TUTORIAL 7

Empirical banking and finance

Matriculation numbers: 3391610

3290592

1.3 QUESTIONS

1. IDENTIFICATION

The authors use the staggered interstate bank branching deregulation after 1994 as a source of exogenous variation in credit supply, see Section I.B of the paper. If they just used the deregulation episode their empirical strategy would be a traditional difference-in-difference setting. But the authors go one step further and modify their estimation strategy to address some of the short-comings of a traditional diff-in-diff setting.

a) On page 959 they write: "Identification [when using state-level deregulation] typically rests on the fact that state-level deregulation is motivated by political, rather than economic, reason." What are [Favara and Imbs, 2015] doing to relax this strong assumption in Section II of the paper, i.e. Table 2?

It is because authors could design 2 control groups. Firstly, they consider thrifts and credit unions (TCUs) and Independent Mortgage Companies (IMC) as placebo sample, because those institutions' lending decisions should not be affected by deregulations. This fact suggests that increase in the lending by commercial banks should come from increase in credit supply, **but not the boom in credit demand.** To summarize, table 2 confirms this fact; the coefficient of estimate for change in lending is significant for commercial banks, but insignificant for TCUs and IMC. In that case, we do not need to assume that state-level deregulation is not motivated by economic reason, because this is already ruled out by the placebo samples.

Secondly, they split commercial banks also according to the geographic reason; basically, for a given deregulated state deregulation should affect out-of-state banks (because now they can open branches inside the state), not the in-state-banks, so the differences in the lending behavior of 2 groups should not be associated with economic reason, but the geographical differences (accordingly deregulation).

b) What is the name of their actual estimation strategy? Note: the name of the strategy is not explicitly stated in the paper.

Basically, what they did different is that although treatment is allocated on the state level, they chose control and treatment group not among the states, but among the individual units (institutions) within the state. So we can call this estimation strategy as mixture of DiD and Fixed Effect method.

c) Consider Table 1, Column 1 of [Favara and Imbs, 2015]. What do you conclude from the differences across institutions for Average loan originated and Average applicant's income concerning [Favara and Imbs, 2015] identification strategy?

For the Average loan originated, we can see that all three categories (commercial banks, TCUs and IMCs) have roughly the same growth rate in credit, which supports their identification strategy that TCUs and IMCs are good control group for the commercial banks (treatment group). Regarding the average applicant's income, the number in the table 1 suggests that TCUs and commercial banks' customer base is similar, so they do **not** have any **systematic differences** in that regard. However, the table suggests that on average, the customers of IMCs are poorer than that of other 2 groups, but the authors claim that such small differences should not lead to systematic differences in the customer base of commercial banks and IMCs. All these numbers support choice of IMCs and TCUs as control group for commercial banks.

2. DIFFERENCE IN DIFFERENCE

We start by studying the impact of deregulation on credit supply using Dl_nloans_b and +Dl_nloans_pl as LHS variables. The main specification, see Section II.A, is as follows:

$$log(yc,t) - log(yc,t-1) = \alpha c + \gamma t + \beta 1 Deregulation\ Indexs,t-1 + \beta 2 Xc,t + ec,t$$

with Xc,t including Dl_her_v, Dl_inc, Dl_pop and a lagged version of each of these controls, additionally a lagged version of the LHS variable.

a) Discuss how the key identifying assumption changes when estimating equation 1 with yc,t as LHS variable instead of log(yc,t) − log(yc,t−1)? Illustrate your argument with a graph, if possible.

Their identification strategy (there is no systematic difference in the growth between treated and untreated counties) then should also be applied to the level of lending.

b) We will use log(yc,t) − log(yc,t−1) as LHS variable, as in the paper. The set of control variables Xc,t includes the lagged LHS variable, i.e. log(yc,t−1) − log(yc,t−2). Why could it be important to include the lagged LHS variable? Provide a concrete economic example.

