Comprehensive Examination
Quantitative Methods
Fall, 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 a random variable X has a uniform distribution on $(0,1)$, i.e., the PDF $f(x)$ of X is $1(0 < x < 1)$. Consider the following sequence of functions defined on $[0,1]$:

$$g_n(x) = \begin{cases} n - n^2 x & 0 \leq x \leq \frac{1}{n} \\ 0 & \frac{1}{n} \leq x \leq 1 \end{cases}$$

Is $\lim_{n\to\infty} E[g_n(X)] = E[\lim_{n\to\infty} g_n(X)]$? Your answer should be either Yes or No.

Question 2 (5 pts.) *No derivation is required; your derivation will not be read anyway.* Suppose that

$$Y = X^2 + \varepsilon$$

where X and ε are independent standard normal random variables. What is $E[1(X > 0)Y]$? Your answer should be a number.

Question 3 (5 pts.) *No derivation is required; your derivation will not be read anyway.* Suppose that a random variable X has a uniform distribution on $(0,1)$, i.e., the PDF $f(x)$ of X is $1(0 < x < 1)$. Let A_n and B_n denote the events defined by

$$A_n = \left\{ X \leq \frac{1}{2} - \frac{1}{n} \right\},$$
$$B_n = \left\{ X < \frac{1}{2} + \frac{1}{n} \right\}.$$

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 $P(B)$? What is $P(A)$? Your answers should be numbers.

Question 4 (5 pts.) *No derivation is required; your derivation will not be read anyway.* Let (X, Y) be a two dimensional random vector with joint pdf

$$f_{X,Y}(x, y) = \begin{cases} e^{-y} & y > x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Let $\varphi(X)$ denote the solution to $\min_g E\left[(Y - g(X))^2\right]$. What is $\varphi(X)$? Your answer should be a mathematical characterization of $\varphi(X)$. Non-mathematical verbal expression will not be accepted as an answer.

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 $N(1, 1)$, and let $\bar{X}$ denote the sample average. It can be shown that $\sqrt{n}\left(\left(\bar{X}\right)^3 - 1\right) \xrightarrow{D} N(\mu, \sigma^2)$ for some μ and σ^2? What are μ and σ^2? Your answer should be numbers.

Question 6 (3 pts.) *This is a Yes-or-No question; no explanation is necessary.* Let X_n denote a sequence of random variables such that the PDF f_n of X_n is given by

$$f_n(x) = \frac{1}{n\sqrt{2\pi}} \exp\left(-\frac{x^2}{2n^2}\right).$$

Is $X_n = O_p(1)$? Your answer should be either Yes or No.

Question 7 (5 pts.) *No derivation is required for the questions below; your derivation will not be read anyway.* Consider $X_1, \ldots, X_4$ *i.i.d.* $N(\mu, \sigma^2)$. We assume that σ^2 is known to be 1. We have $H_0 : \mu = 0$ vs. $H_1 : \mu = 2$, and we decided to use the Neyman-Pearson test such that the probability of type I error is 5%. Suppose that you observe $X_1 = 1$, $X_2 = 1$, $X_3 = 1$, $X_4 = 0.5$. Do you reject H_0 or not? Your answer should be either Reject or Do not reject. Hint: If $Z \sim N(0, 1)$, we have $\Pr(|Z| > 1.96) = 0.05$ and $\Pr(Z > 1.645) = 0.05$.

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 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 $1 \leq t \leq T$. Let $Y_i \equiv (Y_{i1}, \ldots, Y_{iT})$, $X_i \equiv (X_{i1}, \ldots, X_{iT})$, and

$$Y_{it} = X_{it}'\beta_0 + \gamma_t + \varepsilon_{it},$$

with ε_{it} satisfying $E[\varepsilon_{it}|X_i] = 0$. In what follows, assume $\{Y_i, X_i\}_{i=1}^n$ are i.i.d. across i and, for any $1 \leq t \leq T$, let $\bar{Y}_t \equiv \frac{1}{n}\sum_{i=1}^n Y_{it}$ and $\bar{X}_t \equiv \frac{1}{n}\sum_{i=1}^n X_{it}$. All asymptotic statements should be understood as $n \to \infty$ with T fixed.

(a) (5 points) Show that, under the stated assumptions, it follows that

$$E\left[((Y_{it} - \bar{Y}_t) - (X_{it} - \bar{X}_t)'\beta_0)X_{it}\right] = 0.$$

Clearly state what assumption you employ at each step in your derivations.

(b) (10 points) State the rank condition necessary for identifying β_0.

(c) (10 points) Provide a concrete example of a regressor that would cause the rank condition in part (b) to fail and mathematically explain why it fails. (**Hint:** Recall, for instance, that when including individual fixed effects, we cannot include gender as a covariate).

(d) (10 points) Propose an estimator for β_0 and establish its consistency clearly stating what assumptions you need.

(e) (15 points) Can γ_t be consistently estimated? If yes, then provide a consistent estimator and establish its consistency. If not, then explain why not.

Question 2 (50 points) Let $Y_i \in \mathbf{R}$, $X_i \in \mathbf{R}^d$, $D_i \in \mathbf{R}$, $Z_i \in \mathbf{R}$, and suppose that we have

$$Y_i = X_i'\beta_0 + D_i\gamma_0 + \varepsilon_i$$

where β_0 and γ_0 are unknown, $E[\varepsilon_i|X_i, Z_i] = 0$, and D_i and ε_i are potentially correlated. Throughout the problem, assume we have an i.i.d. sample $\{Y_i, X_i, D_i, Z_i\}_{i=1}^n$.

(a) (10 points) State the rank condition for identifying (β_0, γ_0).

