Comprehensive Examination
Quantitative Methods
Spring, 2021

Instruction: This exam consists of three parts. You are required to answer all the questions in all the parts.

Grading policy:

1. Each part will be graded separately, and there are four possible results in each part: H (honor pass), P (PhD pass), M (master pass), and F (fail).

2. Each part contains a precise grade determining algorithm.

3. The grades from the three parts will be summarized in the descending order, after which the overall grade will be determined using the algorithm summarized in the table below:

Highest	Middle	Lowest	**Overall**
H	H	H	**H**
H	H	P	**H**
H	H	M	**P**
H	H	F	**M**
H	P	P	**P**
H	P	M	**P**
H	P	F	**M**
H	M	M	**M**
H	M	F	**M**
H	F	F	**F**
P	P	P	**P**
P	P	M	**P**
P	P	F	**M**
P	M	M	**M**
P	M	F	**M**
P	F	F	**F**
M	M	M	**M**
M	M	F	**F**
M	F	F	**F**
F	F	F	**F**

Part I - 203A

Instructions for Part I: Solve every question. For every question in this part, your answer should be numerical; an abstract formula will not be accepted as an answer.

Grading policy for Part I: Your grade in this part of the exam is based on the total points that you earn. Below is how the grade is determined. In order to avoid any confusion arising from the difference between strict and weak inequalities, the grade assignment algorithm is presented in terms of mathematical inequalities. Note that $<$ denotes a strict inequality, and $\leq$ denotes a weak inequality. Let T denote the total number of points.

1. If $T \geq 25$, you will get H.
2. If $20 \leq T < 25$, you will get P.
3. If $15 \leq T < 20$, you will get M.
4. If $T < 15$, you will get F.

Question 1 (2 pts.) *No derivation is required; your derivation will not be read anyway.* Suppose that $X \sim N(0,1)$, and let

$$f_n(x) = x^2 + \frac{1}{n}\sin(x).$$

What is $\lim_{n\to\infty} E[f_n(X)]$? Your answer should be a number.

Question 2 (5 pts.) *No derivation is required; your derivation will not be read anyway.* Suppose that $X_1, \ldots, X_n$ *i.i.d.* $N(\mu, \sigma^2)$. We assume that σ^2 is known to be equal to 1. We have $H_0 : \mu = 0$ vs. $H_1 : \mu > 0$. Suppose that C is a critical region such that (1) the test of the form "Reject H_0 if $(X_1, \ldots, X_n) \in C$" is uniformly most powerful test; and (2) the probability of rejecting H_0 when $\mu = 0$ is 5%. What is the power of the test (i.e., the probability of rejecting H_0) when $\mu = 0.8225$ and $n = 4$? Your answer should be a number.

Question 3 (5 pts.) *No derivation is required; your derivation will not be read anyway.* Suppose that X has the MGF equal to

$$M(t) = E[\exp(tX)] = \exp\left(\frac{t^2}{2}\right).$$

Let A_n and B_n denote the events defined by

$$A_n = \left\{X \leq -\frac{1}{n}\right\},$$
$$B_n = \left\{X < \frac{1}{n}\right\}.$$

Furthermore, let A and B denote the events defined by

$$A = \lim_{n\to\infty} A_n,$$
$$B = \lim_{n\to\infty} B_n.$$

What is $P(B) - P(A)$? Your answer should be a number.

Question 4 (5 pts.) *No derivation is required; your derivation will not be read anyway.* Consider $X_1, \ldots, X_n$ *i.i.d.* $N(0,1)$. What is $\lim_{n\to\infty} \Pr\left(\overline{X}_n \geq 1\right)$? Your answer should be a number.

Question 5 (5 pts.) *No derivation is required; your derivation will not be read anyway.* Let $X_1, \ldots, X_n$ be a random sample of size n from $N(3,2)$, and let $\bar{X}$ denote the sample average. It can be shown that $\sqrt{n}\left(\left(\bar{X}\right)^2 - 9\right) \xrightarrow{D} N(\mu, \sigma^2)$ for some μ and σ^2? What are μ and σ^2? Your answer should be numbers.

