University of Pennsylvania **Economics 705, Spring 2016**

Prelim Examination

Friday June 10, 2016 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: TRUE or FALSE Questions (8 Points)

You need to explain your answers to get full credit.

(i) (2 Points) TRUE or FALSE? The variance of a sum of two random variables equals the sum of the variances of the two random variables.

(ii) (2 Points) TRUE or FALSE? Suppose that random variables X and Y are jointly distributed with density $p(x, y)$. Then $\mathbb{E}[X|Y = y] = 0$ for each y implies $\mathbb{E}[X] = 0$.

(iii) (2 Points) TRUE or FALSE? Consider a regression of Y on X. Suppose you are interested in estimating the marginal effect on Y of raising X by 1 unit. Regressing Y on X and a constant (intercept) gives you a more precise estimate of this effect than demeaning both Y and X and regressing demeaned Y on demeaned X.

(iv) (2 Points) TRUE or FALSE? Suppose that Y is a strictly positive random variable with density $p(y)$. Now define $X = \ln Y$. Then the probability density function of x is given by $p(e^x)$.

Question 2: Frequentist Testing (8 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) (4 Points) Consider the location model $Y \sim N(\theta, 1)$. Propose a test for the hypothesis $H_0 : \ \theta = 0$ with type-I error α and derive its power against the alternative $\theta = \tilde{\theta} \neq 0$.

Question 3: Linear Regression Model (9 Points)

Consider the linear regression model

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

Moreover, the x_i's are also independent across i. Notice that we assumed that the conditional variance of u given x is known to be one. Moreover, $k = 1$ and both x_i and θ_i are scalar.

(i) (4 Points) Derive the limit distribution of the OLS estimator $\hat{\beta}$.

(ii) (4 Points) Now consider the alternative estimator

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

Derive its limit distribution.

(iii) (1 Points) Which estimator, $\hat{\beta}$ or $\tilde{\beta}$, is preferable? Explain.

Question 4: MLE versus Bayes versus Empirical Bayes (25 Points)

If the d dimensional vector Y is distributed according to a $N(\nu, \Sigma)$, then its probability density function is

$$p(y|\mu, \Sigma) = (2\pi)^{-d/2} |\Sigma|^{-1/2} \exp\left\{ -\frac{1}{2}(y-\mu)'^{-1/2}(y-\mu) \right\}.$$

Also note that for a positive scalar λ and a $d \times d$ matrix A it is the case that $|\lambda A| = \lambda^d |A|$.

Now suppose that $Y \sim N(\theta, I)$ where Y is a $d \times 1$ vector of random variables, θ is a $d \times 1$ vector of unknown means, and I is the $d \times d$ identity matrix. The econometrician observes one realization, y, of the random vector Y.

(i) (3 Points) Derive the likelihood function and the maximum likelihood estimator of θ. Use $p(y|\theta)$ to denote the likelihood function.

(ii) (2 Points) Now consider the prior distribution $\theta \sim N(0, I/\lambda)$, where λ is a scalar. What happens to the prior as $\lambda \longrightarrow \infty$ or $\lambda \longrightarrow 0$?

(iii) (5 Points) Use $p(\theta|\lambda)$ to denote the prior distribution of θ. Derive the posterior distribution $p(\theta|y, \lambda)$ of θ.

(iv) (5 Points) Derive the marginal likelihood (or marginal data density) $p(y|\lambda)$. **Hint:** You can either re-arrange the terms in Bayes theorem and express it as the ratio $p(y|\theta)p(\theta|\lambda)/p(\theta|y, \lambda)$ or you can exploit that $Y = \theta + U$ where both θ and U are Normal random variables.

(v) (5 Points) Derive a data-drive estimator of the prior precision λ by maximizing the marginal likelihood, i.e., $\hat{\lambda} = \text{argmax}_\lambda \, p(y|\lambda)$. Interpret your formula for $\hat{\lambda}$. Suppose θ is zero (very different from zero). Is $\hat{\lambda}$ likely to be small or large? **Hint:** you might want to reparametrize the problem by defining $\phi = f(\lambda)$ where $f(\cdot)$ is a monotone transformation of λ. Once you have $\hat{\phi}$, you can let $\hat{\lambda} = f^{-1}(\hat{\phi})$.

(vi) (5 Points) Suppose that you condition your inference on $\hat{\lambda}$ and compute the posterior mean estimator $\mathbb{E}[\theta|Y, \hat{\lambda}]$. This estimator is called an empirical Bayes estimator because the prior is influenced by the data. How does the empirical Bayes estimator differ from the maximum likelihood estimator in (i) and the posterior mean with a fixed choice of λ, e.g., $\lambda = 1$? Which of the three estimators do you expect to work best in practice? Be specific about your notion of "best."

Part II

Question 5: Linear Model with Endogeneity (15 points)

The model is

$$\begin{aligned} y_i &= x_i'\beta + u_i, \\ E(z_i u_i) &= 0, \end{aligned}$$

where $x_i, \beta \in R^k$, $z_i \in R^\ell, \ell \geq k, y_i, u_i \in R$, $\{(x_i, z_i, y_i, u_i) : i = 1, ..., n\}$ are i.i.d., $E(u_i^2 | X_i, Z_i) = \sigma^2$, and $E x_i x_i'$ and $E x_i z_i'$ both have full rank k.

