



```

1 .
2 . *****
3 . * "Do Shared Auditors Facilitate Follow-on Innovation?"
4 . * Tian, Yan, and Zuo (2025)
5 . * Journal of Accounting Research
6 . * 2025-07
7 . *****
8 .
9 . /* Start of code */
10. * step 1: download and clean data from various sources
11. do ./1_clean_various_datasets.do

12. *****
13. * PART 1. clean datasets
14. *****
15.
16. * code to generate datasets to merge with later
17.
18. * firm-pair controls (Martens and Sextroh, 2021)
19.
20. * Common analysts
21. * use ibes-crsp link from wrds
22. import sas using "ibcrsphist.sas7bdat", case(upper) clear
    (6 vars, 35,308 obs)

23. rename NCUSIP CUSIP

24. drop if PERMNO == . | CUSIP == ""
    (7,061 observations deleted)

25. drop SCORE

26. drop TICKER

27. /*
    > import sas using "det_epsus.sas7bdat", case(upper) clear
    > keep CUSIP ANALYS ACTDATS
    > save "det_epsus.dta", replace
    > */
28. rangejoin ACTDATS SDATE EDATE using "det_epsus.dta", by(CUSIP)
    (using rangestat version 1.1.1)

29. g YEAR = year(ACTDATS)
    (4,196 missing values generated)

```

```
30. drop ACTDATS
31. gdrop duplicates drop PERMNO ANALYS YEAR , force
    Duplicates in terms of PERMNO ANALYS YEAR
    (26,937,697 observations deleted)
32. rename PERMNO PERMNO1
33. preserve
34. rename PERMNO1 PERMNO2
35. tempfile _tmpfile
36. save `'_tmpfile'
    file F:\STATAMP\ST_6974_000002.tmp saved as .dta format
37. restore
38. joinby YEAR ANALYS using `'_tmpfile'
39. quietly unique ANALYS, by(PERMNO1 PERMNO2 YEAR) generate(CommonAnalyst
    > st)
40. drop SDATE EDATE
41. sort PERMNO1 PERMNO2 YEAR
42. drop if CommonAnalyst == .
    (14,399,529 observations deleted)
43. drop if YEAR == .
    (12,138,256 observations deleted)
44. drop ANALYS
45. drop if PERMNO1 == PERMNO2
    (176,576 observations deleted)
46. save "CommonAnalyst.dta", replace
    file CommonAnalyst.dta saved
47.
48. * Common Alliance (Gomes-Casseres et al. 2016)
```

```
49. clear all

50. import sas using "ibcrsphist.sas7bdat", case(upper) clear
    (6 vars, 35,308 obs)

51. rename NCUSIP CUSIP

52. drop if PERMNO == . | CUSIP == ""
    (7,061 observations deleted)

53. drop SCORE

54. drop TICKER

55. rename CUSIP CUSIP1

56. replace CUSIP1= substr(CUSIP1, 1 , 6)
    (28,247 real changes made)

57. rangejoin date SDATE EDATE using "CommonAlliance_CUSIP.dta", by(CUSI
    > P1)
    (using rangestat version 1.1.1)

58. drop if date == .
    (20,684 observations deleted)

59. rename PERMNO PERMNO1

60. drop CUSIP2

61. drop SDATE EDATE

62. gduplicates drop CUSIP1 date, force
    Duplicates in terms of CUSIP1 date
    (7,101 observations deleted)

63. tempfile _cusip1

64. save `_cusip1'
    file F:\STATAMP\ST_6974_000003.tmp saved as .dta format

65.

66. import sas using "ibcrsphist.sas7bdat", case(upper) clear
    (6 vars, 35,308 obs)
```

```

67. rename NCUSIP CUSIP
68. drop if PERMNO == . | CUSIP == ""
    (7,061 observations deleted)
69. drop SCORE
70. drop TICKER
71. rename CUSIP CUSIP2
72. replace CUSIP2= substr(CUSIP2, 1 , 6)
    (28,247 real changes made)
73. rangejoin date SDATE EDATE using "CommonAlliance_CUSIP.dta", by(CUSI
    > P2)
    (using rangestat version 1.1.1)
74. drop if date == .
    (22,951 observations deleted)
75. rename PERMNO PERMNO2
76. drop CUSIP1
77. drop SDATE EDATE
78. gdropuplicates drop CUSIP2 date, force
    Duplicates in terms of CUSIP2 date
    (4,699 observations deleted)
79. tempfile _cusip2
80. save `cusip2'
    file F:\STATAMP\ST_6974_000004.tmp saved as .dta format
81.
82. use "CommonAlliance_CUSIP.dta", clear
83. merge m:1 CUSIP1 date using "`cusip1'"
    (variable CUSIP1 was str6, now str8 to accommodate using data's
    values)

```

| Result | Number of obs | |
|-------------|----------------|--------------|
| Not matched | 137,588 | |
| from master | 137,588 | (_merge==1) |
| from using | 0 | (_merge==2) |
| Matched | 36,835 | (_merge==3) |


```
84. keep if _merge == 3
    (137,588 observations deleted)
```

```
85. drop _merge
```

```
86. merge m:1 CUSIP2 date using "`_cusip2'"
    (variable CUSIP2 was str6, now str8 to accommodate using data's
     values)
```

| Result | Number of obs | |
|-------------|---------------|--------------|
| Not matched | 39,261 | |
| from master | 26,668 | (_merge==1) |
| from using | 12,593 | (_merge==2) |
| Matched | 10,167 | (_merge==3) |

```
87. keep if _merge == 3
    (39,261 observations deleted)
```

```
88. drop _merge
```

```
89. drop CUSIP1 CUSIP2
```

```
90. g year = year(date)
```

```
91. drop date
```

```
92. gdropuplicates drop PERMNO1 PERMNO2 year, force
```

Duplicates in terms of **PERMNO1 PERMNO2 year**

(314 observations deleted)

```
93. save "CommonAlliance PERMNO.dta", replace
    file CommonAlliance_PERMNO.dta saved
```

```
94.
```

```
95. * Common Inventor
```

```
96. clear all
```

```
97. import delimited "KPSS_public.csv", clear delimiter(",") encoding("u
    > tf-8") varnames(1)
    (7 vars, 3,053,011 obs)
```

```

98. g year = substr(filing_date, -4, 4)
    (2 missing values generated)

99. destring year, replace force
    year: all characters numeric; replaced as int
    (2 missing values generated)

100 keep patent_num permno year

101 save "KPSS_public_to_merge.dta" , replace
    file KPSS_public_to_merge.dta saved

102
103 import delimited "patent_inventor.tsv", clear delimiter("\t") encodi
    > ng("utf-8") varnames(1)
    (3 vars, 19,111,181 obs)

104 drop location_id

105 rename inventor_id inventor

106 gegen inventor_id = group(inventor)

107 drop inventor

108 rename patent_id patent_num

109 destring patent_num , replace force
    patent_num: contains nonnumeric characters; replaced as long
    (1335888 missing values generated)

110 drop if patent_num == .
    (1,335,888 observations deleted)

111 merge m:1 patent_num using "KPSS_public_to_merge.dta"

```

| Result | Number of obs | |
|-------------|-------------------|--------------|
| Not matched | 11,747,885 | |
| from master | 11,120,821 | (_merge==1) |
| from using | 627,064 | (_merge==2) |
| Matched | 6,654,472 | (_merge==3) |

```

112 keep if _merge == 3
    (11,747,885 observations deleted)

```

```

113 drop _merge
114 drop patent_num
115 gdeduplicates drop year inventor_id permno, force
    Duplicates in terms of year inventor_id permno
    (2,849,571 observations deleted)
116 rename permno PERMNO1
117 rename year YEAR
118 preserve
119 rename PERMNO1 PERMNO2
120 tempfile _tmpfile
121 save `_tmpfile'
    file F:\STATAMP\ST_6974_000006.tmp saved as .dta format
122 restore
123 rangejoin YEAR -3 0 using `_tmpfile', by(inventor_id)
    (using rangestat version 1.1.1)
124 drop if PERMNO1 == PERMNO2
    (7,203,575 observations deleted)
125 drop YEAR_U
126 quietly unique inventor_id, by(PERMNO1 PERMNO2 YEAR) generate(Common
    > Inventor)
127 drop if CommonInventor == .
    (744,773 observations deleted)
128 gdeduplicates report PERMNO1 PERMNO2 YEAR
    Duplicates in terms of PERMNO1 PERMNO2 YEAR

