University of Pennsylvania **Economics 705, Spring 2013**

Prelim Examination

Friday June 7, 2013, 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 and Regressions (13 Points)

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

$$p(x, y) = \frac{1}{c} y(x+1), \quad 0 \leq x \leq 1, \ 0 \leq y \leq 1.$$

(i) (2 Points) Find the value of the normalization constant c that ensures that $p(x, y)$ is a proper (pdf).

(ii) (2 Points) Derive the marginal densities $p(x)$ and $p(y)$.

(iii) (2 Points) Derive the conditional pdf $p(y|x)$.

(iv) (3 Points) Derive the cumulative density function (cdf) $F(x, y)$.

Now suppose you have a sample of n iid draws (Y_i, X_i) from the above distribution. Based on this sample, you estimate the (misspecified) regression model

$$Y_i = \beta X_i + U_i$$

using the OLS estimator

$$\hat{\beta}_n = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2}.$$

(v) (4 Points) Derive the probability limit *plim* of $\hat{\beta}_n$ as $n \longrightarrow \infty$. Note: you can calculate the *plim* numerically.

Question 2: Do Stopping Rules Matter? (26 Points)

Consider the following experiment:

$$Y_i \sim iidN(\theta, 1), \quad i = 1, \dots, n$$

The goal is to compare frequentist and Bayesians inference with respect to θ. Define

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i.$$

Now consider the following two stopping rules:

1. Fixed sample size: terminate the experiment when the sample size has reached $n = 2$.

2. Random sample size: terminate the experiment when the t-statistic for the hypothesis $H_0 : \theta = 0$ exceeds the 5% critical value:

$$\left| \sqrt{n} \hat{\theta}_n \right| > 1.96.$$

Consider the Stopping Rule 1. The sample size is fixed at $n = 2$.

(i) (3 Points) Derive the distribution of $\hat{\theta}|\theta$.

(ii) (5 Points) Derive the posterior distribution $p(\theta|Y_{1:2})$ under the improper (the probability mass is infinite) prior $p(\theta) = 1$, i.e. the prior is uniform on the real line.

Now consider Stopping Rule 2. Before you start your calculations, think about this: we are terminating the experiment when we reject the null hypothesis $\theta = 0$. At this point the frequentist gag reflex should kick in!

Suppose that the experiment happens to be terminated at $n = 2$. The goal is to derive the pdf of $\hat{\theta}$ conditional on the sampling being terminated at $n = 2$. We also condition on the "true" θ being equal to zero.

You can use $\phi(x)$ and $\Phi(x)$ to denote the pdf and cdf of a standard $N(0, 1)$ random variable. We break the frequentist calculation into several steps:

(iii) (1 Points) Derive $p\big(y_1, y_2 \mid \theta = 0,\ |y_1| \leq 1.96\big)$.

(iv) (3 Points) Using a change of variables $(Y_1, Y_2) \mapsto (\hat{\theta}_1, \hat{\theta}_2)$ derive $p\big(\hat{\theta}_1, \hat{\theta}_2 \mid \theta = 0,\ |\hat{\theta}_1| \leq 1.96\big)$.

(v) (4 Points) Finally, derive $p\big(\hat{\theta}_2 | \theta = 0, n = 2\big)$.

(vi) (1 Points) Does the stopping rule affect the sampling distribution of $\hat{\theta}_n$ and subsequent inference such as hypothesis testing and confidence intervals? Explain.

Now we repeat the Bayesian analysis under Stopping Rule 2.

(vii) (4 Points) Treating the sample size n as random variable that can take values $n = 1, 2, 3, \ldots$, derive the density of $Y_{1:n}$ given θ, denoted by $p(y_1, y_2, \ldots, y_n|\theta)$. Note: this density will take the form of a sum. Each summand corresponds to a size that the sample can take. You only need to write out the first couple of terms.

(viii) (3 Points) In our observed sample $n = 2$. Using the general density derived in (xii) provide an expression for the likelihood function (given the observed sample).

