

Comments on Empirical Exercises

- Presentation
 - Tables
 - * Include description of variable, NOT your Stata variable name
 - * Include standard errors and a note about how they are calculated
 - * The reader should be able to understand what the table shows (what the independent and dependent variables are, what estimation method was used, and how the standard errors were calculated) without referring to the text. It is often useful to include this information in the table notes immediately below the results.
- Standard Errors
- Regression Equations
 - You should show the regression equations of interest in the body of the paper. The tables should correspond to the estimation of these equations.
 - When you discuss a particular equation or the results from estimating that regression equation, make sure you refer to the table and to the appropriate column within the table, if your table has multiple sets of results. This allows the reader to quickly match any results in the table to discussion in the text.
- Selection Bias/Endogeneity
 - You should explain why a given regression is endogenous, not simply that there is an endogeneity problem. Also, what would the sign of the bias be under that story? For example, in the METCO case, it might be that the characteristics of the schools that are selected for METCO participation look different than those schools that are not selected. This would be a problem if we thought that the characteristics of those schools directly affected the performance of white or minority students. Students have pointed out that the scores of white students are higher at participating schools. However, this might not be a function of METCO, but simply of baseline school and school district characteristics. This would introduce upward bias into the regressions.
- OLS
 - You should always show a basic OLS, even if you think that the coefficients are biased.

- Explain why you think the coefficients are biased (conditional on believing they are.)
 - When you write about the OLS and you think that there are selection or OVB issues, try to avoid writing as if the results are causal. You can say things like, there is an association between higher test scores and METCO participation. This association could either be a causal benefit of METCO or could indicate that better schools chose to participate in METCO.
- Instrumental Variables and 2SLS
 - Need to discuss 2 important aspects:
 - * The instrument predicts the endogenous regressor. i.e. there is a first stage. Please show the first stage regressions.
 - * The exclusion restriction holds - i.e. that the instrument is only correlated with the outcome through the endogenous regressor of interest. This is not something that can be shown in a regression. You need to argue why or why not this is a plausible assumption. What are possibly mechanisms through which the exclusion restriction might be invalid. To use a silly example from METCO: The instrument is basically distance to Boston. Suppose that land value is less expensive the farther from Boston. Therefore, far districts may have nicer schools, which may imply higher achievement. This for example, would bias the results downward.
 - If you run 2SLS, you should always show the first stage, the reduced form and the 2SLS output

MIT OpenCourseWare
<http://ocw.mit.edu>

14.33 Economics Research and Communication
Spring 2012

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.