Problem Set 1: Due Thursday, February 5, by 2:30pm.

This problem set is based on Handout 1, Handout 2, the lecture notes from the first three weeks of class, and Chapters 2.1-2.25, 2.27-2.28 of Hansen *Econometrics* textbook. You should review the handouts, the readings, and your class notes before starting the problem set. The problem set is primarily theoretical, though the last question involves some simple calculations using **R**, using **R** functions covered in Handout 2. Future problem sets will involve substantially more **R** coding.

Problem 1. Suppose $\mathbf{X}$ is a $K \times 1$ random vector, and suppose that

$$\mathbb{E}[X_k^2] < \infty \quad \text{for each } k = 1, \ldots, K.$$

Recall that, by definition, the variance–covariance matrix of $\mathbf{X}$ is

$$\mathrm{Var}(\mathbf{X}) \equiv \mathbb{E}\Big[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^T\Big],$$

where the expectation is taken elementwise.

(a) Show, by expanding the outer product and taking expectations term by term, that

$$\mathrm{Var}(\mathbf{X}) = \mathbb{E}[\mathbf{X}\mathbf{X}^T] - \mathbb{E}(\mathbf{X})\mathbb{E}(\mathbf{X})^T.$$

(b) What is the (i, j) element of $\mathrm{Var}(\mathbf{X})$? Show explicitly that it equals $\mathrm{Cov}(X, Y)$.

Problem 2. Let (Y, Z, X) denote a random vector, with Y and Z both scalar random variables. Suppose $\mathbb{E}[Y^2] < \infty$, $\mathbb{E}[Z^2] < \infty$. Let a, b, and c denote constants, and let

$$\tilde{m}(X) = a + b \cdot \mathbb{E}[Y \mid X] + c \cdot \mathbb{E}[Z \mid X].$$

(a) From results from class, we know that

$$\begin{aligned} \mathbb{E}\left[(Y - \mathbb{E}(Y \mid X)) \cdot g(X)\right] = 0 \\ \mathbb{E}\left[(Z - \mathbb{E}(Z \mid X)) \cdot g(X)\right] = 0 \end{aligned} \quad \text{for all } \; g \; \text{ s.t. } \; \mathbb{E}[(g(X))^2] < \infty.$$

Using those equalities, show that

$$\mathbb{E}\left[((a + b \cdot Y + c \cdot Z) - \tilde{m}(X)) \cdot g(X)\right] = 0 \;\text{ for all }\; g \;\text{ s.t. }\; \mathbb{E}[(g(X))^2] < \infty.$$

(b) Why can we now conclude that $\tilde{m}(X)$ is the expected value of $a + b \cdot Y + c \cdot Z$ conditional on X, i.e., that

$$\mathbb{E}[a + b \cdot Y + c \cdot Z \mid X] = a + b \cdot \mathbb{E}[Y \mid X] + c \cdot \mathbb{E}[Y \mid Z]?$$

Problem 3. Suppose that X is a discrete random variable taking the values $-1, 0, 1$, with

$$\Pr[X = -1] = \Pr[X = 0] = \Pr[X = 1] = 1/3.$$

Let $Y = X^2$.

(a) Are Y and X independent? Justify your response by comparing conditional and marginal distributions.

(b) Show that $\mathrm{Cov}(X, Y) = 0$.

(c) What function $g(X)$ minimizes $\mathbb{E}[(Y - g(X))^2]$? Identify this best predictor of Y given X and compute it explicitly.

(d) Let $\beta_0 + \beta_1 X$ denote the best linear predictor of Y given X. What is the value of β_0 and β_1?

(e) Briefly discuss the difference between the best predictor and the best linear predictor in this example, and explain what feature of the joint distribution of (X, Y) drives this difference.

Problem 4. Suppose X and Y are scalar random variables, with $\mathbb{E}[\ |Y|\ ] < \infty$ and X a bernoulli r.v. with $\Pr[X = 1] = p$ and $\Pr[X = 0] = 1 - p$.

(a) Show that, for proper choice of β_0 and β_1, we can write $\mathbb{E}[Y \mid X]$ as

$$\mathbb{E}[Y \mid X] = \beta_0 + \beta_1 X.$$

Express β_0 and β_1 in terms of $\mathbb{E}[Y \mid X = 0]$ and $\mathbb{E}[Y \mid X = 1]$.

(b) Given your answer to (a) above, what is the best linear predictor of Y given X in this example? justify your answer.

