Generalized Least Squares, BLUES & BUES

STA 721: Lecture 6

Merlise Clyde (clyde@duke.edu)
Duke University

Outline

- General Least Squares and MLEs
- Gauss-Markov Theorem & BLUEs
- MVUE

Readings:

- Christensen Chapter 2 and 10 (Appendix B as needed)
- Seber & Lee Chapter 3

Other Error Distributions

Model:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad \textsf{E}[\boldsymbol{\epsilon}] = \mathbf{0}_n$$
$$\textsf{Cov}[\boldsymbol{\epsilon}] = \sigma^2\mathbf{V}$$

where σ^2 is a scalar and $\mathbf{V}$ is a $n \times n$ symmetric matrix

Examples:

- Heteroscedasticity: $\mathbf{V}$ is a diagonal matrix with $[\mathbf{V}]_{ii} = v_i$
  - $v_i = 1/n_i$ if y_i is the mean of n_i observations
  - survey weights or propogation of measurement errors in physics models
- Correlated data:
  - time series; first order auto-regressive model with equally spaced data $\textsf{Cov}[\boldsymbol{\epsilon}] = \sigma^2\mathbf{V}$, where $v_{ij} = \rho^{|i-j|}$.
- Hierarchical models with random effects

OLS under a General Covariance

- Is it still unbiased? What's its variance? Is it still the BLUE?
- **Unbiasedness of $\hat{\boldsymbol{\beta}}$**

$$
\begin{aligned}
\mathsf{E}[\hat{\boldsymbol{\beta}}] &= \mathsf{E}[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}] \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathsf{E}[\mathbf{Y}] = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathsf{E}[\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}] \\
&= \boldsymbol{\beta} + \mathbf{0}_p = \boldsymbol{\beta}
\end{aligned}
$$

- **Covariance of $\hat{\boldsymbol{\beta}}$**

$$
\begin{aligned}
\mathsf{Cov}[\hat{\boldsymbol{\beta}}] &= \mathsf{Cov}[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}] \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathsf{Cov}[\mathbf{Y}]\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}
\end{aligned}
$$

- Not necessarily $\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$ unless $\mathbf{V}$ has a special form

GLS via Whitening

Transform the data and reduce problem to one we have solved!

- For $\mathbf{V} > 0$ use the Spectral Decomposition

$$\mathbf{V} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T = \mathbf{U}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Lambda}^{1/2}\mathbf{U}^T$$

- define the symmetric square root of $\mathbf{V}$ as

$$\mathbf{V}^{1/2} \equiv \mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathbf{U}^T$$

- transform model:

$$\begin{aligned}\mathbf{V}^{-1/2}\mathbf{Y} &= \mathbf{V}^{-1/2}\mathbf{X}\boldsymbol{\beta} + \mathbf{V}^{-1/2}\boldsymbol{\epsilon} \\ \tilde{\mathbf{Y}} &= \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{\epsilon}}\end{aligned}$$

- Since Cov$[\tilde{\boldsymbol{\epsilon}}] = \sigma^2\mathbf{V}^{-1/2}\mathbf{V}\mathbf{V}^{-1/2} = \sigma^2\mathbf{I}_n$, we know that $\hat{\boldsymbol{\beta}}_{\mathbf{V}} \equiv (\tilde{\mathbf{X}}^T\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^T\tilde{\mathbf{Y}}$ is the BLUE for $\boldsymbol{\beta}$ based on $\tilde{\mathbf{Y}}$ ($\mathbf{X}$ full rank)

GLS

- If $\mathbf{V}$ is known, then $\tilde{\mathbf{Y}}$ and $\mathbf{Y}$ are known linear transformations of each other
- any estimator of $\boldsymbol{\beta}$ that is linear in $\mathbf{Y}$ is linear in $\tilde{\mathbf{Y}}$ and vice versa from previous results
- $\hat{\boldsymbol{\beta}}_{\mathbf{V}}$ is the BLUE of $\boldsymbol{\beta}$ based on either $\tilde{\mathbf{Y}}$ or $\mathbf{Y}$!
- Substituting back, we have

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{\mathbf{V}} &= (\tilde{\mathbf{X}}^T\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^T\tilde{\mathbf{Y}} \\
&= (\mathbf{X}^T\mathbf{V}^{-1/2}\mathbf{V}^{-1/2}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1/2}\mathbf{V}^{-1/2}\mathbf{Y} \\
&= (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}
\end{aligned}$$

which is the **Generalized Least Squares Estimator** of $\boldsymbol{\beta}$

▼ **Exercise:** Weighted Regression

Consider the model $\mathbf{Y} = \beta\mathbf{x} + \boldsymbol{\epsilon}$ where Cov$[\boldsymbol{\epsilon}]$ is a known diagonal matrix $\mathbf{V}$. Write out the GLS estimator in terms of sums and interpret.