Question 6 (3 pts.) *No derivation is required; your derivation will not be read anyway.* Let X_n denote a sequence of random variables such that the PDF f_n of X_n is given by

$$f_n(x) = \frac{1}{2n} 1(|x| \leq n).$$

(a) (1 pts.) Let $0 < B < \infty$ be given. What is $\lim_{n\to\infty} \Pr[|X_n| \leq B]$? Your answer should be a number.

(b) (2 pts.) Is $X_n = O_p(1)$? This is a Yes-or-No question; no explanation is necessary

Question 7 (5 pts.) True-false questions. No explanation is necessary, and you are not allowed to provide any explanation.

(a) (1 pt.) If $X_n = 2 + o_p(1)$, then $E[X_n] = 2 + o(1)$

(b) (1 pt.) If $X_n = 2 + o_p(1)$, then $X_n^2 = 4 + o_p(1)$

(c) (1 pt.) If $X_n \xrightarrow{D} N(0,4)$, then $E[X_n] = o(1)$

(d) (1 pt.) If $X_n \xrightarrow{D} N(0,4)$, then $E[X_n^2] = 4 + o(1)$

(e) (1 pt.) If $X_n \xrightarrow{D} N(0,1)$, then $X_n^2 \xrightarrow{D} \chi^2(1)$

Part II - 203B

Instruction for Part II: Solve every question.

Grading policy for Part II: Your grade in this part of the exam is based on the total points that you earn. Below is how the grade is determined. In order to avoid any confusion arising from the difference between strict and weak inequalities, the grade assignment algorithm is presented in terms of mathematical inequalities. Note that $<$ denotes a strict inequality, and $\leq$ denotes a weak inequality. Let T denote the total number of points.

1. If $T \geq 80$, you will get H.
2. $60 \leq T < 80$, you will get P.
3. $45 \leq T < 60$, you will get M.
4. If $T < 45$, you will get F.

Question 1 (50 points) Consider a LATE model where we have potential outcomes $(Y(0), Y(1))$, potential binary treatment decisions $(D(0), D(1))$, a binary instrument Z, covariates $X \in \mathbf{R}^d$ and we observe (Y, D, X, Z), with Y and D given by

$$Y = Y(0) + D(Y(1) - Y(0)) \qquad D = D(0) + Z(D(1) - D(0)).$$

Throughout this problem assume $D(1) \geq D(0)$ and $(Y(0), Y(1), D(0), D(1)) \perp Z|X$ – i.e., Z is independent of $(Y(0), Y(1), D(0), D(1))$ conditional on X. (**Note:** This is *not the same* as Z being independent of $(Y(0), Y(1), D(0), D(1))$).

(a) (5 points) Show that under the assumptions of the problem it follows that

$$P(D = 1|Z = 1, X) = P(D(1) = 1|X)$$
$$P(D = 1|Z = 0, X) = P(D(0) = 1|X)$$

(b) (10 points) Show that under the assumptions of the problem it follows that

$$P(D = 1|Z = 1, X) - P(D = 1|Z = 0, X) = P(D(1) > D(0)|X)$$

(c) (10 points) Show that under the assumptions of the problem it follows that

$$E[Y(1)(D(1) - D(0))|X] = E[YD|Z = 1, X] - E[YD|Z = 0, X]$$
$$E[Y(0)(D(1) - D(0))|X] = E[Y(1 - D)|Z = 0, X] - E[Y(1 - D)|Z = 1, X]$$

(d) (10 points) Let $\text{LATE}(X) = E[Y(1) - Y(0)|D(1) > D(0), X]$. Using the results from parts (b) and (c) show that $\text{LATE}(X)$ is identified – i.e. obtain an expression for $\text{LATE}(X)$ as a function of conditional moments of (Y, D, Z, X). (**Hint:** Recall that under the monotonicity condition $D(1) - D(0) \in \{0, 1\}$.)

(e) (15 points) Suppose that X is also binary so that $X \in \{0,1\}$ and in addition

$$P(X = 1|D(1) > D(0)) = \frac{1}{2}$$

Propose an estimator for LATE $\equiv E[Y(1)-Y(0)|D(1) > D(0)]$ based on a sample $\{Y_i, D_i, Z_i, X_i\}_{i=1}^n$ and establish its consistency. Be explicit about any assumptions you need. (**Note**: LATE may not be the same as LATE(X)!).

