Policy Evaluation: Problem Set 1
Due July 3, 2019

You have been hired by the Vietnam Chamber of Commerce and Industry (VCCI) to evaluate a training program on corporate governance and internal accounts management. The designers of the training course believe that businesses which have higher quality corporate governance and strong procedures for managing internal purchases will have enjoy greater productivity and business growth. Your job is to construct the research design and analyze the data from the experiment.

1. Download the dataset dn2016_t00dn.dta. This data is the 2016 Enterprise Survey, annual survey of businesses in Vietnam conducted by the General Statistical Office. Open the dataset in STATA and start a .do file. At the top of the file, set a seed (**set seed 20190625**). This command is important as it will ensure that the randomization sequence is the same for all students.

2. Limit your population of firms to Vietnamese private businesses, who are the main target for this study. **Loại hình kinh tế của doanh nghiệp (lhdn)=6-10.**

3. What type of experimental design (i.e. encouragement, over-subscription) do you believe is most appropriate for this study? Explain why?

4. Using the data provided perform a power calculation (**power)** in STATA. The managers of the training program believe that firms exposed to the management training program will experience a 4% increase in profitability above the control group. They estimate a standard deviation of 870 in profitability. Profitability is coded as (*kqkd23*) in the dataset. Assuming individual level randomization, how many firms do you need to sample to have sufficient power to identify a statistically significant effect if one exits?

5. BONUS: For logistical reasons, VCCI has suggested performing the training at the provincial level, working with local VCCI branches to conduct the training. They would like to randomly sample 32 provinces and cities to be in the control group and 31 provinces and cities for the treatment group. Using the **loneway** command, calculate the intracluster correlation in profitability before the experiment. Perform a power calculation that takes into account this clustered design suggestion.

6. Using the number of observations from your power calculation in step 4, draw a random sample of firms for the experiment, blocking on firm ownership (*lhdn*). Use the STATA command (**gsample**).

7. Again, blocking on *lhdn*, randomly assign an equal number of firms to the control and treatment group.

8. After the intervention, VCCI reports that there was some non-compliance in the experiment. 30% of firms selected for treatment did not show up for the training. At

the same time, 5% of the control group enrolled in the training, even they were not permitted to do so. Explain the effect of non-compliance on the effect of the program, and whether you estimate the intention to treat effect (ITE) or the Treatment Effect on the Treated (TER) for the experiment.

9. What sorts of limitations (i.e. LATE, attrition) do you expect for this experiment? What can be done to mitigate the effects of those problems?

10. Submit a .do file that correctly runs from the raw data and include annotated answers. Points will be taken off if the .do file does not run correctly from start to finish.