Let $\widehat{\beta}_n$ denote the least-squares estimator obtained by regressing y_i on x_i, and let $\widetilde{\beta}_n$ denote the efficient GMM estimator using the instruments z_i.

(i) (3 points) Define $\widehat{\delta}_n = \widehat{\beta}_n - \widetilde{\beta}_n$. Derive the probability limit $\delta = \lim_{n\to\infty} \widehat{\delta}_n$.

(ii) (3 points) Derive the asymptotic distribution of $\widehat{\delta}_n$ when $\delta = 0$.

(iii) (3 points) Propose an estimator of the asymptotic variance derived above.

(iv) (3 points) Propose a condition under which $\delta = 0$ and interpret it.

(v) (3 points) Propose a test for $H_0 : \delta = 0$ vs $H_0 : \delta \neq 0$. Clearly specify the test statistic and its asymptotic distribution under the null.

Question 6: Linear Model with Binary Endogenous Variable (20 Points)

Consider the following model

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

where

$$X_i = 1\{Z_i\delta + v_i > 0\},$$

$Z_i \in R$ is independent of e_i and v_i, $E(Z_iX_i) \neq 0$, $E(e_i) = E(v_i) = 0$, e_i and v_i are possibly correlated. We have i.i.d. observations $\{X_i, Z_i, Y_i\}_{i=1}^n$.

(i) (5 Points) Suppose v_i has a standard normal distribution. Write down the log-likelihood function for the maximum likelihood estimation of δ.

(ii) (5 Points). What is the asymptotic distribution of this maximum likelihood estimator $\widehat{\delta}$? How to estimate its standard error?

(iii) (4 Points) Consider the IV estimator of β :

$$\widehat{\beta}_{IV} = \frac{\sum_{i=1}^n Z_iY_i}{\sum_{i=1}^n Z_iX_i}.$$

Is this estimator consistent without any distribution assumption on v?

(iv) (6 points) Let $\widehat{\delta}$ be the maximum likelihood estimator of δ defined above and $W_i(\delta) = \mathbb{E}(X_i|Z_i)$. Consider the two stage least squares estimator

$$\widehat{\beta}_{TSLS} = \frac{\sum_{i=1}^n W_i(\widehat{\delta})Y_i}{\sum_{i=1}^n W_i(\widehat{\delta})^2}.$$

Is this estimator consistent when v_i has a standard normal distribution? Is this estimator consistent when v_i does not have a standard normal distribution?

Question 7: Testing in GMM (15 points)

Consider an efficient GMM estimator of θ_0 based on the moment condition $E(\phi(Z_i, \theta_0)) = 0$, where $\{Z_i : 1, ..., n\}$ are i.i.d. data. Denote this efficient GMM estimator by $\widehat{\theta}_n$. We have shown in class that

$$n^{1/2}(\widehat{\theta}_n - \theta_0) \rightarrow_d N(0, \Sigma(\theta_0)),$$

where

$$\begin{aligned}
\Sigma(\theta) &= \left(G(\theta)'\Omega(\theta)G(\theta)\right)^{-1}, \\
G(\theta) &= E\left(\frac{\partial}{\partial\theta'}\phi(Z_i, \theta)\right) \in R^{\ell\times k}, \\
\Omega(\theta) &= E\left(\phi(Z_i, \theta)\phi(Z_i, \theta)\right) \in R^{\ell\times\ell}.
\end{aligned}$$

(i) (5 points) To test the null hypothesis $H_0 : \theta = \theta_0$, the Wald statistic is

$$W_n = n\left(\widehat{\theta}_n - \theta_0\right)' \widehat{\Sigma}_n^{-1}\left(\widehat{\theta}_n - \theta_0\right),$$

where $\widehat{\Sigma}_n \rightarrow_p \Sigma(\theta_0)$ is a consistent estimator of $\Sigma(\theta_0)$. Suppose $G(\theta_0)$ has full rank and $\Omega(\theta_0)$ is positive definite. What is the asymptotic distribution of the Wald statistic W_n? How do you choose critical value for a test of significance level α?

(ii) (5 points) Consider an alternative test called the Anderson-Rubin test. The test statistic is

$$\begin{aligned}
AR_n &= n\overline{\phi}_n(\theta_0)'\widehat{\Omega}_n^{-1}(\theta_0)\overline{\phi}_n(\theta_0), \text{ where} \\
\overline{\phi}_n(\theta) &= n^{-1}\sum_{i=1}^{n}\phi(W_i, \theta)
\end{aligned}$$

and $\widehat{\Omega}_n(\theta_0) \rightarrow_p \Omega(\theta_0)$. Suppose $\Omega(\theta_0)$ is positive definite. What is the asymptotic distribution of the Anderson-Rubin statistic AR_n? How do you choose critical value for a test of significance level α?

(iii) (5 points) What will happen to the tests based on the Wald test and Anderson-Rubin test, respectively, when $G(\theta_0)$ is near singular in practice?

END OF EXAM