```

| copies | observations | surplus |
|--------|---------------|----------|
| 1 | 226702 | 0 |

```
129 drop inventor_id

130 save "CommonInventor.dta", replace
    file CommonInventor.dta saved

131
132 * Common ownership
133 * thank Backus, Conlon, and Sinkinson, 2021 for sharing this data *
134
135 * Relative Size
136 * directly use AT from funda
137
138 * Relative Analysts
139 import sas using "ibcrsphist.sas7bdat", case(upper) clear
    (6 vars, 35,308 obs)

140 rename NCUSIP CUSIP

141 drop if PERMNO == . | CUSIP == ""
    (7,061 observations deleted)

142 drop SCORE

143 drop TICKER

144 rangejoin ACTDATS SDATE EDATE using "det_epsus.dta", by(CUSIP)
    (using rangestat version 1.1.1)

145 g YEAR = year(ACTDATS)
    (4,196 missing values generated)

146 drop ACTDATS

147 drop if YEAR == .
    (4,196 observations deleted)

148 gduplicates drop PERMNO ANALYS YEAR , force
    Duplicates in terms of PERMNO ANALYS YEAR
    (26,936,985 observations deleted)

149 keep PERMNO ANALYS YEAR

150 bys YEAR PERMNO: gegen AnalystNo = count(ANALYS)
    performance warning: -by- prefix may be slower than -by()-
```

```

151 drop ANALYS
152 gdropuplicates drop PERMNO YEAR, force
    Duplicates in terms of PERMNO YEAR
    (1,291,347 observations deleted)
153 save "AnalystNo.dta", replace
    file AnalystNo.dta saved
154
155 * Relative Patent Stock
156 clear all
157 import delimited "KPSS_public.csv", clear delimiter(",") encoding("u
    > tf-8") varnames(1)
    (7 vars, 3,053,011 obs)
158 g year = substr(issue_date, -4, 4)
159 destring year, replace force
    year: all characters numeric; replaced as int
160 keep permno year patent_num
161 bys year permno: gegen PatentPerYear = count(patent_num)
    performance warning: -by- prefix may be slower than -bys()-
162 gdropuplicates drop year permno, force
    Duplicates in terms of year permno
    (2,967,295 observations deleted)
163 sort permno year
164 drop patent_num
165 xtset permno year
    Panel variable: permno (unbalanced)
    Time variable: year, 1926 to 2020, but with gaps
    Delta: 1 unit
166 tsfill, full

```



```
167 replace PatentPerYear = 0 if PatentPerYear == .
    (727,579 real changes made)
```

```
168 rangestat (sum) PatentStock = PatentPerYear, by(permno) interval(yea
    > r, ., 0)
```

```
169 drop PatentPerYear
```

```
170 gduplicates report permno year
```

Duplicates in terms of **permno year**

| copies | observations | surplus |
|--------|---------------|----------|
| 1 | 813295 | 0 |

```
171 rename permno PERMNO
```

```
172 rename year YEAR
```

```
173 save "PatentStock.dta", replace
    file PatentStock.dta saved
```

```
174
```

```
175 * Relative Citation
```

```
176 clear all
```

```
177 frame create fr1
```

```
178 frame change fr1
```

```
179 import delimited "KPSS_public.csv", delimiter(",") clear
    (encoding automatically selected: ISO-8859-1)
    (7 vars, 3,053,011 obs)
```

```
180 rename patent_num patent_id
```

```
181 keep patent_id permno filing_date
```

```
182 frame create fr2
```

```
183 frame change fr2
```

```
184 use "uspatentcitation.dta", clear
```

```

185 destring patent_id, replace force
    patent_id already numeric; no replace

186 drop if patent_id == .
    (0 observations deleted)

187 frlink m:1 patent_id, frame(fr1)
    (63,777,639 observations in frame fr2 unmatched)

188 frget permno filing_date, from(fr1)
    (63,777,639 missing values generated)
    (63,777,639 missing values generated)
    (2 variables copied from linked frame)

189 drop if permno == .
    (63,777,639 observations deleted)

190 drop date

191 g date = date(filing_date, "MDY")

192 g year = year(date)

193 bys year permno: gegen num_citation sum = count(year)
    performance warning: -by- prefix may be slower than -by()-

194 rename num_citation_sum CitationNum

195 keep year permno CitationNum

196 gdeduplicates drop year permno , force

    Duplicates in terms of year permno

    (39,878,305 observations deleted)

197 rename year YEAR

198 rename permno PERMNO

199 save "CitationNum.dta", replace
    file CitationNum.dta saved

200
201
202 * Technology Similarity

```

```

203 clear all

204 import delimited "patent_number_cpc.csv", delimiter(",") clear
    (encoding automatically selected: ISO-8859-1)
    (2 vars, 3,053,011 obs)

205 drop if cpc == ""
    (3,319 observations deleted)

206 split cpc , p(";") limit(10)
    variables created as string:
    cpc1    cpc2    cpc3    cpc4    cpc5    cpc6    cpc7    cpc8    cpc9    cpc10

207 g i = _n

208 drop cpc

209 reshape long cpc , i(i) j(j)
    (j = 1 2 3 4 5 6 7 8 9 10)

```

```

214 drop if _merge != 3
    (3,319 observations deleted)

215 drop _merge

216 bys cpc permno year: gegen patent_num_c = count(patent_num)
    performance warning: -by- prefix may be slower than -by()-

217 gduplicates drop cpc year permno, force

    Duplicates in terms of cpc year permno

    (4,093,686 observations deleted)

218 rangestat (sum) patent_num_c, interval(year . 0) by(cpc permno)

219 replace patent_num_c = patent_num_c_sum
    (871,676 real changes made, 2 to missing)

220 drop patent_num_c_sum

221 drop i j patent_num

222 rename permno PERMNO1

223 gegen cpc_id = group(cpc)

224 drop cpc

225 save "tmp2.dta", replace
    file tmp2.dta saved

226 use "tmp2.dta", clear

227 g Pict2 = patent_num_c * patent_num_c
    (2 missing values generated)

228 bys PERMNO1 year: gegen Pict2_sum = sum(Pict2)
    performance warning: -by- prefix may be slower than -by()-

229 preserve

230 rename PERMNO1 PERMNO2

231 rename Pict2 Pjct2

```

```
232 rename Pict2_sum Pjct2_sum
233 tempfile _tmpfile
234 save `_tmpfile'
    file F:\STATAMP\ST_6974_000008.tmp saved as .dta format
235 restore
236 joinby year cpc_id using `_tmpfile'
237 drop if PERMNO1 == PERMNO2
    (1,067,826 observations deleted)
238 g PictPjct = sqrt(Pict2)*sqrt(Pjct2)
239 bys year PERMNO1 PERMNO2: gegen nom = sum(PictPjct)
    performance warning: -by- prefix may be slower than -by()-
240 gduplicates drop year PERMNO1 PERMNO2 ,force
    Duplicates in terms of year PERMNO1 PERMNO2
    (54,525,052 observations deleted)
241 g dom = sqrt(Pict2_sum*Pjct2_sum)
242 g TechSimilarity = nom/dom
243 keep PERMNO1 PERMNO2 year TechSimilarity
244 rename year YEAR
245 keep if YEAR >= 1980
    (10,295,552 observations deleted)
246 save "TechSimilarity.dta", replace
    file TechSimilarity.dta saved
247
248 * to generate shared_auditor_to_merge.dta
249 clear all
250 * read Linking file
251 frame create linkingtable
```



```

252 frame change linkingtable

253 import sas using "ccmxpf_linktable.sas7bdat", case(upper) clear
    (9 vars, 79,762 obs)

254 replace LINKENDDT = 99999 if LINKENDDT == . /* It uses the SAS missi
    > ng code ".E" if a link is still valid. */
    variable LINKENDDT was int now long
    (11,334 real changes made)

255 replace LINKDT = -99999 if LINKDT == 0 /* If the link was valid befo
    > re CRSP's earliest record, LINKDT is set to be SAS missing code ".B"
    > .*/
    variable LINKDT was int now long
    (404 real changes made)

256 keep if LINKTYPE == "LC" | LINKTYPE== "LU" /* 1) Confirmed Linkage O
    > nly */
    (49,148 observations deleted)

257 keep if LINKPRIM == "P" | LINKPRIM == "C" /* 2) Take Primary Shares
    > Only */
    (455 observations deleted)

258
259 * link with compustat Data
260 rangejoin DATADATE LINKDT LINKENDDT using "funda_allvar.dta", by(GVK
    > EY) /*the raw funda file with all variables from compustat*/
    (using rangestat version 1.1.1)

261 gduplicates drop LPERMNO FYEAR, force

    Duplicates in terms of LPERMNO FYEAR

    (49 observations deleted)

262 rename LPERMNO PERMNO

263 rename FYEAR YEARC

264
265 * link with msenames data
266 save tmp.dta, replace
    (file tmp.dta not found)
    file tmp.dta saved

```

```
267 import sas using "msenames.sas7bdat", clear /*msenames at compustat/
> crsp */
(21 vars, 117,830 obs)

268 rangejoin DATADATE NAMEDT NAMEENDT using tmp.dta, by(PERMNO)
(using rangestat version 1.1.1)

269 rm tmp.dta

270 keep if SHRCDD == 11 | SHRCDD == 10
(75,036 observations deleted)

271 keep if EXCHCD == 1 | EXCHCD == 2 | EXCHCD == 3
(3,668 observations deleted)

272 gduplicates drop PERMNO YEAR, force

Duplicates in terms of PERMNO YEARC

(11,093 observations deleted)

273
274 * get SIC info
275 frame create company
276 frame change company

277 import sas using "company.sas7bdat", case(upper) clear /*company inf
> o at compustat/crsp */
(39 vars, 47,261 obs)

278 deststring GVKEY, replace
GVKEY: all characters numeric; replaced as long

279
280 * link with SIC dataset
281 frame change linkingtable

282 deststring GVKEY, replace
GVKEY: all characters numeric; replaced as long
(14569 missing values generated)

283 frlink m:1 GVKEY, frame(company)
(14,947 observations in frame linkingtable unmatched)

284 frget SIC, from(company)
(14,947 missing values generated)
(1 variable copied from linked frame)
```

```
285 destring SIC, replace
    SIC: all characters numeric; replaced as int
    (14947 missing values generated)

286 capt g SICCD = SIC

287 replace SICCD = SIC if SICCD == .
    (0 real changes made)

288 keep PERMNO SICCD YEAR AU

289 sort PERMNO YEAR

290 rename YEARC year

291 rename PERMNO permno

292 rename SICCD siccd

293 rename AU au

294 drop if permno == . | au == ""
    (62,143 observations deleted)

295 save shared_auditor_to_merge.dta, replace
    file shared_auditor_to_merge.dta saved

296
297 * update: use Audit Analytics data for variable AU
298 clear all

299 * use audit analytics file
300 import sas using "auditopin.sas7bdat", clear
    (225 vars, 336,443 obs)

301 keep if FISCAL_YEAR_OF_OP >= 2000
    (10,373 observations deleted)

302 keep AUDITOR_FKEY FISCAL_YEAR_OF_OP COMPANY_FKEY

303 gduplicates drop FISCAL_YEAR_OF_OP COMPANY_FKEY, force
    Duplicates in terms of FISCAL_YEAR_OF_OP COMPANY_FKEY
    (2,336 observations deleted)
```

```

304 rename COMPANY_FKEY CIK
305 rename AUDITOR_FKEY AUAA
306 rename FISCAL_YEAR_OF_OP YEAR
307 *download cik-gvkey link from wrds
308 merge m:1 CIK using "CIK_GVKEY2.dta"