GLS of $\boldsymbol{\mu}$ (Full Rank Case)†

- the OLS/MLE of $\boldsymbol{\mu} \in C(\mathbf{X})$ with transformed variables is

$$\mathbf{P}_{\tilde{\mathbf{X}}} \tilde{\mathbf{Y}} = \tilde{\mathbf{X}} \hat{\boldsymbol{\beta}}_{\mathbf{V}}$$

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

$$\mathbf{V}^{-1/2} \mathbf{X} (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y} = \mathbf{V}^{-1/2} \mathbf{X} \hat{\boldsymbol{\beta}}_{\mathbf{V}}$$

- since $\mathbf{V}$ is positive definite, multiple thru by $\mathbf{V}^{1/2}$, to show that $\hat{\boldsymbol{\beta}}_{\mathbf{V}}$ is a GLS/MLE estimator of $\boldsymbol{\beta}$ iff

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

- Is $\mathbf{P}_{\mathbf{V}} \equiv \mathbf{X} (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1}$ a projection onto $C(\mathbf{X})$? Is it an orthogonal projection onto $C(\mathbf{X})$?

Projections

We want to show that $\mathbf{P_V} \equiv \mathbf{X}\left(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{V}^{-1}$ is a projection onto $C(\mathbf{X})$

Oblique Projections

Proposition: Projection

The $n \times n$ matrix $\mathbf{P_V} \equiv \mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}$ is a projection onto the $C(\mathbf{X})$

- Show that $\mathbf{P}_\mathbf{V}^2 = \mathbf{P_V}$ (idempotent)
- every vector $\mathbf{y} \in \mathbb{R}^n$ may be written as $\mathbf{y} = \mathbf{m} + \mathbf{n}$ where $\mathbf{P_v y} = \mathbf{m}$ and $(\mathbf{I}_n - \mathbf{P_v})\mathbf{y} = \mathbf{n}$ where $\mathbf{m} \in C(\mathbf{P_V})$ and $\mathbf{u} \in N(\mathbf{P_V})$
- Is $\mathbf{P_V}$ an orthogonal projection onto $C(\mathbf{X})$ for the inner product space $(\mathbb{R}^n, \langle \mathbf{v}, \mathbf{u} \rangle = \mathbf{v}^T\mathbf{u})$?

Definition: Oblique Projection

For the inner product space $(\mathbb{R}^n, \langle \mathbf{v}, \mathbf{u} \rangle = \mathbf{v}^T\mathbf{u})$, a projection $\mathbf{P}$ that is not an orthogonal projection is called an *oblique projection*

Loss Function

The GLS estimator minimizes the following generalized squared error loss:

$$\begin{aligned}\|\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\boldsymbol{\beta}\|^2 &= (\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\boldsymbol{\beta})^T(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\boldsymbol{\beta}) \\ &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{V}^{-1/2}\mathbf{V}^{-1/2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2_{\mathbf{V}^{-1}}\end{aligned}$$

where we can change the inner product to be

$$\langle \mathbf{u}, \mathbf{v} \rangle_{\mathbf{V}^{-1}} \equiv \mathbf{u}^T\mathbf{V}^{-1}\mathbf{v}$$

Orthogonality in an Inner Product Space

▼ **Definition:** Orthogonal Projecton

For an inner product space, $(\mathbb{R}^n, \langle , \rangle)$. The projection $\mathbf{P}$ is an orthogonal projection if for every vector $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$,

$$\langle \mathbf{P}\mathbf{x}, (\mathbf{I}_n - \mathbf{P})\mathbf{y} \rangle = \langle (\mathbf{I}_n - \mathbf{P})\mathbf{x}, \mathbf{P}\mathbf{y} \rangle = 0$$

Equivalently:

$$\langle \mathbf{x}, \mathbf{P}\mathbf{y} \rangle = \langle \mathbf{P}\mathbf{x}, \mathbf{P}\mathbf{y} \rangle = \langle \mathbf{P}\mathbf{x}, \mathbf{y} \rangle$$

▼ Exercise

Show that $\mathbf{P}_{\mathbf{V}}$ is an orthogonal projection under the inner product $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{V}^{-1}} \equiv \mathbf{x}^T \mathbf{V}^{-1} \mathbf{y}$

