University of Pennsylvania **Economics 705, Fall 2015**

Prelim Examination

Friday August 7, 2015 Time limit: 150 minutes

Instructions:

(i) The exam consists of two parts. The total number of points for each part is 50. The number of points for each question is given below.

(ii) The exam is closed book and closed notes.

(iii) To receive full credit for your answers you have to explain your calculations. **You may state additional assumptions.**

Part I

Question 1: Probability Density Functions (14 Points)

Two random variables, Y, and X, have a joint distribution characterized by the following probability density function (pdf):

$$p(y,x) = cx(x+y), \quad 0 < y < 1, \; 0 < x < 1. \tag{1}$$

(i) (3 Points) Find the value of the normalization constant c that ensures that $p(y,x)$ is a proper (pdf).

(ii) (3 Points) Are the random variable Y and X independent? Explain.

(iii) (8 Points) Suppose that (Y_i, X_i) are *iid* with density given by $p(y,x)$ in (1). An econometrician observes a sample of n observations and runs a regression of the form

$$Y_i = \beta X_i + \text{resid}_i \tag{2}$$

Derive the probability limit (as $n \longrightarrow \infty$) of the OLS estimator of β.

Question 2: Statistical Inference (16 Points)

(i) (4 Points) Suppose you have a test $\varphi(Y;\theta_0)$ of a null hypothesis $H_0 : \; \theta = \theta_0$ with type-I error α. Explain how this test can be "inverted" to obtain a $1-\alpha$ confidence set.

(ii) (6 Points) Consider the location model $Y \sim N(\theta, 1)$. Propose a test for the hypothesis $H_0 : \; \theta \geq 0$ with size α and derive its power against the alternative $\theta = \tilde{\theta} < 0$. *Note*: this is a one-sided test. You need to verify that the type-I error of your test is less or equal than α for each $\theta \geq 0$.

(iii) (6 Points) Consider the model $Y \sim N(\theta, 1)$ with the constraint $\theta \geq 1$. Is the maximum likelihood estimator identical to the posterior mean (you can assume an improper prior distribution that is uniform on $\mathbb{R}^+$ with density $p(\theta) \propto 1$ if $\theta \geq 0$; and $p(\theta) = 0$ if $\theta < 0$)? To answer this question you should derive explicity formulas for the MLE and the posterior mean.

Question 3: Linear Regression Model (20 Points)

Consider the linear regression model

$$y_i = x_i\theta + u_i, \quad u_i|x_i \sim iid(0,1), \quad x_i \geq \epsilon > 0, \quad \mathbb{E}[x_i^2] = Q. \tag{3}$$

The x_i's are also independent across i, $k = 1$ and both x_i and θ_i are scalar.

(i) (5 Points) Derive the limit distribution of the OLS estimator.

(ii) (5 Points) Now consider the alternative estimator

$$\tilde{\beta} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{x_i} \tag{4}$$

Is this estimator consistent? Derive its limit distribution.

(iii) (2 Points) How would you rank two point estimators? According to your criterion, which estimator, $\hat{\beta}$ or $\tilde{\beta}$, is preferable? Explain.

For the remaining questions we change the assumption on the error term to

$$u_i|x_t \sim iid(0, x_i^2). \tag{5}$$

(iv) (2 Points) Does the OLS estimator remain consistent if the error term follows the distribution in (5)?

(v) (3 Points) Derive the limit distribution of the OLS estimator $\hat{\beta}$ under the assumption in (5).

(vi) (5 Points) Derive the limit distribution of $\tilde{\beta}$ defined in (4). Which estimator, $\hat{\beta}$ or $\tilde{\beta}$, is preferable?

Part II

Question 4: Linear Instrumental Variable Regression (25 Points)

Take a linear instrumental variable equation

$$\begin{aligned} y_i &= x_i\beta_1 + z_i\beta_2 + e_i, \\ E(e_i|z_i) &= 0, \end{aligned}$$

where both x_i and z_i are scalars and $E(x_ie_i) \neq 0$. Assume $\beta_2 \neq 0$.

(i) (5 Points) Propose a method to estimate β_1.

(ii) (10 Points) Show the estimator you propose is consistent and derive its asymptotic distribution.

(iii) (5 Points) How would you test the null hypothesis: $H_0 : \beta_1 = 1$ and $\beta_2^2 = 1$ versus the alternative that H_0 does not hold? Provide the test statistic, its asymptotic distribution, and the critical value. The nominal level of the test is α.

(iv) (5 Points) Now suppose you doubt the specification $E(e_i|z_i) = 0$. Propose a method to test one of its implications.

Question 5: Regression with Binary Dependent Variable (10 Points)

Consider the following model

$$Y_i^* = X_i'\beta + e_i,$$

where the error has a conditional distribution $e_i|X_i \sim N(0, \sigma^2)$. Only the sign of Y_i^* is reported. That is, the dataset contains the variable

$$Y_i = \begin{cases} 1 \text{ if } Y_i^* \geq 0 \\ 0 \text{ if } Y_i^* < 0. \end{cases}$$

(i) (2 Points) Explain why β is not identified.

(ii) (3 Points) Provide an additional assumption under which β is identified.

(iii) (5 Points) Given the identification of β, explain the maximum likelihood estimation of the parameter.

Question 6: Linear Model with Binary Endogenous Variable (15 Points)

Consider the following model

$$Y_i = X_i\beta + e_i,$$

where $X_i \in R$ is an endogenous dummy variable. We assume

$$X_i = 1\{W_i > v_i\},$$

where $W_i \in R$ is independent of e_i and v_i, $e_i \sim N(0,1)$, $v_i \sim N(0,1)$. We have i.i.d. observations $\{X_i, W_i, Y_i\}_{i=1}^n$.

(i) (5 Points) Suppose e_i and v_i are independent, how would you estimate β?

(ii) (5 Points) Suppose e_i and v_i are correlated, how would you estimate β?

(iii) (5 Points) In both cases, provide the asymptotic distributions of the estimators of β.