

# Best Linear Unbiased Estimators in Prediction, MVUEs and BUEs

STA 721: Lecture 5

Merlise Clyde (clyde@duke.edu)

Duke University



# Outline

- Gauss-Markov Theorem for non-full rank  $\mathbf{X}$  (recap)
- Best Linear Unbiased Estimators for Prediction
- MVUE
- Discussion of recent papers on Best Unbiased Estimators beyond linearity

## Readings:

- Christensen Chapter 2 (Appendix B as needed)
- Seber & Lee Chapter 3
- For the curious:
  - Andersen (1962) [Least squares and best unbiased estimates](#)
  - Hansen (2022) [A modern gauss-markov theorem](#)
  - [What Estimators are Unbiased for Linear Models](#) (2023) and references within



# Identifiability

## ▼ Definition: Identifiable

$\beta$  and  $\sigma^2$  are identifiable if the distribution of  $\mathbf{Y}$ ,  $f_{\mathbf{Y}}(\mathbf{y}; \beta_1, \sigma_1^2) = f_{\mathbf{Y}}(\mathbf{y}; \beta_2, \sigma_2^2)$  implies that  $(\beta_1, \sigma_1^2)^T = (\beta_2, \sigma_2^2)^T$

- For linear models, equivalent definition is that  $\beta$  is identifiable if for any  $\beta_1$  and  $\beta_2$ ,  $\mu(\beta_1) = \mu(\beta_2)$  or  $\mathbf{X}\beta_1 = \mathbf{X}\beta_2$  implies that  $\beta_1 = \beta_2$ .
- If  $r(\mathbf{X}) = p$  then  $\beta$  is identifiable
- If  $\mathbf{X}$  is not full rank, there exists  $\beta_1 \neq \beta_2$ , but  $\mathbf{X}\beta_1 = \mathbf{X}\beta_2$  and hence  $\beta$  is not identifiable!
- identifiable linear functions of  $\beta$ ,  $\mathbf{\Lambda}^T \beta$  that have an unbiased estimator are historically referred to as **estimable** in linear models.



# BLUE of $\Lambda^T \beta$

If  $\Lambda^T = \mathbf{B}\mathbf{X}$  for some matrix  $\mathbf{B}$  (or  $\Lambda = \mathbf{X}^T \mathbf{B}$  then

- $E[\mathbf{B}\mathbf{P}\mathbf{Y}] = E[\mathbf{B}\mathbf{X}\hat{\beta}] = E[\Lambda^T \hat{\beta}] = \Lambda^T \beta$
- identifiable as it is a function of  $\mu$ , linear and unbiased
- The unique OLS estimate of  $\Lambda^T \beta$  is  $\Lambda^T \hat{\beta}$
- $\mathbf{B}\mathbf{P}\mathbf{Y} = \Lambda^T \hat{\beta}$  is the BLUE of  $\Lambda^T \beta$

$$E[\|\mathbf{B}\mathbf{P}\mathbf{Y} - \mathbf{B}\mu\|^2] \leq E[\|\mathbf{A}\mathbf{Y} - \mathbf{B}\mu\|^2]$$

$\Leftrightarrow$

$$E[\|\Lambda^T \hat{\beta} - \Lambda^T \beta\|^2] \leq E[\|\mathbf{L}^T \tilde{\beta} - \Lambda^T \beta\|^2]$$

for LUE  $\mathbf{A}\mathbf{Y} = \mathbf{L}^T \tilde{\beta}$  of  $\Lambda^T \beta$



# Non-Identifiable Example

One-way ANOVA model

$$\mu_{ij} = \mu + \tau_j \quad \boldsymbol{\mu} = (\mu_{11}, \dots, \mu_{n_1 1}, \mu_{12}, \dots, \mu_{n_2 2}, \dots, \mu_{1J}, \dots, \mu_{n_J J})^T$$

- Let  $\boldsymbol{\beta}_1 = (\mu, \tau_1, \dots, \tau_J)^T$
- Let  $\boldsymbol{\beta}_2 = (\mu - 42, \tau_1 + 42, \dots, \tau_J + 42)^T$
- Then  $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$  even though  $\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$
- $\boldsymbol{\beta}$  is not identifiable
- yet  $\boldsymbol{\mu}$  is identifiable, where  $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$  (a linear combination of  $\boldsymbol{\beta}$ )