```

| Result | Number of obs | |
|-------------|----------------|--------------|
| Not matched | 146,895 | |
| from master | 130,965 | (_merge==1) |
| from using | 15,930 | (_merge==2) |
| Matched | 192,769 | (_merge==3) |

```

309 drop if _merge != 3
    (146,895 observations deleted)

```

```
310 drop _merge
```

```
311 drop CIK
```

```

312 destring GVKEY, replace
    GVKEY: all characters numeric; replaced as long

```

```

313 *download gvkey-permno link from wrds
314 merge m:1 GVKEY YEAR using "GVKEY2PERMNO.dta"
    (variable YEAR was int, now float to accommodate using data's
      values)

```

| Result | Number of obs | |
|-------------|----------------|--------------|
| Not matched | 293,623 | |
| from master | 67,490 | (_merge==1) |
| from using | 226,133 | (_merge==2) |
| Matched | 125,279 | (_merge==3) |

```

315 keep if _merge == 3
    (293,623 observations deleted)

```

```
316 drop _merge
```



```

317 drop GVKEY
318 gdropuplicates drop PERMNO YEAR , force
    Duplicates in terms of PERMNO YEAR
    (2,203 observations deleted)
319 save "AUAA.dta", replace
    file AUAA.dta saved
320
321 * create auditor office level data to merge
322 /*link Audit Analytics to permno, see http://kaichen.work/?p=358 for
    > codes
    > the output file is AA.dta*/
323 use "AA.dta", clear
324 keep if FORM_FKEY == "10-K"
    (187,715 observations deleted)
325 drop if permno == .
    (60,505 observations deleted)
326 rename FISCAL_YEAR_OF_OP year
327 gdropuplicates drop permno year, force
    Duplicates in terms of permno year
    (19 observations deleted)
328 keep AUDITOR_NAME AUDITOR_CITY AUDITOR_STATE year permno
329 save "auditor_office_to_merge.dta" , replace
    file auditor_office_to_merge.dta saved
330
331 * to generate "INNOVATION_AM50.dta"
332 import sas using "company.sas7bdat", case(upper) clear
    (39 vars, 47,261 obs)
333 keep SIC GVKEY
334 merge 1:m GVKEY using "FUNDA.dta"

```

| Result | Number of obs | |
|-------------|----------------|--------------|
| Not matched | 22,114 | |
| from master | 21,376 | (_merge==1) |
| from using | 738 | (_merge==2) |
| Matched | 311,484 | (_merge==3) |


```

335 keep if _merge == 3
    (22,114 observations deleted)

336 drop _merge

337 destring SIC, replace
    SIC: all characters numeric; replaced as int

338 g SIC2 = int(SIC/100)

339 * AM - amortization
340 foreach i of varlist AM {
    2. g `i'2AT = `i'/AT
    3. * cut at year-ind median
341 foreach v of numlist 50 {
    4. bys YEAR SIC2: gegen p`v' = pctlile(`i'2AT), p(`v')
    5. g I_`i'2AT_`v'_1 = (`i'2AT > p`v') if `i'2AT != .
    6. drop p`v'
    7. }
    8. bys YEAR SIC2: gegen `i'2AT_m = mean(`i'2AT)
    9. g `i'2AT2 = `i'2AT
    10. replace `i'2AT2 = `i'2AT_m if `i'2AT2 == .
    11. foreach v of numlist 50 {
    12. bys YEAR SIC2: gegen p`v' = pctlile(`i'2AT2), p(`v')
    13. g I_`i'2AT2_`v'_2 = (`i'2AT2 > p`v') if `i'2AT2 != .
    14. drop p`v'
    15. }
    16. drop `i'2AT_m
    17. }
    (126,915 missing values generated)
    performance warning: -by- prefix may be slower than -by()-
    (126,915 missing values generated)
    performance warning: -by- prefix may be slower than -by()-
    (126,915 missing values generated)
    (110,212 real changes made)
    performance warning: -by- prefix may be slower than -by()-
    (16,703 missing values generated)

342 gdropuplicates drop PERMNO YEAR , force

    Duplicates in terms of PERMNO YEAR

    (0 observations are duplicates)

343 keep PERMNO YEAR I_*

```

```

344 capt save "INNOVATION_AM50.dta", replace

345
346 import delimited "KPSS_public.csv", delimiter(",") clear
    (encoding automatically selected: ISO-8859-1)
    (7 vars, 3,053,011 obs)

347 g date = date(filing_date, "MDY")
    (2 missing values generated)

348 drop if date == .
    (2 observations deleted)

349 g year = year(date)

350 rename permno PERMNO

351 rename year YEAR

352 bys PERMNO YEAR: gegen PATENTS = count(patent)
    performance warning: -by- prefix may be slower than -by()-

353 duplicates drop PERMNO YEAR, force

    Duplicates in terms of PERMNO YEAR

    (2,960,790 observations deleted)

354 keep PERMNO YEAR PATENTS

355 save "num_patents.dta", replace
    file num_patents.dta saved

356
357
    end of do-file

358 * step 2: construct the main sample
359 do ./2_construct_main.do

360 *****
361 * PART 2. construct main sample
362 *****
363
364 * download patent data files from patentsview.org
365 * https://patentsview.org/download/data-download-tables

```

```
366 * first convert raw tsv data to Stata dta format

367 * import delimited "uspatentcitation.tsv", delimiter("\t") clear
368 * drop uuid
369 * destring patent_id citation_id, replace force
370 * drop if patent_id == .
371 * drop if citation_id == .
372 * save "uspatentcitation.dta", replace
373 * import delimited "g_application.tsv", delimiter("\t") clear
374 * save "g_application.dta", replace .

375 * start from patentsview citation data
376 clear all

377 use "uspatentcitation.dta", clear

378 drop name kind country category sequence

379 destring patent_id citation_id , replace force
    patent_id already numeric; no replace
    citation_id already numeric; no replace

380 drop if patent_id == . | citation_id == .
    (0 observations deleted)

381
382 * load patent info from patentsview
383 frame create patentview

384 frame change patentview

385 use "g_application.dta", clear

386 duplicates drop patent_id, force

    Duplicates in terms of patent_id

    (0 observations are duplicates)

387 frame change default

388
389 * load KPSS data
390 capt frame drop KPSS

391 frame create KPSS
```

```
392 frame change KPSS

393 import delimited "KPSS_public.csv"
    (encoding automatically selected: ISO-8859-1)
    (7 vars, 3,053,011 obs)

394 keep patent_num permno

395 frame change default

396
397 * match permno for citing
398 g patent_num = patent_id

399 frlink m:1 patent_num, frame(KPSS)
    (63,777,639 observations in frame default unmatched)

400 frget permno , from(KPSS)
    (63,777,639 missing values generated)
    (1 variable copied from linked frame)

401 rename permno permno1

402
403 * match permno for cited
404 replace patent_num = citation_id
    (103,721,205 real changes made)

405 drop KPSS

406 frlink m:1 patent_num, frame(KPSS)
    (62,492,908 observations in frame default unmatched)

407 frget permno , from(KPSS)
    (62,492,908 missing values generated)
    (1 variable copied from linked frame)

408 rename permno permno2

409 drop patent_num KPSS

410 drop if permno1 == . | permno2 == .
    (82,655,077 observations deleted)

411
412 * get filing year of each patent
```



```

413 frlink m:1 patent_id, frame(patentview)
    (all observations in frame default matched)

414 frget filing_date, from(patentview)
    (1 variable copied from linked frame)

415 drop date

416 g year = substr(filing_date, 1, 4)

417 destring year, replace force
    year: all characters numeric; replaced as int

418 drop if year == .
    (0 observations deleted)

419 drop filing_date

420 save "citation_pair_public_firms_20231210.dta", replace
    file citation_pair_public_firms_20231210.dta saved

421
422 * aggregate into company-pair-year
423 bys year permno1 permno2: gegen citations = count(patent_id)
    performance warning: -by- prefix may be slower than -by()-

424 gdropuplicates drop year permno1 permno2 , force

    Duplicates in terms of year permno1 permno2

    (19,126,458 observations deleted)

425 keep year permno1 permno2 citations

426
427 * validity check
428 drop if year >= 2030 /* drop date format errors*/
    (24 observations deleted)

429 save "PERMNO2PERMNO_PatentsView20231210.dta", replace
    file PERMNO2PERMNO_PatentsView20231210.dta saved

430 egen permno12 = group(permno1 permno2)

431 drop if year <= 1925
    (51 observations deleted)

