PhD in Economics (15th Cycle)
Econometrics test (2014-06-06)

Name: ______________________________

Below, you'll find three exercises; number 1 is obligatory. Then, you have the choice between number 2 and number 3.

1. Given the model:

$$\begin{aligned} y_i &= x_i'\beta + \epsilon_i \\ \epsilon_i|x_i, z_i &\sim N\left(0, \sigma^2 \cdot \exp\left\{z_i'\gamma\right\}\right) \end{aligned}$$

where

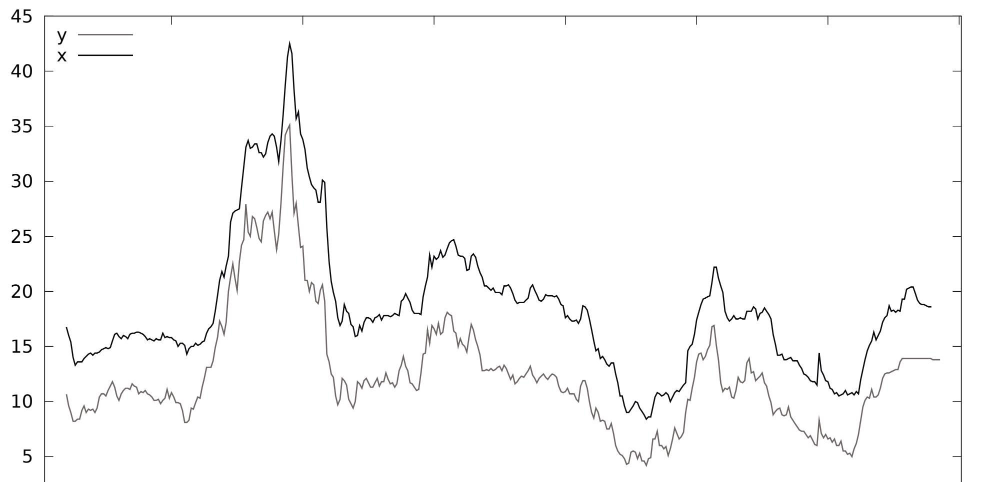
- Observations $i = 1 \ldots n$ are iid;
- x_i is a vector of exogenous variables that may contain a constant term;
- z_i is a vector of exogenous variables that does not contain a constant term;
- β, γ and σ^2 are unknown parameters.

(a) Write the first-order conditions for ML estimation of the parameters

(b) Find the parameter restriction that makes the model homoskedastic

(c) Devise a test for the homoskedasticity hypothesis

2. The following graph shows the time series plot of two prices of lard in USA:



(a) wholesale price (x_t),

(b) retail price (y_t).

Both series are expressed in U$ cents/LB and the available sample contains monthly data ranging from January, 1911 to December 1943 ($T = 396$).

An OLS regression of y_t on x_t and a constant yields the following results:

```
Dependent variable: y

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const        4.50097     0.175446      25.65    4.66e-86  ***
  x            1.11767     0.0129082     86.59    1.55e-258 ***

Mean dependent var   18.42045   S.D. dependent var   6.250230
Sum squared resid    770.4499   S.E. of regression   1.398376
R-squared            0.950071   Adjusted R-squared   0.949944
rho                  0.792423   Durbin-Watson        0.416560
```

A DF test on the residuals gives:

```
Dickey-Fuller test for uhat

  model: (1-L)y = (a-1)*y(-1) + e
  estimated value of (a - 1): -0.207577
  test statistic: tau_c(2) = -6.73207 (p-value 3.886e-08)
```

Moreover, the estimation output of a Johansen test for cointegration is given below.

```
------------------------------------------------------------
Number of equations = 2, Lag order = 3
Estimation period: 1911:04 - 1943:12 (T = 393)

Log-likelihood = 409.127 (including constant term: -706.159)

Rank Eigenvalue Trace test  p-value  Lmax test  p-value
   0   0.1114     52.287    0.0000       L1      0.0000
   1   0.0148     5.8581    0.2091       L2      0.2088

beta                                alpha
y         1.2980     -0.0427        y        0.07330    -0.1031
x        -1.1027      0.1979        x        0.16156    -0.0181
const     4.1548     -3.2179

Pi=alpha*beta'
           y            x          const
y       0.09954     -0.10123      0.63626
x       0.21047     -0.18174      0.72964
------------------------------------------------------------
```

- what is the deterministic part for the Johansen test?

