Difference-in-Differences and OLS Regressions

Sid Satya

This is a notebook that runs the regressions for my honors thesis project.

Difference-in-Differences Regressions

Importing Relevant DiD Datasets:

```
# Wage Datasets
pa_wage <-read.csv(file = "/Users/sidsatya/Desktop/berkeley/senior_year/thesis/aithesis/data/physicians
sa_wage <-read.csv(file = "/Users/sidsatya/Desktop/berkeley/senior_year/thesis/aithesis/data/secretary_a

# Employment Datasets
pa_emp <-read.csv(file = "/Users/sidsatya/Desktop/berkeley/senior_year/thesis/aithesis/data/physicians_a
sa_emp <-read.csv(file = "/Users/sidsatya/Desktop/berkeley/senior_year/thesis/aithesis/data/secretary_ai
```

Physicians DiD Regressions:

```
# Wage DiD for Physicians vs. Aides, reg1 = without controls and with FE, reg2 = with controls and with
pa_wage_did_reg1 <- lm(LOGWAGE ~ AI + AIPost + factor(YEAR) + factor(IND1990), data = pa_wage)
pa_wage_did_reg2 <- lm(LOGWAGE ~ AI + AIPost + FRACCOLLEGE + FRACWHITE + FRACFEMALE + AGE + factor(YEAR)

# Employment DiD for Secretaries vs. Aides
pa_emp_did_reg1 <- lm(LOGEMP ~ AI + AIPost + factor(YEAR) + factor(IND1990), data = pa_emp)
pa_emp_did_reg2 <- lm(LOGEMP ~ AI + AIPost + FRACCOLLEGE + FRACWHITE + FRACFEMALE + AGE + factor(YEAR) +
```

```
stargazer(pa_wage_did_reg1,
          pa_wage_did_reg2,
          pa_emp_did_reg1,
          pa_emp_did_reg2,
          header = FALSE,
          type = "latex",
          title = "Difference-in-Differences: Physicians and Surgeons as Treatment Group, Nursing, Psycl
          omit = c("IND1990", "YEAR"),
          column.labels =c("No Controls", "With Controls", "No Controls", "With Controls"),
          covariate.labels = c("AI", "AI x Post", "Frac. College", "Frac. White", "Frac. Female", "Age"
          dep.var.labels = c("log(wage)", "log(employment)"),
          omit.stat =c("LL", "ser", "f"),
          add.lines = list(c("Industry FE?", "Yes", "Yes", "Yes", "Yes"),
                           c("Year FE?", "Yes", "Yes", "Yes", "Yes"))
          )
```

Table 1: Difference-in-Differences: Physicians and Surgeons as Treatment Group, Nursing, Psychiatric, and Home Health Aides as Control Group

	Dependent variable:			
	log(wage)		log(employment)	
	No Controls	With Controls	No Controls	With Controls
	(1)	(2)	(3)	(4)
AI	1.776^{***}	1.366^{***}	-1.284^{**}	5.576^{***}
	(0.030)	(0.120)	(0.569)	(1.002)
AI x Post	0.061	0.081^{*}	0.143	-1.254^{***}
	(0.049)	(0.046)	(0.917)	(0.381)
Frac. College		−0.036		-9.916^{***}
		(0.100)		(0.834)
Frac. White		0.645^{***}		−1.772
		(0.169)		(1.403)
Frac. Female		-0.259^{***}		-2.637^{***}
		(0.089)		(0.742)
Age		0.425^{*}		-3.987^{**}
		(0.225)		(1.875)
Constant	10.308^{***}	8.883^{***}	11.809^{***}	17.457^{**}
	(0.049)	(0.843)	(0.910)	(7.013)
Industry FE?	Yes	Yes	Yes	Yes
Year FE?	Yes	Yes	Yes	Yes
Observations	78	78	78	78
R^2	0.990	0.993	0.343	0.910
Adjusted R^2	0.987	0.991	0.170	0.878
Note:				*p<0.1; **p<0.05; ***p<0.01

