

STAT 714 fa 2025 Lec 03

Gauss-Markov model, BLUE, Aitken model, generalized least-squares

Karl B. Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

- 1 Gauss-Markov model
- 2 Best linear unbiased estimator
- 3 Variance estimation
- 4 Underfitting and overfitting
- 5 Aitken model and generalized least squares
- 6 Best linear unbiased estimation in the restricted model

Gauss-Markov model

The *Gauss-Markov model* assumes $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$, where $\mathbb{E}\mathbf{e} = \mathbf{0}$ and $\text{Cov } \mathbf{e} = \sigma^2 \mathbf{I}_n$.

We consider the mean and variance of the LS estimator $\mathbf{c}^T \hat{\mathbf{b}}$ in this model.

We also consider how to estimate the error term variance σ^2 .

Result (Must know)

For a random vector $\mathbf{y}$, vectors $\mathbf{a}$ and $\mathbf{b}$, and a matrices $\mathbf{A}$ and $\mathbf{B}$, we have

- ① $\mathbb{E}\mathbf{a}^T\mathbf{y} = \mathbf{a}^T\mathbb{E}\mathbf{y}.$
- ② $\text{Var}\mathbf{a}^T\mathbf{y} = \mathbf{a}^T(\text{Cov}\mathbf{y})\mathbf{a}.$
- ③ $\text{Cov}(\mathbf{a}^T\mathbf{y}, \mathbf{b}^T\mathbf{y}) = \mathbf{a}^T(\text{Cov}\mathbf{y})\mathbf{b}.$
- ④ $\text{Cov}\mathbf{A}\mathbf{y} = \mathbf{A}(\text{Cov}\mathbf{y})\mathbf{A}^T.$
- ⑤ $\text{Cov}(\mathbf{A}\mathbf{y}, \mathbf{B}\mathbf{y}) = \mathbf{A}(\text{Cov}\mathbf{y})\mathbf{B}^T.$

Result (Variance of estimable contrast)

Let $\mathbf{c}^T\mathbf{b}$ be an estimable contrast in $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$ with LS estimator $\mathbf{c}^T\hat{\mathbf{b}}$. Then

$$\text{Var}\mathbf{c}^T\hat{\mathbf{b}} = \sigma^2\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-}\mathbf{c}.$$

Exercise: Derive and show invariance to the choice of generalized inverse.

- 1 Gauss-Markov model
- 2 Best linear unbiased estimator
- 3 Variance estimation
- 4 Underfitting and overfitting
- 5 Aitken model and generalized least squares
- 6 Best linear unbiased estimation in the restricted model

Best linear unbiased estimator (BLUE)

A *best linear unbiased estimator* (BLUE) of an estimable contrast $\mathbf{c}^T \mathbf{b}$ is an estimator of the form $a_0 + \mathbf{a}^T \mathbf{y}$ such that $\mathbb{E}[a_0 + \mathbf{a}^T \mathbf{y}] = \mathbf{c}^T \mathbf{b}$ for all $\mathbf{b}$ and such that no other such estimator has smaller variance for any $\mathbf{b}$.

A BLUE achieves the smallest possible variance among linear unbiased estimators.

Theorem (Gauss-Markov Theorem)

Let $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$ with $\mathbb{E}\mathbf{e} = \mathbf{0}$ and $\text{Cov } \mathbf{e} = \sigma^2 \mathbf{I}_n$ and let $\mathbf{c}^T \mathbf{b}$ be an estimable contrast. If $\hat{\mathbf{b}}$ satisfies $\mathbf{X}^T \mathbf{X} \hat{\mathbf{b}} = \mathbf{X}^T \mathbf{y}$ then $\mathbf{c}^T \hat{\mathbf{b}}$ is the BLUE for $\mathbf{c}^T \mathbf{b}$.

See Thm 4.1 of Monahan (2008).

Prove the result.

- 1 Gauss-Markov model
- 2 Best linear unbiased estimator
- 3 Variance estimation**
- 4 Underfitting and overfitting
- 5 Aitken model and generalized least squares
- 6 Best linear unbiased estimation in the restricted model

Result (Expected value of a quadratic form)

Let $\mathbf{z}$ be a random vector with $\mathbb{E}\mathbf{z} = \boldsymbol{\mu}$ and $\text{Cov } \mathbf{z} = \boldsymbol{\Sigma}$. Then

$$\mathbb{E}\mathbf{z}^T \mathbf{A} \mathbf{z} = \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu} + \text{tr}(\mathbf{A} \boldsymbol{\Sigma}).$$

See Lem 4.1 of Monahan (2008).