Question 2 (50 points) Consider a panel data model in which we observe $\{Y_{it}, X_{it}\}$ with $Y_{it} \in \mathbf{R}$, $X_{it} \in \mathbf{R}^d$, $1 \leq i \leq n$, and $1 \leq t \leq 2$ (i.e. two time periods). Further suppose that

$$Y_{it} = X_{it}'\beta + \alpha_i + \gamma_t + \varepsilon_{it},$$

and in addition $E[\varepsilon_{it}|X_{i1}, X_{i2}, \gamma_1, \gamma_2] = 0$. Throughout the problem, let $\Delta Y_{i2} = Y_{i2} - Y_{i1}$, $\Delta X_{i2} = X_{i2} - X_{i1}$, $\Delta\gamma = \gamma_2 - \gamma_1$, and $\Delta\varepsilon_2 = \varepsilon_2 - \varepsilon_1$.

(a) (5 points) Show that under the assumptions of the problem $E[\Delta\varepsilon_2] = 0$.

(b) (10 points) Show that under the assumptions of the problem $\Delta\gamma$ and ΔX_2 satisfy

$$E\left[(\Delta Y_{i2} - \Delta\gamma - \Delta X_{i2}'\beta)\begin{pmatrix} 1 \\ \Delta X_{i2} \end{pmatrix}\right] = 0.$$

(c) (10 points) Based on the moment restrictions from part (b) a researcher proposes estimating β by regressing ΔY_2 on ΔX_2 and a constant – i.e. the estimator solves

$$(\hat{\beta}_n, \hat{c}_n) \equiv \arg \min_{b \in \mathbf{R}^d, c \in \mathbf{R}} \frac{1}{n}\sum_{i=1}^n (\Delta Y_{i2} - c - \Delta X_{i2}'b)^2.$$

For $\Delta\bar{Y} = \frac{1}{n}\sum_{i=1}^n \Delta Y_{i2}$ and $\Delta\bar{X} = \frac{1}{n}\sum_{i=1}^n \Delta X_{i2}$ show that $\hat{\beta}_n$ also solves

$$\hat{\beta}_n \equiv \arg\min_{b \in \mathbf{R}^d} \frac{1}{n}\sum_{i=1}^n ((\Delta Y_{i2} - \Delta\bar{Y}) - (\Delta X_{i2} - \Delta\bar{X})'b)^2.$$

(**Hint:** You can solve out for $\hat{c}_n$ or use results for partitioned regression)

(d) (10 points) What is the appropriate rank condition for ensuring that the estimator $\hat{\beta}_n$ from part (c) is indeed consistent for β?

(e) (15 points) Formally establish that $\hat{\beta}_n$ is consistent. Carefully state any assumptions you need for this end. (**Hint:** Recall $T = 2$ is fixed and asymptotics are as $n \to \infty$, so be careful with terms such as $\Delta\bar{X}_2$).

Part III - 203C

Instruction for Part III: Solve every question. The total number of points is 100.

Grading policy for Part III: Your grade in this part of the exam is based on the total points that you earn. Below is how the grade is determined. In order to avoid any confusion arising from the difference between strict and weak inequalities, the grade assignment algorithm is presented in terms of mathematical inequalities. Note that $<$ denotes a strict inequality, and $\leq$ denotes a weak inequality. Let T denote the total number of points.

1. If $T \geq 85$, you will get H.
2. $60 \leq T < 85$, you will get P.
3. $45 \leq T < 60$, you will get M.
4. If $T < 45$, you will get F.