```



```
432 tsset permno12 year
```

```
Panel variable: permno12 (unbalanced)
Time variable: year, 1931 to 2020, but with gaps
Delta: 1 unit
```

```
433 tsfill, f
```

```
434 * restore permno after tsfill
```

```
435 bys permno12: egen permno1_m = mean(permno1)
```

```
436 replace permno1 = permno1_m if permno1 == .
(41,676,565 real changes made)
```

```
437 drop permno1_m
```

```
438 bys permno12: egen permno2_m = mean(permno2)
```

```
439 replace permno2 = permno2_m if permno2 == .
(41,676,565 real changes made)
```

```
440 drop permno2_m
```

```
441 replace citations = 0 if citations == .
(41,676,565 real changes made)
```

```
442 g Icitation = (citations>0)
```

```
443
```

```
444 * cleaning sample
```

```
445 * at least one cross-firm citation
```

```
446 * (This is already satisfied, if no citation, firm pairs would not e
> xist)
```

```
447 capt rm "output\t1.rtf"
```

```
448 unique permno12 year if year>=2000 & year <=2018
Number of unique values of permno12 year is 9207856
Number of records is 9207856
```

```
449 /* asdoc stop Stata log, commented out */
```

```
450 //asdoc unique permno12 year if year>=2000 & year <=2018, save(outpu
> t\t1.rtf) append
```

```
451
```

```
452 * drop if cited firm has zero patent until and including year t
```

```
453 frame create fr1
```

```

454 frame change fr1

455 import delimited "KPSS_public.csv", delimiter(",") clear
    (encoding automatically selected: ISO-8859-1)
    (7 vars, 3,053,011 obs)

456 g date = date(filing_date, "MDY")
    (2 missing values generated)

457 drop if date == .
    (2 observations deleted)

458 g year = year(date)

459 bys year permno: egen num_patent = count(year)

460 duplicates drop year permno , force

    Duplicates in terms of year permno

    (2,960,790 observations deleted)

461 xtset permno year

    Panel variable: permno (unbalanced)
    Time variable: year, 1834 to 2020, but with gaps
    Delta: 1 unit

462 tsfill, f

463 keep permno year num_patent

464 replace num_patent = 0 if num_patent == .
    (1,508,688 real changes made)

465 rangestat (sum) num_patent, by(permno) interval(year . 0)

466 rename permno permno2

467 frame change default

468 frlink m:1 permno2 year, frame(fr1)
    (all observations in frame default matched)

469 frget num_patent_sum , from(fr1)
    (1 variable copied from linked frame)

```

```

470 drop if num_patent_sum == 0
    (16,747,358 observations deleted)

471 unique permno12 year if year>=2000 & year <=2018
    Number of unique values of permno12 year is  9172345
    Number of records is  9172345

472 //asdoc unique permno12 year if year>=2000 & year <=2018, save(outpu
    > t\t1.rtf) append
473
474 * drop if citing firm's total number of citations in year t is 0
475 bys permno1 year: egen num_citation_sum1 = sum(citations)

476 drop if num_citation_sum1 == 0
    (15,566,912 observations deleted)

477 unique permno12 year if year>=2000 & year <=2018
    Number of unique values of permno12 year is  5465202
    Number of records is  5465202

478 //asdoc unique permno12 year if year>=2000 & year <=2018, save(outpu
    > t\t1.rtf) append
479
480 * drop self-citations
481 drop if permno1 == permno2
    (55,266 observations deleted)

482 unique permno12 year if year>=2000 & year <=2018
    Number of unique values of permno12 year is  5441330
    Number of records is  5441330

483 //asdoc unique permno12 year if year>=2000 & year <=2018, save(outpu
    > t\t1.rtf) append
484
485 * now have firm-pair-year observations
486 * dep. var
487 tsset permno12 year

    Panel variable: permno12 (unbalanced)
    Time variable: year, 1931 to 2020, but with gaps
    Delta: 1 unit

488 g citations_raw = citations

489 winsor2 citations, cuts(1 99) by(year) replace

```

```
490 g lcitation=log(1+citations)
```

```
491
```

```
492 * indep. var
```

```
493 * datasets used to merge are detailed below
```

```
494 * shared_auditor_to_merge.dta records audit firm from compustat
```

```
495 g permno = permno1
```

```
496 merge m:1 permno year using "shared_auditor_to_merge.dta"
```

```
(variable permno was float, now double to accommodate using data's  
values)
```

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 3,134,174 | |
| from master | 2,976,179 | (_merge==1) |
| from using | 157,995 | (_merge==2) |
| Matched | 8,270,445 | (_merge==3) |

```
497 drop if _merge != 3
```

```
(3,134,174 observations deleted)
```

```
498 drop _merge
```

```
499 rename siccd siccd1
```

```
500 rename au aul
```

```
501
```

```
502 replace permno = permno2
```

```
(8,270,445 real changes made)
```

```
503 merge m:1 permno year using "shared_auditor_to_merge.dta"
```

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 4,208,403 | |
| from master | 4,091,506 | (_merge==1) |
| from using | 116,897 | (_merge==2) |
| Matched | 4,178,939 | (_merge==3) |

```
504 drop if _merge != 3
```

```
(4,208,403 observations deleted)
```



```

505 drop _merge
506 rename siccd siccd2
507 rename au au2
508 g I_shared_auditor = (au1 == au2)
509
510 * drop if both au = `others', which is coded as 9
511 g YEAR = year
512 drop if au1 == "9" & au2 == "9" & YEAR<2000
    (4,079 observations deleted)
513
514 * update: after 2000, use audit analytics data to identify audit fir
    > m
515 * AUAA.dta records audit firm from audit analytics
516 g PERMNO1 = permno1
517 g PERMNO2 = permno2
518 capt drop PERMNO
519 g PERMNO = PERMNO1
520 merge m:1 PERMNO YEAR using "AUAA.dta"
    (variable PERMNO was float, now double to accommodate using data's
    values)

```

| Result | Number of obs | |
|-------------|------------------|----------------------|
| Not matched | 2,708,841 | |
| from master | 2,604,807 | (_merge==1) |
| from using | 104,034 | (_merge==2) |
| Matched | 1,570,053 | (_merge==3) |

```

521 drop if _merge == 2
    (104,034 observations deleted)
522 drop _merge
523 rename AUAA AUAA1

```

```

524
525 capt drop PERMNO

526 g PERMNO = PERMNO2

527 merge m:1 PERMNO YEAR using "AUAA.dta"
    (variable PERMNO was float, now double to accommodate using data's
      values)

```

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 2,685,861 | |
| from master | 2,597,822 | (_merge==1) |
| from using | 88,039 | (_merge==2) |
| Matched | 1,577,038 | (_merge==3) |

```

528 drop if _merge == 2
    (88,039 observations deleted)

529 drop _merge

530 rename AUAA AUAA2

531
532 capt drop I_shared_auditor

533 destring au1 au2, replace
    au1: all characters numeric; replaced as byte
    au2: all characters numeric; replaced as byte

534 replace au1 = AUAA1 if AUAA1!= . & AUAA2 != . & YEAR>=2000
    variable au1 was byte now int
    (1,522,982 real changes made)

535 replace au2 = AUAA2 if AUAA1!= . & AUAA2 != . & YEAR>=2000
    variable au2 was byte now int
    (1,522,409 real changes made)

536 g I_shared_auditor = (au1 == au2)

537
538 * shared audit office
539 * auditor_office_to_merge.dta records audit office from audit analyt.
    > ics

```

540 capt drop permno

541 g permno = permno1

542 merge m:1 permno year using "auditor_office_to_merge.dta"
 (variable **year** was **int**, now **double** to accommodate using data's
 values)
 (variable **permno** was **float**, now **double** to accommodate using data's
 values)

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 2,777,857 | |
| from master | 2,707,618 | (_merge==1) |
| from using | 70,239 | (_merge==2) |
| Matched | 1,467,242 | (_merge==3) |

543 drop if _merge == 2
 (70,239 observations deleted)

544 drop _merge

545 rename AUDITOR_NAME AUDITOR_NAME1

546 rename AUDITOR_CITY AUDITOR_CITY1

547 rename AUDITOR_STATE AUDITOR_STATE1

548

549 replace permno = permno2
 (4,174,860 real changes made)

550 merge m:1 permno year using "auditor_office_to_merge.dta"

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 2,772,621 | |
| from master | 2,717,362 | (_merge==1) |
| from using | 55,259 | (_merge==2) |
| Matched | 1,457,498 | (_merge==3) |

551 drop if _merge == 2
 (55,259 observations deleted)

```

552 drop _merge
553 rename AUDITOR_NAME AUDITOR_NAME2
554 rename AUDITOR_CITY AUDITOR_CITY2
555 rename AUDITOR_STATE AUDITOR_STATE2
556
557 capt drop I_shared_office
558 g I_shared_office = (AUDITOR_NAME1==AUDITOR_NAME2 & AUDITOR_CITY1 ==
> AUDITOR_CITY2 & AUDITOR_STATE1 == AUDITOR_STATE2) & I_shared_audito
> r == 1
559 replace I_shared_office = 0 if AUDITOR_NAME1 == "" | AUDITOR_NAME2
> == "" | AUDITOR_CITY1 == "" | AUDITOR_CITY2 == "" | AUDITOR_STATE1 =
> = "" | AUDITOR_STATE2 == ""
(384,098 real changes made)
560
561 * merge controls variables
562 * see Martens and Sextroh 2021 for detailed definition of each var
563 capt drop PERMNO1
564 capt drop PERMNO2
565 capt drop YEAR
566 g PERMNO1 = permno1
567 g PERMNO2 = permno2
568 g YEAR = year
569
570 * merge common analyst
571 merge 1:1 PERMNO1 PERMNO2 YEAR using "CommonAnalyst.dta"
(variable PERMNO1 was float, now double to accommodate using data's
values)
(variable PERMNO2 was float, now double to accommodate using data's
values)

