180.633 Econometrics
Final Exam, 1pm – 4pm, May 6th, 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) Take the model

$$\begin{aligned} y_i &= \boldsymbol{x}_{1i}'\boldsymbol{\beta}_1 + \boldsymbol{x}_{2i}'\boldsymbol{\beta}_2 + e_i \\ \mathbb{E}[e_i|\boldsymbol{x}_i] &= 0 \\ \mathbb{E}[e_i^2|\boldsymbol{x}_i] &= \sigma^2 \end{aligned}$$

where $\boldsymbol{x}_i = (\boldsymbol{x}_{1i}, \boldsymbol{x}_{2i})$ with $\boldsymbol{x}_{1i}$ is $k_1 \times 1$ and $\boldsymbol{x}_{2i}$ is $k_2 \times 1$. Consider the short regression

$$y_i = \boldsymbol{x}_{1i}'\widehat{\boldsymbol{\beta}}_1 + \widehat{e}_i$$

and define the error variance estimator

$$s^2 = \frac{1}{n-k}\sum_{i=1}^{n} \widehat{e}_i^2$$

Find $\mathbb{E}[s^2 \,|\, \boldsymbol{X}]$.

2. (10 points) The model is

$$\begin{aligned} y_i &= x_i\beta + e_i \\ \mathbb{E}[e_i|x_i] &= 0 \end{aligned}$$

where $x_i \in \mathbb{R}$. Consider the two estimators

$$\widehat{\beta} = \frac{\sum_i^n x_i y_i}{\sum_{i=1}^n x_i^2}$$
$$\widetilde{\beta} = \frac{1}{n}\sum_{i=1}^n \frac{y_i}{x_i}.$$

(a) Under the stated assumptions, are both estimators consistent for β?

(b) Are there conditions under which either estimators is efficient?
(Hint: consider the *Cauchy-Schwarz inequality*)

3. (10 points) In the linear projection model $y_i = \alpha + \boldsymbol{x}_i'\boldsymbol{\beta} + e_i$, consider the restriction $\boldsymbol{\beta} = \mathbf{0}$.

(a) Find the constrained least-squares (CLS) estimator of α under the restriction $\boldsymbol{\beta} = \mathbf{0}$.

(b) Find an expression for the efficient minimum distance estimator of α under the restriction $\boldsymbol{\beta} = \mathbf{0}$.

4. (10 points) Take the model

$$y_i = x_i\beta_1 + x_i^2\beta_2 + e_i$$
$$\mathbb{E}[e_i|x_i] = 0$$

where y_i is wages (dollars per hour) and x_i is age. Describe how you would test the hypothesis that the expected wage for a 40-year-old worker is \$20 an hour.

5. (10 points) You want to test $\mathbb{H}_0 : \boldsymbol{\beta}_2 = 0$ against $\mathbb{H}_1 : \boldsymbol{\beta}_2 \neq \mathbf{0}$ in the model

$$y_i = \boldsymbol{x}_{1i}'\boldsymbol{\beta}_1 + \boldsymbol{x}_{2i}'\boldsymbol{\beta}_2 + e_i$$
$$\mathbb{E}[\boldsymbol{x}_i e_i] = 0$$

You read a paper which estimates model

$$y_i = \boldsymbol{x}_{1i}'\widehat{\gamma}_1 + (\boldsymbol{x}_{2i} - \boldsymbol{x}_{1i})'\widehat{\gamma}_2 + \widehat{e}_i$$

and reports a test of $\mathbb{H}_0 : \boldsymbol{\gamma}_2 = \mathbf{0}$ against $\mathbb{H}_1 : \boldsymbol{\gamma}_2 \neq \mathbf{0}$. Is this related to the test you wanted to conduct?

6. (20 points) You have n i.i.d. observations (y_i, x_{1i}, x_{2i}), and consider two alternative regression models

(1)
$$y_i = \boldsymbol{x}_{1i}'\boldsymbol{\beta}_1 + e_{1i}$$
$$\mathbb{E}[\boldsymbol{x}_{1i}e_{1i}] = 0$$

(2)
$$y_i = \boldsymbol{x}_{2i}'\boldsymbol{\beta}_2 + e_{2i}$$
$$\mathbb{E}[\boldsymbol{x}_{2i}e_{2i}] = 0$$

where $\boldsymbol{x}_{1i}$ and $\boldsymbol{x}_{2i}$ have at least some different regressors. (For example, Eq. (1) is a wage regression on geographic variables and Eq. (2) is a wage regression on personal appearance measurements.) You want to know if model 1 or model 2 fits the data better. Define $\sigma_1^2 = \mathbb{E}[e_{1i}^2]$ and $\sigma_2^2 = \mathbb{E}[e_{2i}^2]$. You decide that the model with the smaller variance fits (e.g. , model 1 fits better if $\sigma_1^2 < \sigma_2^2$.) You decide to test for this by testing the hypothesis of equal fit $\mathbb{H}_0 : \sigma_1^2 = \sigma_2^2$ against the alternative of unequal fit $\mathbb{H}_1 : \sigma_1^2 \neq \sigma_2^2$. For simplicity, suppose that e_{1i} and e_{2i} are observed.

