Introductory Econometrics

Tutorial 12

<u>PART A:</u> To be done before you attend the tutorial. The solutions will be made available at the end of the week.

1. (c) is the right answer. First differencing removes both the linear trend and the unit root.

2. Following the instructions in the question:

$$
\begin{aligned}
y_t &= \alpha + \beta\ t + u_t \\
u_t &= u_{t-1} + e_t \\
y_{t-1} &= \alpha + \beta\ (t-1) + u_{t-1} = \alpha + \beta\ t\ - \beta + u_{t-1} \\
y_t - y_{t-1} &= \alpha + \beta\ t + u_t - \alpha - \beta\ t\ + \beta - u_{t-1} \\
&= \beta + u_t - u_{t-1} = \beta + e_t \\
&\Rightarrow\ \triangle y_t = \beta + e_t
\end{aligned}
$$

$\triangle y_t$ is a constant plus a white noise, so it neither has a trend nor a unit root.

3. Define each of the following and provide an example:

(a) an estimator: An estimator is a function (e.g. a formula) of sample observations that produces and estimate for a population parameter. For example, sample average is an estimator for the population mean. Another example is that in the multiple regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, the OLS estimator $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is an estimator $\boldsymbol{\beta}$.

(b) an unbiased estimator: An unbiased estimator is an estimator whose expected value is the population parameter. For example, the sample average is an unbiased estimator of the population mean when the sample is randomly selected. Another example is that in the multiple regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, as long as $E(\mathbf{u} \mid \mathbf{X}) = \mathbf{0}$, and columns of $\mathbf{X}$ are linearly independent, the OLS estimator $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is an unbiased estimator of $\boldsymbol{\beta}$.

(c) the best linear unbiased estimator: The best linear unbiased estimator (BLUE) is an estimator that (i) is a linear function of the dependent variable, (ii) is unbiased, and (iii) has the smallest variance among all linear and unbiased estimators. For example, in the multiple regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, as long as $E(\mathbf{u} \mid \mathbf{X}) = \mathbf{0}$, columns of $\mathbf{X}$ are linearly independent, and $Var(\mathbf{u} \mid \mathbf{X}) = \sigma^2\mathbf{I}_n$, then the OLS estimator $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is BLUE.

(d) a consistent estimator: A consistent estimator is an estimator that converges in probability to the population parameter as the sample size goes to infinity. For example, sample average is a consistent estimator of the population mean when the sample is randomly selected. Another example is that in the multiple regression model $y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + u_i$, as long as u_i has mean zero and is uncorrelated with x_{i1} to x_{ik}, and the explanatory variables are linearly independent, the OLS estimator $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is a consistent estimator of $\boldsymbol{\beta}$.

(e) an asymptotically normal estimator: An asymptotically normal estimator is an estimator whose distribution becomes close to a normal distribution as the sample size gets large (this is not a precise definition, but OK for this unit). For example, sample mean is asymptotically normal.

Do not forget to bring your answers to PART A and a copy of the tutorial questions to your tutorial.

<u>Part B:</u> This part will be covered in the tutorial. It is still a good idea to attempt these questions before the tutorial.

1. Assume that y_t is a **stationary** AR(1) process given by

$$y_t = c + \varphi_1 y_{t-1} + e_t, \tag{1}$$

where

$$e_t \sim i.i.d.(0, \sigma^2).$$

(a) Show that

$$E(y_t) = \mu = \frac{c}{(1-\phi_1)}, \ \forall t.$$

$$\begin{aligned} E(y_t) &= E(c + \varphi_1 y_{t-1} + e_t) \\ &= c + E(\varphi_1 y_{t-1}) + E(e_t) \\ &= c + \varphi_1 E(y_{t-1}) + 0 \end{aligned} \tag{2}$$

From the stationarity assumption, we have

$$E(y_t) = E(y_{t-1}) = \mu.$$

Therefore, (2) becomes

$$\mu = c + \varphi_1 \mu,$$

or

$$\mu = \frac{c}{(1-\phi_1)}.$$

(b) Show that

$$Var(y_t) = \gamma_0 = \frac{\sigma^2}{(1-\varphi_1^2)}, \ \forall t.$$

$$\begin{aligned} Var(y_t) &= Var(c + \varphi_1 y_{t-1} + e_t) \\ &= Var(\varphi_1 y_{t-1} + e_t) \\ &= Var(\varphi_1 y_{t-1}) + Var(e_t) + 2\varphi_1 Cov\left(e_t, y_{t-1}\right) \\ &= \varphi_1^2 Var(y_{t-1}) + \sigma^2 + 2\varphi_1 Cov\left(e_t, y_{t-1}\right). \end{aligned} \tag{3}$$