Can be used to prove the following result.

Result (Unbiased estimator of the variance)

Let $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$ with $\mathbb{E}\mathbf{e} = \mathbf{0}$ and $\text{Cov } \mathbf{e} = \sigma^2 \mathbf{I}_n$ and let $\mathbf{X}$ have rank r . Then

$$\hat{\sigma}^2 = \frac{\|\hat{\mathbf{e}}\|^2}{n - r}, \quad \text{where } \hat{\mathbf{e}} = (\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{y},$$

is an unbiased estimator of σ^2 .

See Res 4.2 of Monahan (2008).

Prove both results.

- 1 Gauss-Markov model
- 2 Best linear unbiased estimator
- 3 Variance estimation
- 4 Underfitting and overfitting**
- 5 Aitken model and generalized least squares
- 6 Best linear unbiased estimation in the restricted model

A setup for “underfitting” or omitting important covariates

Suppose $\mathbf{y}$ is truly given by

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \boldsymbol{\eta} + \mathbf{e}, \quad \mathbb{E}\mathbf{e} = 0, \quad \text{Cov } \mathbf{e} = \sigma^2 \mathbf{I}_n,$$

but one assumes the model $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$ and finds $\hat{\mathbf{b}}$ satisfying $\mathbf{X}^T \mathbf{X} \hat{\mathbf{b}} = \mathbf{X}^T \mathbf{y}$.

So $\boldsymbol{\eta}$ represents the effects of omitted covariates.

Exercise: In the above setup find the expected value of

- ① $\mathbf{c}^T \hat{\mathbf{b}}$, where $\mathbf{c} \in \text{Col } \mathbf{X}^T$
- ② $\hat{\sigma}^2 = \|\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}\|^2 / (n - r)$, where $r = \text{rank } \mathbf{X}$.

In each case give a condition on $\boldsymbol{\eta}$ under which the estimator will be unbiased.

A setup for “overfitting” or including unimportant covariates

Suppose $\mathbf{y}$ is truly given by

$$\mathbf{y} = \mathbf{X}_1 \mathbf{b}_1 + \mathbf{e}, \quad \mathbb{E} \mathbf{e} = 0, \quad \text{Cov } \mathbf{e} = \sigma^2 \mathbf{I}_n,$$

but one assumes $\mathbf{y} = \mathbf{X}_1 \mathbf{b}_1 + \mathbf{X}_2 \mathbf{b}_2 + \mathbf{e}$ and finds $\hat{\mathbf{b}} = [\mathbf{b}_1^T \ \mathbf{b}_2^T]^T$ satisfying

$$\mathbf{X}^T \mathbf{X} \hat{\mathbf{b}} = \mathbf{X}^T \mathbf{y}, \quad \text{where} \quad \mathbf{X} = [\mathbf{X}_1 \ \mathbf{X}_2].$$

In truth $\mathbf{b}_2 = \mathbf{0}$, so $\mathbf{X}_2$ contains unimportant (extra) covariate information.

Exercise: In the above setup, assume (for simplicity) that $\mathbf{X}$ is full-rank. Then:

- 1 Find $\mathbb{E} \mathbf{c}_1^T \hat{\mathbf{b}}_1$ and $\text{Var } \mathbf{c}_1^T \hat{\mathbf{b}}_1$
- 2 In what situation does adding “extra” covariates have no effect on $\text{Var } \mathbf{c}_1^T \hat{\mathbf{b}}_1$?
- 3 Find $\mathbb{E} \|(\mathbf{I} - \mathbf{P}_{\mathbf{X}}) \mathbf{y}\|^2 / (N - \text{rank } \mathbf{X})$ and $\mathbb{E} \|(\mathbf{I} - \mathbf{P}_{\mathbf{X}_1}) \mathbf{y}\|^2 / (N - \text{rank } \mathbf{X}_1)$.