Variance of GLS

- Variance of the GLS estimator $\hat{\boldsymbol{\beta}}_{\mathbf{V}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}$ is much simpler

$$\begin{aligned}
\textsf{Cov}[\hat{\boldsymbol{\beta}}_{\mathbf{V}}] &= (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\textsf{Cov}[\mathbf{Y}]\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1} \\
&= (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{V}\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}
\end{aligned}$$

▼ **Theorem:** Gauss-Markov-Aitkin

Let $\tilde{\boldsymbol{\beta}}$ be a linear unbiased estimator of $\boldsymbol{\beta}$ and $\hat{\boldsymbol{\beta}}_{\mathbf{V}}$ be the GLS estimator of $\boldsymbol{\beta}$ in the linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ with $\textsf{E}[\boldsymbol{\epsilon}] = \mathbf{0}_n$ and $\textsf{Cov}[\boldsymbol{\epsilon}] = \sigma^2\mathbf{V}$ with $\mathbf{X}$ and $\mathbf{V} > 0$ known. Then $\hat{\boldsymbol{\beta}}_{\mathbf{V}}$ is the BLUE where

$$\textsf{Cov}[\tilde{\boldsymbol{\beta}}] \geq \sigma^2(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1} = \textsf{Cov}[\hat{\boldsymbol{\beta}}_{\mathbf{V}}]$$

When will OLS and GLS be Equal?

- For what covariance matrices $\mathbf{V}$ will the OLS and GLS estimators be the same?
- Figuring this out can help us understand why the GLS estimator has a lower variance in general.

▼ Theorem

The estimators $\hat{\beta}$ (OLS) and $\hat{\beta}_{\mathbf{V}}$ (GLS) are the same for all $\mathbf{Y} \in \mathbb{R}^n$ iff

$$\mathbf{V} = \mathbf{X}\mathbf{\Psi}\mathbf{X}^T + \mathbf{H}\mathbf{\Phi}\mathbf{H}^T$$

for some positive definite matrices $\mathbf{\Psi}$ and $\mathbf{\Phi}$ and a matrix $\mathbf{H}$ such that $\mathbf{H}^T\mathbf{X} = \mathbf{0}$.

Outline of Proof

We need to show that $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_{\mathbf{V}}$ are the same for all $\mathbf{Y}$. Since both $\mathbf{P}$ and $\mathbf{P}_{\mathbf{V}}$ are projections onto $C(\mathbf{X})$, $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_{\mathbf{V}}$ will be the same iff

1. Show that $\mathbf{P} = \mathbf{P}_V$ iff $C(\mathbf{X}) = C(\mathbf{V}^{-1}\mathbf{X})$
2. Show that $C(\mathbf{X}) = C(\mathbf{V}^{-1}\mathbf{X})$ iff $C(\mathbf{X}) = C(\mathbf{V}\mathbf{X})$
3. Show that $C(\mathbf{X}) = C(\mathbf{V}\mathbf{X})$ iff $\mathbf{V}$ can be written as

$$\mathbf{V} = \mathbf{X}\boldsymbol{\Psi}\mathbf{X}^T + \mathbf{H}\boldsymbol{\Phi}\mathbf{H}^T$$

See Proposition 2.7.5 and Proof in Christensen

▼ Corollary

Suppose $\mathbf{V} = \mathbf{X}\boldsymbol{\Psi}\mathbf{X}^T + \boldsymbol{\Omega}$. Then $\hat{\boldsymbol{\beta}}_{\mathbf{V}} = \hat{\boldsymbol{\beta}}_{\boldsymbol{\Omega}}$

Some Intuition

For the linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ with $\textsf{E}[\boldsymbol{\epsilon}] = \mathbf{0}_n$ and $\textsf{Cov}[\boldsymbol{\epsilon}] = \sigma^2\mathbf{V}$, we can always write

$$\begin{aligned}\boldsymbol{\epsilon} &= \mathbf{P}\boldsymbol{\epsilon} + (\mathbf{I} - \mathbf{P})\boldsymbol{\epsilon} \\ &= \boldsymbol{\epsilon}_{\mathbf{X}} + \boldsymbol{\epsilon}_N\end{aligned}$$