```

| Result | Number of obs | |
|-------------|-------------------|--------------|
| Not matched | 18,031,068 | |
| from master | 3,504,457 | (_merge==1) |
| from using | 14,526,611 | (_merge==2) |
| Matched | 670,403 | (_merge==3) |


```
572 drop if _merge == 2
    (14,526,611 observations deleted)
```

```
573 drop _merge
```

```
574 replace CommonAnalyst = 0 if CommonAnalyst == .
    (3,504,457 real changes made)
```

```
575
```

```
576 * merge common alliance
```

```
577 merge m:1 PERMNO1 PERMNO2 year using "CommonAlliance_PERMNO.dta"
```

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 4,179,613 | |
| from master | 4,172,310 | (_merge==1) |
| from using | 7,303 | (_merge==2) |
| Matched | 2,550 | (_merge==3) |

```
578 g CommonAlliance = (_merge == 3)
```

```
579 drop if _merge ==2
    (7,303 observations deleted)
```

```
580 drop _merge
```

```
581
```

```
582 * merge common investor
```

```
583 merge m:1 PERMNO1 PERMNO2 YEAR using "CommonInvestor.dta"
```

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 4,221,808 | |
| from master | 4,084,983 | (_merge==1) |
| from using | 136,825 | (_merge==2) |
| Matched | 89,877 | (_merge==3) |

```
584 drop if _merge == 2
    (136,825 observations deleted)
```

```
585 drop _merge
```

```
586 replace CommonInventor = 0 if CommonInventor == .  
    (4,084,983 real changes made)  
  
587  
588 * merge with size  
589 frame create funda  
  
590 frame change funda  
  
591 use "FUNDA.dta"  
    (  
        )  
  
592 frame change default  
  
593 capt drop PERMNO  
  
594 g PERMNO = PERMNO1  
  
595 frlink m:1 YEAR PERMNO, frame(funda) generate(FUNDA_PERMNO1)  
    (all observations in frame default matched)  
  
596 frget AT_PERMNO1 = AT, from(FUNDA_PERMNO1)  
    (14 missing values generated)  
    (1 variable copied from linked frame)  
  
597 replace PERMNO = PERMNO2  
    (4,174,860 real changes made)  
  
598 frlink m:1 YEAR PERMNO, frame(funda) generate(FUNDA_PERMNO2)  
    (all observations in frame default matched)  
  
599 frget AT_PERMNO2 = AT, from(FUNDA_PERMNO2)  
    (5 missing values generated)  
    (1 variable copied from linked frame)  
  
600 g RelativeSize = AT_PERMNO1 / AT_PERMNO2 / 1000  
    (19 missing values generated)  
  
601 winsor2 RelativeSize, cuts(1 99) replace by(YEAR)  
  
602 drop AT_PERMNO1 AT_PERMNO2  
  
603  
604 * merge with AnalystNo  
605 capt drop PERMNO
```

606 g PERMNO = PERMNO1

607 merge m:1 PERMNO YEAR using "AnalystNo.dta"
 (variable **PERMNO** was **float**, now **double** to accommodate using data's values)

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 937,915 | |
| from master | 796,187 | (_merge==1) |
| from using | 141,728 | (_merge==2) |
| Matched | 3,378,673 | (_merge==3) |

608 drop if _merge == 2
 (141,728 observations deleted)

609 drop _merge

610 replace AnalystNo = 0 if AnalystNo == .
 (796,187 real changes made)

611 rename AnalystNo AnalystNo_PERMNO1

612 replace PERMNO = PERMNO2
 (4,174,860 real changes made)

613 merge m:1 PERMNO YEAR using "AnalystNo.dta"

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 1,040,887 | |
| from master | 923,445 | (_merge==1) |
| from using | 117,442 | (_merge==2) |
| Matched | 3,251,415 | (_merge==3) |

614 drop if _merge == 2
 (117,442 observations deleted)

615 drop _merge

616 replace AnalystNo = 0 if AnalystNo == .
 (923,445 real changes made)

```

617 rename AnalystNo AnalystNo_PERMNO2
618 g RelativeAnalyst = (AnalystNo_PERMNO1+1) / (AnalystNo_PERMNO2+1) /
    > 1000
619 drop AnalystNo_PERMNO1 AnalystNo_PERMNO2
620
621 * merge with PatentStock
622 capt drop PERMNO
623 g PERMNO = PERMNO1
624 merge m:1 PERMNO YEAR using "PatentStock.dta"
    (variable PERMNO was float, now double to accommodate using data's
      values)

```

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 766,505 | |
| from master | 0 | (_merge==1) |
| from using | 766,505 | (_merge==2) |
| Matched | 4,174,860 | (_merge==3) |

```

625 drop if _merge == 2
    (766,505 observations deleted)
626 drop _merge
627 rename PatentStock PatentStock_permno1
628 replace PERMNO = PERMNO2
    (4,174,860 real changes made)
629 merge m:1 PERMNO YEAR using "PatentStock.dta"

```

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 725,263 | |
| from master | 0 | (_merge==1) |
| from using | 725,263 | (_merge==2) |
| Matched | 4,174,860 | (_merge==3) |


```

630 drop if _merge == 2
    (725,263 observations deleted)

631 drop _merge

632 rename PatentStock PatentStock_permno2

633 replace PatentStock_permno1 = 0 if PatentStock_permno1 == .
    (0 real changes made)

634 replace PatentStock_permno2 = 0 if PatentStock_permno2 == .
    (0 real changes made)

635 g RelativePatentStock = (1+PatentStock_permno1) / (1+PatentStock_per
    > mno2) / 1000

636 winsor2 RelativePatentStock, cuts(1 99) replace by(YEAR)

637 drop PatentStock_permno1 PatentStock_permno2

638
639 * merge with Relative Citation
640 capt drop PERMNO

641 g PERMNO = PERMNO1

642 merge m:1 PERMNO YEAR using "CitationNum.dta"
    (variable PERMNO was float, now double to accommodate using data's
    values)

```

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 18,523 | |
| from master | 48 | (_merge==1) |
| from using | 18,475 | (_merge==2) |
| Matched | 4,174,812 | (_merge==3) |

```

643 drop if _merge == 2
    (18,475 observations deleted)

644 drop _merge

645 replace CitationNum = 0 if CitationNum == .
    (48 real changes made)

```

```
646 rename CitationNum CitationNum_permnol
```

```
647 replace PERMNO = PERMNO2
      (4,174,860 real changes made)
```

```
648 merge m:1 PERMNO YEAR using "CitationNum.dta"
```

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 606,514 | |
| from master | 589,469 | (_merge==1) |
| from using | 17,045 | (_merge==2) |
| Matched | 3,585,391 | (_merge==3) |

```
649 drop if _merge == 2
      (17,045 observations deleted)
```

```
650 drop _merge
```

```
651 replace CitationNum = 0 if CitationNum == .
      (589,469 real changes made)
```

```
652 rename CitationNum CitationNum_permno2
```

```
653 g RelativeCitation = (CitationNum_permnol+1)/(CitationNum_permno2+1)
      > /1000
```

```
654 winsor2 RelativeCitation, cuts(1 99) replace by(YEAR)
```

```
655 drop CitationNum_permnol CitationNum_permno2
```

```
656 drop PERMNO
```

```
657
```

```
658 * merge tech similarity
```

```
659 merge 1:1 PERMNO1 PERMNO2 YEAR using "TechSimilarity.dta"
```

| Result | Number of obs | |
|-------------|-------------------|--------------|
| Not matched | 25,596,540 | |
| from master | 1,567,674 | (_merge==1) |
| from using | 24,028,866 | (_merge==2) |
| Matched | 2,607,186 | (_merge==3) |

```

660 drop if _merge == 2
    (24,028,866 observations deleted)

661 replace TechSimilarity = 0 if TechSimilarity == .
    (1,567,674 real changes made)

662 drop _merge

663
664 * update common ownership - the old one ends in 2012.
665 /* thank Backus, Conlon, and Sinkinson, 2021 for sharing this data *
    > /
666 capt rename CommonOwnership CommonOwnership_old

667 merge 1:1 permno1 permno2 year using "official-kappas.dta"

      Result                                Number of obs
      -----                                -
      Not matched                          12,508,398
        from master                        3,273,605  (_merge==1)
        from using                         9,234,793  (_merge==2)

      Matched                              901,255  (_merge==3)
      -----                                -

668 drop if _merge == 2
    (9,234,793 observations deleted)

669 drop _merge

670 replace CommonOwnership = 0 if CommonOwnership == .
    (3,273,605 real changes made)

671 save "merged_controls_20231210.dta", replace
    file merged_controls_20231210.dta saved

672 global Control1 = "CommonAnalyst CommonAlliance CommonInventor Commo
    > nOwnership TechSimilarity RelativeAnalyst RelativeSize RelativePaten
    > tStock RelativeCitation"

673 winsor2 $Control1, cuts(1 99) by(YEAR) replace

674 unique permno12 year if year>=2000 & year <=2018
    Number of unique values of permno12 year is 1568167
    Number of records is 1568167

```

```

675 //asdoc unique permno12 year if year>=2000 & year <=2018, save(outpu
> t\tl.rtf) append
676
677 * merge Cx cut var - amortization computed from compustat/crsp
678 capt drop PERMNO

679 g PERMNO = permno1

680 merge m:1 PERMNO YEAR using "INNOVATION_AM50.dta"
(variable PERMNO was float, now double to accommodate using data's
values)

