Panteion University Gregory Kordas
Applied Econometrics July 2021

# Final Exam

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

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

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)** $\Pr(\frac{1}{3} < X < \frac{2}{3}|Y = \frac{2}{3})$.
**(d)** Find $Cov(X,Y)$, 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), $\beta$ 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.** [**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, $\beta_1$ is a $k_1 \times 1$ vector of coefficients, $\beta_2$ 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 $\beta_1$ 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 idempotent matrix that projects into the space of $X_2$ residuals, $S^{\perp}(X_2)$. Now take the expectation using the true model for $y$.

**4.** 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 $Y = -b \log X$.

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

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

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

**5.** Suppose that $(X_1, ..., X_n)$ are independent and identically distributed $Poisson(\lambda)$ random variables.

**(a)** Find the maximum likelihood (ML) estimator, and an asymptotic normal distribution for the estimator, of $\lambda$.

**(b)** Suppose that, rather than observing the random variables in (b) precisely, only the events $X_i = 0$ or $X_i > 0$ for $i = 1, ..., n$ are observed. (i) Find the ML estimator of $\lambda$ under this new observation scheme. (ii) In this new scheme, when does the ML estimator not exist (at a finite value in the parameter space)? Justify your answer. (iii) Compute the probability that the ML estimator does not exist for a finite sample of size $n$, assuming that the true value of $\lambda$ is $\lambda_0$. (iv) Construct a modified estimator that is consistent for $\lambda$.

**(c)** How would you measure the "loss of information" when we only observe the binary variable in (b) relative to observing the uncensored variable in (a)?

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

**(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. Explain your answer.

**(b)** How would you construct a 95% CI for your estimate in (a)? (Hint: Use the bootstrap.)

**7.** Consider the Demand and Supply model we discussed in class

**(a)** Explain why observations $(q_t, p_t, t = 1, ..., T)$ from the model

$$\begin{aligned} q_t^d &= \alpha_0 + \alpha_1 p_t + u_t, && \text{(demand equation)} \\ q_t^s &= \beta_0 + \beta_1 p_t + \upsilon_t, && \text{(supply equation)} \\ q_t^d &= q_t^s, && \text{(market equilibrium)} \end{aligned}$$

do not identify either the demand or the supply equations.

**(b)** Assume now that we have observations $(q_t, p_t, w_t, t = 1, ..., T)$ from the model

$$\begin{aligned} q_t^d &= \alpha_0 + \alpha_1 p_t + u_t, && \text{(demand equation)} \\ q_t^s &= \beta_0 + \beta_1 p_t + \beta_2 w_t + \upsilon_t, && \text{(supply equation)} \\ q_t^d &= q_t^s, && \text{(market equilibrium).} \end{aligned}$$

Are any of the two equations now identified, and how would you estimate it/them?