11 Exercises - Times Series: Autocorrelation and Heteroskedasticity

11.1 Consider the following model: $y_t = \beta_0 + \beta_1 x_t + u_t$, where the errors follow an AR(1) process: $u_t = \rho u_{t-1} + e_t$. Assume that the sample average of x_t is zero ($\bar{x} = 0$). Show that the variance of the OLS estimator, $\hat{\beta}_1$, is given by:

$$Var(\hat{\beta}_1) = \sigma^2 / SST_x + 2(\sigma^2 / SST_x^2) \sum_{t=1}^{n-1} \sum_{j=1}^{n-t} \rho^j x_t x_{t+j}$$

[Hint: start by writing the OLS estimator as $\hat{\beta}_1 = \beta_1 + 1/SST_x \sum_{t=1}^{n} x_t u_t$, where $SST_x = \sum_{t=1}^{n} x_t^2$. Also, note that $E(u_t u_{t+j}) = Cov(u_t u_{t+j}) = \rho^j \sigma^2$.]

11.2 Consider a sample of observations $Y_1, Y_2, ..., Y_n$ generated by the process $Y_t = \beta + u_t, t = 1, 2, ..., n$, where β is an unknown parameter and u_t is a random variable with zero mean and variance σ^2. u_t follows an AR(1) process given by $u_t = \rho u_{t-1} + \epsilon_t$, where $0 < \rho < 1$ and $\epsilon_1, \epsilon_2, ..., \epsilon_n$ is a sequence of i.i.d random variables with zero mean and variance λ^2.

a) Show that the OLS estimator for β is given by $\hat{\beta} = \sum_{t=1}^{n} Y_t / n$.

b) Prove that the variance of the OLS estimator is larger than σ^2/n. What is the importance of this result?

c) Show that the usual estimator of σ^2, given by $\sum_{t=1}^{n} u_t^2/(n-1)$, will not be unbiased in this case.

d) What are the implications of the results from b) and c) for the confidence intervals of β?

11.3 Fair (1996) explains the proportion of the two-party vote going to the Democratic candidate using data for every four years from 1916 through 1992 (20 observations). Consider the following simplified version of Fair's model:

$$\widehat{demvote} = \underset{(.012)}{.481} - \underset{(.0405)}{.0435} partyWH + \underset{(.0234)}{.0544} incum + \underset{(.0041)}{.0108} partyWH.gnews - \underset{(.0033)}{.0077} partyWH.inf$$

where $partyWH$ takes value 1 if a Democrat is in the White House and -1 if a Republican is in the White House, $incum$ takes value 1 if a Democratic incumbent is running, -1 if a Republican incumbent is running, and zero otherwise, $gnews$ is the number of quarters, during the current administration's first 15 quarters, where the quarterly growth in real per capita output was above

2.9% (annualized rate), and inf is the average annual inflation rate over the first 15 quarters of the administration.

a) What argument can be made for the error term in this equation being serially uncorrelated? (Hint: How often do presidential elections take place?)

b) When the OLS residuals from the regression above are regressed on the lagged residuals, we obtain $\hat{\rho} = -.068$ and $se(\hat{\rho}) = .240$. What do you conclude about serial correlation in the u_t?

11.4 Consider the estimates for the simple regression of growth in real per capita consumption (gc) on growth in real income per capita (gy), using data from 1973 to 2009.

$$\widehat{gc} = \underset{(.0019)}{.0081} + \underset{(.0673)}{.5708} gy$$

a) Interpret the regression coefficient on gy. Is it statistically significant?

b) The Durbin-Watson (DW) statistic for the previous regression is 2.12. Consider now the estimation results below, where $resid$ are the residuals from the model above. What do you conclude from these results? Why was gy also included in this regression? How do you compare the coefficient on $resid_{t-1}$ with the value of the DW statistic?

Source	SS	df	MS
Model	.000015029	2	7.5143e-06
Residual	.001777543	33	.000053865
Total	.001792572	35	.000051216

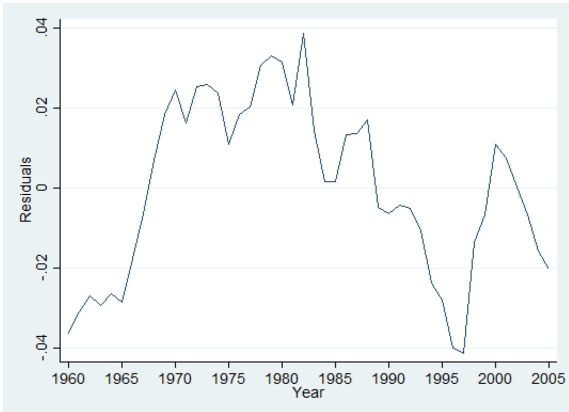
Number of obs	= 35
F(2, 33)	= 0.14
Prob > F	= 0.8703
R-squared	= 0.0084
Adj R-squared	= -0.0517
Root MSE	= .00734

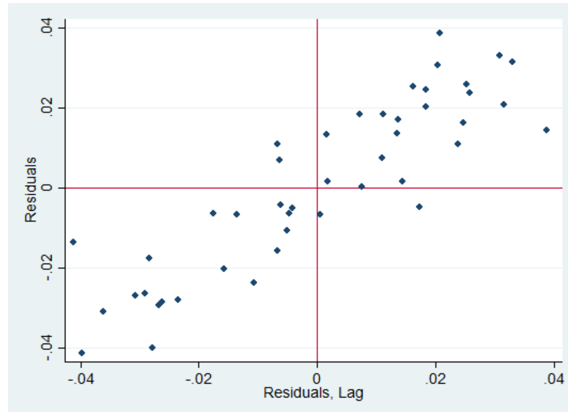
resid	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
resid L1.	-.0995161	.1884401	-0.53	0.601	-.4829003	.2838681
gy	.0075731	.0460046	0.16	0.870	-.086024	.1011702

