PhD in Economics (17th Cycle)
Econometrics test (2016-06-06)

Name: ______________________________

1. A non-negative random variable X follows the *Lomax distibution* if its density function can be written as
$$f(x) = \frac{\alpha}{\lambda}\left[1 + \frac{x}{\lambda}\right]^{-(\alpha+1)}$$
where α and λ are positive reals, called the *shape* and *scale* parameters, respectively. A Lomax-distributed r.v. has no finite moments $E(X^\nu)$ for $\nu \geq \alpha$; here are a few special values:
$$E(X) = \frac{\lambda}{\alpha - 1} \quad (1)$$
$$V(X) = \frac{\lambda^2\alpha}{(\alpha-1)^2(\alpha-2)} \quad (2)$$
Given an iid sample of Lomax random variates, define
$$\hat{r} = \left[\frac{1}{n}\cdot\sum_i x_i^2\right]^{-1}\left[\frac{1}{n}\sum_i x_i\right]^2;$$
   (a) Find the score with respect to α;
   (b) Find $\text{plim}\,\hat{r}$;
   (c) Prove that $\hat{\alpha} = \frac{2\hat{r}-2}{2\hat{r}-1}$ is a consistent estimator of α if $\alpha > 2$;
   (d) explain why $\hat{\alpha}$ is inconsistent for $\alpha \leq 2$;

2. Consider the bivariate stochastic process $\mathbf{z}_t = [y_t, x_t]$ given by

$$\begin{aligned} y_t &= 3 - 1.2x_{t-1} + 0.6y_{t-2} + \varepsilon_t \\ x_t &= 1 + 0.5y_{t-1} + \eta_t \end{aligned}$$

where $\mathbf{u}_t = [\varepsilon_t, \eta_t]$ is a bivariate white noise process with covariance matrix Σ.

(a) Prove that $\mathbf{z}_t$ is a VAR(2) process and find the numerical values of μ, Φ_1 and Φ_2 in the representation

$$\mathbf{z}_t = \mu + \Phi_1 \mathbf{z}_{t-1} + \Phi_2 \mathbf{z}_{t-2} + \mathbf{u}_t$$

$$\mu = \begin{bmatrix} \\ \\ \end{bmatrix} \quad \Phi_1 = \begin{bmatrix} & \\ & \end{bmatrix} \quad \Phi_2 = \begin{bmatrix} & \\ & \end{bmatrix}$$

(b) Prove that $\mathbf{z}_t$ is a VMA(2) process and find the numerical values of $\mathbf{m}$, Θ_1 and Θ_2 in the representation

$$\mathbf{z}_t = \mathbf{m} + \mathbf{u}_t + \Theta_1 \mathbf{u}_{t-1} + \Theta_2 \mathbf{u}_{t-2}$$

$$\mathbf{m} = \begin{bmatrix} \\ \\ \end{bmatrix} \quad \Theta_1 = \begin{bmatrix} & \\ & \end{bmatrix} \quad \Theta_2 = \begin{bmatrix} & \\ & \end{bmatrix}$$

(c) Prove that $\mathbf{z}_t$ is stationary

(d) Find $E(\mathbf{z}_t)$.

$$E(\mathbf{z}_t) = \begin{bmatrix} \\ \\ \end{bmatrix}$$

3. Assume that you are interested in understanding the effect of smoking (cigarettes) on earnings. Smokers might indeed have worse health than non-smokers, take too many breaks on the workplace to smoke, be discriminated by the employer for this bad habit, resulting therefore in lower productivity and lower earnings.

   You have data on the labor market status, earnings, some family characteristics and smoking behavior of 2,049 employees and 1,115 not-employed individuals in 2009.[1] You plan to estimate the following wage equation to understand the impact of smoking on wages:

$$\text{lhwage}_i = \alpha + \beta \text{smoke}_i + \gamma \text{female}_i + \eta \text{head}_i + \delta \text{age}_i + \eta \text{age}_i^2 + \theta \text{educ}_i + u_i, \quad (3)$$

   where

   - lhwage_i is the natural logarithm of the hourly wage of individual i;
   - smoke_i is a dummy equal to 1 if individual i is a smoker;
   - female_i is a dummy equal to 1 if individual i is a woman;
   - head_i is a dummy equal to 1 if individual i is the household head;
   - age_i is the age of individual i;
   - educ_i is the number of years of education of individual i;
   - u_i is the error term.