Question 1 (50 points)

Consider a model where the value of a variable Y is determined from variables $X_0,\ X_1$, and Z according to the model

$$Y = \begin{cases} Y_0^* & if \quad Z = 0 \\ \\ Y_1^* & if \quad Z = 1 \end{cases}$$

where for some unknown, differentiable functions $g : R \to R$ and $s : R \to R$, and for a vector of unobservable variables $(\varepsilon_0, \varepsilon_1)$

$$Y_0^* = g(X_0) + \varepsilon_0$$

and

$$Y_1^* = s(X_1) + \varepsilon_1$$

Assume that (i) the distribution of the random vector (Y, X_0, X_1, Z) is known, (ii) the conditional densities of $(\varepsilon_0, \varepsilon_1)$ are such that for all (x_0, x_1) at which they are defined and for all $(\varepsilon_0, \varepsilon_1)$

$$f_{\varepsilon_0,\varepsilon_1|(X_0,X_1,Z)=(x_0,x_1,0)}(\varepsilon_0, \varepsilon_1) = f_{\varepsilon_0,\varepsilon_1|Z=0}(\varepsilon_0, \varepsilon_1)$$

and

$$f_{\varepsilon_0,\varepsilon_1|(X_0,X_1,Z)=(x_0,x_1,1)}\left(\varepsilon_0,\varepsilon_1\right)=f_{\varepsilon_0,\varepsilon_1|Z=1}\left(\varepsilon_0,\varepsilon_1\right)$$

(iii) $f_{\varepsilon_0,\varepsilon_1|Z=0}$ and $f_{\varepsilon_0,\varepsilon_1|Z=1}$ are everywhere strictly positive but of unknown values, and (iv) the support of Z is $\{0,1\}$.

(a; 10 points) Obtain an expression for $\Pr\left(Y\leq y\mid X_0,X_1\right)$ in terms of the functions g and s and the conditional densities of $\left(\varepsilon_0,\varepsilon_1\right)$.

(b; 5 points) Obtain an expression for the expectation of Y, in terms of the functions g and s and the conditional densities of $\left(\varepsilon_0,\varepsilon_1\right)$.

(c; 15 points) Is the function g identified? If your answer is YES, provide a proof. If your answer is NO, provide a counter example, specify any features of g that are identified, and prove that these features are indeed identified. (Since no assumptions have been made about $\left(X_0,X_1\right)$, you should consider different cases.)

(d; 10 points) Is the conditional density $f_{\varepsilon_0,\varepsilon_1|Z=0}$ identified? If your answer is YES, provide a proof. If your answer is NO, provide a counter example, specify any features of $f_{\varepsilon_0,\varepsilon_1|Z=0}$ that are identified, and prove that these features are indeed identified.

(e; 10 points) Suppose that, in addition to (i)-(iv), it is also known that

$$Med\left(\varepsilon_0|Z=0\right)=0$$

Answer (c) and (d) under this new situation.

Question 2 (50 points)

Consider again the model in Question 1, except that now for some parameters α_0, α_1, β_0, and β_1,

$$g\left(X_0\right)=\alpha_0+\beta_0\ X_0$$

and

$$s\left(X_1\right)=\alpha_1+\beta_1\ X_1$$

and for η distributed $N\left(0,\sigma^2\right)$ independently of $\left(X_0,X_1\right)$.

$$Z = \begin{cases} 0 & if \quad \gamma_0 + \gamma_1 X_0 \geq \eta \\ & \\ 1 & \qquad otherwise \end{cases}$$

Assume that (X_0, X_1) is continuously distributed with support R^2 and the distribution of $(\varepsilon_0, \varepsilon_1)$ is $N(\mu(z), \Omega(z))$, where for $z = 0, 1, \quad \mu(z) = (\mu_1(z), \mu_2(z))$ and

$$\Omega(z) = \begin{pmatrix} \omega_{11}(z) & \omega_{12}(z) \\ \omega_{21}(z) & \omega_{22}(z) \end{pmatrix}$$

are parameters of unknown values.

(a; 10 points) Determine what functions of the parameters are identified and provide proofs for your claim.

(b; 10 points) Suppose that you are given a random sample $\{Y^i, X_0^i, X_1^i, Z^i\}_{i=1}^N$ generated from the above model. After re-parameterizing the model in terms of parameters that can be identified, describe how to estimate the identified parameters by a Maximum Likelihood method.

(c; 10 points) Specify in detail the asymptotic properties of the estimator you proposed in (b).

(d; 20 points) Describe in detail how to use $\{Y^i, X_0^i, X_1^i, Z^i\}_{i=1}^N$ to test the hypothesis that $\beta_0 = 1$ and $\beta_1 = 1$ versus the alternative that at least one of the equalities is not satisfied.