



# ECON 6200 : Section 1

1/24/25

## TA Info

Name: Serena Leung

Email: sl2862@cornell.edu

Office: Uris 473

Office: 9AM - 11AM Wed  
Hours

Email me questions in advance!

## Homework

- graded by the TA
- will be graded generously
- work in groups!

Note: HW 1 due Jan 31

## Sections

- Upload notes in advance
- Resubmit annotated notes after class

# Data Matrix & Vector Notation

$$\begin{aligned}\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\ &= (\mathbb{E}_n \mathbf{X}\mathbf{X}')^{-1}\mathbb{E}_n \mathbf{X}\mathbf{Y} \\ &= \left( \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{Y}_i \\ &= \left( \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i' \right)^{-1} \sum_{i=1}^n \mathbf{X}_i \mathbf{Y}_i.\end{aligned}$$

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & \cdots & x_{1k} \\ x_{21} & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ x_{n1} & \cdots & \cdots & \cdots & x_{nk} \end{bmatrix}_{n \times k}$$

$$= \begin{bmatrix} \text{--- } x_1 \text{ ---} \\ \text{--- } x_2 \text{ ---} \\ \vdots \\ \text{--- } x_k \text{ ---} \end{bmatrix}$$

$$WTS: \sum_{i=1}^n x_i x_i' = \mathbf{x}' \mathbf{x}$$

$$\sum_{i=1}^n x_i x_i' = \left[ \begin{array}{c|ccccc} 1 & & & & & \\ x_1 & -x_1 & - & & & \\ \vdots & & I \times k & & & \\ \hline k \times 1 & & & & & \end{array} \right] + \left[ \begin{array}{c|ccccc} 1 & & & & & \\ x_2 & -x_2 & - & & & \\ \vdots & & I \times k & & & \\ \hline k \times 1 & & & & & \end{array} \right] + \dots$$

where  $\begin{bmatrix} 1 \\ x_i \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ik} \end{bmatrix}$

$$= \begin{bmatrix} x_{11}^2 & x_{11}x_{12} & \dots & x_{11}x_{ik} \\ x_{12}x_{11} & \vdots & & \\ \vdots & & & \\ x_{ik}x_{11} & \dots & x_{kk}^2 \end{bmatrix}_{k \times k} + \begin{bmatrix} & & & \\ & & & \\ & & & \\ & & & \end{bmatrix} + \dots$$

$$= \begin{bmatrix} \sum_{in}^n x_{ii}^2 & \sum x_{ii}x_{i2} & \dots & \sum x_{ii}x_{ik} \\ \sum x_{i2}x_{ii} & \dots & & \vdots \\ \vdots & & & \\ \sum x_{ik}x_{ii} & \sum x_{ik}x_{i2} & \dots & \sum x_{kk}^2 \end{bmatrix}$$

$$\mathbf{X}' \mathbf{X} = \begin{bmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_k \\ | & | & & | \end{bmatrix} \begin{bmatrix} -x_1- \\ -x_2- \\ \vdots \\ -x_k- \end{bmatrix}$$

kxn                          nxk

$$= \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} \\ \vdots \\ x_{1k} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} \\ \vdots \\ x_{n1} & \cdots & x_{nk} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}x_{i2} & \cdots & \sum_{i=1}^n x_{i1}x_{ik} \\ \sum_{i=1}^n x_{i2}x_{i1} & \ddots & & \\ \vdots & & & \\ \sum_{i=1}^n x_{ik}x_{i1} & \cdots & \sum_{i=1}^n x_{ik}^2 \end{bmatrix}$$

## Linear Model

$n$  : # of observations / sample size

$k$  : # of regressors / covariates

$X, Y, Z$  : vectors (capitalized letters)

$\mathbf{X}, \mathbf{Y}, \mathbf{Z}$  : Matrix (bold)

$E_n, \hat{Q}$  : sample analogs

In data matrix notation:

$$Y = X\beta + e$$

$n \times 1 \quad n \times k \quad k \times 1 \quad n \times 1$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & \cdots & \cdots & \cdots & X_{1k} \\ X_{21} & - & \cdots & \cdots & \cdots & X_{2k} \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ X_{n1} & - & \cdots & \cdots & \cdots & X_{nk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ \vdots \\ e_n \end{bmatrix}$$

\* show them  
matrix diff rules  
(Prop 6, 9)

## \* Derive OLS Estimator in Matrix Form

$$\min_{\beta} (Y - X\beta)' (Y - X\beta)$$

Recall

$$(AB)^T = B^T A^T$$

$$\begin{aligned} &= Y'Y - Y'X\beta - \beta'X'Y + \beta'X'X\beta \\ &= Y'Y - Y'X\beta - (Y'X\beta) + \beta'X'X\beta \\ &= Y'Y - 2Y'X\beta + \beta' \underbrace{X'X\beta}_{\text{symmetric}} \end{aligned}$$