   The estimation of Equation (3) by OLS, with heteroskedasticity robust standard errors, returns the estimated coefficients reported in Table 1.

Table 1: OLS estimation results of demand of bio apples

```
Model 1: OLS, using observations 1-2049
Dependent variable: lhwage
Heteroskedasticity-robust standard errors, variant HC1

             coefficient    std. error    t-ratio    p-value
  ---------------------------------------------------------
  const       1.50759       0.116874      12.90     1.18e-036
  smoke      -0.0612259     0.0150658     -4.064    5.01e-05
  fem         0.0186565     0.0160140      1.165    0.2441
  age         0.0137258     0.00552825     2.483    0.0131
  age2       -5.62159e-05   6.60319e-05   -0.8513   0.3947
  edu         0.0349194     0.00234650    14.88     1.24e-047
  head        0.0932289     0.0162540      5.736    1.12e-08

Mean dependent var   2.418424   S.D. dependent var   0.307963
Sum squared resid    154.2588   S.E. of regression   0.274851
R-squared            0.205811   Adjusted R-squared   0.203478
F(6, 2042)           81.04129   P-value(F)           3.78e-91
Log-likelihood      -257.5612   Akaike criterion     529.1224
Schwarz criterion    568.4982   Hannan-Quinn         543.5627
```

   (a) What is the effect of smoking on wages? Is the effect significant?

   (b) What is the effect of age on wages?

[1] Data comes from the LISS panel collected by CentERData of Tilburg University, The Netherlands.

Since you have data on personal characteristics and smoking behavior also on 1,115 not-employed individuals, you decide to correct for endogenous sample selection by estimating a standard two-step Heckman sample selection model. Table 2 displays the estimation results, where `nkids` is the number of dependent kids and `single` is a dummy equal to 1 if the individual is single.

Table 2: Heckman sample selection corrected estimation of wage equation

```
Model 2: Two-step Heckit, using observations 1-3164
Dependent variable: lhwage
Selection variable: emp
Heteroskedasticity-robust standard errors, variant HC1


             coefficient    std. error        z        p-value
  -------------------------------------------------------------
  const        1.73881       0.252723        6.880     5.97e-012
  smoke       -0.0608027     0.0150721      -4.034     5.48e-05
  fem          0.0250409     0.0162538       1.541     0.1234
  age          0.00356860    0.0112040       0.3185    0.7501
  age2         7.32492e-05   0.000140825     0.5201    0.6030
  edu          0.0331569     0.00297855     11.13      8.78e-029
  head         0.0819395     0.0191686       4.275     1.91e-05
  lambda      -0.0634715     0.0634451

             Selection equation into employment (probit)

  const       -6.16411       0.340066      -18.13      1.98e-073
  smoke       -0.00425866    0.0657869      -0.06473   0.9484
  fem         -0.143583      0.0705755      -2.034     0.0419
  age          0.342356      0.0160260      21.36      2.99e-101
  age2        -0.00438636    0.000186195   -23.56      1.04e-122
  edu          0.0715932     0.0113062       6.332     2.42e-010
  head         0.635628      0.0805482       7.891     2.99e-015
  nkids       -0.0874453     0.0269453      -3.245     0.0012
  single      -0.465686      0.0799261      -5.826     5.66e-09

Mean dependent var   2.418424   S.D. dependent var   0.307963
sigma                0.277239   rho                 -0.228942

Total observations: 3164
Censored observations: 1115 (35.2%)
```

(c) Test the presence of sample selection bias for the estimated coefficients reported in Table 1:

(a) write the null and the alternative hypotheses for this test

(b) give the formula for the test statistic

(c) calculate the test statistic.

Is the sample selection correction needed?

(d) What are the exclusion restrictions in the selection equations? Why are they needed? What are the assumptions for the validity of these exclusion restrictions? Are they credible in this framework?