(c) Use the law of iterated expectations to show that

$$\mathbb{E}[Y] = p \cdot \mathbb{E}[Y \mid X = 1] + (1 - p) \cdot \mathbb{E}[Y \mid X = 0].$$

(d) Use the law of iterated expectations to show that

$$\mathbb{E}[XY] = p \cdot \mathbb{E}[Y \mid X = 1].$$

(e) Use your answers to (a)-(d) above to show that

$$\beta_1 = \mathrm{Cov}(X, Y)/\mathrm{Var}(X).$$

Problem 5. Consider the following model, $Y = X'\beta + e$, where $X = (1, X_1)'$, $\beta = (\beta_0, \beta_1)'$, and Y and X_1 are scalar random variables. Suppose $\mathbb{E}[Y^2] < \infty$, $\mathbb{E}[X_1^2] < \infty$. Suppose that we are defining $X'\beta$ as the best linear predictor of Y given X.

(a) Why does $\mathbb{E}(Xe)$ necessarily equal zero in this example?

(b) Why does $\mathbb{E}(Xe) = 0$ imply $\mathbb{E}(e) = 0$ and $\mathbb{E}(X_1 e) = 0$?

(c) Show that $\mathbb{E}(Xe) = 0$ implies $\mathrm{Cov}(X_1, e) = 0$.

(d) Why is no perfect collinearity in X equivalent to $\mathrm{Var}(X_1) > 0$?

(e) Suppose that $\mathrm{Var}(X_1) > 0$. Recall that $\mathbb{E}(Xe) = 0$ implies that

$$\beta = \left(E(XX')\right)^{-1} E(XY).$$

Use this expression to solve for β_0 and β_1 in terms of $\mathbb{E}(X_1), \mathbb{E}(Y)$, $\mathrm{Cov}(X_1, Y)$, and $\mathrm{Var}(X_1)$.

Problem 6. Consider the following model, $Y = X'\beta + e$, with Y a bernoulli random variable, $Y \in \{0, 1\}$ and X a $K+1$-dimensional random vector. Suppose $\mathbb{E}[e \mid X] = 0$. This model is called a *linear probability model*, and is frequently used to predict binary outcomes. Let $\mu_X = \mathbb{E}[X]$.

(a) Find an expression for $\Pr[Y = 1]$ in terms of μ_X and β.

(b) Find an expression for $\mathrm{Var}[Y]$ in terms of μ_X and β.

(c) Find an expression for $\Pr[Y = 1 \mid X]$ in terms of X and β.

(d) Using your answer to part (c), find an expression for $\mathrm{Var}[Y \mid X]$ in terms of X and β.

(e) Using your answer to part (d), find an expression for $\mathrm{Var}[e \mid X]$ in terms of X and β.

(f) Under what conditions (if any) will e be heteroskedastic?

Problem 7. Suppose you are interested in investing in a mutual fund B with return $r_{B,t}$. Suppose $r_{B,t} \sim N(\mu, \sigma^2)$ and that returns are i.i.d. over time. The fund's manager tells you that $\mathbb{E}[r_{B,t}] = 0.12$ and $\text{Var}(r_{B,t}) = 0.01$, though you worry that the actual expected return might be lower. Suppose you are interested in the asset's excess return above the risk-free rate r_f, where $r_f = 0.02$. In the following, take r_f to be a constant. Let $X_t = r_{B,t} - r_f$ denote the fund's excess return in year t.

(a) Show that, if what the manager told you is true, then:

$$\frac{X_1 - 0.10}{0.1} \sim N(0,1).$$

(b) Suppose you only observe one year of returns, with a zero excess return in that year, $X_1 = 0$. How unlikely is it to have zero excess return or less in one year if what the manager told you is true? In particular, if what she said is true, what is $\Pr[X_1 \leq 0]$?

(i) Express your answer in terms of $\Phi(\cdot)$.

(ii) Using **R**, express your answer as a number, and discuss its magnitude.

(iii) Is what the manager told you implausible given the evidence of zero excess return in one year?

(c) Suppose you observe four years of returns, and that $\bar{X}_4 = 0$. How unlikely is it to have zero average excess return across four years if what the manager told you is true? In particular, if what she said is true, what is $\Pr[\bar{X}_4 \leq 0]$?

(i) Express your answer in terms of $\Phi(\cdot)$.

(ii) Using **R**, express your answer as a number, and discuss its magnitude.

(iii) Is what the manager told you implausible given the evidence of zero average excess return across four years?

(d) Repeat question (c), but now find probability that $\bar{X}_4$ would be that far from asserted expected excess return, i.e., find
$\Pr[\ |\bar{X}_4 - 0.10| \geq 0.10]$.

(i) Express your answer in terms of $\Phi(\cdot)$.

(ii) Using **R**, express your answer as a number, and discuss its magnitude.