FINAL REVIEW PROBLEMS 2

1. This problem extends the linear predictor concept to allow Y_i to be a vector. In addition, the dimension of Y_i may depend upon i. We shall develop inference that corresponds to the cluster option in Stata. Allowing the dimension of Y_i to depend upon i corresponds to allowing the number of observations within a cluster to vary across the clusters.

The key is random sampling of clusters:

$$(Y_i, X_i) \quad \text{i.i.d.} \qquad (i = 1, \ldots, n),$$

where Y_i is $H_i \times 1$ and X_i is $H_i \times K$. Here H_i, the size of cluster i, is a random variable. Due to the random sampling of clusters, H_i is i.i.d. Define the $K \times 1$ vector β to solve the following linear predictor problem:

$$\beta = \arg \min_{a \in \mathcal{R}^K} E[(Y_i - X_i a)'(Y_i - X_i a)].$$

(a) Define the prediction error

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

Show that

$$E(X_i' U_i) = 0.$$

(b) Let $\hat{\beta}$ denote the least-squares estimator:

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

where

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}, \quad Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}.$$

The number of rows in X and Y is $\sum_{i=1}^n H_i$; X has K columns, Y has one. Show that

$$\hat{\beta} \xrightarrow{p} \beta \quad \text{as} \quad n \to \infty.$$

(c) Use the arguments in Note 9 to show that

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, \Lambda).$$

Provide a formula for Λ.

(d) Explain how to use the limit distribution in (c) to obtain standard errors and confidence intervals corresponding to the ones Stata reports with the cluster option.

2. In class, we first developed the following special case of Note 9:

$$Q_i = R_i\gamma + V_i, \quad E(R_i'V_i) = 0 \qquad (i = 1, \ldots, n), \tag{1}$$

where Q_i is $H \times 1$ and R_i is $H \times K$. Then we developed the general case:

$$Q_i = R_i\gamma + V_i, \quad E(W_iV_i) = 0 \quad (i = 1, \ldots, n), \tag{2}$$

where W_i is $L \times H$ with $L \geq K$. Based on problem 1, it appears that we can regard the orthogonality condition $E(R_i'V_i)$ in (1) as holding by construction. But we used this framework to do inference with the Mundlak method, based on (within group) deviations. The justification was based on Note 4, where we developed the complete conditioning restrictions. If $E(R_i'V_i) = 0$ can hold by construction, it might seem that the complete conditioning restrictions are not needed. Are they needed? Explain.

3. It turns out that Vera's method (Final Review Problems 1, question 1) works fine. It is remarkable that all the benefits of the exact t-based confidence interval can be obtained simply by recognizing that

$$(b - \beta)/\sqrt{\text{SSR}}$$

is a pivot. Vera has been studying Note 8. She sees

$$\frac{l'(b - \beta)}{\text{SE}} \xrightarrow{d} \mathcal{N}(0, 1)$$

and wonders whether she could construct a useful confidence interval just by treating

$$\frac{l'(b - \beta)}{\text{SE}}$$

as a pivot. When working with the normal linear model, she could form a population distribution by choosing values for β and σ. But in Note 8, the population distribution is left very general. So Vera tries using the empirical distribution of her sample data as a population. She forms a random sample of size n as follows. Draw a number at random from the set $\{1, \ldots, n\}$ (assigning probability $1/n$ to each of those integers); the data on that cross-section unit supplies the first observation. Take an independent draw from the same set $\{1, \ldots, n\}$; the data on that cross-section unit supplies the second observation. Repeat, sampling with replacement, to get a sample of size n. Use this sample to construct

$$\frac{l'(b^{(1)} - b)}{\text{SE}^{(1)}}.$$

She draws J independent samples in this way, forms the absolute values

$$\left|\frac{l'(b^{(j)} - b)}{\text{SE}^{(\text{j})}}\right| \qquad (j = 1, \ldots, J),$$

and sorts them to find the 95^{th} percentile (.95 quantile) of their empirical distribution; i.e., she finds a value, which she labels CRIT, such that $.95 \cdot J$ of the absolute values are below CRIT and $.05 \cdot J$ of them are above CRIT. Then she argues that

$$[(l'b - \text{CRIT} \cdot \text{SE}),\ (l'b + \text{CRIT} \cdot \text{SE})]$$

can be used as an approximate .95 confidence interval for $l'\beta$. Do you think this procedure can be justified, in the sense that as $J \to \infty$ and $n \to \infty$ the probability that the interval covers the true value converges to .95? Explain your reasoning.