EXAMINATION QUESTION PAPER - Take-home examination

DRE 70061
Panel Data/Microeconometrics

Department of Economics

Start date:	17.06.2019	Time 09.00
Finish date:	18.06.2019	Time 15.00

Weight:	100% of DRE 7006
Total no. of pages:	3 incl. front page
No. of attachments files to question paper:	0
To be answered:	Individually

Answer paper size:	10 pages. excl. attachments
Max no. of answer paper attachment files:	1
Allowed answer paper file types:	pdf
Allowed answer paper attachment file types:	pdf

The exam paper must also include a table of contents, a summary at the beginning and a bibliography/reference list at the conclusion of the paper. These pages are counted separately from the main paper. You will find a thorough explanation of how to use quotes and references on the BI library webpage: https://portal.bi.no/en/services/library/write-and-cite/. For information on formal requirements and template paper; see www.bi.edu/templatepaper

Question 1 (counts 70 %)

In this problem, we will look closely into some of the analyses in the paper "The causal effect of competition on prices and quality: evidence from a field experiment" by Busso and Galliani (2019) and the data and programs that are available with the paper.

a) Give a short description (one page) of the empirical strategy applied in the paper.
b) The authors suggest that they estimate average treatment effects through an instrumental variable strategy. Under what assumptions do they estimate an average treatment effect?
c) Try to replicate the analysis of Table 5 - panel A. (You will need to find the datafile "Retailers_panel.dta". The dependent variable is log_P_index, the endogenous covariate is hfgt0 and the instrument is hgt0. It is necessary to condition on round==2. The lines in the table correspond to dummy variables sample1 sample2 and sample3. It is necessary to cluster on the variable mercado to get the right standard errors. It should be fairly easy to find the list of controls needed for columns (4) and (6) in the file "tables.do".) The analysis uses an endogenous covariate that is originally a variable taking values 0-3, but appears as a dummy variable in the analysis. Explain why this may be problematic and redo the analysis using the endogenous covariate in unreduced form (the variable entrants). Explain why it is less problematic to use a simplified version of the instrument (they do), but try to also to change the instrument to the unreduced version (look for a variable that takes values 0 to 3 in the data set). Comment on the results. In the following subproblems, stick to the version of the instrument and the endogenous covariate used in the paper.
d) The authors report that first stage F-statistics are in most cases larger than 10. What are the first stage F-statistics for columns (5) and (6) in Table 5 panel A? It is a bit unclear if all the entries in column (6) of Table 5 are statistically significant (significantly different from zero) at the 5 percent level. In there a simple way to test whether these coefficients are different from zero using methods that are robust to the weak instruments problem?
e) Construct 95 percent confidence intervals for the results in columns (5)-(6) by inverting the tests suggested by Chernozhukov and Hansen (2008). (That is, include a value in the 95 percent confidence interval if you cannot reject the value at a 5 percent significance level.)
f) Table 5 reports the average effect on log prices. Can you expand the analysis in Table 5 columns (3) and (4) to also report the "intention to treat"-effect on the median log price and the first and third quartiles of the log price? Comment on the results.
g) We are interested in the external validity of the results. Can you, based on the same data you have used for analyzing Table 5 - Panel A - column (5), estimate the number of always-takers, never-takers and compliers. In addition, try to estimate the average log price for always-takers, never-takers, treated compliers and untreated compliers (based on Imbens and Rubin, 1997) and compare the results to the IV estimates in column 5.

Question 2 (counts 30 percent)

In panel data analyses or difference-in-differences analyses we often need to use clustered standard errors to take into account potential time series correlation in the data. This question asks you to perform a small simulation study to assess the performance of different standard error procedures. The simulation study mimics a difference-in-differences problem.

a) Simulate a panel data set with 12 units and 6 time periods. Let the outcome be independent standard normal for each unit in period 1. In the following periods, let $y_{it} = 0.9 + y_{i,t-1} + e_{it}$, where e_{it} are independent normal with mean zero and variance 0.19. The exception is units 7-12 in periods 4-6, where you can add 1 to the outcome after generating the data. The idea is to use this dataset for difference-in-differences analyses with periods 1-3 as pretreatment period, 4-6 as treatment period, units 1-6 as control units and units 7-12 as treatment units.
b) We are interested in the following inference procedures. We want to measure two things: Whether or not we reject a null hypothesis that the treatment effect is 1 at the five percent significance level and the p-value of the same test. Implement the following five procedures and report the results from a single run.
    1. Aggregated differences-in-differences, where observations from periods 1-3 and periods 4-6 are added together before performing a standard difference in difference analysis.
    2. Standard regression based difference-in-difference analysis without any unit or time controls (except treated/untreated and pre- and post-treatment) and non-clustered standard errors.
    3. An analysis like 2. but with clustered standard errors.
    4. An analysis like 2. but with unit fixed effects.
    5. An analysis like 3. but with unit fixed effects.
c) Simulate data 1000 times and assess the performance of each of the procedures from b) based on how often we reject the (true) null hypothesis and the distribution of the p-values over the 1000 simulations. What should the distribution of p-values ideally look like? Comment on the results.
d) Redo the analysis in c), with 6 units and with 24 units. Comment on the results.

References

Busso and Galliani (2019), The causal effect of competition on prices and quality: evidence from a field experiment, American Economic Journal: Applied.

Chernozhukov and Hansen (2008), The reduced form: A simple approach to inference with weak instruments. Economics Letters.

Imbens and Rubin (1997), Estimating outcome distributions for compliers in instrumental variable models. Review of Economic Studies.