Panteion University
Applied Econometrics

Gregory Kordas
June 30, 2025

Final Exam

<u>Instructions:</u> The in-class final examination will include 3 of the following questions. You may attach to your exam any computer output you may have prepared.

1. [**Joint and Conditional Probabilities**]. Let

$$f_{XY}(x,y) = c(2x+3y), \qquad 0 < x < 1,\ 0 < y < 1.$$

be the joint probability density function of X and Y.

(a) Find c that makes $f(x,y)$ a valid probability density function.
(b) Find $g_{Y|X}(y|x)$, the conditional probability density function of $Y|X$.
(c) Find $\Pr(\frac{1}{3} < X < \frac{2}{3} | Y = \frac{2}{3})$.
(d) Find $\text{Cov}(X,Y)$, the covariance of X and Y.
(e) Are X and Y stochastically independent? Justify your answer.
(f) Let $Z = X^2 + Y^2$. Find $E(Z)$, the expected value of Z.

2. [**Linear Regression Model under Endogeneity**]. Consider the linear regression model

$$y = X\beta + u$$

where, y is an $n \times 1$ vector, X is an $n \times k$ matrix of regressors (including an intercept), β is a $k \times 1$ vector of coefficients, and u is an $n \times 1$ vector of errors.

(a) (10 points) State the classical assumptions and briefly explain them.
(b) (10 points) Which of the above assumptions is violated when a regressor is endogenous? Give an example of a regression in which the problem is likely to arise.
(c) (10 points) What are the properties of the OLS estimates under endogeneity?
(d) (10 points) Which estimator should you use in this case, and what are its properties?

3. [**Least Squares Identities**]. Prove that in the linear regression model $y = X\beta + u$ where X includes an intercept (a column of 1's as the first regressor), the OLS plane $\widehat{y} = X\widehat{\beta}$ has the following mathematical properties:

(a)

$$\overline{x}'\widehat{\beta} = \overline{y}.$$

where $\overline{x} = (1, \overline{x}_1, ..., \overline{x}_k)'$ is the $k \times 1$ vector of means of the independent variables x_j, $j = 1, ..., k$. This means that the point $(\overline{y}, \overline{x}) \in \mathbb{R}^{k+1}$ satisfies the normal equations, and therefore the OLS plane always passes through the sample means when the regression includes a conscant term. We say that OLS passes through the "center-of-gravity" $(\overline{y}, \overline{x})$ of the sample.

(b)

$$\overline{\widehat{y}} = \overline{y},$$

that is, the mean of the fitted values $\widehat{y}$ equals the mean of y.

(c)

$$1'\widehat{u} = \overline{u} = 0,$$

that is, the sum and the mean of the OLS residuals is zero.

(d)

$$\widehat{y}'\widehat{u} = 0 \qquad \text{or} \qquad \widehat{y} \perp \widehat{u}.$$

that is, the OLS fitted values $\widehat{y}$ and the OLS residulas $\widehat{u}$ are orthogonal vectors.

[Hint: See Stavrinos, ch.3.]

4. [**Long and Short Regressions**].

(a) (15 points) Assume that the true linear regression model explaining y is given by

$$y = X_1\beta_1 + X_2\beta_2 + u$$

where, y is an $n \times 1$ vector, X_1 is a $n \times k_1$ matrix of regressors (including an intercept), X_2 is a $n \times k_2$ matrix of regressors, β_1 is a $k_1 \times 1$ vector of coefficients, β_2 is a $k_2 \times 1$ vector of coefficients, and u is an $n \times 1$ vector of errors. Instead of estimating the true model, we estimate by OLS the *short* model

$$y = X_1\beta_1 + u.$$

What are the properties of the OLS estimate $\widehat{\beta}_1$?

Hint: Write the OLS estimator for β_1 and compute its expectation using the true model for y.

(b) (15 points) Now consider the opposite situation where the true model for y is given by

$$y = X_1\beta_1 + u$$

we estimate by OLS the *long* model

$$y = X_1\beta_1 + X_2\beta_2 + u$$

What are the properties of the OLS estimate $\widehat{\beta}_1$ in this case?

Hint: We can write $\widehat{\beta}_1 = (X_1' M_2 X)^{-1} X_1 M_2 y$, where $M_2 = I - X_2(X_2'X_2)^{-1}X_2'$ is an indempotent matrix that projects into the space of X_2 residuals, $S^{\perp}(X_2)$. Now take the expectation using the true model for y.

5. Consider the logit model for the survival of the passengers on the Titanic, as we discussed it in class.

TABLE 1. Logit Model 1

Variable	Coefficient	Std. Error	Odds Ratio	Std. Error
Child	1.062	.277	2.8908	.705
Female	2.420	.136	11.247	1.579
1st Class	-0.376	.126	0.6864	.093
2nd Class	-1.394	.129	0.2480	.039
3rd Class	-2.154	.144	0.1160	.015
Crew	-1.234	.080	0.2912	.023

(a) Based on the model in the lecture notes, compute the survival odds of a passenger traveling 1st class relative to a passenger traveling 3rd class. Prove any formulas you use.

(b) Give a 95% CI for the survival odds estimate in (a) (Hint: Use the bootstrap.)

6. Let $X \sim U[0,1]$ be uniformly distributed on the interval $[0,1]$.

(a) Find the probability distribution function, the cumulative distribution function, and the quantile function of $Y = -b \log X$.

(b) Find the median of the distribution in (a).

(c) Find the moment generating function of the distribution in (a).

(d) Let $(Y_1, ..., Y_n)$ be a random sample from the distribution in (a). Find the mle of b and its asymptotic distribution.

7. Consider a random variable X from the Gumbel(a, b) distribution with pdf

$$f(x) = (1/b)\exp[-(x-a)/b]\exp[-\exp(-(x-a)/b)], \qquad x \in \mathbb{R},$$

where, $a \in \mathbb{R}$ is a location parameter, and $b > 0$ is a shape paramater.

(a) Plot the pdf for $(a,b) = (0,1)$, $(a,b) = (0,2)$, and $(a,b) = (0,3)$.

(b) Find the cdf and quantile function (qf) of X.

(c) Find $E(X)$ and $Var(X)$.

- **(d)** Show that if X_1 and X_2 are independent standard Gumbel$(0, 1)$ then their difference $X_1 - X_2$ follows the logistic distribution.
- **(e)** Let $X_1, ..., X_n$ be random sample from the Gumbel(a, b) distribution. Find the mles for a and b. Is the asymptotic distribution of these mles normal? Justify your answer.
- **(f)** What kind of data are expected to be Gumbel distributed? The file **blooming.zip** contains a sample of $n = 5,692$ scores from the game *Blooming Gardens* (the scores a player achieved in $5,962$ games). Fit a Gumbel distribution to these scores using the R code in the file. Discuss the fit.
- **(g)** Discuss the **Elo rating system for Chess** and its connection to topic of this exercise.