```

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 264,696 | |
| from master | 1 | (_merge==1) |
| from using | 264,695 | (_merge==2) |
| Matched | 4,174,859 | (_merge==3) |

```

681 drop if _merge == 2
(264,695 observations deleted)

```

```

682 drop _merge

```

```

683 foreach i in "AM" {
2. foreach v of numlist 50{
3. rename I_`i'2AT_`v'_1 I_`i'2AT_`v'_1_citing
4. rename I_`i'2AT2_`v'_2 I_`i'2AT_`v'_2_citing
5. }
6. }

```

```

684 replace PERMNO = permno2
(4,174,860 real changes made)

```

```

685 merge m:1 PERMNO YEAR using "INNOVATION_AM50.dta"

```

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 223,500 | |
| from master | 43 | (_merge==1) |
| from using | 223,457 | (_merge==2) |
| Matched | 4,174,817 | (_merge==3) |


```

686 drop if _merge == 2
    (223,457 observations deleted)

687 drop _merge

688 foreach i in "AM"{
    2. foreach v of numlist 50{
    3. rename I_`i'2AT_`v'_1 I_`i'2AT_`v'_1_cited
    4. rename I_`i'2AT2_`v'_2 I_`i'2AT_`v'_2_cited
    5. }
    6. }

689 foreach i in "AM"{
    2. foreach v of numlist 50{
    3. capt drop I
    4. g I = (I_`i'2AT_`v'_1_citing == 1 & I_`i'2AT_`v'_1_cited == 1 )
    5. }
    6. }

690 capt drop officeid1 officeid2

691 gegen officeid1 = group(AUDITOR_NAME1 AUDITOR_CITY1 AUDITOR_STATE1)

692 gegen officeid2 = group(AUDITOR_NAME2 AUDITOR_CITY2 AUDITOR_STATE2)

693 keep permno1 permno2 year citations citations_raw permno12 Icitation
    > lcitation I_shared auditor I_shared_office YEAR $Control1 I AM2AT_5
    > 0_1_citing I_AM2AT_50_1_cited siccd1 siccd2 officeid1 officeid2 AUAA
    > 1 AUAA2

694 keep if YEAR>=2000 & YEAR <=2018
    (2,606,693 observations deleted)

695 count if YEAR != .
    1,568,167

696 save "main.dta", replace
    file main.dta saved

697

698 use "merged_controls_20231210.dta", clear

699 keep if YEAR>=2000 & YEAR <=2018
    (2,606,693 observations deleted)

700 bys PERMNO2 YEAR: gegen CITES = sum(citations)
    performance warning: -by- prefix may be slower than -by()-

```

```
701 keep PERMNO2 YEAR CITES
702 rename PERMNO2 PERMNO
703 duplicates drop PERMNO YEAR, force
      Duplicates in terms of PERMNO YEAR
      (1,533,574 observations deleted)
704 save "num_cites.dta", replace
      file num_cites.dta saved
705
706 * repeat the above code & to extend to 1980-2018 sample
707 clear all
708 use "PERMNO2PERMNO_PatentsView20231210.dta"
709 egen permno12 = group(permno1 permno2)
710 drop if year <= 1925
      (51 observations deleted)
711 tsset permno12 year
      Panel variable: permno12 (unbalanced)
      Time variable: year, 1931 to 2020, but with gaps
      Delta: 1 unit
712 tsfill, f
713 * restore permno after tsfill
714 bys permno12: egen permno1_m = mean(permno1)
715 replace permno1 = permno1_m if permno1 == .
      (41,676,565 real changes made)
716 drop permno1_m
717 bys permno12: egen permno2_m = mean(permno2)
718 replace permno2 = permno2_m if permno2 == .
      (41,676,565 real changes made)
719 drop permno2_m
```

```

720 replace citations = 0 if citations == .
    (41,676,565 real changes made)

721 g Icitation = (citations>0)

722
723 * cleaning sample
724 * at least one cross-firm citation
725 * (This is already satisfied, if no citation, firm pairs would not e
    > xist)
726 * drop if cited firm has zero patent until and including year t
727 frame create fr1

728 frame change fr1

729 import delimited "KPSS_public.csv", delimiter(",") clear
    (encoding automatically selected: ISO-8859-1)
    (7 vars, 3,053,011 obs)

730 g date = date(filing_date, "MDY")
    (2 missing values generated)

731 drop if date == .
    (2 observations deleted)

732 g year = year(date)

733 bys year permno: egen num_patent = count(year)

734 duplicates drop year permno , force

    Duplicates in terms of year permno

    (2,960,790 observations deleted)

735 xtset permno year

    Panel variable: permno (unbalanced)
    Time variable: year, 1834 to 2020, but with gaps
    Delta: 1 unit

736 tsfill, f

737 keep permno year num_patent

738 replace num_patent = 0 if num_patent == .
    (1,508,688 real changes made)

```

```

739 rangestat (sum) num_patent, by(permno) interval(year . 0)
740 rename permno permno2
741 frame change default
742 frlink m:1 permno2 year, frame(frl)
    (all observations in frame default matched)
743 frget num_patent_sum , from(frl)
    (1 variable copied from linked frame)
744 drop if num_patent_sum == 0
    (16,747,358 observations deleted)

745
746 * drop if citing firm's total number of citations in year t is 0
747 bys permno1 year: egen num_citation_sum1 = sum(citations)

748 drop if num_citation_sum1 == 0
    (15,566,912 observations deleted)

749
750 * drop self-citations
751 drop if permno1 == permno2
    (55,266 observations deleted)

752
753 * now have firm-pair-year observations
754 * dep. var
755 tsset permno12 year

    Panel variable: permno12 (unbalanced)
    Time variable: year, 1931 to 2020, but with gaps
    Delta: 1 unit

756 g citations_raw = citations
757 winsor2 citations, cuts(1 99) by(year) replace
758 g lcitation=log(1+citations)

759
760 * indep. var
761 * datasets used to merge are detailed below
762 * shared_auditor_to_merge.dta records audit firm from compustat

```



```
763 g permno = permno1
```

```
764 merge m:1 permno year using "shared_auditor_to_merge.dta"
      (variable permno was float, now double to accommodate using data's
       values)
```

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 3,134,174 | |
| from master | 2,976,179 | (_merge==1) |
| from using | 157,995 | (_merge==2) |
| Matched | 8,270,445 | (_merge==3) |

```
765 drop if _merge != 3
      (3,134,174 observations deleted)
```

```
766 drop _merge
```

```
767 rename siccd siccd1
```

```
768 rename au au1
```

```
769
```

```
770 replace permno = permno2
      (8,270,445 real changes made)
```

```
771 merge m:1 permno year using "shared_auditor_to_merge.dta"
```

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 4,208,403 | |
| from master | 4,091,506 | (_merge==1) |
| from using | 116,897 | (_merge==2) |
| Matched | 4,178,939 | (_merge==3) |

```
772 drop if _merge != 3
      (4,208,403 observations deleted)
```

```
773 drop _merge
```

```
774 rename siccd siccd2
```

```

775 rename au au2
776 g I_shared_auditor = (au1 == au2)
777
778 * drop if both au = `others', which is coded as 9
779 g YEAR = year
780 drop if au1 == "9" & au2 == "9" & YEAR<2000
      (4,079 observations deleted)
781
782 * update: after 2000, use audit analytics data to identify audit fir
      > m
783 * AUAA.dta records audit firm from audit analytics
784 g PERMNO1 = permno1
785 g PERMNO2 = permno2
786 capt drop PERMNO
787 g PERMNO = PERMNO1
788 merge m:1 PERMNO YEAR using "AUAA.dta"
      (variable PERMNO was float, now double to accommodate using data's
        values)

```

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 2,708,841 | |
| from master | 2,604,807 | (_merge==1) |
| from using | 104,034 | (_merge==2) |
| Matched | 1,570,053 | (_merge==3) |

```

789 drop if _merge == 2
      (104,034 observations deleted)
790 drop _merge
791 rename AUAA AUAA1
792
793 capt drop PERMNO

```

```
794 g PERMNO = PERMNO2
```

```
795 merge m:1 PERMNO YEAR using "AUAA.dta"
      (variable PERMNO was float, now double to accommodate using data's
        values)
```

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 2,685,861 | |
| from master | 2,597,822 | (_merge==1) |
| from using | 88,039 | (_merge==2) |
| Matched | 1,577,038 | (_merge==3) |

```
796 drop if _merge == 2
      (88,039 observations deleted)
```

```
797 drop _merge
```

```
798 rename AUAA AUAA2
```

```
799
```

```
800 capt drop I_shared_auditor
```

```
801 destring au1 au2, replace
      au1: all characters numeric; replaced as byte
      au2: all characters numeric; replaced as byte
```

```
802 replace au1 = AUAA1 if AUAA1!= . & AUAA2 != . & YEAR>=2000
      variable au1 was byte now int
      (1,522,982 real changes made)
```

```
803 replace au2 = AUAA2 if AUAA1!= . & AUAA2 != . & YEAR>=2000
      variable au2 was byte now int
      (1,522,409 real changes made)
```

```
804 g I_shared_auditor = (au1 == au2)
```

```
805
```

```
806 * shared audit office
```

```
807 * auditor_office_to_merge.dta records audit office from audit analyt
      > ics
```

```
808 capt drop permno
```

```
809 g permno = permno1
```

```

810 merge m:1 permno year using "auditor_office_to_merge.dta"
    (variable year was int, now double to accommodate using data's
      values)
    (variable permno was float, now double to accommodate using data's
      values)