From the stationarity assumption, we have

$$Var(y_t) = Var(y_{t-1}) = \gamma_0.$$

Also, we have that the regressor is uncorrelated with the error term such that

$$Cov\left(e_t, y_{t-1}\right) = 0.$$

Therefore, (3) becomes

$$\gamma_0 = \varphi_1^2 \gamma_0 + \sigma^2,$$

which implies that

$$\gamma_0 = \frac{\sigma^2}{(1-\varphi_1^2)}.$$

(c) Show that (1) may be written in mean deviation form as

$$y_t - \mu = \varphi_1(y_{t-1} - \mu) + e_t. \tag{4}$$

From the result of (a) we have

$$c = \mu\left(1 - \varphi_1\right).$$

Therefore, (1) can be written as

$$\begin{aligned} y_t &= \mu(1-\varphi_1)+\varphi_1 y_{t-1}+e_t \\ &= \mu+\varphi_1(y_{t-1}-\mu)+e_t, \end{aligned}$$

or

$$y_t-\mu=\varphi_1(y_{t-1}-\mu)+e_t.$$

(d) Use (4) to show that

$$\gamma_1 = Cov(y_t, y_{t-1}) = \frac{\sigma^2}{(1-\varphi_1^2)}\varphi_1.$$

Multiplying both sides of (4) by $(y_{t-1}-\mu)$ and taking expectations we obtain

$$E\left[(y_t-\mu)(y_{t-1}-\mu)\right]=\varphi_1 E(y_{t-1}-\mu)^2+E\left[(y_{t-1}-\mu)e_t\right],$$

or

$$\begin{aligned} \gamma_1 &= \varphi_1\gamma_0+Cov\left(e_t, y_{t-1}\right) \\ &= \varphi_1\gamma_0. \end{aligned} \tag{5}$$

or

$$\gamma_1=\varphi_1\frac{\sigma^2}{(1-\varphi_1^2)}.$$

(e) Show that

$$\rho_1 = Corr\left(y_t, y_{t-1}\right)=\varphi_1.$$

By definition,

$$\rho_1 = Corr\left(y_t, y_{t-1}\right)=\frac{Cov(y_t, y_{t-1})}{\sqrt{Var\left(y_t\right)}\sqrt{Var\left(y_{t-1}\right)}}=\frac{\gamma_1}{\sqrt{\gamma_0}\sqrt{\gamma_0}}=\frac{\gamma_1}{\gamma_0}.$$

Then, using (5) we get

$$\rho_1=\frac{\varphi_1\gamma_0}{\gamma_0}=\varphi_1.$$

2. **This is based on a question from S1, 2018 final exam:** A researcher wants to test the Efficient Market Hypothesis (EMH) using weekly percentage returns, denoted by r_t, on the New York Stock Exchange composite index. In its strict form the EMH states that information observable to the market prior to week t should not help to predict the return during week t. If we use only past information on r, the EMH is stated as

$$E\left(r_t|r_{t-1}, r_{t-2}, \ldots\right)=E\left(r_t\right). \tag{6}$$

One simple way to test that (6) holds is to specify the following alternative AR(1) model to describe r_t:

$$r_t=\beta_0+\beta_1 r_{t-1}+u_t, \tag{7}$$

where $E\left(u_t|r_{t-1}, r_{t-2}, \ldots\right)=0$ and $Var\left(u_t|r_{t-1}, r_{t-2}, \ldots\right)=\sigma^2$. Using data from the first week of January 2004 to the third week of April 2018, estimation of (7) gives:

$$\begin{aligned} \hat{r}_t &= \underset{(0.096)}{0.086}-\underset{(0.038)}{0.059}r_{t-1}, \\ n &= 689,\ R^2=0.0035,\ \bar{R}^2=0.0020. \end{aligned} \tag{8}$$

(standard errors are reported in parentheses underneath the parameter estimates).