- we can recover $\boldsymbol{\epsilon}_N$ from the data $\mathbf{Y}$ but not $\boldsymbol{\epsilon}_{\mathbf{X}}$:

$$\begin{aligned}\mathbf{P}\mathbf{Y} &= \mathbf{P}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}_{\mathbf{X}} + \boldsymbol{\epsilon}_n) \\ &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}_{\mathbf{X}} = \mathbf{X}\hat{\boldsymbol{\beta}} \\ (\mathbf{I}_n - \mathbf{P})\mathbf{Y} &= \boldsymbol{\epsilon}_N = \hat{\boldsymbol{\epsilon}} = \mathbf{e}\end{aligned}$$

- Can $\boldsymbol{\epsilon}_{\textsf{N}}$ help us estimate $\mathbf{X}\boldsymbol{\beta}$? What if $\boldsymbol{\epsilon}_N$ could tell us something about $\boldsymbol{\epsilon}_X$?
- Yes if they were highly correlated! But if they were independent or uncorrelated then knowing $\boldsymbol{\epsilon}_{\textsf{N}}$ doesn't help us!

Intuition Continued

- For what matrices are $\boldsymbol{\epsilon}_{\mathbf{X}}$ and $\boldsymbol{\epsilon}_N$ uncorrelated?
- Under $\mathbf{V} = \mathbf{I}_n$:

$$\begin{aligned}\mathsf{E}[\boldsymbol{\epsilon}_X \boldsymbol{\epsilon}_N] &= \mathbf{P}\mathsf{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T](\mathbf{I} - \mathbf{P}) \\ &= \sigma^2 \mathbf{P}(\mathbf{I} - \mathbf{P}) = \mathbf{0}\end{aligned}$$

  so they are uncorrelated

- For the $\mathbf{V}$ in the theorem, introduce
  - $\mathbf{Z}_{\mathbf{X}}$ where $\mathsf{E}[\mathbf{Z}_{\mathbf{X}}] = \mathbf{0}_n$ and $\mathsf{Cov}[\mathbf{Z}_{\mathbf{X}}] = \boldsymbol{\Psi}$
  - $\mathbf{Z}_{\mathsf{N}}$ where $\mathsf{E}[\mathbf{Z}_{\mathsf{N}}] = \mathbf{0}_n$ and $\mathsf{Cov}[\mathbf{Z}_{\mathsf{N}}] = \boldsymbol{\Phi}$
  - $\mathbf{Z}_{\mathbf{X}}$ and $\mathbf{Z}_{\mathsf{N}}$ are uncorrelated, $\mathsf{E}[\mathbf{Z}_{\mathbf{X}}\mathbf{Z}_{\mathsf{N}}] = \mathbf{0}$
  - $\boldsymbol{\epsilon} = \mathbf{X}\mathbf{Z}_{\mathbf{X}} + \mathbf{H}\mathbf{Z}_{\mathsf{N}}$ so that $\boldsymbol{\epsilon}$ has the desired mean and covariance $\mathbf{V}$ in the theorem

Intuition Continued

As a consequence we have
- $\boldsymbol{\epsilon}_{\mathbf{X}} = \mathbf{P}\boldsymbol{\epsilon} = \mathbf{X}\mathbf{Z}_{\mathbf{X}}$ - $\boldsymbol{\epsilon}_{\mathsf{N}} = (\mathbf{I}_n - \mathbf{P})\boldsymbol{\epsilon} = \mathbf{H}\mathbf{Z}_{\mathsf{N}}$ - $\boldsymbol{\epsilon}_{\mathbf{X}}$ and $\boldsymbol{\epsilon}_{\mathsf{N}}$ are uncorrelated

$$\begin{aligned}
\mathsf{E}[\boldsymbol{\epsilon}_{\mathbf{X}}\boldsymbol{\epsilon}_{\mathsf{N}}] &= \mathsf{E}[\mathbf{X}\mathbf{Z}_{\mathbf{X}}\mathbf{Z}_{\mathsf{N}}^T\mathbf{H}^T] \\
&= \mathbf{X}0\mathbf{H}^T \\
&= \mathbf{0}
\end{aligned}$$

- so that $\boldsymbol{\epsilon}_{\mathbf{X}}$ and $\boldsymbol{\epsilon}_{\mathsf{N}}$ are uncorrelated with $\mathbf{V} = \mathbf{X}\boldsymbol{\Psi}\mathbf{X}^T + \mathbf{H}\boldsymbol{\Phi}\mathbf{H}$ ^T$
- Alternative Statement of Theorem: $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{\mathbf{V}}$ for all $\mathbf{Y}$ under $\mathsf{Cov}[\mathbf{Y}] = \sigma^2\mathbf{V}$ iff $\mathbf{P}\mathbf{Y}$ and $(\mathbf{I} - \mathbf{P})\mathbf{Y}$ are uncorrelated