CS112 Final Project Proposal

Siyu Wang, Yuhao Chen
Fall 2018

Proposal

The paper we decide to replicate is “Replication Data for: Do Businesses Pay to Do Science? The Effects of Hiring PhDs on Firms' Patenting Process”.

We have the raw data at hand, and several files of R codes for creating tables and performing matching and regression analysis. In the downloaded files, there are also stata scripts that the authors use to run the generation of the sample used, descriptive statistics and bivariate probit estimations, and robustness checks. Currently we’re wondering if we need to use stata to run the script. What’s more, there are several lines of code that we don’t understand, and will try to understand in the following weeks.

The original paper applies coarsened exact matching to group the data and performs probit regression analysis to answer two questions: the impact of PhD employment in the firm on the firm’s scientific reliance on conferences and tech workshops, and its commercialization for patents.

We’ll replicate Table 3-5 in the original paper, as following:

Table 3: Probit Estimation of Unmatched Data:Effect of PhD Degree on Conferences

	Science		*Conferences*	
	Model 1	Model 2	Model 3	Model 4
Near	0.461***	0.500***	0.467***	0.488***
	(0.057)	(0.06)	(0.049)	(0.051)
Far	0.348***	0.348***	0.555***	0.574***
	(0.05)	(0.053)	(0.044)	(0.047)
Age	-0.013***	-0.013***	-0.007***	-0.007***
	(0.002)	(0.002)	(0.002)	(0.002)
Gender	-0.163	-0.124	-0.198	-0.133
	(0.162)	(0.17)	(0.127)	(0.13)
PhD Degree	0.395***	0.442***	0.306***	0.350***
	(0.06)	(0.065)	(0.049)	(0.053)
Far Past Coinventors	0.115	0.112	0.001	0.014
	(0.075)	(0.081)	(0.067)	(0.072)
Coinventors Dummy	0.005	0.025	0.161**	0.148**
	(0.075)	(0.081)	(0.07)	(0.074)
Experience	-0.005	-0.007	-0.005	-0.006
	(0.006)	(0.006)	(0.005)	(0.005)
Country of Degree Dummy	0.190*	0.215*	0.192*	0.178*
	(0.113)	(0.124)	(0.099)	(0.104)
Mobility in Region	0.110	0.132	0.070	0.112
	(0.085)	(0.09)	(0.077)	(0.08)
Mobility out Region	-0.006	-0.010	0.059	0.066
	(0.054)	(0.058)	(0.049)	(0.051)
Experience Herfindahl	-0.222**	-0.103	0.032	0.080
	(0.107)	(0.114)	(0.094)	(0.099)
Previous Patents Dummy	0.163	0.058	-0.057	-0.101
	(0.123)	(0.132)	(0.109)	(0.115)
log Employees	0.013	0.015	0.011	0.009
	(0.011)	(0.012)	(0.008)	(0.009)
R&D Intensity	2.442***	3.019***	1.507**	2.078**
	(0.889)	(0.962)	(0.745)	(0.801)
Inventor Past Patents	0.088***	0.099***	0.068***	0.072***
	(0.016)	(0.017)	(0.013)	(0.014)
Application Year FE	YES	YES	YES	YES
NUTS2 Region FE	YES		YES	
NUTS3 Region FE		YES		YES
Regional Controls		YES		YES

N=6122,*$p < 0.1$,* ∗ $p < 0.05$,* ∗∗$p < 0.01$

Table 4: Probit Estimation of Matched Data:Effect of PhD Degree on Conferences

	Science		*Conferences*	
	Model 1	Model 2	Model 3	Model 4
Near	0.495***	0.540***	0.477***	0.498***
	(0.072)	(0.078)	(0.062)	(0.067)
Far	0.354***	0.368***	0.548***	0.591***
	(0.063)	(0.069)	(0.057)	(0.063)
Age	-0.014***	-0.013***	-0.007**	-0.007**
	(0.003)	(0.003)	(0.003)	(0.003)
Gender	0.350	0.488	0.133	0.344
	(0.294)	(0.325)	(0.261)	(0.282)
PhD Degree	0.499***	0.532***	0.393***	0.434***
	(0.074)	(0.08)	(0.06)	(0.064)
Far Past Coinventors	-0.091	-0.058	-0.106	-0.071
	(0.109)	(0.119)	(0.095)	(0.103)
Coinventors Dummy	0.096	0.134	0.180	0.154
	(0.115)	(0.126)	(0.101)	(0.107)
Experience	-0.002	-0.003	-0.002	-0.004
	(0.008)	(0.009)	(0.007)	(0.007)
Country of Degree Dummy	0.334	0.422	0.013	0.005
	(0.269)	(0.301)	(0.202)	(0.213)
Mobility in Region	-0.011	-0.060	0.019	0.087
	(0.166)	(0.177)	(0.152)	(0.159)
Mobility out Region	-0.030	-0.038	0.068	0.074
	(0.074)	(0.081)	(0.067)	(0.072)
Experience Herfindahl	-0.330**	-0.233	0.038	0.044
	(0.159)	(0.173)	(0.138)	(0.148)
Previous Patents Dummy	0.274	0.154	0.009	0.029
	(0.187)	(0.203)	(0.165)	(0.176)
log Employees	0.019	0.026	0.011	0.006
	(0.013)	(0.016)	(0.011)	(0.013)
R&D Intensity	1.764*	2.517**	1.560	2.294**
	(0.965)	(1.109)	(0.97)	(1.075)
Inventor Past Patents	0.076***	0.090***	0.062***	0.069***
	(0.019)	(0.021)	(0.016)	(0.017)
Application Year FE	YES	YES	YES	YES
NUTS2 Region FE	YES		YES	
NUTS3 Region FE		YES		YES
Regional Controls		YES		YES