(a) i. Given (7), we can formulate the null hypothesis that EMH holds as $H_0 : \beta_1 = 0$. Under the null, a zero coefficient on the lagged value of returns suggests that past information observable in the market prior to week t and included in r_{t-1} is not helpful in predicting the return in week t.

ii. The alternative hypothesis is $H_0 : \beta_1 \neq 0$. We compute a t-stastitic as:

$$t_{calc} = \frac{\hat{\beta}_1 - \beta_1}{se\left(\hat{\beta}_1\right)} = \frac{-0.059}{0.038} = 1.553.$$

We compare the t-statistic with the critical value obtained from the t-distribution under the null with $n - k - 1 = 689 - 1 - 1 = 687$ degrees of freedom, or $t_{crit} = 1.96$. Since $|t_{calc}| < t_{crit}$, this suggests that we cannot reject the null that $\beta_1 = 0$, which implies that the empirical evidence is in favour of the EMH.

(b) i. If (7) is the correct specification for describing returns, one would expect the errors, u_t, to follow a white noise process. The properties of a white noise process are the following:
- $E\left(u_t\right) = 0 < \infty$, for all t, i.e. the error term has zero mean
- $Var\left(u_t\right) = \sigma^2 < \infty$, for all t, i.e. the error term has constant variance
- $Cov\left(u_t, u_{t-j}\right) = 0$ for all t and $j \neq 0$, i.e. there is no serial correlation in the error term

ii. If the researcher suspects that lags 3 and 4 of r_t help predict r_t but have not been included in (7) then this would imply that the error term is no longer white noise but exhibits some time dependence. The researcher would be worried about the problem of serial correlation in the error term as this would affect the validity of OLS standard errors and by consequence inference based on this OLS regression.

(c) The researcher is also interested in the behaviour of the squared residuals from regression (8) because he is concerned that the variance given past information might not be constant. For this purpose he runs a regression of $\hat{u}_t^2$ on r_{t-1} and obtains the following results:

$$\begin{aligned} \hat{u}_t^2 &= \underset{(0.43)}{4.66} - \underset{(0.201)}{1.104} r_{t-1} + \hat{v}_t, \qquad (9) \\ n &= 689,\ R^2 = 0.042. \end{aligned}$$

i. In this case the researcher is worried about the problem of heteroskedasticity. Given regression (9) that the researcher is running he is implementing the Breusch-Pagan (BP) test. The original regression of interest is (7), and the null hypothesis of the BP test is stated as: $H_0 : E\left(u_t^2 | r_{t-1}\right) = \sigma^2$, i.e. the variance of the error term is constant over time. The alternative hypothesis given regression (9), is $H_1 : E\left(u_t^2 | r_{t-1}\right) = a_0 + a_1 r_{t-1}$. The BP test follows the steps below:
We evaluate OLS regression (7) and obtain residuals, $\hat{u}_t$.
We square these and obtain, $\hat{u}_t^2$, and subsequently run auxiliary regression (9).
We obtain the R^2 of (9) and construct the BG statistic $n \times R^2$. Under the null this has a chi-squared distribution. We compare the statistic that we obtain with the critical value obtained from the chi-squared distribution with $q = 1$ degrees of freedom in this case to reject or not reject the null hypothesis.

ii. Given the information provided in (9) we have that $n \times R^2 = 689 \times 0.042 = 28.94$. The corresponding critical value from the chi-squared distribution with 1 degree of freedom at 5% significance level is 3.84. Since $28.94 > 3.84$ this indicates that the null of constant variance is rejected and hence heteroskedasticity is evident in the errors. In this case our advice would be to use heteroskedasticity robust standard errors instead of the usual OLS standard errors The option of log transformation is not valid here because returns can be negative. Also weighted least squares here is tricky because variance is not proportional to a single variable (it cannot be proportional to r_{t-1} because variance is always positive, but r_{t-1} can be negative). So, among the remedies that we have learnt in this unit, robust standard errors is the only one feasible.

3. **This is based on a question from S2, 2016 final exam:**

(a) A researcher who wished to study the behavior of a stationary time series $\{y_t\}$ estimated both an AR(2) model,

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t,$$

and an ADL(2,1) model,

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \beta_1 x_{t-1} + u_t.$$

The researcher obtained the results reported in Table 1 below (standard errors are reported below the estimated coefficients). Based on the information in Table 1, which model do you prefer? Briefly explain.