Secretaries & Administrative Assistants DiD Regressions

Secretaries DiD Regressions:

```
# Wage DiD for Secretaries vs. Aides, reg1 = without controls and with FE, reg2 = with controls and wit
sa_wage_did_reg1 <- lm(LOGWAGE ~ AI + AIPost + factor(YEAR) + factor(IND1990), data = sa_wage)
sa_wage_did_reg2 <- lm(LOGWAGE ~ AI + AIPost + FRACCOLLEGE + FRACWHITE + FRACFEMALE + AGE + factor(YEAR

# Employment DiD for Secretaries vs. Aides
sa_emp_did_reg1 <- lm(LOGEMP ~ AI + AIPost + factor(YEAR) + factor(IND1990), data = sa_emp)
sa_emp_did_reg2 <- lm(LOGEMP ~ AI + AIPost + FRACCOLLEGE + FRACWHITE + FRACFEMALE + AGE + factor(YEAR)
```

```
stargazer(sa_wage_did_reg1,
          sa_wage_did_reg2,
          sa_emp_did_reg1,
          sa_emp_did_reg2,
          header = FALSE,
          type = "latex",
          title = "Difference-in-Differences: Secretaries and Administrative Assistants as Treatment Gr
          omit = c("IND1990", "YEAR"),
          column.labels =c("No Controls", "With Controls", "No Controls", "With Controls"),
          covariate.labels = c("AI", "AI x Post", "Frac. College", "Frac. White", "Frac. Female", "Age"
          dep.var.labels = c("log(wage)", "log(employment)"),
          omit.stat =c("LL", "ser", "f"),
          add.lines = list(c("Industry FE?", "Yes", "Yes", "Yes", "Yes"),
                           c("Year FE?", "Yes", "Yes", "Yes", "Yes"))
          )
```

OLS Regressions

Importing relevant OLS datasets:

```
p_wage_ols <-read.csv(file = "/Users/sidsatya/Desktop/berkeley/senior_year/thesis/aithesis/data/p_wage_
s_wage_ols <-read.csv(file = "/Users/sidsatya/Desktop/berkeley/senior_year/thesis/aithesis/data/s_wage_
a_wage_ols <-read.csv(file = "/Users/sidsatya/Desktop/berkeley/senior_year/thesis/aithesis/data/a_wage_

p_emp_ols <-read.csv(file = "/Users/sidsatya/Desktop/berkeley/senior_year/thesis/aithesis/data/p_emp_ol
s_emp_ols <-read.csv(file = "/Users/sidsatya/Desktop/berkeley/senior_year/thesis/aithesis/data/s_emp_ol
a_emp_ols <-read.csv(file = "/Users/sidsatya/Desktop/berkeley/senior_year/thesis/aithesis/data/a_emp_ol
```

Wage OLS Regressions:

```
# reg1: without controls, reg2: with controls
pwagereg1 <- lm(LOGWAGE ~ num_patents, data=p_wage_ols)
pwagereg2 <- lm(LOGWAGE ~ num_patents + FRACCOLLEGE + FRACWHITE + FRACFEMALE + AGE + factor(IND1990), da

swagereg1 <- lm(LOGWAGE ~ num_patents, data=s_wage_ols)
swagereg2 <- lm(LOGWAGE ~ num_patents + FRACCOLLEGE + FRACWHITE + FRACFEMALE + AGE + factor(IND1990), da

awagereg1 <- lm(LOGWAGE ~ num_patents, data=a_wage_ols)
awagereg2 <- lm(LOGWAGE ~ num_patents + FRACCOLLEGE + FRACWHITE + FRACFEMALE + AGE + factor(IND1990), da
```

Table 2: Difference-in-Differences: Secretaries and Administrative Assistants as Treatment Group, Nursing, Psychiatric, and Home Health Aides as Control Group

	Dependent variable:			
	log(wage)		log(employment)	
	No Controls	With Controls	No Controls	With Controls
	(1)	(2)	(3)	(4)
AI	0.076^{**}	-0.363^{***}	-0.820^{***}	2.674^{***}
	(0.035)	(0.080)	(0.303)	(0.452)
AI x Post	-0.018	-0.033	-0.063	0.029
	(0.056)	(0.037)	(0.489)	(0.207)
Frac. College		0.400^{***}		-3.868^{***}
		(0.107)		(0.603)
Frac. White		0.588^{***}		-5.914^{***}
		(0.195)		(1.095)
Frac. Female		0.255		0.288
		(0.234)		(1.313)
Age		1.452^{**}		-9.446^{***}
		(0.617)		(3.467)
Constant	10.160^{***}	5.422^{**}	11.098^{***}	40.268^{***}
	(0.056)	(2.277)	(0.485)	(12.800)
Industry FE?	Yes	Yes	Yes	Yes
Year FE?	Yes	Yes	Yes	Yes
Observations	78	78	78	78
R^2	0.511	0.850	0.434	0.927
Adjusted R^2	0.382	0.798	0.285	0.901
Note:				*p<0.1; **p<0.05; ***p<0.01

```
stargazer(pwagereg2,
          swagereg2,
          awagereg2,
          header = FALSE,
          type = "latex",
          title = "Wage OLS for Each Occupation Against Number of AI-Related Patents by Year",
          omit = c("IND1990"),
          dep.var.caption = "Dependent variable: log(wage)",
          dep.var.labels = "",
          column.labels = c('\\shortstack{Physicians and \\\\ Surgeons}',
                            '\\shortstack{Secretaries and \\\\ Administrative Assistants}',
                            '\\shortstack{Nursing, Psychiatric, and \\\\ Home Health Aides}'),
          covariate.labels = c("Num. Patents", "Frac. College", "Frac. White", "Frac. Female", "Age"),
          omit.stat = c("LL", "ser", "f"),
          add.lines = list(c("Industry Fixed Effects?", "  Yes", "  Yes", "  Yes"))
          )
```