  ◯ Case 1: no constant ◯ Case 2: restricted constant
  ◯ Case 3: unrestricted constant ◯ Case 4: restricted trend
  ◯ Case 5: unrestricted trend

- what is the estimated cointegration rank? $r =$ ___________
- what is the estimated cointegration relationship (if any)?

  ◯ There is no cointegration relationship because ______________________________

  ____

  ____

  ◯ $y_t =$ ____

- what are the values of the Λ-max test statistics?

L1: ______________ L2: ______________

(a) Do OLS and the Johansen test give contradictory indications?

(b) Is there any way to attach any economic meaning to the OLS coefficients? If so, which?

3. You have a dataset providing information on the smoking behaviour of 1,191 pregnant women. Let s_i be a binary variable equal to 1 if woman i smokes during pregnancy and 0 otherwise. You want to estimate the causal impact of cigarette price (p) and the log of the family income (y) on the smoking behaviour. The ethnicity dummy variable w (1=white) is a further control variable. Summary statistics of the variables are reported below:

	mean	std. dev.	min.	max.	observations
s	0.135	0.342	0.000	1.000	1,191
y	3.277	0.715	-0.693	4.174	1,191
p	130.710	10.351	103.800	152.500	1,191
w	0.844	0.363	0.000	1.000	1,191

Consider the probit model

$$P(s_i = 1|y_i,\ p_i,\ w_i) = \Phi(\alpha + \beta_1 y_i + \beta_2 p_i + \beta_3 w_i); \quad (1)$$

ML estimates of equation (1) are reported in Table 1.

Table 1: Probit estimation results of smoking behaviour

```
Model 1: Probit, using observations 1-1191
Dependent variable: s

            coefficient   std. error      z       p-value
  -------------------------------------------------------
  const     -0.897130     0.591429     -1.517   0.1293
  y         -0.347181     0.062315     -5.571   2.53e-08 ***
  p          0.005340     0.004453      1.199   0.2305
  w          0.238372     0.135261      1.762   0.0780   *

Mean dependent var   0.135181   S.D. dependent var   0.342060
McFadden R-squared   0.033625   Adjusted R-squared   0.025146
Log-likelihood      -455.9123   Akaike criterion     919.8246
Number of cases 'correctly predicted' = 1028 (86.3%)
```

Someone pointed out to you that the variable y may be endogenous. Fortunately, your datasets contain information about the years of education of the mother (med) and the father (fed) of each unit i and you are willing to assume that they are exogenous. Table 2 reports the first and second steps of the Rivers-Vuong (1988) procedure.

(a) Comment on the estimation results reported in Table 1. How do the covariates affect smoking behaviour?

(b) Estimate the partial effect at the average (PEA) of the variable y from model 1.

(c) Comment on the results presented in Table 2.

Table 2: Two-step Rivers-Vuong (1988) IVprobit

```
First step: OLS, using observations 1-1191
Dependent variable: y

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const      0.754922      0.246262      3.066    0.0022    ***
  p          0.00390528    0.00176794    2.209    0.0274    **
  w          0.336673      0.0504840     6.669    3.94e-011 ***
  med        0.0708678     0.00981779    7.218    9.38e-013 ***
  fed        0.0604283     0.00871174    6.936    6.60e-012 ***

Mean dependent var   3.276758   S.D. dependent var   0.715159
Sum squared resid    465.7746   S.E. of regression   0.626680
R-squared            0.234713   Adjusted R-squared   0.232132
F(4, 1186)           90.93652   P-value(F)           1.80e-67
Log-likelihood      -1130.873   Akaike criterion     2271.745
F-test of excluded instruments: F(2, 1186)            104.420

***************************************************************
***************************************************************
***************************************************************

Second step: Probit, using observations 1-1191
Dependent variable: s

             coefficient   std. error     z       p-value
  --------------------------------------------------------
  const       1.18159      0.684617      1.726   0.0844    *
  y          -1.38135      0.169251     -8.162   3.31e-016 ***
  p           0.0120645    0.00471865    2.557   0.0106    **
  w           0.667202     0.153874      4.336   1.45e-05  ***
  vhat_2      1.21865      0.182109      6.692   2.20e-011 ***

Mean dependent var   0.135181   S.D. dependent var   0.342060
McFadden R-squared   0.083619   Adjusted R-squared   0.073021
Log-likelihood      -432.3263   Akaike criterion     874.6525
Number of cases 'correctly predicted' = 1023 (85.9%)
```