Problem Set 1

Panel Data and Program Evaluation

Due on 2020/12/02

I. Econometrics

1. Let's think about the problem of serial correlation using a simple example. Consider the following sample-mean model for the panel data:

$$Y_{it} = \beta + u_{it}$$

where $t = 1, \cdots, T$ for all i (the panel is balanced). The structure of u_{it} is $u_{it} = \theta_i + \epsilon_{it}$. θ_i is uncorrelated with ϵ_{it}; $Var(\theta) = \sigma_\theta^2$; $Var(\epsilon_{it}) = \sigma_\epsilon^2$.

   (a) Assume that u_{it} is i.i.d, write down the OLS estimate for β (denote it as $\hat{\beta}$) and the OLS estimate of its variance, $\widehat{Var}\left(\hat{\beta}\right)$.

   (b) Write down the actual variance of $Var\left(\hat{\beta}\right)$ and compare it with $\widehat{Var}\left(\hat{\beta}\right)$,which one is larger?

(c) What is the estimate of the variance of $\hat{\beta}$ if we cluster the standard error by each individual i?

2. Suppose we use following model to evaluate a certain policy:

$$Y_{it} = \alpha_0 + \alpha\tau_{g(i)t} + \delta t + \theta_i + \varepsilon_{it}$$

where $g(i) = 1$ is i is treated and zero otherwise, and $\tau_{g(i)t} = 1$ if $g(i)$=1 and the policy is enacted before t.

(a) What is the FE estimator for α? You may work out the case of two periods and then consider the general case with multiple periods.

(b) Compare the FE estimator with the FD (first-difference) estimator for α.

II. Software exercises