ECON-GA 2100 Econometrics I
Fall 2025
Problem Set 2
Due: Monday, November 17.

Problem 1 (Partitioned Inverse Formula) Let A be a nonsingular matrix. Consider the partition of A and a matrix B

$$B = \begin{bmatrix} W^{-1} & -W^{-1}A_{12}A_{22}^{-1} \\ -A_22^{-1}A_{21}W^{-1} & A_{22}^{-1} + A_{22}^{-1}A_{21}W^{-1}A_{12}A_{22}^{-1} \end{bmatrix}, \qquad \text{where } A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

is a partition of A such that A_{11} and A_{22} are square matrices, and we define $W = A_{11} - A_{12}A_{22}^{-1}A_{21}$.

(a) Show that $B = A^{-1}$, i.e. the expression for B gives a formula for the four blocks of the matrix inverse of A. *Hint:* verify that the matrix product $BA = I$.

(b) Use your result in (a) to prove that for the linear regression model

$$y = X_1\beta_1 + X_2\beta_2 + e$$

the LS estimator for β_1 is given by $\hat{\beta}_1 = (X_1'(I - P_{X_2})X_1)^{-1}X_1'(I - P_{X_2})y$. *Hint:* note that we already derived this in the lecture when we proved the Frisch-Waugh Theorem. This problem is a constructive derivation of the same result.

In the second part of this problem, you are supposed to use the partitioned inverse formula to analyze the variance of a linear forecast for y based on a least-square regression. More specifically, you have a sample of n i.i.d. observations stacked in a vector y and matrix of regressors X, respectively, where the first column of X is an n-vector of ones, i.e. the regression includes an intercept. You make a forecast for the value x_{n+1} for the covariates using your LS estimate from the first n observations:

$$\hat{y}_{n+1} = x_{n+1}'\hat{\beta}_{LS} = x_{n+1}'(X'X)^{-1}X'y$$

Suppose that the assumptions of the homoskedastic LRM hold with $\mathrm{Var}(e|X) = \sigma^2 I_n$.

(c) Derive a formula for the variance of the forecast for a fixed value of x_{n+1}.

(d) Now denote $x_{n+1}' = [1, z']$ where $z \in \mathbb{R}^{k-1}$. Derive an expression for the forecast variance of the form $\mathrm{Var}(\hat{y}_{n+1}) = z'Az + b'z + c$ where A is a $(k-1) \times (k-1))$ matrix, $b \in \mathbb{R}^{k-1}$, and c is a scalar. *Hint:* use the partitioned inverse formula where $A_{11} = X_2'X_2$, the block corresponding to the *non-constant* regressors.

(e) Use your result from (d) to find the value of z^* that minimizes the forecast variance. Interpret.

(f) Suppose $k = 2$, i.e. the regression includes only a scalar regressor and the intercept. Draw a conceptual graph in the (x, y)-plane that includes the estimated regression line and 95% confidence bands (i.e. the regression line plus/minus 1.96 standard deviations of the linear forecast at x) for the linear forecast given $x_{n+1} = x$. Be sure to indicate the value of x^* corresponding to your answer in (e).

Problem 2 Consider the linear regression model

$$y_i = \beta_0 + \beta_1 x_i + e_i, \qquad i = 1, \dots, n$$

where e_i is drawn independently according to

$$e_i = \begin{cases} -1 & \text{with probability } \frac{1}{2} \\ 1 & \text{with probability } \frac{1}{2} \end{cases}$$

independent of x_i. Also assume for simplicity that the scalar x_i only takes the values 0 and 1. If it makes your proofs easier, you can also assume that $x_1, x_2, \dots$ is a deterministic sequence, alternating between $x_i = 0$ for i even, and $x_j = 1$ for j odd.

(a) Verify that this model satisfies the assumptions of the homoskedastic LRM.

(b) What is the asymptotic variance $\lim_{n\to\infty} \mathrm{Var}\left(\sqrt{n}(\hat{\beta}_{LS} - \beta)\right)$ of the least squares estimator for $\beta = (\beta_0, \beta_1)'$?

(c) What is a maximum likelihood estimator for β? Will the MLE be unique for any realization of the sample? How many different values can your estimator possibly take? *Hint:* draw a diagram with the regression line and all possible values for (x_i, y_i). No need for formal derivations, just state verbally what the estimator does.

(d) Give the probability distribution of your estimator from (c). What is the probability $P(\hat{\beta}_{ML} \neq \beta)$ given the sample size n? Argue the the MLE is consistent.

(e) Given your answer to (d), what is the asymptotic variance of the MLE, $\lim_{n\to\infty} \mathrm{Var}\left(\sqrt{n}(\hat{\beta}_{ML} - \beta)\right)$? How does it compare to the asymptotic variance of the least squares estimator in (b)? How do you reconcile your result with the Gauss-Markov Theorem? *Hint:* No need for formal calculations, the answer should be clear from your results in (d).

Problem 3 (Ruud, exercise 13.8) Suppose the assumptions for the homoskedastic linear projection model and consistent variance estimation hold. Show that for the least-squares estimator of the residual variance,

$$\hat{s}_n^2 = \frac{1}{n-k} \sum_{i=1}^{n} (y_i - x_i'\hat{\beta})^2$$

we have $\sqrt{n}(\hat{s}_n^2 - \sigma^2) \xrightarrow{p} N(0, \mu_4 - \sigma^4)$ as $n \to \infty$, where $\mu_4 := \mathbb{E}[(y_i - x_i'\beta)^4]$, using the following steps:

(a) Show that

$$\hat{s}_n^2 - \sigma^2 = \sigma^2 \frac{k}{n-k} + \frac{n}{n-k}\left[\frac{1}{n}\sum_{i=1}^{n}[(y_i - x_i'\beta)^2 - \sigma^2] - (\hat{\beta} - \beta)'\frac{1}{n}\sum_{i=1}^{n} x_i x_i'(\hat{\beta} - \beta)\right]$$

(b) Show that

$$\sqrt{n}(\hat{s}_n^2 - \sigma^2) - \sqrt{n}\frac{1}{n}\sum_{i=1}^{n}[(y_i - x_i'\beta)^2 - \sigma^2] \xrightarrow{p} 0$$

(c) Show that

$$\sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^{n}[(y_i - x_i'\beta)^2 - \sigma^2]}{\sqrt{\mu_4 - \sigma^4}} \xrightarrow{d} N(0, 1)$$