It might be because current trend in the county-level activity in the mortgage market may depend also on its past values. It might be case that before and after regulation there were some differences across counties affecting mortgage activity. Maybe before the treatment happened, there was some favorable factors (correlated with deregulation) in some counties leading to higher mortgage loans, so this factor may still be there after treatment, creating bias in case of being omitted. However, if we add lagged values, we can control the factors affecting past values in the mortgage activity in the counties, as well. For example, if one of county has a vision to be as "Silicon Valley" of the country, government is motivated to introduce deregulation for those counties and because of higher inflow of labor force (e.g. "technologers") into the county, mortgage loans would be more supplied in the county. Also change in loan supply would be also time-varying (like depending on the immigrated labor force), so that is why there can be some correlation between current and past values of mortgage loan, driven by omitted variable.

c) Replicate column (1) of Table 2 of the paper, both for Panel A and B. What is the precise interpretation of β^ 1 and the economic magnitude of the effect found in Panel A. Briefly comment on the different estimates found for β1 in Panel A vs Panel B.

For Panel A

Dl_nloans_b	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
Linter_bra	.0277614	.0104958	2.65	0.011	.0066694	.0488535
LDl_nloans_b	-.1069449	.0193096	-5.54	0.000	-.145749	-.0681409
Dl_her_v	-.151598	.033396	-4.54	0.000	-.2187098	-.0844862
LDl_her_v	-.0551171	.0169132	-3.26	0.002	-.0891055	-.0211288
Dl_hpi	.3443466	.0791661	4.35	0.000	.1852563	.5034369
LDl_hpi	.1117969	.1189747	0.94	0.352	-.1272916	.3508855
Dl_pop	3.482108	.7728152	4.51	0.000	1.929078	5.035138
LDl_pop	.5998989	.5515352	1.09	0.282	-.5084526	1.70825
Dl_inc	.1813692	.0954081	1.90	0.063	-.0103606	.373099
LDl_inc	-.1131316	.2094422	-0.54	0.592	-.5340214	.3077581
_cons	.0126887	.0184543	0.69	0.495	-.0243965	.0497739

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs	
year	11	0	11	
county	1018	1018	0	*

For Panel B

Dl_nloans_pl	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
Linter_bra	.0020663	.0081301	0.25	0.800	-.0142718	.0184043
LDl_nloans_pl	-.2357139	.0124923	-18.87	0.000	-.2608181	-.2106096
Dl_her_v	-.1794272	.0203818	-8.80	0.000	-.220386	-.1384684
LDl_her_v	-.0991887	.0131297	-7.55	0.000	-.1255737	-.0728037
Dl_hpi	.5094958	.124693	4.09	0.000	.258916	.7600757
LDl_hpi	-.2911076	.1449932	-2.01	0.050	-.5824823	.0002671
Dl_pop	2.04582	.4082875	5.01	0.000	1.225335	2.866304
LDl_pop	-.3813009	.4455738	-0.86	0.396	-1.276715	.5141132
Dl_inc	.0496651	.0364077	1.36	0.179	-.0234988	.122829
LDl_inc	.431744	.179432	2.41	0.020	.0711619	.792326
_cons	.0239467	.0155285	1.54	0.129	-.007259	.0551524

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs	
county	1014	1014	0	*
year	11	0	11	

Interpretation for panel A: When the Linter index increases by 1 (meaning fewer restrictions), log difference in the number of loan originations between 2 years increases by 0.028. That means,

deregulation increases **proportion of tomorrow's mortgage loans to today's loans** by 2.8%. Economic magnitude of the coefficient is found by dividing 0.028 by 0.45 (standard deviation of y), which makes equal to 0.06. That means on average, deregulation increases mortgage lending by 6 of the standard deviation of log of change in lending.

In the panel B, the coefficient is not statistically significant with p-value equal to 0.8. This is consistent with their assumption that IMC and TCU can be good control (placebo) for commercial banks because deregulation did not affect their mortgage lending.

