180.633 Econometrics
Midterm, 10am-12pm, March 25, Spring 2020
Note:
1 This is an open-book exam. But please finish it independently.
2 The TA will email you the exam 5 minutes before the exam.
3 Students should scan their answers and email back to tzhou11@jhu.edu within 20 minutes after the exam.
4 The TA will be available for questions on Skype or Zoom during the exam.

Directions:

1. Any theorem in Hansen's book may be invoked without proof, but should be cited in your proofs.
2. Show all of your work and explain your reasoning: partial credit is given for partial solutions.

QUESTIONS:
1. (10 points) Show that the conditional variance can be written as

$$\sigma^2(\boldsymbol{x}) = \mathbb{E}[y^2|\boldsymbol{x}] - \left(\mathbb{E}[y|\boldsymbol{x}]\right)^2.$$

2. (20 points) Suppose that

$$\boldsymbol{x} = \begin{pmatrix} 1 \\ x_2 \\ x_3 \end{pmatrix}$$

and $x_3 = \alpha_1 + \alpha_2 x_2$ is a linear function of x_2.

(a) Show that $\boldsymbol{Q}_{xx} = \mathbb{E}[\boldsymbol{x}\boldsymbol{x}']$ is not invertible.

(b) Use a linear transformation of $\boldsymbol{x}$ to find an expression for the best linear predictor of y given $\boldsymbol{x}$. (Be explicit, but do not just use the generalized inverse formula.)

3. (10 points) Show that if $\boldsymbol{X} = [\boldsymbol{X}_1 \ \boldsymbol{X}_2]$ and $\boldsymbol{X}_1'\boldsymbol{X}_2 = 0$, then the corresponding projection matrices satisfy $\boldsymbol{P} = \boldsymbol{P}_1 + \boldsymbol{P}_2$.

4. (5 points) Using the notations in Chapter 3, for which observations will $\widehat{\boldsymbol{\beta}}_{(-i)} = \widehat{\boldsymbol{\beta}}$?

5. (20 points) Consider an i.i.d. sample $\{y_i, \boldsymbol{x}_i\}$, $i = 1, \dots, n$, where $\boldsymbol{x}_i$ is $k \times 1$. Assume the linear conditional expectation model

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + e_i$$
$$\mathbb{E}[e_i|\boldsymbol{x}_i] = 0$$

Assume that $\boldsymbol{X}'\boldsymbol{X}/n = \boldsymbol{I}_k$ (orthonormal regressors). Consider the OLS estimator $\widehat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$.

(a) Find $\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}} = \mathsf{Var}\left(\widehat{\boldsymbol{\beta}}\right)$.

(b) In general, are $\widehat{\beta}_j$ and $\widehat{\beta}_\ell$ for $j \neq \ell$ correlated or uncorrelated? Explain.

(c) Find a sufficient condition so that $\widehat{\beta}_j$ and $\widehat{\beta}_\ell$ for $j \neq \ell$ are uncorrelated.

6. (20 points) The model is

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + e_i$$
$$\mathbb{E}[e_i|\boldsymbol{x}_i] = 0$$
$$\mathbb{E}[e_i^2|\boldsymbol{x}_i] = \sigma_i^2$$
$$\boldsymbol{\Omega} = \text{diag}\left\{\sigma_1^2, \dots, \sigma_n^2\right\}.$$

The parameter $\boldsymbol{\beta}$ is estimated both by OLS $\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y}$ and GLS $\widetilde{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{y}$. Let $\widehat{\boldsymbol{e}} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{e}} = \boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}$ denote the residuals. Let $\hat{R}^2 = 1 - \widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}/(\boldsymbol{y}^{*\prime}\boldsymbol{y}^*)$ and $\tilde{R}^2 = 1 - \tilde{\boldsymbol{e}}'\tilde{\boldsymbol{e}}/(\boldsymbol{y}^{*\prime}\boldsymbol{y}^*)$ denote the equation R^2 where $\boldsymbol{y}^* = \boldsymbol{y} - \bar{y}$.

If the error e_i is truly heteroskedastic,

(a) Show that $\mathsf{Cov}\left(\widehat{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\beta}} \,\middle|\, \boldsymbol{X}\right) = \mathsf{Var}\left(\widetilde{\boldsymbol{\beta}} \mid \boldsymbol{X}\right)$. Also find $\mathsf{Cov}\left(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}} \,\middle|\, \boldsymbol{X}\right)$.

(b) Will $\hat{R}^2$ or $\tilde{R}^2$ be smaller? Prove it.

7. (5 points) For the regression in-sample predicted value $\widehat{y}_i$ show that $\widehat{y}_i|\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{x}_i'\boldsymbol{\beta}, \sigma^2 h_{ii})$ where h_{ii} are the leverage values as in equation (3.41) in the textbook.

8. (10 points)

(a) In the *normal* regression model, let s^2 be the unbiased estimator of the error variance σ^2 from equation (4.26) in the textbook. Find $\mathsf{Var}(s^2)$. Show that $\mathsf{Var}(s^2)$ is strictly larger than the Cramér-Rao Lower Bound for σ^2.

(b) In the *linear* regression model (Assumption 4.2), provide a condition under which the variance of OLS $\widehat{\boldsymbol{\beta}}$ hits the Cramér-Rao lower bound for $\boldsymbol{\beta}$.