Smart use of OLS

In spite of a huge expansion in econometrics literature, OLS is still the most ubiquitous method of estimation in Micro data econometrics. This simple but resilient tool can be ended to falsified results if one uses it without a deep insight of how it works and how it should be adapted to different problems. For most of the students econometrics is either reading some heavy cookbook textbooks or just generating a table of results using econometrics packages that only needed to have significant coefficients and high R-square! No one can use this apparatus correctly unless he has cracked this black box beforehand.

This mini-class aims to build an intuitive but rigorous understanding of OLS which is very useful in doing empirical research in applied micro constituting a solid percentage of researches and theses that is done in GSME. The content of this course is an organized summary of what I have learned from a group of best econometricians in the world at the University of Chicago. This "behind of the scene" knowledge of econometrics is something that barely can be found in conventional textbooks, so I like to share this valuable and extraordinarily useful knowledge with my fellow friends in GSME. After this course it will be expected that the participants have a solid understanding of how OLS works at *their mind* not just a weak memories of some formulas in a burdensome textbook!

The materials that are scheduled to cover in this mini-course are as follows:

1. The mechanics of OLS
2. Statistical meaning of OLS
3. Three different interpretations of OLS
4. Endogeneity
   4.1. How endogeneity makes the estimations inconsistent?
   4.2. Why endogeneity is the most important concern in doing any regression to the point that every problem that yields to inconsistency can be considered as a kind of endogeneity?
   4.3. How should we deal with endogeneity?
5. Measurement error VS proxy variable
6. Long regressions VS short regressions: Why adding more regressors in order to add more controls could be a detrimental mistake?
7. Treatment effect: Why measuring treatment effect is not as easy as it seems?
   7.1. Dif in Dif approach
   7.2. Local treatment effect VS average treatment effect
   7.3. Treatment with heterogeneous returns
8. Sample selection: Why most of the datasets in economics are not really a random sample?
   8.1. Selection VS Endogeneity
   8.2. Fixing sample selection bias
      8.2.1. Heckman's two stage Logit method
      8.2.2. Nonparametric probability score method
9. Missing data: How should we deal with the fact that most of the datasets in economics have missing data?