

## 2. A Review of Some Key Linear Models Results

## A General Linear Model (GLM)

Suppose  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ , where

- $\mathbf{y} \in \mathbb{R}^n$  is the response vector,
- $\mathbf{X}$  is an  $n \times p$  matrix of known constants,
- $\boldsymbol{\beta} \in \mathbb{R}^p$  is an unknown parameter vector, and
- $\boldsymbol{\epsilon}$  is a vector of random “errors” satisfying  $E(\boldsymbol{\epsilon}) = \mathbf{0}$ .

This GLM says simply that  $\mathbf{y}$  is a random vector satisfying  $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$  for some  $\boldsymbol{\beta} \in \mathbb{R}^p$ .

The distribution of  $\mathbf{y}$  is left unspecified.

We know only that  $\mathbf{y}$  is random and its mean is in the column space of  $\mathbf{X}$ ; i.e.,  $E(\mathbf{y}) \in \mathcal{C}(\mathbf{X})$ .

## Ordinary Least Squares (OLS) Estimation of $E(\mathbf{y})$

Because we know  $E(\mathbf{y}) \in \mathcal{C}(\mathbf{X})$ , a natural estimator of  $E(\mathbf{y})$  is the *Ordinary Least Squares Estimator* (OLSE), which is the unique point in  $\mathcal{C}(\mathbf{X})$  that is closest to  $\mathbf{y}$  in terms of Euclidean distance.

The OLSE of  $E(\mathbf{y})$  is given by  $\hat{\mathbf{y}} \equiv \mathbf{P}_\mathbf{X}\mathbf{y}$ , where

$$\mathbf{P}_\mathbf{X} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}',$$

because  $\mathbf{P}_\mathbf{X}\mathbf{y} \in \mathcal{C}(\mathbf{X})$  and

$$\|\mathbf{y} - \mathbf{P}_\mathbf{X}\mathbf{y}\|^2 < \|\mathbf{y} - \mathbf{z}\|^2 \quad \forall \mathbf{z} \in \mathcal{C}(\mathbf{X}) \setminus \{\mathbf{P}_\mathbf{X}\mathbf{y}\}.$$

## The Orthogonal Projection Matrix

$P_X = X(X'X)^-X'$  is known as the *orthogonal projection matrix* (a.k.a. the perpendicular projection operator) onto the column space of  $X$  and has the following properties:

- $P_X$  is symmetric (i.e.,  $P_X = P_X'$ ).
- $P_X$  is idempotent (i.e.,  $P_X P_X = P_X$ ).
- $P_X X = X$  and  $X' P_X = X'$ .
- $\text{rank}(X) = \text{rank}(P_X) = \text{tr}(P_X)$ .
- $P_X = X(X'X)^-X'$  is the same matrix for all generalized inverses  $(X'X)^-$  of  $X'X$ .

## The OLSE of a Linear Function of $E(\mathbf{y})$

For any  $q \times n$  matrix  $A$ ,  $AE(\mathbf{y})$  is a linear function of  $E(\mathbf{y})$ .

For any  $q \times n$  matrix  $A$ , the OLSE of  $AE(\mathbf{y}) = \mathbf{A}\mathbf{X}\boldsymbol{\beta}$  is

$$A[\text{OLSE of } E(\mathbf{y})] = \mathbf{A}\hat{\mathbf{y}} = \mathbf{A}\mathbf{P}_X\mathbf{y} = \mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

## The OLSE of an estimable function of $\beta$

For any  $q \times n$  matrix  $A$ ,  $AE(y) = AX\beta$  is a linear function of  $\beta$  of the form  $C\beta$ , where  $C = AX$ .

From the previous slide, we know

OLSE of  $C\beta = AX\beta = AE(y)$  is  $AX(X'X)^{-1}X'y = C(X'X)^{-1}X'y$ .

Now if  $C$  is any  $q \times p$  matrix, we say that the linear function of  $\beta$  given by  $C\beta$  is *estimable* if and only if  $C = AX$  for some matrix  $q \times n$  matrix  $A$ .

The OLSE of an estimable linear function  $C\beta$  is  $C(X'X)^{-1}X'y$ .