In Chapter 5.2. of their book, [Angrist and Pischke, 2009] write that OLS does not consistently estimate coefficients when both individual fixed effects and lagged LHS variables are included in a specification. They also show that a specification including fixed effects (but no lagged LHS variable) provides an upper-, whereas a specification with a lagged dependent variable (but without individual FE) provides a lower bound of the treatment effect.

d) Estimate these two modified versions of Equation 1 for Dl_nloans_b as a LHS variable to provide an upper and a lower bound and comment on your results.

Specification including fixed effects (no lagged LHS):

Dl_nloans_b	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
Linter_bra	.0342796	.0106577	3.22	0.002	.0128621	.0556971
Dl_her_v	-.1663759	.0368777	-4.51	0.000	-.2404845	-.0922674
LDl_her_v	-.0326246	.0189897	-1.72	0.092	-.0707859	.0055366
Dl_hpi	.3509678	.0836242	4.20	0.000	.1829188	.5190169
LDl_hpi	.1399373	.1238263	1.13	0.264	-.1089009	.3887755
Dl_pop	3.449961	.7107679	4.85	0.000	2.02162	4.878303
Dl_pop	0	(omitted)				
LDl_pop	.471264	.6232588	0.76	0.453	-.7812215	1.72375
Dl_inc	.2140695	.1177043	1.82	0.075	-.0224661	.450605
LDl_inc	-.1271644	.2607127	-0.49	0.628	-.6510861	.3967573
_cons	-.0019658	.017437	-0.11	0.911	-.0370069	.0330752

Specification with lagged independent variable (no FE):

Dl_nloans_b	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
Linter_bra	-.0209956	.0038998	-5.38	0.000	-.0288325	-.0131587
LDl_nloans_b	-.0441979	.0203477	-2.17	0.035	-.0850881	-.0033078
Dl_her_v	-.261487	.0359964	-7.26	0.000	-.3338245	-.1891496
LDl_her_v	-.1037333	.0178421	-5.81	0.000	-.1395883	-.0678783
Dl_hpi	.2260719	.0941171	2.40	0.020	.0369365	.4152072
LDl_hpi	.1560545	.1497746	1.04	0.303	-.1449289	.457038
Dl_pop	2.6337	.5921528	4.45	0.000	1.443724	3.823675
LDl_pop	-.8371977	.5038038	-1.66	0.103	-1.849629	.1752339
Dl_inc	.217989	.0992497	2.20	0.033	.0185392	.4174388
LDl_inc	-.3562873	.2185607	-1.63	0.109	-.7955014	.0829269
_cons	.0932329	.0103972	8.97	0.000	.0723389	.1141269

So upper bound for the coefficient is 0.034 and lower bound for the coefficient is -0.021.

3. DIFFERENCE IN DIFFERENCE AND INSTRUMENTAL VARIABLE

Now, we will turn to estimating the impact of a credit supply shock on asset prices, using house price growth as a LHS variable. On page 958, [Favara and Imbs, 2015] write: "There is every reason to expect that credit supply depends on the price of assets, which may be used as collateral. Credit also responds endogenously to current and expected economic conditions."

$$log(pc,t) - log(pc,t-1) = \alpha c + \gamma t + \delta 1(log(yc,t) - log(yc,t-1))s,t-1 + \delta 2Xc,t + ec,t$$

with log(pc,t) − log(pc,t−1) local house price growth and log(yc,t) − log(yc,t−1) local loan growth.

a) Discuss how you expect the two channels mentioned in the above quote affect δ1 in Equation 2, a regression of asset prices on a measure of credit supply.

First concern is the reverse causality issue that assets prices may affect lending level, which is not easy to predict how it is going to affect the coefficient. The second concern is that there can be omitted variable (economic condition) that would create bias. In that case, we expect the coefficient to be upwardly biased because we expect correlation between lending and economic condition to be positive and also correlation between asset price and economic conditions to be positive.