Table 1		
	AR(2)	ADL(2,1)
$\widehat{c}$	1.28 (0.53)	1.30 (0.44)
$\widehat{\phi}_1$	-0.31 (0.09)	-0.42 (0.08)
$\widehat{\phi}_2$	-0.39 (0.08)	-0.37 (0.08)
$\widehat{\beta}_1$	–	-2.64 (0.46)
$\overline{R}^2$	0.55	0.71
SSR	475	462
AIC	1.08	1.04
BIC	1.09	1.11
n	200	200

Answer: There are wrong answers, but there is no unique correct answer. $\bar{R}^2$ and AIC favour ADL(2,1), while the BIC prefers the AR(2). Also, $\widehat{\beta}_1$ has a t-statistic of -5.74, which means x_{t-1} is a highly significant predictor of y_t. So overall, there is more evidence to support ADL(2,1). However, an answer that says something like "I choose AR(2) because it has lower BIC even though other criteria favour ADL(2,1) because smaller models often do better in forecasting" will be fine also. However, an answer that says "I choose ADL(2,1) because it has smaller SSR" would be incorrect, because ADL(2,1) would necessarily smaller SSR because it has one more explanatory variable.

(b) i.

$$\widehat{E}(y_t|y_{t-1}, y_{t-2}, x_{t-1}) = \begin{cases} 1.62 - 0.39y_{t-1} - 0.31y_{t-2} + 2.30x_{t-1} & t = 1, 2, \ldots, 100 \\ 1.32 - 0.39y_{t-1} - 0.31y_{t-2} + 2.15x_{t-1} & t = 101, 102, \ldots, 200 \end{cases}$$

ii. The immediate impact of a one unit increase in x on y is zero (note that there is no x_t in the equation) before and after time 100. The long-run impact is $\frac{2.30}{1-0.39-0.31} = 7.67$ before time 100, and it is $\frac{2.15}{1-0.39-0.31} = 7.17$ after time 100.

iii.

$$\begin{aligned}
H_0 &: \beta_0 = \beta_2 = 0 \\
H_1 &: \text{(at least one of } \beta_0 \text{ or } \beta_2) \neq 0 \\
F &= \frac{(SSR_r - SSR_{ur})/2}{SSR_{ur}/(200-6)} \sim F_{2,194} \quad \text{under } H_0 \\
F_{calc} &= \frac{(462-450)/2}{450/194} = 2.59 \\
F_{crit} &\approx 3.07 \\
F_{calc} &< F_{crit} \Rightarrow \text{ We fail to reject } H_0
\end{aligned}$$

There is no evidence of structural break in either the intercept or the coefficient of x_{t-1}. Some students may sample size to be 198 because of the two lags, which is fine. The degrees of freedom of the F will be $2, 192$, and $F_{calc} = 2.56$, which gives us the same conclusion.

(c) Augment the model with an equation for the dynamics of the error term

$$\begin{aligned}
y_t &= c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \beta_1 x_{t-1} + u_t \\
u_t &= \rho_1 u_{t-1} + \rho_2 u_{t-2} + e_t
\end{aligned}$$

where $e_t \sim WN\left(0, \sigma^2\right)$. The null of no serial correlation in errors against the the alternative that the error process is autoregressive of up to order 2 can be written as:

$$\begin{aligned}
H_0 &: \rho_1 = \rho_2 = 0 \\
H_1 &: \text{at least one of } \rho_1 \text{ or } \rho_2 \text{ is not zero}
\end{aligned}$$

To obtain the test statistic,

- we estimate the first equation using OLS and save its residuals $\hat{u}_t$
- we estimate a regression of $\hat{u}_t$ on a constant $y_{t-1}, y_{t-2}, x_{t-1}$ and $\hat{u}_{t-1}$ and $\hat{u}_{t-2}$. We call this auxiliary regression below and denote its R^2 by $R^2_{\hat{u}}$
- The test statics is

$$BG = n_{\hat{u}} \times R^2_{\hat{u}} \overset{a}{\sim} \chi^2_2 \text{ under } H_0$$

  where $n_{\hat{u}}$ is the number of observations in the auxiliary regression, which is $200-2 = 198$ (or $198 - 2 = 196$. Both are correct!)
- We calculate BG_{calc} and we reject the null if $BG_{calc} > BG_{crit} = 5.99$, and otherwise we do not reject the null. Rejecting the null means that there is evidence of serial correlation in errors.
- Please emphasise the importance of expressing the answer in the context of the question for full marks. Copying the procedure from BG test without any reference to the question from the formula sheet only gets half marks, and if the copied material has the wrong alternative hypothesis, it gets zero marks.