Table 3: Wage OLS for Each Occupation Against Number of AI-Related Patents by Year

	Dependent variable: log(wage)		
	Physicians and Surgeons	Secretaries and Administrative Assistants	Nursing, Psychiatric, and Home Health Aides
	(1)	(2)	(3)
Num. Patents	0.077***	0.0003	0.015
	(0.026)	(0.012)	(0.016)
Frac. College	−0.073	0.184*	−0.030
	(0.956)	(0.094)	(0.110)
Frac. White	0.737***	−0.178	−0.385*
	(0.211)	(0.234)	(0.221)
Frac. Female	−0.142	−0.113	−0.565
	(0.188)	(0.390)	(0.462)
Age	−0.669	−0.006	2.639***
	(0.744)	(0.403)	(0.714)
Constant	14.503***	10.245***	0.348
	(3.064)	(1.538)	(2.683)
Industry Fixed Effects?	Yes	Yes	Yes
Observations	39	39	39
R^2	0.788	0.255	0.938
Adjusted R^2	0.741	0.087	0.924
Note:			*p<0.1; **p<0.05; ***p<0.01

Employment OLS Regressions:

```
# reg1: without controls, reg2: with controls
pempreg1 <- lm(LOGEMP ~ num_patents, data=p_emp_ols)
pempreg2 <- lm(LOGEMP ~ num_patents + FRACCOLLEGE + FRACWHITE + FRACFEMALE + AGE + factor(IND1990), data

sempreg1 <- lm(LOGEMP ~ num_patents, data=s_emp_ols)
sempreg2 <- lm(LOGEMP ~ num_patents + FRACCOLLEGE + FRACWHITE + FRACFEMALE + AGE + factor(IND1990), data

aempreg1 <- lm(LOGEMP ~ num_patents, data=a_emp_ols)
aempreg2 <- lm(LOGEMP ~ num_patents + FRACCOLLEGE + FRACWHITE + FRACFEMALE + AGE + factor(IND1990), data
```

```
stargazer(pempreg2,
          sempreg2,
          aempreg2,
          header = FALSE,
          type = "latex",
          title = "Employment OLS for Each Occupation Against Number of AI-Related Patents by Year",
          omit = c("IND1990"),
          dep.var.caption = "Dependent variable: log(employment)",
          dep.var.labels = "",
          column.labels = c('\\shortstack{Physicians and \\\\ Surgeons}',
                            '\\shortstack{Secretaries and \\\\ Administrative Assistants}',
                            '\\shortstack{Nursing, Psychiatric, and \\\\ Home Health Aides}'),
          covariate.labels = c("Num. Patents", "Frac. College", "Frac. White", "Frac. Female", "Age"),
          omit.stat = c("LL", "ser", "f"),
          add.lines = list(c("Industry Fixed Effects?", "  Yes", "  Yes", "  Yes"))
          )
```

Table 4: Employment OLS for Each Occupation Against Number of AI-Related Patents by Year

	Dependent variable: log(employment)		
	Physicians and Surgeons	Secretaries and Administrative Assistants	Nursing, Psychiatric, and Home Health Aides
	(1)	(2)	(3)
Num. Patents	0.084	-0.065^{***}	-0.022
	(0.060)	(0.023)	(0.024)
Frac. College	3.256	0.166	0.119
	(2.194)	(0.183)	(0.160)
Frac. White	0.088	0.030	0.531
	(0.485)	(0.455)	(0.322)
Frac. Female	0.181	-0.890	-1.576^{**}
	(0.432)	(0.757)	(0.674)
Age	1.563	-0.954	0.618
	(1.708)	(0.783)	(1.041)
Constant	6.340	15.245^{***}	7.932^{*}
	(7.031)	(2.984)	(3.909)
Industry Fixed Effects?	Yes	Yes	Yes
Observations	39	39	39
R^2	0.992	0.995	0.997
Adjusted R^2	0.991	0.994	0.996
Note:			*p<0.1; **p<0.05; ***p<0.01