Preliminary Examination: LINEAR MODELS

Answer all questions and show all work.
Q1 is 30 points; Q2 is 35 points, and Q3 is 35 points.

1. Let $\mathbf{X}_1$ and $\mathbf{X}_2$ be $n \times p_1$ and $n \times p_2$ matrices of predictors whose columns are linearly independent to each other. We consider the linear regression model below:

$$\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are p_1- and p_2-dimensional vectors, respectively, and $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$.

a. Express the ordinary least square (OLS) estimator for $\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}$ using $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{Y}$.

b. Let

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} \\ \mathbf{G}_{21} & \mathbf{G}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{X}_1'\mathbf{Y} \\ \mathbf{X}_2'\mathbf{Y} \end{pmatrix}$$

be the OLS estimator found above in part (a). Find the explicit forms of $\mathbf{G}_{11}$, $\mathbf{G}_{12}$, $\mathbf{G}_{21}$, and $\mathbf{G}_{22}$.

c. Based on the results in part (b), show that $\hat{\boldsymbol{\beta}}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y}$ when $\mathbf{X}_1'\mathbf{X}_2 = \mathbf{0}$.

2. The theory of optimal linear *prediction* parallels closely the theory of optimal linear *estimation*. In this question you are asked to derive a fundamental result pertaining to the former.

Let $Y \in \mathbb{R}$ be a random response, and let $\mathbf{x} \in \mathbb{R}^p$ be a *random* predictor. We seek a linear predictor $\alpha + \mathbf{x}'\boldsymbol{\beta}$ that minimizes $E[(Y - \alpha - \mathbf{x}'\boldsymbol{\beta})^2]$ over all choice of $\alpha \in \mathbb{R}$ and $\boldsymbol{\beta} \in \mathbb{R}^p$. Such a predictor $\alpha + \mathbf{x}'\boldsymbol{\beta}$ is called a *best linear predictor* of Y.

Define $\mu_y = E(Y)$, $\boldsymbol{\mu}_x = E(\mathbf{x})$, $\sigma_y^2 = var(Y)$, $\mathbf{V}_{xx} = Cov(\mathbf{x})$, and $\mathbf{V}_{xy} = Cov(\mathbf{x}, Y) = V_{yx}' = Cov(Y, \mathbf{x})'$. Without loss of generality, we will write an arbitrary linear predictor in the form $\alpha + (\mathbf{x} - \boldsymbol{\mu}_x)'\boldsymbol{\beta}$.

a. Show that the optimal choice of α is simply $\hat{\alpha} = \mu_y$.

b. Show that if $\boldsymbol{\beta}_*$ is a solution to the linear system

$$\mathbf{V}_{xx}\boldsymbol{\beta} = \mathbf{V}_{xy},$$

then $\mu_y + (\mathbf{x} - \boldsymbol{\mu}_x)'\boldsymbol{\beta}_*$ is the best linear predictor of Y. [Hint: Without loss of generality, you may assume $\mu_y = 0$ and $\boldsymbol{\mu}_x = \mathbf{0}$ for this part.]

c. Derive an expression for a 95% prediction interval for Y based on the best linear predictor in part (b). State any assumptions necessary for the validity of your interval.

3. Given the dataset $\{\mathbf{x}_i, y_i\}_{i=1}^n$, we wish to obtain a bootstrapped least-squares estimate for $\boldsymbol{\beta}$ under the model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},\ E[\boldsymbol{\varepsilon}|\mathbf{X}] = \mathbf{0},\ Var[\boldsymbol{\varepsilon}|\mathbf{X}] = \sigma^2\mathbf{I}_n, \tag{1}$$

with $\mathbf{X} = [\mathbf{x}_1', \ldots, \mathbf{x}_n']'$ and $\mathbf{Y} = [y_1, \ldots, y_n]'$. Here, $\mathbf{X}$ is $n \times p$ with $rank(\mathbf{X}) = p$.

Bootstrapping is a type of resampling where samples of the *same size* n are drawn, *with replacement*, from a single original sample. Let $\mathbf{u}_i \equiv (\delta_{ij})_{j=1}^n$ be a $1 \times n$ vector with zero entries apart from $u_{ii} = 1$, then we can select the i-th row of $\mathbf{X}$ by $\mathbf{x}_i = \mathbf{u}_i\mathbf{X}$, or the i-th entry in $\mathbf{Y}$ by $y_i = \mathbf{u}_i\mathbf{Y}$. Thus, if $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$ are the corresponding bootstrap samples of $\mathbf{X}$ and $\mathbf{Y}$, respectively, we can represent the bootstrap sample by a matrix of $\mathbf{B}$ with rows that are similar to $\mathbf{u}_i$, but with i being sampled, and such that $\tilde{\mathbf{X}} = \mathbf{B}\mathbf{X}$ and $\tilde{\mathbf{Y}} = \mathbf{B}\mathbf{Y}$.

a. For the bootstrap sample $(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}})$, we will calculate the *OLS* estimator which is called the bootstrapped least square estimator for $\boldsymbol{\beta}$. Show that this bootstrapped least square estimator for $\boldsymbol{\beta}$, say $\tilde{\boldsymbol{\beta}}(\mathbf{B})$, is actually a *weighted* least square estimator from the original data. More specifically, show that $\tilde{\boldsymbol{\beta}}(\mathbf{B})$ is the generalized least square estimator for $\boldsymbol{\beta}$ in

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},\ E[\boldsymbol{\varepsilon}|\mathbf{X}, \mathbf{B}] = \mathbf{0},\ Var[\boldsymbol{\varepsilon}|\mathbf{X}, \mathbf{B}] = \sigma^2\mathbf{W}(\mathbf{B})^{-1},$$

where the matrix $\mathbf{W}(\mathbf{B})^{-1}$ is a function of the bootstrap sampling scheme, i.e., thus a function of $\mathbf{B}$, and show that $\mathbf{W}(\mathbf{B})^{-1}$ is a diagonal matrix.

b. Provide an expression for $\mathbf{w}(\mathbf{B}) \equiv \texttt{Diag}(\mathbf{W}(\mathbf{B})^{-1})$, the vector consisting of diagonal elements of $\mathbf{W}(\mathbf{B})^{-1}$. In addition, show that for its i-th element, we have

$$w(B)_i \sim Binomial\left(n, \frac{1}{n}\right),\ i = 1, \ldots, n.$$

What does $w(\mathbf{B})_i$ represent?

c. Since $\tilde{\boldsymbol{\beta}}(\mathbf{B})$ depends on $\mathbf{w}(\mathbf{B})$, let us denote $\mathbf{w}(\mathbf{B})$ as $\mathbf{w}$ and set $\tilde{\boldsymbol{\beta}}(\mathbf{w}) \equiv \tilde{\boldsymbol{\beta}}(\mathbf{B})$. Now regard $\tilde{\boldsymbol{\beta}}(\mathbf{w})$ as a function of $\mathbf{w}$ and use the delta method with a second order approximation around $E(\mathbf{w})$ to conclude that $E_w[\tilde{\boldsymbol{\beta}}(\mathbf{w})] \approx \hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is the OLS estimator for $\boldsymbol{\beta}$ in the original model in (1).

d. Argue that with a positive probability $\tilde{\boldsymbol{\beta}}(\mathbf{w})$ might not exist. What can go wrong? Be as specific as possible.