- 1 Gauss-Markov model
- 2 Best linear unbiased estimator
- 3 Variance estimation
- 4 Underfitting and overfitting
- 5 Aitken model and generalized least squares**
- 6 Best linear unbiased estimation in the restricted model

Aitken model

The *Aitken model* assumes $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$, where $\mathbb{E}\mathbf{e} = \mathbf{0}$ and $\text{Cov}\mathbf{e} = \sigma^2\mathbf{V}$, $\mathbf{V}$ pd.

The ordinary least-squares estimator $\mathbf{c}^T\mathbf{b}$ may not be the BLUE anymore. . .

Generalized least-squares estimator of an estimable contrast

Under the Aitken model, the *GLS estimator* of a contrast is $\mathbf{c}^T\hat{\mathbf{b}}_{\text{glS}}$, where $\hat{\mathbf{b}}_{\text{glS}}$ is any vector such that $(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})\hat{\mathbf{b}}_{\text{glS}} = \mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$.

Exercise: Show that $\hat{\mathbf{b}}_{\text{glS}}$ minimizes $Q_{\mathbf{V}}(\mathbf{b}) = (\mathbf{y} - \mathbf{X}\mathbf{b})^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\mathbf{b})$.

We will need these results for establishing some properties of the GLS estimator.

Result (“Generalized cool result” and another gen. inverse of $\mathbf{X}$)

- 1 If $\mathbf{V}$ is pd there exists a nonsingular, symm. matrix $\mathbf{R}$ such that $\mathbf{R}\mathbf{R} = \mathbf{V}^{-1}$.
- 2 $\text{Nul } \mathbf{X}^T \mathbf{V}^{-1} \mathbf{X} = \text{Nul } \mathbf{X}$.
- 3 $\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X} \mathbf{A} = \mathbf{X}^T \mathbf{V}^{-1} \mathbf{X} \mathbf{B} \iff \mathbf{X} \mathbf{A} = \mathbf{X} \mathbf{B}$
- 4 $(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-} \mathbf{X}^T \mathbf{V}^{-1}$ is a generalized inverse of $\mathbf{X}$.

Prove the results.

Note: Estimability of $\mathbf{c}^T \mathbf{b}$ is still equivalent to $\mathbf{c} \in \text{Col } \mathbf{X}^T$.

Result (Properties of the GLS estimator of an estimable contrast)

Let $\mathbf{c}^T \mathbf{b}$ be an estimable contrast. Then the GLS estimator $\mathbf{c}^T \hat{\mathbf{b}}_{\text{GLS}}$

- 1 is invariant to the choice of $\hat{\mathbf{b}}_{\text{GLS}}$ which satisfies $(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}) \hat{\mathbf{b}}_{\text{GLS}} = \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y}$.
- 2 has expected value equal to $\mathbf{c}^T \mathbf{b}$ for all $\mathbf{b}$.
- 3 has variance $\text{Var } \mathbf{c}^T \hat{\mathbf{b}}_{\text{GLS}} = \sigma^2 \mathbf{c}^T (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-} \mathbf{c}$.

Prove the results.

Theorem (Aitken's Theorem)

Let $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$, where $\mathbb{E}\mathbf{e} = \mathbf{0}$ and $\text{Cov } \mathbf{e} = \sigma^2\mathbf{V}$, $\mathbf{V}$ pd, and let $\mathbf{c}^T\mathbf{b}$ be an estimable contrast. Then $\mathbf{c}^T\hat{\mathbf{b}}_{\text{gls}}$ is the BLUE for $\mathbf{c}^T\mathbf{b}$.

See Thm 4.2 of Monahan (2008).

Prove the result.

Result (Unbiased estimator of the variance in the Aitken model)

Let $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$ with $\mathbb{E}\mathbf{e} = \mathbf{0}$ and $\text{Cov } \mathbf{e} = \sigma^2\mathbf{V}$ and let $\mathbf{X}$ have rank r . Then