## The OLSE of Estimable Functions of $\beta$

An equivalent definition for the OLSE of an estimable linear function of  $\beta$ , given by  $C\beta = AX\beta$ , can be stated in terms of solutions to the

$$\text{Normal Equations:} \quad X'Xb = X'y.$$

The OLSE of estimable  $C\beta$  is  $C\hat{\beta}$ , where  $\hat{\beta}$  is any solution for  $b$  in the Normal Equations.



# Solutions to the Normal Equations

## The Normal Equations

$$X'X\mathbf{b} = X'\mathbf{y}$$

have  $(X'X)^{-1}X'\mathbf{y}$  as the unique solution for  $\mathbf{b}$  if  $\text{rank}(X) = p$ .

The Normal Equations have infinitely many solutions for  $\mathbf{b}$  if  $\text{rank}(X) < p$ .

$\hat{\beta} = (X'X)^{-}X'\mathbf{y}$  is always a solution to the Normal Equations for any  $(X'X)^{-}$ , a generalized inverse of  $X'X$ .

## Uniqueness of the OLSE of an Estimable $C\beta$

If  $C\beta$  is estimable, then  $C\hat{\beta}$  is the same for all solutions  $\hat{\beta}$  to the Normal Equations.

In particular, the unique OLSE of  $C\beta$  is

$$C\hat{\beta} = C(X'X)^-X'y = AX(X'X)^-X'y = AP_Xy,$$

where  $C = AX$ .

## The OLSE is a Linear Unbiased Estimator

If  $C\beta$  is estimable, then  $C\hat{\beta}$  is a *linear unbiased estimator* of  $C\beta$ .

The OLSE is a *linear estimator* because it's a linear function of  $y$ :

$$C\hat{\beta} = C(X'X)^{-1}X'y = My, \text{ where } M = C(X'X)^{-1}X'.$$

The OLSE is *unbiased* because, for all  $\beta \in \mathbb{R}^p$ ,

$$\begin{aligned} E(C\hat{\beta}) &= E(C(X'X)^{-1}X'y) = C(X'X)^{-1}X'E(y) = AX(X'X)^{-1}X'E(y) \\ &= AP_XE(y) = AP_XX\beta = AX\beta = C\beta. \end{aligned}$$

## The Gauss-Markov Model (GMM)

Suppose  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ , where

- $\mathbf{y} \in \mathbb{R}^n$  is the response vector,
- $\mathbf{X}$  is an  $n \times p$  matrix of known constants,
- $\boldsymbol{\beta} \in \mathbb{R}^p$  is an unknown parameter vector, and
- $\boldsymbol{\epsilon}$  is a vector of random “errors” satisfying  $E(\boldsymbol{\epsilon}) = \mathbf{0}$  and  $\text{Var}(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}$  for some unknown variance parameter  $\sigma^2 \in \mathbb{R}^+$ .

## The GMM is a Special Case of the GLM

The GMM is a special case of the GLM presented previously.

We have added the assumption  $\text{Var}(\epsilon) = \sigma^2 \mathbf{I}$ ; i.e., we assume the errors are uncorrelated and have constant variance.

All the results presented for the GLM hold for the GMM.

# The Gauss-Markov Theorem

For the GMM, we have an additional result provided by the *Gauss-Markov Theorem*:

The OLSE of an estimable function  $C\beta$  is the

*Best Linear Unbiased Estimator (BLUE)* of  $C\beta$

in the sense that the OLSE  $C\hat{\beta}$  has the smallest variance among all linear unbiased estimators of  $C\beta$ .

## Unbiased Estimation of $\sigma^2$

An unbiased estimator of  $\sigma^2$  under the GMM is given by

$$\hat{\sigma}^2 \equiv \frac{\mathbf{y}'(\mathbf{I} - \mathbf{P}_X)\mathbf{y}}{n - r}, \text{ where } r = \text{rank}(\mathbf{X}).$$