```

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 2,777,857 | |
| from master | 2,707,618 | (_merge==1) |
| from using | 70,239 | (_merge==2) |
| Matched | 1,467,242 | (_merge==3) |

```

811 drop if _merge == 2
    (70,239 observations deleted)

```

```
812 drop _merge
```

```
813 rename AUDITOR_NAME AUDITOR_NAME1
```

```
814 rename AUDITOR_CITY AUDITOR_CITY1
```

```
815 rename AUDITOR_STATE AUDITOR_STATE1
```

```
816
```

```
817 replace permno = permno2
    (4,174,860 real changes made)

```

```
818 merge m:1 permno year using "auditor_office_to_merge.dta"
```

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 2,772,621 | |
| from master | 2,717,362 | (_merge==1) |
| from using | 55,259 | (_merge==2) |
| Matched | 1,457,498 | (_merge==3) |

```

819 drop if _merge == 2
    (55,259 observations deleted)

```

```
820 drop _merge
```



```

821 rename AUDITOR_NAME AUDITOR_NAME2
822 rename AUDITOR_CITY AUDITOR_CITY2
823 rename AUDITOR_STATE AUDITOR_STATE2
824
825 capt drop I_shared_office
826 g I_shared_office = (AUDITOR_NAME1==AUDITOR_NAME2 & AUDITOR_CITY1 ==
> AUDITOR_CITY2 & AUDITOR_STATE1 == AUDITOR_STATE2) & I_shared_audito
> r == 1
827 replace I_shared_office = 0 if AUDITOR_NAME1 == "" | AUDITOR_NAME2
> -- "" | AUDITOR_CITY1 -- "" | AUDITOR_CITY2 -- "" | AUDITOR_STATE1 -
> = "" | AUDITOR_STATE2 == ""
(384,098 real changes made)
828
829 * merge controls variables
830 * see Martens and Sextroh 2021 for detailed definition of each var
831 capt drop PERMNO1
832 capt drop PERMNO2
833 capt drop YEAR
834 g PERMNO1 = permno1
835 g PERMNO2 = permno2
836 g YEAR = year
837
838 * merge common analyst
839 merge 1:1 PERMNO1 PERMNO2 YEAR using "CommonAnalyst.dta"
(variable PERMNO1 was float, now double to accommodate using data's
values)
(variable PERMNO2 was float, now double to accommodate using data's
values)

```

| Result | Number of obs | |
|-------------|-------------------|--------------|
| Not matched | 18,031,068 | |
| from master | 3,504,457 | (_merge==1) |
| from using | 14,526,611 | (_merge==2) |
| Matched | 670,403 | (_merge==3) |

```
840 drop if _merge == 2
      (14,526,611 observations deleted)
```

```
841 drop _merge
```

```
842 replace CommonAnalyst = 0 if CommonAnalyst == .
      (3,504,457 real changes made)
```

```
843
```

```
844 * merge common alliance
```

```
845 merge m:1 PERMNO1 PERMNO2 year using "CommonAlliance_PERMNO.dta"
```

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 4,179,613 | |
| from master | 4,172,310 | (_merge==1) |
| from using | 7,303 | (_merge==2) |
| Matched | 2,550 | (_merge==3) |

```
846 g CommonAlliance = (_merge == 3)
```

```
847 drop if _merge ==2
      (7,303 observations deleted)
```

```
848 drop _merge
```

```
849
```

```
850 * merge common investor
```

```
851 merge m:1 PERMNO1 PERMNO2 YEAR using "CommonInvestor.dta"
```

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 4,221,808 | |
| from master | 4,084,983 | (_merge==1) |
| from using | 136,825 | (_merge==2) |
| Matched | 89,877 | (_merge==3) |

```
852 drop if _merge == 2
      (136,825 observations deleted)
```

```
853 drop _merge
```

```
854 replace CommonInventor = 0 if CommonInventor == .  
    (4,084,983 real changes made)  
  
855  
856 * merge with size  
857 frame create funda  
  
858 frame change funda  
  
859 use "FUNDA.dta"  
    (  
        )  
  
860 frame change default  
  
861 capt drop PERMNO  
  
862 g PERMNO = PERMNO1  
  
863 frlink m:1 YEAR PERMNO, frame(funda) generate(FUNDA_PERMNO1)  
    (all observations in frame default matched)  
  
864 frget AT_PERMNO1 = AT, from(FUNDA_PERMNO1)  
    (14 missing values generated)  
    (1 variable copied from linked frame)  
  
865 replace PERMNO = PERMNO2  
    (4,174,860 real changes made)  
  
866 frlink m:1 YEAR PERMNO, frame(funda) generate(FUNDA_PERMNO2)  
    (all observations in frame default matched)  
  
867 frget AT_PERMNO2 = AT, from(FUNDA_PERMNO2)  
    (5 missing values generated)  
    (1 variable copied from linked frame)  
  
868 g RelativeSize = AT_PERMNO1 / AT_PERMNO2 / 1000  
    (19 missing values generated)  
  
869 winsor2 RelativeSize, cuts(1 99) replace by(YEAR)  
  
870 drop AT_PERMNO1 AT_PERMNO2  
  
871  
872 * merge with AnalystNo  
873 capt drop PERMNO
```

874 g PERMNO = PERMNO1

875 merge m:1 PERMNO YEAR using "AnalystNo.dta"
 (variable **PERMNO** was **float**, now **double** to accommodate using data's values)

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 937,915 | |
| from master | 796,187 | (_merge==1) |
| from using | 141,728 | (_merge==2) |
| Matched | 3,378,673 | (_merge==3) |

876 drop if _merge == 2
 (141,728 observations deleted)

877 drop _merge

878 replace AnalystNo = 0 if AnalystNo == .
 (796,187 real changes made)

879 rename AnalystNo AnalystNo_PERMNO1

880 replace PERMNO = PERMNO2
 (4,174,860 real changes made)

881 merge m:1 PERMNO YEAR using "AnalystNo.dta"

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 1,040,887 | |
| from master | 923,445 | (_merge==1) |
| from using | 117,442 | (_merge==2) |
| Matched | 3,251,415 | (_merge==3) |

882 drop if _merge == 2
 (117,442 observations deleted)

883 drop _merge

884 replace AnalystNo = 0 if AnalystNo == .
 (923,445 real changes made)


```

885 rename AnalystNo AnalystNo_PERMNO2
886 g RelativeAnalyst = (AnalystNo_PERMNO1+1) / (AnalystNo_PERMNO2+1) /
    > 1000
887 drop AnalystNo_PERMNO1 AnalystNo_PERMNO2
888
889 * merge with PatentStock
890 capt drop PERMNO
891 g PERMNO = PERMNO1
892 merge m:1 PERMNO YEAR using "PatentStock.dta"
    (variable PERMNO was float, now double to accommodate using data's
      values)

```

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 766,505 | |
| from master | 0 | (_merge==1) |
| from using | 766,505 | (_merge==2) |
| Matched | 4,174,860 | (_merge==3) |

```

893 drop if _merge == 2
    (766,505 observations deleted)
894 drop _merge
895 rename PatentStock PatentStock_permno1
896 replace PERMNO = PERMNO2
    (4,174,860 real changes made)
897 merge m:1 PERMNO YEAR using "PatentStock.dta"

```

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 725,263 | |
| from master | 0 | (_merge==1) |
| from using | 725,263 | (_merge==2) |
| Matched | 4,174,860 | (_merge==3) |

```

898 drop if _merge == 2
    (725,263 observations deleted)

899 drop _merge

900 rename PatentStock PatentStock_permno2

901 replace PatentStock_permno1 = 0 if PatentStock_permno1 == .
    (0 real changes made)

902 replace PatentStock_permno2 = 0 if PatentStock_permno2 == .
    (0 real changes made)

903 g RelativePatentStock = (1+PatentStock_permno1) / (1+PatentStock_per
    > mno2) / 1000

904 winsor2 RelativePatentStock, cuts(1 99) replace by(YEAR)

905 drop PatentStock_permno1 PatentStock_permno2

906
907 * merge with Relative Citation
908 capt drop PERMNO

909 g PERMNO = PERMNO1

910 merge m:1 PERMNO YEAR using "CitationNum.dta"
    (variable PERMNO was float, now double to accommodate using data's
    values)

```

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 18,523 | |
| from master | 48 | (_merge==1) |
| from using | 18,475 | (_merge==2) |
| Matched | 4,174,812 | (_merge==3) |

```

911 drop if _merge == 2
    (18,475 observations deleted)

912 drop _merge

913 replace CitationNum = 0 if CitationNum == .
    (48 real changes made)

```

```
914 rename CitationNum CitationNum_permnol
```

```
915 replace PERMNO = PERMNO2
      (4,174,860 real changes made)
```

```
916 merge m:1 PERMNO YEAR using "CitationNum.dta"
```