N=6122,*$p < 0.1$,* ∗ $p < 0.05$,* ∗∗$p < 0.01$

Table 5: Probit Estimation of Unmatched and Matched Data:Effect of PhD Degree on Commercialization

	Unmatched		*Matched*	
	Model 1	Model 2	Model 3	Model 4
Near	-0.034	-0.029	-0.033	-0.021
	(0.045)	(0.046)	(0.055)	(0.058)
Far	-0.005	-0.005	-0.008	-0.009
	(0.042)	(0.043)	(0.052)	(0.055)
Age	0.004***	0.004***	0.006***	0.005***
	(0.002)	(0.002)	(0.003)	(0.003)
Gender	0.015	0.021	0.001	0.037
	(0.116)	(0.118)	(0.234)	(0.249)
PhD Degree	0.002	0.005	0.001	0.002
	(0.045)	(0.047)	(0.054)	(0.057)
Far Past Coinventors	0.004	0.006	0.021	0.022
	(0.063)	(0.067)	(0.088)	(0.093)
Coinventors Dummy	-0.009	-0.015	-0.044	-0.042
	(0.065)	(0.068)	(0.092)	(0.096)
Experience	-0.002	-0.001	-0.002	-0.001
	(0.005)	(0.005)	(0.006)	(0.007)
Country of Degree Dummy	0.018	0.024	-0.003	-0.006
	(0.087)	(0.09)	(0.176)	(0.183)
Mobility in Region	-0.017	-0.016	-0.057	-0.066
	(0.072)	(0.074)	(0.142)	(0.146)
Mobility out Region	0.007	0.007	0.043	0.041
	(0.045)	(0.047)	(0.062)	(0.065)
Experience Herfindahl	-0.020	-0.032	0.011	0.008
	(0.087)	(0.09)	(0.127)	(0.133)
Previous Patents Dummy	-0.009	0.012	-0.012	-0.013
	(0.102)	(0.106)	(0.151)	(0.158)
log Employees	-0.009	-0.007	-0.012	-0.012
	(0.008)	(0.008)	(0.01)	(0.011)
R&D Intensity	-0.467	-0.531	-0.383	-0.396
	(0.59)	(0.626)	(0.721)	(0.787)
Inventor Past Patents	0.005	0.005	0.003	0.004
	(0.012)	(0.012)	(0.015)	(0.015)
Application Year FE	YES	YES	YES	YES
NUTS2 Region FE	YES		YES	
NUTS3 Region FE		YES		YES
Regional Controls		YES		YES

N=6122,*$p < 0.1$,* ∗ $p < 0.05$,* ∗∗$p < 0.01$

As extension, we’ll apply genetic matching to reach higher balance in the matched groups, and visualize the results in the table through graphs. We’ll also try to do a sensitivity test on whether

the hidden bias from unobserved variables are significant enough to change the qualitative conclusion from the data.

What's more, the paper only discussed the scientific orientation of the firm from the theoretical perspective, using the variable of "intention to commercialize", but we'll take a step further and test out the treatment effect on the firm's resultant commercialization rate as well as its revenues and costs. Since those data are not provided, we're still considering whether or not to estimate and how to estimate them.

[Word count: 281]

Reference

Sappenfield, James M.;Shi, Dongbo;Mushtaq, Shawn J., 2015, "Mushtaq Sappenfield Shi, Do Businesses Pay to Do Science, 2015.pdf", *Replication Data for: Do Businesses Pay to Do Science? The Effects of Hiring PhDs on Firms' Patenting Process*, https://doi.org/10.7910/DVN/MREYRZ/TBGXS8, Harvard Dataverse, V1