Because  $\mathbf{I} - \mathbf{P}_X = (\mathbf{I} - \mathbf{P}_X)(\mathbf{I} - \mathbf{P}_X) = (\mathbf{I} - \mathbf{P}_X)'(\mathbf{I} - \mathbf{P}_X)$ ,

$$\begin{aligned}\mathbf{y}'(\mathbf{I} - \mathbf{P}_X)\mathbf{y} &= \mathbf{y}'(\mathbf{I} - \mathbf{P}_X)'(\mathbf{I} - \mathbf{P}_X)\mathbf{y} = \{(\mathbf{I} - \mathbf{P}_X)\mathbf{y}\}'\{(\mathbf{I} - \mathbf{P}_X)\mathbf{y}\} \\ &= ||(\mathbf{I} - \mathbf{P}_X)\mathbf{y}||^2 = ||\mathbf{y} - \mathbf{P}_X\mathbf{y}||^2 \\ &= ||\mathbf{y} - \hat{\mathbf{y}}||^2 = \text{“Sum of Squared Errors” (SSE)}.\end{aligned}$$

# Gauss-Markov Model with Normal Errors (GMMNE)

Suppose

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where

- $\mathbf{y} \in \mathbb{R}^n$  is the response vector,
- $\mathbf{X}$  is an  $n \times p$  matrix of known constants,
- $\boldsymbol{\beta} \in \mathbb{R}^p$  is an unknown parameter vector, and
- $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$  for some unknown variance parameter  $\sigma^2 \in \mathbb{R}^+$ .



The GMMNE is a special case of the GMM.

We have added the assumption  $\epsilon$  is multivariate normal.

The GMMNE implies  $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$ .

The GMMNE is useful for drawing statistical inferences regarding estimable  $\mathbf{C}\boldsymbol{\beta}$ .

Throughout the remainder of these slides we will assume

- the GMMNE model holds,
- $C$  is a  $q \times p$  matrix such that  $C\beta$  is estimable,
- $\text{rank}(C) = q$ , and
- $d$  is a known  $q \times 1$  vector.

These assumptions imply  $H_0 : C\beta = d$  is a *testable hypothesis*.

## The Distribution of $\mathbf{C}\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$

- $\mathbf{C}\hat{\boldsymbol{\beta}} \sim N(\mathbf{C}\boldsymbol{\beta}, \sigma^2 \mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}')$ .
- $\frac{(n-r)\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-r}^2 \iff \frac{\hat{\sigma}^2}{\sigma^2} \sim \frac{\chi_{n-r}^2}{n-r} \iff \hat{\sigma}^2 \sim \frac{\sigma^2}{n-r} \chi_{n-r}^2$ .
- $\mathbf{C}\hat{\boldsymbol{\beta}}$  and  $\hat{\sigma}^2$  are independent.

## The $F$ Test of $H_0 : C\beta = d$

The test statistic

$$\begin{aligned} F &\equiv (C\hat{\beta} - d)'[\widehat{\text{Var}}(C\hat{\beta})]^{-1}(C\hat{\beta} - d)/q \\ &= (C\hat{\beta} - d)'[\hat{\sigma}^2 C(X'X)^{-1}C']^{-1}(C\hat{\beta} - d)/q \\ &= \frac{(C\hat{\beta} - d)'[C(X'X)^{-1}C']^{-1}(C\hat{\beta} - d)/q}{\hat{\sigma}^2} \end{aligned}$$

has a non-central  $F$  distribution with non-centrality parameter

$$\frac{(C\beta - d)'[C(X'X)^{-1}C']^{-1}(C\beta - d)}{2\sigma^2}$$

and degrees of freedom  $q$  and  $n - r$ .

## The $F$ Test of $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{d}$ (continued)

The non-negative non-centrality parameter

$$\frac{(\mathbf{C}\boldsymbol{\beta} - \mathbf{d})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}(\mathbf{C}\boldsymbol{\beta} - \mathbf{d})}{2\sigma^2}$$

is equal to zero if and only if  $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{d}$  is true.

If  $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{d}$  is true, the statistic  $F$  has a central  $F$  distribution with  $q$  and  $n - r$  degrees of freedom ( $F_{q,n-r}$ ).

## The $F$ Test of $H_0 : C\beta = d$ (continued)

Thus, to test  $H_0 : C\beta = d$ , we compute the test statistic  $F$  and compare the observed value of  $F$  to the  $F_{q,n-r}$  distribution.

If  $F$  is so large that it seems unlikely to have been a draw from the  $F_{q,n-r}$  distribution, we reject  $H_0$  and conclude  $C\beta \neq d$ .

The  $p$ -value of the test is the probability that a random variable with distribution  $F_{q,n-r}$  matches or exceeds the observed value of the test statistic  $F$ .