# LUEs of Individual $\beta_j$

## ▼ Proposition

For  $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} = \sum_j \mathbf{X}_j\beta_j$   $\beta_j$  is **not identifiable** if and only if there exists  $\alpha_j$  such that  $\mathbf{X}_j = \sum_{i \neq j} \mathbf{X}_i\alpha_i$

One-way Anova Model:  $Y_{ij} = \mu + \tau_j + \epsilon_{ij}$

$$\boldsymbol{\mu} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \cdots & \mathbf{0}_{n_1} \\ \mathbf{1}_{n_2} & \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & \cdots & \mathbf{0}_{n_2} \\ \vdots & \vdots & \ddots & \vdots & \\ \mathbf{1}_{n_J} & \mathbf{0}_{n_J} & \mathbf{0}_{n_J} & \cdots & \mathbf{1}_{n_J} \end{bmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \vdots \\ \tau_J \end{pmatrix}$$

- Are any parameters  $\mu$  or  $\tau_j$  identifiable?



# Examples:

- A  $j$ th element of  $\beta$ :  $\lambda = (0, 0, \dots, 1, 0, \dots, 0)^T$ ,

$$\lambda^T \beta = \beta_j$$

- Difference between two treatments:  $\tau_1 - \tau_2$ :  $\lambda = (0, 1, -1, \dots, 0, \dots, 0)^T$ ,

$$\lambda^T \beta = \tau_1 - \tau_2$$

- Estimation at observed  $\mathbf{x}_i$ :  $\lambda = \mathbf{x}_i$

$$\mu_i = \mathbf{x}_i^T \beta$$

- Estimation or prediction at a new point  $\mathbf{x}_*$ :  $\lambda = \mathbf{x}_*$ ,

$$\mu_* = \mathbf{x}_*^T \beta$$



# Another Non-Full Rank Example

```

1  x1 = -4:4
2  x2 = c(-2, 1, -1, 2, 0, 2, -1, 1, -2)
3  x3 = 3*x1 - 2*x2
4  x4 = x2 - x1 + 4
5  Y = 1+x1+x2+x3+x4 + c(-.5,.5,.5,-.5,0,.5,-.5,-.5,.5)
6  dev.set = data.frame(Y, x1, x2, x3, x4)
7
8  # Order 1
9  lm1234 = lm(Y ~ x1 + x2 + x3 + x4, data=dev.set)
10 round(coefficients(lm1234), 4)

```

(Intercept)	x1	x2	x3	x4
5	3	0	NA	NA

```

1  # Order 2
2  lm3412 = lm(Y ~ x3 + x4 + x1 + x2, data = dev.set)
3  round(coefficients(lm3412), 4)

```

(Intercept)	x3	x4	x1	x2
-19	3	6	NA	NA





# In Sample Predictions

```
1 cbind(dev.set, predict(lm1234), predict(lm3412))
```

	Y	x1	x2	x3	x4	predict(lm1234)	predict(lm3412)
1	-7.5	-4	-2	-8	6	-7	-7
2	-3.5	-3	1	-11	8	-4	-4
3	-0.5	-2	-1	-4	5	-1	-1
4	1.5	-1	2	-7	7	2	2
5	5.0	0	0	0	4	5	5
6	8.5	1	2	-1	5	8	8
7	10.5	2	-1	8	1	11	11
8	13.5	3	1	7	2	14	14
9	17.5	4	-2	16	-2	17	17

- Both models agree for estimating the mean at the observed **X** points!



# Out of Sample

```
1 out = data.frame(test.set,
2           Y1234=predict(lm1234, new=test.set),
3           Y3412=predict(lm3412, new=test.set))
4 out
```

	x1	x2	x3	x4	Y1234	Y3412
1	3	1	7	2	14	14
2	6	2	14	4	23	47
3	6	2	14	0	23	23
4	0	0	0	4	5	5
5	0	0	0	0	5	-19
6	1	2	3	4	8	14

- Agreement for cases 1, 3, and 4 only!
- Can we determine that without finding the predictions and comparing?
- Conditions for general  $\mathbf{\Lambda}$  or  $\mathbf{\lambda}$  without finding  $\mathbf{B}$  ( $\beta^T$ )?