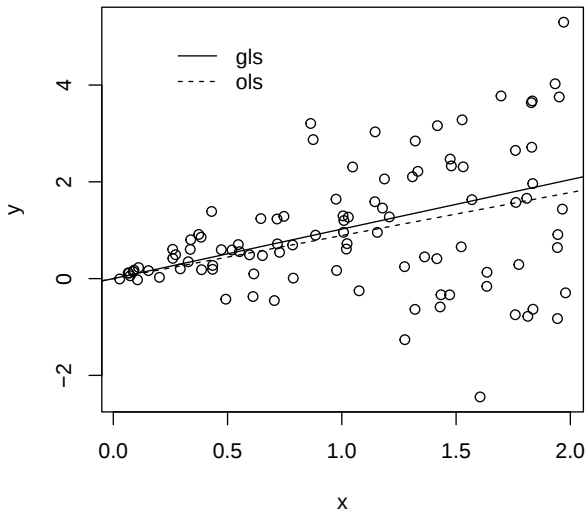
$$\hat{\sigma}_{\text{gls}}^2 = \frac{1}{n-r}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}_{\text{gls}})^T \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}_{\text{gls}})$$

is an unbiased estimator of σ^2 .

Prove the result.

Exercise: Let $Y_i = x_i\beta + \varepsilon_i|x_i|$, $i = 1, \dots, n$, where $\varepsilon_1, \dots, \varepsilon_n \stackrel{\text{ind}}{\sim} \text{Normal}(0, \sigma^2)$.

- 1 Write the model in matrix form as an Aitken model.
- 2 Give $\hat{\beta}_{\text{gls}}$ as well as $\hat{\beta}_{\text{ols}}$.
- 3 Give $\hat{\sigma}_{\text{gls}}^2$.
- 4 Give $\mathbb{E}\hat{\sigma}_{\text{gls}}^2$ as well as $\mathbb{E}\hat{\sigma}_{\text{ols}}^2$.
- 5 Give $\mathbb{E}\hat{\beta}_{\text{gls}}$ as well as $\mathbb{E}\hat{\beta}_{\text{ols}}$.
- 6 Give $\text{Var} \hat{\beta}_{\text{gls}}$ as well as $\text{Var} \hat{\beta}_{\text{ols}}$.
- 7 Show that $\text{Var} \hat{\beta}_{\text{ols}} \geq \text{Var} \hat{\beta}_{\text{gls}}$ using the Cauchy-Schwarz inequality.



- 1 Gauss-Markov model
- 2 Best linear unbiased estimator
- 3 Variance estimation
- 4 Underfitting and overfitting
- 5 Aitken model and generalized least squares
- 6 Best linear unbiased estimation in the restricted model

Go back to the Gauss-Markov model, impose $\mathbf{P}^T \mathbf{b} = \delta$, and consider BLUE...

Theorem (Best linear unbiased estimator in the restricted G-M model)

Let $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$, where $\mathbb{E}\mathbf{e} = \mathbf{0}$ and $\text{Cov } \mathbf{e} = \sigma^2 \mathbf{I}_n$, with the restriction $\mathbf{P}^T \mathbf{b} = \delta$.

Let $\hat{\mathbf{b}}_H$ and $\hat{\mathbf{u}}$ satisfy
$$\begin{bmatrix} \mathbf{X}^T \mathbf{X} & \mathbf{P} \\ \mathbf{P}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}}_H \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T \mathbf{y} \\ \delta \end{bmatrix}$$
 and let $\mathbf{c}^T \mathbf{b}$ be estimable.

Then $\mathbf{c}^T \hat{\mathbf{b}}_H$ is the BLUE for $\mathbf{c}^T \mathbf{b}$ in the restricted model.

See Res 4.5 of Monahan (2008).

Prove the above in the steps:

- 1 Show consistency of the equations
$$\begin{bmatrix} \mathbf{X}^T \mathbf{X} & \mathbf{P} \\ \mathbf{P}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 \\ \hat{\mathbf{v}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{c} \\ \mathbf{0} \end{bmatrix}$$
. Lemma 4.2 of M.
- 2 Show $\mathbb{E} \mathbf{c}^T \hat{\mathbf{b}}_H = \mathbf{c}^T \mathbf{b}$ for all $\mathbf{b}$. Lemma 4.3 of M.
- 3 Show that any unbiased estimator $a_0 + \mathbf{a}^T \mathbf{y}$ has variance $\geq \text{Var } \mathbf{c}^T \hat{\mathbf{b}}_H$.

Monahan, J. F. (2008). *A primer on linear models*. CRC Press.