

EC2203: Econometrics Project Answer Key

1. Well motivated and presented
2. Good use of basic economic theory and/or existing literature to motivate topic and variables used in study
3. Interesting idea
4. Interesting, rich and relevant data set used
5. Good use of descriptive tables and figures to summarise main features of the data.
6. Three decimal places is enough in reporting either sample means or regression estimates (since these are estimates any more decimal places is just spurious accuracy)
7. Good use of tables to present regression results – standard errors below coefficients, different columns for different specifications
8. Good use of **relevant** diagnostic tests on the model
9. Endogeneity can exist because of **3** factors – measurement error, interdependence and omitted variable bias. Make sure you consider each of these features in formulating your model and testing your results. If you think there is endogeneity try instrumenting (2SLS)
10. Presentation and motivation could have been better. Read journals for hints on topic motivation and for ideas on how to motivate and present your work. Give references of the relevant articles you have read.
11. You should **always** discuss the estimated effect of any **significant** variables in your model – this is the fundamental purpose of any regression
Do **not** discuss the effect of any insignificant variables (insignificant t value means that the effect of the variable is indistinguishable from zero)
12. Your model may suffer from interdependence. Causality should always run from the right hand side to the left and never from left to right. Think about lagging the right hand side variables to get round this problem.
13. When presenting the results of diagnostic tests, either add them to the foot of a Table (again read journals for examples) or include them as a footnote in the relevant part of the text
14. When using time series data you should **always** check for
 - i) stationarity
 - ii) autocorrelation
 - iii) endogeneity

If any of these problems exist do something about it (see lecture notes)

If in addition you test for heteroskedasticity, make sure you use the ARCH test

15. **Always** discuss the units of measurement of your data. This helps with the interpretation of the results
16. When using cross-section data you must test formally for the existence of heteroskedasticity (using the Breusch-Pagan version of the test). If heteroskedasticity is found to exist do something about it.
17. When using dummy variable interaction terms, be careful you interpret the effect correctly (it gives a differential slope effect: see lecture notes). Think about doing a Chow test of a sample split across the two groups defined by the dummy variable.
18. The number of observations is rather small (less than or equal to the minimum recommended) or smaller than it need be (because of random sampling from a larger data set). Always try to get as many observations as you can since the higher the sample size the more precise the estimation and the more things can be tested.
19. You were asked to start with a multiple regression. Simple regressions are just correlations and say little about causal relationships.
20. Think how you might use dummy variables and interaction terms in your analysis (see lecture notes)
21. In a lin log (or log-lin) model make sure you interpret the estimated effects of the variables correctly (see lecture notes). Don't confuse percentage changes with percentage point changes.
22. Make sure you understand what moving from 5% to 1% level of significance does (increases critical value, so makes it harder to reject null)
23. Your model may suffer from spurious regression problem. It is hard to think of a reason why the variables would be related other than the fact that they have a common trend.
24. Your variables are likely to be highly correlated (collinear) so they shouldn't appear in the same model.
25. If using seasonal (eg monthly or quarterly) time series data you should always add seasonal dummies to try and account for any seasonality in the data
26. Econometrics is about trying to establish a causal relationship **not** a correlation