(ix) (1 Points) Combine the likelihood function in (xiii) with the improper prior $p(\theta) = 1$ and derive the posterior $p(\theta|y_1, y_2, n = 2)$.

(x) (1 Points) For better or worse, is Bayesian inference affect by the stopping rule?

Question 3: Inference for Variance Parameters (11 Points)

Consider the model $Y_i \sim iidN(0, \theta)$, $i = 1, \dots, n$.

(i) (2 Points) Derive the score $s(\theta) = \partial \ln p(Y_{1:n}|\theta)/\partial\theta$.

(ii) (2 Points) Derive the maximum likelihood estimator $\hat{\theta}$ for θ.

(iii) (4 Points) Assume that the "true" value is θ_0 and derive the limit distribution of the (properly normalized) score evaluated at $\theta = \theta_0$.

(iv) (3 Points) Construct the Lagrange multiplier/score test for the hypothesis $H_0 : \theta = \theta_0$ and state 95% critical value as well as the acceptance and rejection region.

Part II

Question 4: Linear Instrumental Variable Regression (10 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.

(i) (2 Points) Can the coefficients (β_1, β_2) be estimated by two-stage-least-squares (2SLS) estimator using z_i as an instrument for x_i?

(ii) (2 Points) Can the coefficients (β_1, β_2) be estimated using z_i and z_i^2 as instruments?

(iii) (2 Points) For the 2SLS estimator suggested in (ii), what is the exclusion restriction on instrument exogeneity?

(iv) (4 points) For the 2SLS estimator suggested in (ii), what is the implicit assumption about instrument relevance? Write down the implicit reduced form equation for x_i to explain the instrument relevance.

Question 5: Inference for Linear Instrumental Variable Regression (10 Points)

Take a linear equation with endogeneity and a just-identified linear reduced form

$$\begin{aligned} y_i &= x_i\beta + e_i, \\ x_i &= z_i\gamma + u_i, \end{aligned}$$

where both x_i and z_i are scalars. Assume that

$$\begin{aligned} E(z_i e_i) &= 0, \\ E(z_i u_i) &= 0. \end{aligned}$$

(i) (4 Points) Write down the standard 2SLS estimator $\widehat{\beta}_{2SLS}$ for β using z_i as an instrument for x_i.

(ii) (6 points) Derive the asymptotic distribution for $\widehat{\beta}_{2SLS}$. Write the asymptotic variance as a function of $\Omega = E(z_i^2 e_i^2)$, $Q = E(z_i^2)$, and γ. Be clear when you use the weak law of large numbers and the central limit theorem.

Question 6: Hypothesis Testing (15 Points)

Suppose you have two independent i.i.d. samples $\{y_{1i}, x_{1i}, z_{1i} : i = 1, ..., n\}$ and $\{y_{2i}, x_{2i}, z_{2i} : i = 1, ..., n\}$. The dependent variables y_{1i} and y_{2i} are scalars and the regressors x_{1i} and x_{2i} and the instruments z_{1i} and z_{2i} are $k \times 1$ vectors. The model is standard just-identified linear instrumental regression

$$\begin{aligned} y_{1i} &= x'_{1i}\beta_1 + e_{1i}, \\ E(z_{1i}e_{1i}) &= 0. \\ y_{2i} &= x'_{2i}\beta_2 + e_{2i}, \\ E(z_{2i}e_{2i}) &= 0. \end{aligned}$$

You want to test $H_0 : \beta_1 = \beta_2$, that the two samples have the same coefficients.

(i) (5 Points) Develop a test statistic for H_0.

(ii) (6 Points) Derive the asymptotic distribution of the test.

(iii) (4 Points) Describe the testing procedure.

Question 7: Regression with Binary Dependent Variable (15 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 and σ cannot be identified separately.

(ii) (3 Points) Suppose σ is normalized to be 1 here and below. Write $E(Y_i|X_i)$ as a function of X_i and β.

(iii) (5 Points) Write down the log-likelihood function for the maximum likelihood estimation of β.

(iv) (5 Points). What is the asymptotic distribution of this maximum likelihood estimator? How to estimate its standard error?

END OF EXAM