11.5 Using U.S. data for the period 1960-2005, the following regression of wages on productivity in the business sector was estimated:

$$\ln \widehat{Comp_t} = \underset{(.0547)}{1.6067} + \underset{(.0124)}{.6522} \ln Prod_t$$

a) Below you have the plot of the residuals from the above regression along time (left hand-side) and the plot of the residuals against their first lag (right hand-side). What can you conclude regarding serial correlation?





b) The Durbin-Watson (DW) for the above regression is 0.218. What do you conclude? State the null and the appropriate alternative for this test. What are the critical values that you consider for this test? Why?

c) A higher order of autocorrelation (6) was tested using the Breusch-Godfrey test. What do you conclude from the output of this test, provided below?

Source	SS	df	MS
Model	.01232495	7	.001760707
Residual	.003817629	33	.000115686
Total	.016142578	40	.000403564

Number of obs =	40
F(7, 33) =	15.22
Prob > F =	0.0000
R-squared =	0.7635
Adj R-squared =	0.7133
Root MSE =	.01076

resid_ln	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lnprod	.0001753	.0003875	0.45	0.654	-.000613	.0009636
resid_ln						
L1.	1.013937	.1735067	5.84	0.000	.660935	1.366939
L2.	-.0314635	.247091	-0.13	0.899	-.5341738	.4712469
L3.	-.2853235	.2448779	-1.17	0.252	-.7835314	.2128843
L4.	.1454108	.2451024	0.59	0.557	-.3532538	.6440754
L5.	-.0743406	.2463726	-0.30	0.765	-.5755894	.4269082
L6.	.0406316	.1696337	0.24	0.812	-.3044907	.3857539

d) Suspecting model misspecification, a trend was added to the initial model. Considering the reported estimates below, what do you conclude? Is there still evidence for autocorrelation?

$$\ln \widehat{Comp}_t = \underset{(.307)}{.121} - \underset{(.002)}{.008}\, t + \underset{(.0776)}{1.0283} \ln Prod_t \qquad DW = .45$$

e) Three alternatives were considered in order to correct for autocorrelation:

i) Initial model in first-differences: $\Delta \ln \widehat{Comp}_t = \underset{(.057)}{.654} \Delta \ln Prod_t \quad DW = 1.74$

ii) Model with trend in first-differences: $\Delta \ln \widehat{Comp}_t = \underset{(.003)}{.001} + \underset{(.109)}{.619} \Delta \ln Prod_t \quad DW = 1.71$

iii) Quasi-differenced initial model (Cochrane-Orcutt estimation):

$\ln \widehat{Comp}^*_t = \underset{(.192)}{1.955} + \underset{(.042)}{.577} \ln Prod^*_t \quad , \hat{\rho} = .869 \qquad DW = 1.70$

Is the problem of autocorrelation solved?

f) Consider the following heteroskedasticity test for the initial model:

$$\widehat{u_t}^2 = \underset{(23.64)}{94.27} - \underset{(.514)}{1.836} \widehat{\ln(y_t)} + \underset{(.003)}{.009} \widehat{\ln(y_t)}^2 \qquad n = 46 \quad R^2 = .296$$

What do you conclude? Is there a solution for both serial correlation and heteroskedasticity?

g) Consider the following estimates for the initial model with Newey-West standard errors. How do they compare with the usual standard errors?

$$\ln \widehat{Comp}_t = \underset{(.079)^*}{1.607} + \underset{(.018)^*}{.652} \ln Prod_t$$

11.6 Consider the following estimated model:

$$\widehat{return}_t = \underset{(.081)}{.180} + \underset{(.038)}{.059} return_{t-1}, \qquad n = 689, \quad R^2 = 0.004$$

where $return_t$ is the weekly return using NYSE data from January 1976 through March 1989.

a) The Efficient Market Hypothesis (EMH) states that information observable to the market prior to week t should not help predict the return during week t, i.e., EMH presumes that the expected return given past observable information should be constant, otherwise such investment opportunities will be noticed and will disappear almost instantaneously. Is the EMH verified in the above model?

b) Consider now the Breusch-Pagan test for heteroskedasticity:

$$\widehat{u_t}^2 = \underset{(.43)}{4.66} - \underset{(.201)}{1.104} return_{t-1}, \qquad n = 689, \quad R^2 = 0.04$$

What do you conclude regarding the presence of heteroskedasticity in the initial regression? How do you interpret the sign of the coefficient on $return_{t-1}$?

c) Consider now the following estimates:

$$\widehat{u_t}^2 = \underset{(.44)}{2.95} + \underset{(.036)}{.337}\,\hat{u}_{t-1}^2, \qquad R^2 = 0.11$$

What can you conclude regarding autoregressive conditional heteroskedasticity? What is this model useful for?

d) The initial model was estimated by FGLS accounting for heteroskedasticity and AR(1) serial correlation:

$$\widehat{return}_t = \underset{(.078)}{.169} + \underset{(.043)}{.033}\,return_{t-1} \qquad R^2 = 0.0008$$

Is the EMH more evident now?