FOC

$$[\beta] \quad -2Y'X + 2X'X\beta = 0$$

If  $(X'X)^{-1}$  exists, then  $\hat{\beta} = (X'X)^{-1}(X'Y)$

## Simple Linear Regression

$$y = \beta_0 + \beta_1 x + \mu$$

The DLS estimator is defined as

$$(\hat{\beta}_0, \hat{\beta}_1) \equiv \arg \min_{b_0, b_1} \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2$$

FOC

$$[b_0] \quad -2 \sum_{i=1}^n (y_i - b_0 - b_1 x_i) = 0 \quad \textcircled{1}$$

$$[b_1] \quad -2 \sum_{i=1}^n (y_i - b_0 - b_1 x_i) x_i = 0 \quad \textcircled{2}$$

If we simplify \textcircled{1}, we get

$$\sum_{i=1}^n (y_i - b_0 - b_1 x_i) = 0$$

$$\frac{1}{n} \sum_{i=1}^n (y_i - b_0 - b_1 x_i) = 0$$

multiply  
both sides  
by  $\frac{1}{n}$

$$\bar{y} - b_0 - b_1 \bar{x} = 0$$

$$\Rightarrow \hat{\beta}_0 = \bar{y} - b_1 \bar{x}$$

Substituting  $\hat{\beta}_0$  into ②

$$\sum_{i=1}^n (y_i - b_0 - b_1 x_i) x_i = 0$$

$$\sum_{i=1}^n ((y_i - \bar{y}) + b_1 \bar{x} - b_1 x_i) x_i = 0$$

$$\sum_i x_i (y_i - \bar{y}) + b_1 \sum x_i (\bar{x} - x_i) = 0$$

$$\Rightarrow \hat{\beta}_1 = \frac{\sum x_i (y_i - \bar{y})}{\sum x_i (x_i - \bar{x})}$$

OR you could do some more work and get

$$\hat{\beta}_1 = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}$$

# Frisch - Waugh - Lovell Thm

Intuition : FWL Thm simplifies how we interpret multiple regression models

- To get the effect of  $X_1$  on  $Y$  (holding  $X_2$  fixed), we must remove the correlation between  $X_1$  and  $X_2$ .

(Hansen, pg 82)

### Theorem 3.5 Frisch-Waugh-Lovell (FWL)

In the model (3.31), the OLS estimator of  $\beta_2$  and the OLS residuals  $\hat{e}$  may be computed by either the OLS regression (3.32) or via the following algorithm:

1. Regress  $Y$  on  $X_1$ , obtain residuals  $\tilde{e}_1$ ;
2. Regress  $X_2$  on  $X_1$ , obtain residuals  $\tilde{X}_2$ ;
3. Regress  $\tilde{e}_1$  on  $\tilde{X}_2$ , obtain OLS estimates  $\hat{\beta}_2$  and residuals  $\hat{e}$ .

Suppose we want to find  $\hat{\beta}_2$  where the model

$$Y = \beta_1 X_1 + \beta_2 X_2$$

i)  $Y = \tilde{\beta}_1 X_1 + \tilde{e}_1 ; \quad \tilde{e}_1 = \text{effect of } Y \text{ not explained by } X_1$

$$2) \quad X_2 = \tilde{\sigma} \tilde{X}_1 + \tilde{X}_2 \quad ; \quad \tilde{X}_2 = \text{effect of } X_2 \text{ not explained by } X_1$$

$$3) \quad \tilde{e}_1 = \hat{\beta}_2 \tilde{X}_2 + \bar{e}$$

Since we have eliminated all the variation of  $Y$  explained by  $X_1$  and any correlation between  $X_1$  and  $X_2$ , our  $\hat{\beta}_2$  coefficient is equivalent to that of our original model

$$\hat{Y} = \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + e$$

**2. (from an old exam)** A researcher separately regresses an outcome  $Y$  on two different sets of covariates. Specifically, consider separate estimation by OLS of the models

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon, \\ Y &= \gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2 + \eta, \end{aligned}$$

where the outcome variable  $Y$  is the same in both regressions and where the covariates are deterministically linked by the equalities

$$\begin{aligned} Z_1 &= X_1 - 2X_2, \\ Z_2 &= X_1 + 4X_2. \end{aligned}$$

**2.1** Show that (in data matrix notation) one can write

$$Z = XA.$$

What is  $A$ ? (IE, please give the exact matrix.) Is  $A$  invertible? If so, give  $A^{-1}$ . (Solution of this question does not require mechanical inversion of a  $3 \times 3$  matrix.)

**2.2** How are  $\hat{\beta}$  and  $\hat{\gamma}$  related? Let  $\hat{Y}_Z$  denote the vector of fitted values from the regression of  $Y$  on  $Z$  and  $\hat{Y}_X$  the analogous vector but for the regression of  $Y$  on  $X$ . Prove:  $\hat{Y}_Z = \hat{Y}_X$ .

**2.3** What can you say about  $SST$ ,  $SSR$ ,  $SSE$ , and  $R^2$  from the regression on  $X$ , compared to the same quantities from the regression on  $Z$ ?