# Conditions for LUE of $\lambda$

- GM requires that  $\lambda^T = \mathbf{b}^T \mathbf{X} \Leftrightarrow \lambda = \mathbf{X}^T \mathbf{b}$  therefore  $\lambda \in C(\mathbf{X}^T)$
- Suppose we have an arbitrary  $\lambda = \lambda_* + \mathbf{u}$ , where  $\lambda_* \in C(\mathbf{X}^T)$  and  $\mathbf{u} \in C(\mathbf{X}^T)^\perp$  (orthogonal complement)
- Let  $\mathbf{P}_{\mathbf{X}^T}$  denote an orthogonal projection onto  $C(\mathbf{X}^T)$  then  $\mathbf{I} - \mathbf{P}_{\mathbf{X}^T}$  is an orthogonal projection onto  $C(\mathbf{X}^T)^\perp$
- $(\mathbf{I} - \mathbf{P}_{\mathbf{X}^T})\lambda = (\mathbf{I} - \mathbf{P}_{\mathbf{X}^T})\lambda_* + (\mathbf{I} - \mathbf{P}_{\mathbf{X}^T})\mathbf{u} = \mathbf{0}_p + \mathbf{u}$
- so if  $\lambda \in C(\mathbf{X}^T)$  we will have  $(\mathbf{I} - \mathbf{P}_{\mathbf{X}^T})\lambda = \mathbf{0}_p$ !
- Note this is really just a generalization of Proposition 2.1.6 in Christensen that  $\beta_j$  is identifiable iff there exist scalars such that  $\mathbf{X}_j \neq \sum_{i \neq j} \mathbf{X}_i \alpha_i$



## ▼ Exercise

- a. Is  $\mathbf{P}_{\mathbf{X}^T} = (\mathbf{X}^T \mathbf{X})(\mathbf{X}^T \mathbf{X})^{-}$  a projection onto  $C(\mathbf{X}^T)$ ?
- b. Is the expression for  $\mathbf{P}_{\mathbf{X}^T}$  unique?
- c. Is  $\mathbf{P}_{\mathbf{X}^T}$  an orthogonal projection in general?
- d. Is  $\mathbf{P}_{\mathbf{X}^T}$  using the Moore-Penrose generalized inverse an orthogonal projection?



# Prediction Example Again

For prediction at a new  $\mathbf{x}_*$ , this is implemented in the R package `estimability`

```
1 require("estimability" )
2 cbind(epredict(lm1234, test.set), epredict(lm3412, test.set))
```

	[,1]	[,2]
1	14	14
2	NA	NA
3	23	23
4	5	5
5	NA	NA
6	NA	NA

Rows 2, 5, and 6 do not have a unique best linear unbiased estimator,  $\mathbf{x}_*^T \boldsymbol{\beta}$



# MVUE: Minimum Variance Unbiased Estimators

- Gauss-Markov Theorem says that OLS has minimum variance in the class of all Linear Unbiased estimators for  $\mathbf{E}[\boldsymbol{\epsilon}] = \mathbf{0}_n$  and  $\text{Cov}[\boldsymbol{\epsilon}] = \sigma^2 \mathbf{I}_n$
- Requires just first and second moments
- Additional assumption of normality and full rank, OLS of  $\boldsymbol{\beta}$  is the same as MLEs and have minimum variance out of **ALL** unbiased estimators (MVUE); not just linear estimators (section 2.5 in Christensen)
- requires Complete Sufficient Statistics and Rao-Blackwell Theorem - next semester in STA732)
- so Best Unbiased Estimators (BUE) not just BLUE!



# What about ?

- are there nonlinear estimators that are better than OLS under the assumptions ?
- [Anderson \(1962\)](#) showed OLS is not generally the MVUE with  $E[\epsilon] = \mathbf{0}_n$  and  $\text{Cov}[\epsilon] = \sigma^2 \mathbf{I}_n$
- pointed out that linear-plus-quadratic (LPQ) estimators can outperform the OLS estimator for certain error distributions.
- Other assumptions on  $\text{Cov}[\epsilon] = \Sigma$ ?
  - Generalized Least Squares are BLUE (not necessarily equivalent to OLS)
- more recently [Hansen \(2022\)](#) concludes that OLS is BUE over the broader class of linear models with  $\text{Cov}[\epsilon]$  finite and  $E[\epsilon] = \mathbf{0}_n$
- lively ongoing debate! - see [What Estimators are Unbiased for Linear Models \(2023\)](#) and references within



# Next Up

- GLS under assumptions  $\mathbf{E}[\boldsymbol{\epsilon}] = \mathbf{0}_n$  and  $\text{Cov}[\boldsymbol{\epsilon}] = \boldsymbol{\Sigma}^2$
- Oblique projections and orthogonality with other inner products on  $\mathbb{R}^n$
- MLEs in Multivariate Normal setting
- Gauss-Markov

