^* &= \mathbf{Z}^*\boldsymbol{\beta} + \mathbf{e} \qquad (1) \\
\mathbf{e} &\sim N(\mathbf{0}, \sigma^2\mathbf{I})
\end{aligned}
$$

Since

$$
\begin{aligned}
(\mathbf{C}^\top)^{-1}\mathbf{C}^\top &= \mathbf{I} \\
(\mathbf{C}^{-1})^\top\mathbf{C}^\top &= (\mathbf{C}\mathbf{C}^{-1})^\top = \mathbf{I}
\end{aligned}
$$

i.e.

$$(\mathbf{C}^\top)^{-1} = (\mathbf{C}^{-1})^\top$$

Therefore,

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{gls} &= \arg\min_{\boldsymbol{\beta}}(\mathbf{Y}-\mathbf{Z}\boldsymbol{\beta})^\top\mathbf{V}^{-1}(\mathbf{Y}-\mathbf{Z}\boldsymbol{\beta}) \\
&= \arg\min_{\boldsymbol{\beta}}(\mathbf{Y}^*-\mathbf{Z}^*\boldsymbol{\beta})^\top(\mathbf{Y}^*-\mathbf{Z}^*\boldsymbol{\beta}) \\
&= \hat{\boldsymbol{\beta}}^*_{ols}
\end{aligned}
$$

$$= (\mathbf{Z}^{*\top}\mathbf{Z}^{*})^{-1}\mathbf{Z}^{*\top}\mathbf{Y}^{*}$$
$$= (\mathbf{Z}^{\top}\mathbf{V}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\top}\mathbf{V}^{-1}\mathbf{Y}$$

where $\hat{\boldsymbol{\beta}}^{*}_{ols}$ is the ordinary least squares estimator of (1)

$\mathbb{E}\hat{\boldsymbol{\beta}}_{gls}$ and $Var\hat{\boldsymbol{\beta}}_{gls}$ are abtained the same as method 1. □

2. What would be the consequences of incorrectly using the ordinary least-squares estimator $\hat{\boldsymbol{\beta}}_{ols} = (\mathbf{Z}^{\top}\mathbf{Z})^{-1}\mathbf{Z}^{\top}\mathbf{Y}$, of $\boldsymbol{\beta}$ when $Var(\mathbf{e}) = \sigma^2\mathbf{V}$?

Proof.

In the case of $Var(\mathbf{e}^*) = \sigma^2\mathbf{I}$, the Gauss–Markov theorem applies $\hat{\boldsymbol{\beta}}^{*}_{ols}$ is the best linear unbiased estimator (BLUE) for $\boldsymbol{\beta}$. And therefore, $\hat{\boldsymbol{\beta}}_{gls}$ is the BULE for $\boldsymbol{\beta}$.

However, if incorrectly using the ordinary least-squares estimator $\hat{\boldsymbol{\beta}}_{ols} = (\mathbf{Z}^{\top}\mathbf{Z})^{-1}\mathbf{Z}^{\top}\mathbf{Y}$, of $\boldsymbol{\beta}$ when $Var(\mathbf{e}) = \sigma^2\mathbf{V}$, although it is still unbiased, its varaince will larger than $\hat{\boldsymbol{\beta}}_{gls}$ as following.

Since both $\hat{\boldsymbol{\beta}}_{ols}$ and $\hat{\boldsymbol{\beta}}_{ols}$ are linear w.r.t. $\mathbf{Y}$,

$$\hat{\boldsymbol{\beta}}_{ols} = \hat{\boldsymbol{\beta}}_{gls} + \mathbf{A}\mathbf{Y} + \mathbf{b}$$
$$\mathbb{E}\hat{\boldsymbol{\beta}}_{ols} = \mathbb{E}\hat{\boldsymbol{\beta}}_{gls} + \mathbf{A}\mathbf{Z}\boldsymbol{\beta} + \mathbf{b}$$
$$\boldsymbol{\beta} = \boldsymbol{\beta} + \mathbf{A}\mathbf{Z}\boldsymbol{\beta} + \mathbf{b}$$

Therefore,

$$\mathbf{A}\mathbf{Z} = \mathbf{0}$$
$$\mathbf{b} = \mathbf{0}$$

Therefore,

$$\begin{aligned}
Var\hat{\boldsymbol{\beta}}_{ols} &= Var(\hat{\boldsymbol{\beta}}_{gls} + \mathbf{A}\mathbf{Y} + \mathbf{b}) \\
&= Var\{[(\mathbf{Z}^{\top}\mathbf{V}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\top}\mathbf{V}^{-1} + \mathbf{A}]\mathbf{Y}\} \\
&= [(\mathbf{Z}^{\top}\mathbf{V}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\top}\mathbf{V}^{-1} + \mathbf{A}]Var(\mathbf{Y})[(\mathbf{Z}^{\top}\mathbf{V}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\top}\mathbf{V}^{-1} + \mathbf{A}]^{\top} \\
&= \sigma^2[(\mathbf{Z}^{\top}\mathbf{V}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\top}\mathbf{V}^{-1} + \mathbf{A}]\mathbf{V}[(\mathbf{Z}^{\top}\mathbf{V}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\top}\mathbf{V}^{-1} + \mathbf{A}]^{\top} \\
&= Var\hat{\boldsymbol{\beta}}_{gls} + \sigma^2\mathbf{A}\mathbf{V}\mathbf{A}^{\top} \\
&= Var\hat{\boldsymbol{\beta}}_{gls} + \sigma^2(\mathbf{A}\mathbf{C})(\mathbf{A}\mathbf{C})^{\top}
\end{aligned}$$

Since $(\mathbf{A}\mathbf{C})(\mathbf{A}\mathbf{C})^{\top}$ is a positive semi-definite matrix, $Var\hat{\boldsymbol{\beta}}_{ols}$ exceeds $Var\hat{\boldsymbol{\beta}}_{gls}$ by a positive semi-definite matrix. □