KING'S COLLEGE LONDON
(University of London)

King's Business School

6SSMN961: APPLIED ECONOMETRICS

2019-20

Problem Set 6

Hand-in instructions:

The deadline for submission is **Monday 25 November at 10.00 am via KEATS**.

You need to submit **ONE file** containing answers to the questions and the Stata output. You can use an online pdf merger to merge your answers with the Stata output file. For example: https://www.pdfmerge.com/

You should name the files with your student number as follows: Studentnumber.pdf. This is very important to ensure that your work can be identified.

General information:

This problem set is based on the paper:

Sá, F. (2018), "The Effect of University Tuition Fees on Applications and Course Choice: Evidence from a Natural Experiment in the UK", *Economica*.

The paper can be found on the course page on KEATS. You will be asked to reproduce some of the results in the paper.

Questions:

Download the data *TuitionFees.dta* from the course page. This dataset contains data on the number of university applications for students domiciled in England and Scotland, for different universities and subjects for the period from 2008 to 2015.

In 2012, tuition fees for students domiciled in England increased from £3,375 to £9,000 a year. At the same time, students domiciled in Scotland do not have to pay any fees to attend university in Scotland. The article uses this variation across countries in the UK to estimate the following differences-in-differences (DD) model:

$$\ln(y_{djt}) = \gamma_d + \lambda_t + \delta D_{dt} + X'_{djt}\beta + \varepsilon_{djt}$$

The subscript *d* denotes country of domicile, *j* denotes gender, age group, institution and subject group and *t* denotes year. The dependent variable is the log of the number of applications, γ_d is a dummy for each country of domicile and λ_t is a dummy for each year. The vector of controls

(X_{djt}) includes dummies for gender and age group and the log of population living in country d in group j in year t. The regressor of interest is D_{dt} and indicates observations for students domiciled in England in the period after the increase in tuition fees.

1. Use the data provided to estimate the DD model. What type of standard errors are you using and why? Interpret the DD coefficient.
2. Under what assumption can you interpret the DD estimate from part 1 as the causal effect of the increase in tuition fees on university applications? Use the data to construct a graph that provides a visual check of this identification assumption. To do this, calculate the total number of applications in each year by students domiciled in England and Scotland (use the command *collapse (sum)*). Calculate the log number of applications and create a graph with the evolution of the log number of applications over time for students domiciled in England and Scotland. What do you conclude from this graph?
3. Repeat the analysis in part 1 but extend the model to include a country-specific linear trend. How do your results change?
4. To check whether the effect of the increase in tuition fees on applications is different for STEM subjects (science, technology, engineering and mathematics) and non-STEM subjects, the model is modified as follows:

$$\ln(y_{djt}) = \gamma_d + \lambda_t + \alpha STEM_j + \delta_{STEM} D_{dt} \times STEM_j + \delta_{non-STEM} D_{dt} \times NON - STEM_j + X'_{djt}\beta + \varepsilon_{djt}$$

   The variable $STEM_j$ is an indicator equal to 1 for STEM subjects and 0 for non-STEM subjects. Conversely, the variable $NON - STEM_j$ is an indicator equal to 1 for non-STEM subjects and 0 for STEM subjects.

   Use the data provided to estimate this modified version of the model. Interpret the estimates δ_{STEM} and $\delta_{NON-STEM}$. What do you conclude about the effect of the increase in tuition fees on applications for different types of subjects?
5. The dataset contains information on the average salary of graduates six months after graduation for each subject, institution and gender (this is measured in 2011). It also has a variable indicating the quartile of the distribution of salaries for each group, where quartile 4 denotes higher salaries and quartile 1 denotes lower salaries. To check whether the increase in tuition fees affects applications differently depending on the expected salary after graduation, the model is modified as follows:

$$\ln(y_{djt}) = \gamma_d + \lambda_t + \sum_{k=1}^{4} \alpha_k Q_j + \sum_{k=1}^{4} \delta_k D_{dt} \times Q_j + X'_{djt}\beta + \varepsilon_{djt}$$

   Q_j is a set of indicator variables for each quartile of the distribution of average salaries.

Use the data provided to estimate this modified version of the model. Interpret the estimates $\delta_1, \delta_2, \delta_3$ and δ_4. Does the effect of the increase in tuition fees on applications depend on the expected salary after graduation?