PhD in Economics (16th Cycle)
Econometrics test (2015-06-04)

Name: ____

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

1. Suppose you have a large number of individuals $i = 1 \dots N$. For each individual, we observe a time series which is a realisation of an AR(1) process:

$$y_{i,t} = \mu + \phi_i y_{i,t-1} + \varepsilon_{i,t}$$

Individuals differ between each other in the parameter ϕ_i, which is distributed, over the population, as a random variable with support $A = [0, 1)$ and density function $f(\phi_i) = 2(1 - \phi_i)$. Assume that

(a) $\varepsilon_{i,t} \perp\!\!\!\perp \varepsilon_{j,s}$ for all i, j, t and s.
(b) $E(\varepsilon_{i,t}) = 0$
(c) $0 < V(\varepsilon_{i,t}) = \sigma^2 < \infty$.

Now, consider the stochastic process $Y_t = E(y_{i,t})$:

(a) Calculate $E(Y_t)$; (*hint: start from* $E(y_{i,t}|\phi_i)$ *and use the law of iterated expectations*)
(b) Calculate $V(Y_t)$; (*hint: start from the special case* $\mu = 0$)
(c) Calculate the autocorrelation function of Y_t; (*hint: use the law of iterated expectations*)
(d) (bonus question) Show that it's impossible to represent Y_t as an autoregressive process of any order.

2. Imagine you have a dataset providing information on the demand of bio apples (*ecokg*) in kilograms of 660 individuals. You wish to study the impact of the price per kilogram of bio apples (*ecoprc*) and the price per kilogram of regular apples (*regprc*) on the consumer demand of bio apples. The dataset contains also other variables that can be used as controls: the natural log of family income, *ln(faminc)*, years of education, *educ*, the number of family members younger than 5, *numlt5*, and the number of family members between 5 and 17 years of age (*num5_17*). Summary statistics of the variables are reported below:

	mean	std. dev.	min.	max.	observations
ecokg	0.669	1.146	0.000	19.051	660
ln(ecoprc)	0.038	0.291	-0.528	0.464	660
ln(regprc)	-0.165	0.288	-0.528	0.174	660
ln(faminc)	3.773	0.672	1.609	5.521	660
educ	14.382	2.274	8.000	20.000	660
numlt5	0.286	0.643	0.000	4.000	660
num5_17	0.621	0.994	0.000	6.000	660

The estimation of a linear regression model (with heteroskedasticity-robust standard errors) of *ecokg*, with explanatory variables *ln(ecoprc)*, *ln(regprc)*, *ln(faminc)*, *educ*, *numlt5*, and *num5_17*, returns the figures shown in table 1:

Table 1: OLS estimation results of demand of bio apples

	coefficient	std. err.	t-ratio	p-value
lnecoprc	-1.161849	0.212018	-5.48	0.000 ***
lnregprc	0.995814	0.300556	3.31	0.001 ***
lnfaminc	0.093939	0.059689	1.57	0.116
educ	0.011534	0.017763	0.65	0.516
numlt5	-0.040750	0.053292	-0.76	0.445
num5_17	0.052885	0.047172	1.12	0.263
constant	0.335623	0.227824	1.47	0.141

(a) Are the signs of the coefficients for *ln(ecoprc)* and *ln(regprc)* the expected ones? Provide an economic interpretation of the estimated coefficients of *ln(ecoprc)* and *ln(regprc)*, both from the qualitative and the quantitative viewpoints.

Upon closer inspection, you realize that for 248 individuals the demand of apples is 0. This suggests you to estimate a Tobit model for *ecokg* with censoring from below at 0. Table 2 reports the Tobit estimation results, where σ is the standard deviation of the error term (assumed Gaussian).

Table 2: Tobit estimation results of demand of bio apples

	coefficient	std. err.	t-ratio	p-value
lnecoprc	-2.370665	0.396938	-5.97	0.000 ***
lnregprc	1.925765	0.403327	4.77	0.000 ***
lnfaminc	0.194658	0.105598	1.84	0.066 *
educ	0.045040	0.030914	1.46	0.146
numlt5	-0.033816	0.102033	-0.33	0.740
num5_17	0.118766	0.065027	1.83	0.068 *
constant	-0.843511	0.515402	-1.64	0.102
σ	1.562614	0.057506	27.17	0.000 ***

(b) What do you make of the fact that the Tobit estimate on *lnecoprc* is about twice the size of the OLS estimate in the linear model?

(c) Estimate the partial effect at the mean (PEA) of *lnecoprc* on the unconditional mean of the dependent variable. Interpret the estimated PEA.

(d) Estimate the PEA of *lnecoprc* on the probability that an individual demand a strictly positive quantity of bio apples. Interpret the estimated PEA.

Note: when answering points (c) and (d), there's no need to numerically compute the cumulative standard normal distribution $\Phi(\cdot)$ or the standard normal density $\phi(\cdot)$, if your calculator doesn't offer these functions. Expressions like $\Phi(0.123)$ will be enough.

3. Given the following ARMA (2,1) model for the time series y_t

$$y_t = 2 + 1.8y_{t-1} - 0.8y_{t-2} + \theta\varepsilon_{t-1} + \varepsilon_t,$$

where $\theta > 0$ and ε_t is a Gaussian white noise with zero mean and unit variance,

(a) prove that $y_t \sim I(1)$.

____

____

____

____

____

(b) Write the equation for the first differenced process Δy_t as an ARMA model:

$\Delta y_t =$ ____

(c) Given the hypotheses above, provide an interval of values for the parameter θ so that the process Δy_t is invertible (possesses an infinite order AR representation) and its ARMA representation contains no common factors:

$\theta \in$ ______________________________

(d) Write the Wold representation for Δy_t polynomial including the constant term and assuming $\theta = 0.5$. (truncate at the 3rd order)

$\Delta y_t =$ ____

(e) Now let $\theta = 0$ and $x_t = 1 + y_{t-1} + u_t$, where u_t is a Gaussian white noise with zero mean, unit variance and not correlated with ε_t; write the VAR(2) specification for the bivariate process $\begin{bmatrix} y_t & x_t \end{bmatrix}'$.

$$\begin{bmatrix} y_t \\ \\ x_t \end{bmatrix} =$$

(f) Select the right statement:

○ the above VAR(2) model is a bivariate $I(1)$ process, not cointegrated because

____

____

____

○ the above VAR(2) model is a cointegrated system, and its VECM specification is

$$\begin{bmatrix} \Delta y_t \\ \\ \Delta x_t \end{bmatrix} =$$