Therefore [Favara and Imbs, 2015] instrument local loan growth with the interstate branching deregulation index. As explained in Section III.B ((log(yc,t)d− log(yc,t−1))s,t−1 is obtained from a first stage regression using Equation 1. The IV results are reported in Table 5.

b) Replicate column 1 of Table 5. Note that i) you have to add [aw = w1] at the end of the regression, before the comma, because the authors weigh states depending on the number of counties ii) you will not be able to get exactly the same standard errors.

```
                                                Number of obs =    11107
                                                F(  8, 10071) =   246.50
                                                Prob > F      =   0.0000
Total (centered) SS     =  19.25015325          Centered R2   =  -0.7227
Total (uncentered) SS   =  19.25015325          Uncentered R2 =  -0.7227
Residual SS             =  33.16245116          Root MSE      =   .05738
```

Dl_hpi	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Dl_nloans_b	.141466	.0316845	4.46	0.000	.079358	.203574
LDl_hpi	.5066948	.0160799	31.51	0.000	.4751749	.5382147
Dl_pop	-.0776535	.1558312	-0.50	0.618	-.3831137	.2278068
LDl_pop	.2211889	.0814594	2.72	0.007	.0615123	.3808656
Dl_inc	.0212589	.0156631	1.36	0.175	-.009444	.0519617
LDl_inc	.076526	.023116	3.31	0.001	.0312141	.1218379
LDl_her_v	.0047718	.0026721	1.79	0.074	-.000466	.0100097
Dl_her_v	.0216794	.0061075	3.55	0.000	.0097074	.0336514

c) What is the economic magnitude of the IV estimate of δ1?

In order to get an economic magnitude, we divide the coefficient 0.14 by the standard deviation of **log change** in house prices (0.044), which makes it equal to 3.18.

d) Formally test whether the instrument is weak or not. Provide H0, HA, the test statistic and its distribution. Clearly state the result.

Actually, we could not get first-stage estimation results (e.g. partial R-squared) using estat firststage. Hence, we run the first stage regression ourselves by regressing loan growth with Linter_bra index, with country and year fixed effects and then tested Linter_bra index.

```
 ( 1)  Linter_bra = 0

       F(  1, 11280) =   49.50
            Prob > F =    0.0000
```

Here null hypothesis states that Linter_bra index is zero, while alternative hypothesis states that it is non-zero. Thanks to very small p-value, we can reject the null hypothesis, which refutes that instruments are weak.

e) Read the footnotes of Table 5 carefully. If you had to write a referee report for an scientific journal, what would be your main criticism concerning the IV estimation in Table 5?

Our criticism would be that authors illustrate only second-stage regression, but not the first-stage one. It would be nicer if the reader could see first-stage regression table and be informed how well IV performs in terms of explaining instrumented variable.

f) Address the main concern raised in question b). How does the result change?

Tutorial 7 do-file.do*

```
*** Question 2
reghdfe Dl_nloans_b Linter_bra LDl_nloans_b Dl_inc LDl_inc Dl_pop LDl_pop Dl_hpiLDl_hpi Dl_her_v LDl_her_v , absorb( year county) cluster( state_n)
reghdfe Dl_nloans_pl Linter_bra LDl_nloans_pl Dl_inc LDl_inc Dl_pop LDl_pop Dl_hpi LDl_hpi Dl_her_v LDl_her_v , absorb( year county) cluster( state_n)
summarize
reghdfe Dl_nloans_b Linter_bra Dl_inc LDl_inc Dl_pop LDl_pop Dl_hpi LDl_hpi Dl_her_v LDl_her_v , absorb( year county) cluster( state_n)
reg Dl_nloans_b Linter_bra LDl_nloans_b Dl_inc LDl_inc Dl_pop LDl_pop Dl_hpi LDl_hpi Dl_her_v LDl_her_v, cluster( state_n)
*** Question 3
ssc install ivreghdfe
ssc install ranktest, replace
ivreghdfe Dl_hpi ( Dl_nloans_b=Linter_bra) LDl_hpi Dl_pop LDl_pop Dl_inc LDl_inc Dl_her_v LDl_her_v [aw= w1], absorb( year county)
summarize
reghdfe Dl_nloans_b Linter_bra, absorb ( year county)
test Linter_bra

```