(b) (10 points) Suppose that $\gamma_0 \in \mathbf{R}$ is known, and define the estimator $\hat{\beta}_n$ as

$$\hat{\beta}_n \equiv \arg\min_{b \in \mathbf{R}^d} \frac{1}{n}\sum_{i=1}^n (Y_i - X_i'b - D_i\gamma_0)^2. \qquad (1)$$

Establish that $\hat{\beta}_n$ is consistent for β_0. Clearly state what assumptions you need.

(c) (10 points) Carefully derive the probability limit for the following statistic

$$T_n \equiv \frac{1}{n}\sum_{i=1}^n (Y_i - X_i'\hat{\beta}_n - D_i\gamma_0)Z_i, \qquad (2)$$

where $\hat{\beta}_n$ was defined in (1). Clearly state any assumptions you need to establish the result. (**Note:** We are still maintaining that γ_0 is known).

(d) (15 points) Clearly stating any assumptions that you need, establish the following asymptotic expansion for the statistic $\sqrt{n}T_n$:

$$\sqrt{n}T_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(Z_i\varepsilon_i - E[Z_iX_i'](E[X_iX_i'])^{-1}X_i\varepsilon_i\right) + o_P(1).$$

(**Note**: We are still maintaining that γ_0 is known).

(e) (5 points) Now drop the assumption that γ_0 is known, and instead suppose we want to test the following null and alternative hypotheses

$$H_0 : \gamma_0 = 0 \qquad\qquad H_1 : \gamma_0 \neq 0.$$

Explain how result (d) could be used to build a test for this hypothesis. No formal results are necessary, but a clear explanation should be provided. (**Hint:** Suppose we plug in $\gamma_0 = 0$ in (2): What happens to $\sqrt{n}T_n$ asymptotically if $\gamma_0 = 0$? What happens to T_n asymptotically if $\gamma_0 \neq 0$?)

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.contin
4. If $T < 45$, you will get F.

Question 1 (50 points)

Consider a model specified as

$$Y_1 = \begin{cases} 1 & if \qquad X_1 + g(X_2) - \varepsilon_1 \geq 0 \\ 0 & otherwise \end{cases}$$

and if $Y_1 = 1$,

$$Y_2 = m(X_2) + \varepsilon_2$$

where the random variables Y_1, Y_2, X_1, and X_2 are observed and the random variables ε_1 and ε_2 are unobserved. Let $F_{(\varepsilon_1,\varepsilon_2)|(X_1,X_2)}(\varepsilon_1, \varepsilon_2 \mid x_1, x_2)$ denote the cumulative distribution function of $(\varepsilon_1, \varepsilon_2)$, conditional on $(X_1, X_2) = (x_1, x_2)$. Assume that, when defined, $F_{(\varepsilon_1,\varepsilon_2)|(X_1,X_2)}$ is differentiable and strictly increasing in $(\varepsilon_1, \varepsilon_2)$, but otherwise unknown. Assume also that the functions m and g are continuous but otherwise unknown, and that the support of X_1 is R. Denote by $p(x_1, x_2)$ the probability density of (X_1, X_2).

(a; 10) Obtain an expression, in terms of the unknown functions and distributions, for the joint probability density of (Y_2, Y_1) conditional on $(X_1, X_2) = (x_1, x_2)$ when $Y_1 = 1$.

(b; 10) Obtain an expression for $E\left[Y_2 \middle| \, Y_1 = 1, \ (X_1, X_2) = (x_1, x_2)\right]$ in terms of the unknown functions and distributions.

(b; 5) Obtain an expression for $E\left[Y_1 \middle| \, X_1 = x_1\right]$ in terms of the unknown functions and distributions.

(d; 25) Assuming that $(\varepsilon_1, \varepsilon_2)$ is distributed independently of (X_1, X_2), determine what features, if any, of the unknown functions and distributions are identified. Provide proofs.

Question 2 (50 points)

Consider a parametric and slightly different version of the model in Question 1, where

$$Y_1 = \begin{cases} 1 & if \qquad \alpha_1 + \beta_1 \ X_1 + \gamma \ X_2 - \varepsilon_1 \geq 0 \\ 0 & otherwise \end{cases}$$

and if $Y_1 = 1$,

$$Y_2 = \begin{cases} 1 & if \qquad \alpha_2 + \beta_2 \ X_2 - \varepsilon_2 \geq 0 \\ 0 & if \qquad \alpha_2 + \beta_2 \ X_2 - \varepsilon_2 < 0 \end{cases}$$

Assume that $(\varepsilon_1, \varepsilon_2)$ is distributed independently of (X_1, X_2) with a $N(\mu, \Omega)$ distribution, where $\mu = (\mu_1, \mu_2)$ and

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

are parameters of unknown values and that (X_1, X_2) is distributed $N\left(\widetilde{\mu}, \widetilde{\Omega}\right)$, where $\widetilde{\mu} = (\widetilde{\mu}_1, \widetilde{\mu}_2)$ and

$$\widetilde{\Omega} = \begin{pmatrix} \widetilde{\omega}_{11} & \widetilde{\omega}_{12} \\ \widetilde{\omega}_{21} & \widetilde{\omega}_{22} \end{pmatrix}$$

are also parameters of unkown 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 $\{Y_1^i, Y_2^i, X_1^i, X_2^i\}_{i=1}^N$ generated from the above model. After re-parameterizing the model, if needed, in terms of parameters that can be identified, describe how to estimate the identified parameters by a Maximum Likelihood method.

(c; 15 points) What are the asymptotic properties of the estimators that you proposed in (b)? Sketch a proof.

(d; 15 points) Describe in detail how to use $\{Y_1^i, Y_2^i, X_1^i, X_2^i\}_{i=1}^N$ to test the hypothesis that $\gamma = 0$ versus $\gamma \neq 0$.