

## Gov 2001: Problem Set 9

Spring 2026

### Instructions:

- The Problem set is due on **April 21, 11:59 PM Eastern Time**.
- Please upload a PDF of your solutions to Gradescope. Make sure to assign to each question all the pages with your work on that question.
- **Do not use AI assistants (ChatGPT, Claude, Copilot, etc.) on this problem set.** Work with each other instead. The struggle is where learning happens.
- Remember: 70% of your grade comes from in-class exams. Use problem sets to *learn*, not just to get answers.

1. Let

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 5 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}.$$

Compute the following and show your steps:

- (a)  $A + B$
- (b)  $AB$
- (c)  $BA$
- (d)  $\det(A)$
- (e)  $A^{-1}$  and  $B^{-1}$ . Infer a general expression of  $X^{-1}$  for an  $n \times n$  diagonal matrix:

$$X = \begin{pmatrix} x_1 & 0 & \cdots & 0 \\ 0 & x_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_n \end{pmatrix}$$

Hint: Let  $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ , then solve for linear equations.

2. In matrix form, the OLS estimator in a regression with an intercept and two regressors is

$$\hat{\beta} = (X'X)^{-1}X'Y, \quad X = \begin{pmatrix} 1 & x_{11} & x_{21} \\ \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} \end{pmatrix}.$$

in which  $X$  is called design matrix and  $x_{ij}$  is the crystallization of regressor  $X_i$  for observation  $j$ .

Consider the following two possible design matrices with three observations:

$$X_1 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & -2 \\ 1 & -1 & 1 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 2 \\ 1 & 2 & 4 \end{pmatrix}.$$

- (a) Compute  $X_1'X_1$  and determine whether it is invertible. If so, find  $X_1'X_1$ .
- (b) A theorem in matrix algebra states: a square matrix  $M$  is not invertible if there exists a nonzero vector  $a$  such that  $Ma = 0$ . Use this theorem to prove that  $X_2'X_2$  is not invertible.
- (c) Take a close look at  $X_2$ . Briefly explain how your answers in (a) and (b) relates to an assumption of Gauss-Markov theorem.

3. Suppose the population model is

$$Y = X\beta + \epsilon, \quad \mathbb{E}[\epsilon | X] = 0.$$

Draw an iid sample  $(Y_1, X_1), \dots, (Y_n, X_n)$  from this population, and define

$$\epsilon_i = Y_i - \beta X_i.$$

Consider the OLS regression of  $Y_i$  on  $X_i$  *without an intercept*. The OLS estimator, fitted values, and residuals are

$$\hat{\beta} = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2}, \quad \hat{Y}_i = \hat{\beta} X_i, \quad e_i = Y_i - \hat{Y}_i.$$

Decide whether each of the following statements is true or false, and explain your reasoning.

- (a)  $\mathbb{E}[\epsilon_i] = 0$ .
- (b)  $\mathbb{E}[X_i^3 \epsilon_i] = 0$ .
- (c)  $\mathbb{E}[e_i] = 0$ .
- (d)  $\sum_{i=1}^n X_i e_i = 0$ .
- (e)  $\sum_{i=1}^n X_i \epsilon_i = 0$ .

4. Often our data are collected with error, which we refer to as measurement error. For instance, for a dependent variable  $Y$  measured in a survey, respondents may randomly mis-click, or they may systematically misreport a socially undesirable trait. In this question, we study the most favorable case, where the measurement error is independent of the true values.

Suppose the CEF is linear:

$$\mathbb{E}[Y | X] = \beta_0 + \beta_1 X,$$

and define the CEF error as

$$e = Y - \mathbb{E}[Y | X].$$

(This is another way to state the same population setting in Q3)

Let  $\text{Var}(X) = \sigma_X^2$ . Unfortunately, we do not observe  $Y$  or  $X$  directly. Instead, we observe noisy proxies  $\tilde{Y}$  and  $\tilde{X}$ , where

$$\tilde{Y} = Y + V, \quad \tilde{X} = X + W,$$

with

$$V \sim N(0, \sigma_v^2), \quad W \sim N(0, \sigma_w^2),$$

and  $V$  and  $W$  are independent of  $X$  and  $Y$ . This implies

$$\text{Cov}(V, X) = \text{Cov}(V, e) = \text{Cov}(V, W) = \text{Cov}(W, X) = \text{Cov}(W, e) = 0.$$

Assume also that the CEF of the observable variables is linear:

$$\mathbb{E}[\tilde{Y} | \tilde{X}] = \alpha_0 + \alpha_1 \tilde{X}.$$

- (a) Derive  $\alpha_1$  in terms of  $\beta_1$ ,  $\sigma_X^2$ ,  $\sigma_w^2$ , and  $\sigma_v^2$ . Hint: first write  $\alpha_1$  in terms of  $\text{Cov}(\tilde{X}, \tilde{Y})$  and  $\text{Var}(\tilde{X})$ .
- (b) Using your expression in part (a), explain in one or two sentences the effect of measurement error in  $X$  on the sign and magnitude of  $\alpha_1$  relative to  $\beta_1$ . Which parameter governs the amount of measurement error in  $X$ ?
- (c) Using your expression in part (a), explain in one or two sentences the effect of measurement error in  $Y$  on the sign and magnitude of  $\alpha_1$  relative to  $\beta_1$ . Which parameter governs the amount of measurement error in  $Y$ ?