## The $t$ Test of $H_0 : \mathbf{c}'\boldsymbol{\beta} = d$ for Estimable $\mathbf{c}'\boldsymbol{\beta}$

The test statistic

$$\begin{aligned} t &\equiv \frac{\mathbf{c}'\hat{\boldsymbol{\beta}} - d}{\sqrt{\widehat{\text{Var}}(\mathbf{c}'\hat{\boldsymbol{\beta}})}} \\ &= \frac{\mathbf{c}'\hat{\boldsymbol{\beta}} - d}{\sqrt{\hat{\sigma}^2 \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}} \end{aligned}$$

has a non-central  $t$  distribution with non-centrality parameter

$$\frac{\mathbf{c}'\boldsymbol{\beta} - d}{\sqrt{\sigma^2 \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}}$$

and degrees of freedom  $n - r$ .

## The $t$ Test (continued)

The non-centrality parameter

$$\frac{\mathbf{c}'\boldsymbol{\beta} - d}{\sqrt{\sigma^2 \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}}$$

is equal to zero if and only if  $H_0 : \mathbf{c}'\boldsymbol{\beta} = d$  is true.

If  $H_0 : \mathbf{c}'\boldsymbol{\beta} = d$  is true, the statistic  $t$  has a central  $t$  distribution with  $n - r$  degrees of freedom ( $t_{n-r}$ ).



## The $t$ Test (continued)

Thus, to test  $H_0 : \mathbf{c}'\boldsymbol{\beta} = d$ , we compute the test statistic  $t$  and compare the observed value of  $t$  to the  $t_{n-r}$  distribution.

If  $t$  is so far from zero that it seems unlikely to have been a draw from the  $t_{n-r}$  distribution, we reject  $H_0$  and conclude  $\mathbf{c}'\boldsymbol{\beta} \neq d$ .

The  $p$ -value of the test is the probability that a random variable with distribution  $t_{n-r}$  would be as far or farther from 0 than the observed value of the  $t$  test statistic.

## A $100(1 - \alpha)\%$ Confidence Interval for Estimable $\mathbf{c}'\boldsymbol{\beta}$

$$\left( \mathbf{c}'\hat{\boldsymbol{\beta}} - t_{n-r, 1-\alpha/2} \sqrt{\hat{\sigma}^2 \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{c}}, \quad \mathbf{c}'\hat{\boldsymbol{\beta}} + t_{n-r, 1-\alpha/2} \sqrt{\hat{\sigma}^2 \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{c}} \right)$$

$$\mathbf{c}'\hat{\boldsymbol{\beta}} \pm t_{n-r, 1-\alpha/2} \sqrt{\hat{\sigma}^2 \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{c}}$$

estimate  $\pm$  (distribution quantile)  $\times$  (standard error of estimate)

## Form of the $t$ Statistic for Testing $H_0 : \mathbf{c}'\boldsymbol{\beta} = d$

$$t = \frac{\text{estimate} - d}{\text{standard error of estimate}} = \frac{\text{estimate} - d}{\sqrt{\widehat{\text{Var}}(\text{estimator})}}$$

$$\begin{aligned} t^2 &= \frac{(\text{estimate} - d)^2}{\widehat{\text{Var}}(\text{estimator})} \\ &= (\text{estimate} - d) \left[ \widehat{\text{Var}}(\text{estimator}) \right]^{-1} (\text{estimate} - d) / 1 \end{aligned}$$

## Revisiting the $F$ Statistic for Testing $H_0 : C\beta = d$

$$\begin{aligned} F &= (\mathbf{estimate} - d)' \left[ \widehat{\text{Var}}(\mathbf{estimator}) \right]^{-1} (\mathbf{estimate} - d) / q \\ &= (C\hat{\beta} - d)' [\widehat{\text{Var}}(C\hat{\beta})]^{-1} (C\hat{\beta} - d) / q \\ &= (C\hat{\beta} - d)' [\hat{\sigma}^2 C(X'X)^{-1} C']^{-1} (C\hat{\beta} - d) / q \\ &= \frac{(C\hat{\beta} - d)' [C(X'X)^{-1} C']^{-1} (C\hat{\beta} - d) / q}{\hat{\sigma}^2} \end{aligned}$$