(a) Construct an estimator $\widehat{\theta}$ of $\theta = \sigma_1^2 - \sigma_2^2$.

(b) Find the asymptotic distribution of $\sqrt{n}\left(\widehat{\theta} - \theta\right)$ as $n \to \infty$.

(c) Find an estimator of the asymptotic variance of $\widehat{\theta}$.

(d) Propose a test of asymptotic size α of $\mathbb{H}_0$ against $\mathbb{H}_1$.

(e) Suppose the test accepts $\mathbb{H}_0$. Briefly, what is your interpretation?

7. (15 points) Consider the structural equation

(3)
$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + e_i$$

with x_i treated as endogenous so that $\mathbb{E}[x_i e_i] \neq 0$. Assume y_i and x_i are scalar. Suppose we also have a scalar instrument z_i which satisfies

$$\mathbb{E}[e_i | z_i] = 0$$

so in particular $\mathbb{E}[e_i] = 0$, $\mathbb{E}[z_i e_i] = 0$ and $\mathbb{E}[z_i^2 e_i] = 0$.

(a) Should x_i^2 be treated as endogenous or exgonenous?

(b) Suppose we have a scalar instrument z_i which satisfies

$$x_i = \gamma_0 + \gamma_1 z_i + u_i \tag{4}$$

with u_i independent of z_i and mean zero.
Consider using $(1, z_i, z_i^2)$ as instruments. Is this a sufficient number of instruments? (Would this be just-identified, over-identified, or under-identified)?

(c) Write out the reduced form equation for x_i^2. Under what condition on the reduced form parameters in Eq. (4) are the parameters in Eq. (3) identified?

8. (15 points) Consider the model

$$y_{1i} = \boldsymbol{x}_i'\boldsymbol{\beta} + e_i \tag{5}$$

$$\boldsymbol{x}_i = \boldsymbol{\Gamma}'\boldsymbol{z}_i + \boldsymbol{u}_i$$

$$\mathbb{E}[\boldsymbol{z}_i e_i] = \boldsymbol{0} \tag{6}$$

$$\mathbb{E}[\boldsymbol{z}_i \boldsymbol{u}_i'] = \boldsymbol{0} \tag{7}$$

with y_{1i} scalar and $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$ each a k vector. You have a random sample $(y_{1i}, \boldsymbol{x}_i, \boldsymbol{z}_i : i = 1, \dots, n)$. Take the control function equation

$$e_i = \boldsymbol{u}_i'\boldsymbol{\gamma} + \varepsilon_i$$

$$\mathbb{E}[\boldsymbol{u}_i \varepsilon_i] = \boldsymbol{0} \tag{8}$$

Inserting into the structural equation we find

$$y_{1i} = \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{u}_i'\boldsymbol{\gamma} + \varepsilon_i.$$

If $\boldsymbol{u}_i$ is observed, the control function estimator $\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}\right)$ is OLS estimation of this equation.

(a) If $\boldsymbol{u}_i$ is observed, show that $\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}\right)$ is consistent.

(b) If $\boldsymbol{x}_i$ is endogenous, write down the formula of the IV estimator $\widehat{\boldsymbol{\beta}}_{\mathsf{IV}}$.

(c) Suppose now $\boldsymbol{u}_i$ is not observed. Assume $\boldsymbol{x}_i$ is *endogenous*, i.e. $\mathbb{E}[\boldsymbol{x}_i e_i] \neq 0$ with assumptions in (6) and (7) still maintained. Mimicking the control function procedure stated in the question, explain how to conduct a control function approach to obtain a consistent estimator $\widehat{\boldsymbol{\beta}}_{\mathsf{CF}}$.

(d) Show that $\widehat{\boldsymbol{\beta}}_{\mathsf{IV}}$ in (b) and $\widehat{\boldsymbol{\beta}}_{\mathsf{CF}}$ in (c) are numerically identical, i.e. $\widehat{\boldsymbol{\beta}}_{\mathsf{IV}} = \widehat{\boldsymbol{\beta}}_{\mathsf{CF}}$.