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 606,514 | |
| from master | 589,469 | (_merge==1) |
| from using | 17,045 | (_merge==2) |
| Matched | 3,585,391 | (_merge==3) |

```
917 drop if _merge == 2
      (17,045 observations deleted)
```

```
918 drop _merge
```

```
919 replace CitationNum = 0 if CitationNum == .
      (589,469 real changes made)
```

```
920 rename CitationNum CitationNum_permno2
```

```
921 g RelativeCitation = (CitationNum_permnol+1)/(CitationNum_permno2+1)
    > /1000
```

```
922 winsor2 RelativeCitation, cuts(1 99) replace by(YEAR)
```

```
923 drop CitationNum_permnol CitationNum_permno2
```

```
924 drop PERMNO
```

```
925
```

```
926 * merge tech similarity
```

```
927 merge 1:1 PERMNO1 PERMNO2 YEAR using "TechSimilarity.dta"
```

| Result | Number of obs | |
|-------------|-------------------|--------------|
| Not matched | 25,596,540 | |
| from master | 1,567,674 | (_merge==1) |
| from using | 24,028,866 | (_merge==2) |
| Matched | 2,607,186 | (_merge==3) |

```

928 drop if _merge == 2
    (24,028,866 observations deleted)

929 replace TechSimilarity = 0 if TechSimilarity == .
    (1,567,674 real changes made)

930 drop _merge

931
932 * update common ownership
933 /* thank Backus, Conlon, and Sinkinson, 2021 for sharing this data *
    > /
934 capt rename CommonOwnership CommonOwnership_old

935 merge 1:1 permno1 permno2 year using "official-kappas.dta"

      Result                                Number of obs
      -----                                -
      Not matched                          12,508,398
        from master                        3,273,605  (_merge==1)
        from using                         9,234,793  (_merge==2)

      Matched                              901,255  (_merge==3)
      -----                                -

936 drop if _merge == 2
    (9,234,793 observations deleted)

937 drop _merge

938 replace CommonOwnership = 0 if CommonOwnership == .
    (3,273,605 real changes made)

939
940 global Controll = "CommonAnalyst CommonAlliance CommonInventor Commo
    > nOwnership TechSimilarity RelativeAnalyst RelativeSize RelativePaten
    > tStock RelativeCitation"

941 winsor2 $Controll, cuts(1 99) by(YEAR) replace

942 keep permno1 permno2 year lcitation I_shared_auditor I_shared_office
    > YEAR $Controll

943 keep if YEAR>=1980 & YEAR<=2018
    (608,438 observations deleted)

```



```

944 save "main_1980_2018.dta", replace
    file main_1980_2018.dta saved

945
    end of do-file

946 * step 3: construct other datasets for additional analyses
947 do ./3_construct_other.do

948 *****
949 * PART 3. construct other datasets
950 *****
951 * put all data into the data folder
952 * observation per step - table 1
953 * full permutation
954 /* very large, over 1.8b obs.
    > need to run on a server with ~700gb RAM
    > thank Cornell JCB for providing computing resources
    > use "permno_year1.dta", clear /*unique permno-year pairs in full sam
    > ple*/
    > tsset permno year
    > tsfill, f
    > save "permno_year2.dta", replace
    > use "permno_year2.dta", clear
    > rename permno permno2
    > joinby year using "permno_year2.dta"
    > rename permno permno1
    > keep if year>=2000 & year<=2018
    > unique permno1 permno2 year
    > save "company-pair-year_2000-2018.dta", replace
    > */
955 use "company-pair-year_2000-2018.dta", clear

956 count if year != .
    1,812,305,918

957 unique permno1 permno2 year
    Number of unique values of permno1 permno2 year is 1.812e+09
    Number of records is 1812305918

958 //asdoc unique permno1 permno2 year, save(output\t1.rtf) append
959
960 * demise of anderson
961 use main.dta, clear

```

```

962 capt drop SAMPLE
963 g SAMPLE = 1
964 capt drop YEAR
965 g YEAR = year
966 keep if year >=2000 & year <=2006
    (810,681 observations deleted)
967 capt drop POST
968 g POST = 0
969 replace POST = 1 if year > 2002.
    (408,187 real changes made)
970 g au1_anderson = AUAA1 == 5
971 g au2_anderson = AUAA2 == 5
972 by permno12: gegen au1_anderson_max = max(au1_anderson)
    performance warning: -by- prefix may be slower than -by()-
973 by permno12: gegen au2_anderson_max = max(au2_anderson)
    performance warning: -by- prefix may be slower than -by()-
974 drop au1_anderson au2_anderson
975 capt drop TREAT
976 g TREAT = 0
977 replace TREAT = 1 if (au1_anderson_max ==1 | au2_anderson_max ==1) &
    > (I_shared_office == 1 & year>2002)
    (1,117 real changes made)
978 g missing_aaaa1 = AUAA1 == .
979 g missing_aaaa2 = AUAA2 == .
980 by permno12: gegen missing_aaaa1_max = max(missing_aaaa1)
    performance warning: -by- prefix may be slower than -by()-
981 by permno12: gegen missing_aaaa2_max = max(missing_aaaa2)
    performance warning: -by- prefix may be slower than -by()-

```

```

982 * avoid if possible have shared office in pre
983 replace TREAT = 0 if (aul_anderson_max ==1 & au2_anderson_max ==1)
    (88 real changes made)

984 bys permno12: gegen TREAT_max = max(TREAT)
    performance warning: -by- prefix may be slower than -by()-

985 replace TREAT = TREAT_max
    (1,308 real changes made)

986 drop TREAT_max

987 replace SAMPLE = 0 if missing_auaa1_max == 1 | missing_auaa2_max ==
    > 1
    (96,215 real changes made)

988 drop missing_auaa1_max missing_auaa2_max

989 bys permno12: gegen I_shared_office_sd = sd(I_shared_office) if TREA
    > T == 0
    performance warning: -by- prefix may be slower than -by()-

990 bys permno12: gegen I_shared_office_m = mean(I_shared_office) if TRE
    > AT == 0
    performance warning: -by- prefix may be slower than -by()-

991 replace SAMPLE = 0 if I_shared_office_sd > 0 & I_shared_office_sd !=
    > .
    (14,189 real changes made)

992 replace SAMPLE = 0 if I_shared_office_m > 0 & I_shared_office_m != .
    >
    (2,435 real changes made)

993 capt drop num*

994 bys permno12: gegen num = sd(I_shared_office) if year <= 2001
    performance warning: -by- prefix may be slower than -by()-

995 bys permno12: gegen num_sd = mean(num)
    performance warning: -by- prefix may be slower than -by()-

996 replace SAMPLE = 0 if (num_sd>0 & num_sd!=.)
    (42 real changes made)

997 bys permno12: gegen year_min = min(year)
    performance warning: -by- prefix may be slower than -by()-

```

```

998 bys permno12: gegen year_max = max(year)
    performance warning: -by- prefix may be slower than -by()-

999 replace SAMPLE = 0 if year_max <=2002 | year_min > 2002
    (66,101 real changes made)

1000 drop year_min year_max

1001 g YEAR2000 = year == 2000

1002 g YEAR2001 = year == 2001

1003 g YEAR2002 = year == 2002

1004 g YEAR2003 = year == 2003

1005 g YEAR2004 = year == 2004

1006 keep if SAMPLE == 1 & YEAR <= 2004
    (328,930 observations deleted)

1007 keep lcitation YEAR2000 YEAR2001 YEAR2002 YEAR2003 YEAR2004 POST TR
    > EAT lcitation $Control1 year permno1 permno2

1008 save "table4.dta", replace
    file table4.dta saved

1009
1010 *oa figrue - a similar approach
1011 use "main.dta", clear

1012 global Control1 = "CommonAnalyst CommonAlliance CommonInventor Commo
    > nOwnership TechSimilarity RelativeAnalyst RelativeSize RelativePaten
    > tStock RelativeCitation"

1013 keep if year >=2000 & year <= 2006
    (810,681 observations deleted)

1014 capt drop SAMPLE

1015 g SAMPLE = 1

1016 * post var.
1017 capt drop POST

1018 g POST = 0

```



```

1019replace POST = 1 if year>2002
    (408,187 real changes made)

1020* treat var.
1021capt drop TREAT

1022g TREAT = 0

1023g au1_anderson = AUAA1 == 5 & AUAA1 != .

1024g au2_anderson = AUAA2 == 5 & AUAA2 != .

1025bys permno12: gegen au1_anderson_max = max(au1_anderson)
    performance warning: -by- prefix may be slower than -by()-

1026bys permno12: gegen au2_anderson_max = max(au2_anderson)
    performance warning: -by- prefix may be slower than -by()-

1027drop au1_anderson au2_anderson

1028bys permno12: gegen I_shared_office_max = max(I_shared_office) if ye
    > ar<2002
    performance warning: -by- prefix may be slower than -by()-

1029bys permno12: gegen I_shared_office_max_m = mean(I_shared_office_max
    > )
    performance warning: -by- prefix may be slower than -by()-

1030replace TREAT = 1 if (au1_anderson_max ==1 | au2_anderson_max ==1) &
    > (I_shared_office_max_m == 0 & (I_shared_office == 1 & year>2002) )
    (1,070 real changes made)

1031capt drop I_shared_office_max I_shared_office_max_m

1032bys permno12: gegen TREAT_max = max(TREAT)
    performance warning: -by- prefix may be slower than -by()-

1033replace TREAT = TREAT_max
    (1,314 real changes made)

1034drop TREAT_max

1035* treated - exclude if having other auditor change afterwards
1036* not clean treated as shared auditor status may change
1037bys permno12: gegen I_shared_office_sd = sd(I_shared_office) if TREA
    > T == 1 & year>2003
    performance warning: -by- prefix may be slower than -by()-

```

```

1038bys permno12: gegen I_shared_office_m = max(I_shared_office_sd)
    performance warning: -by- prefix may be slower than -by()-
1039replace SAMPLE = 0 if I_shared_office_m > 0 & I_shared_office_m != .
    (605 real changes made)
1040* control - no change in I_shared_office = 0
1041capt drop I_shared_office_sd I_shared_office_m
1042bys permno12: gegen I_shared_office_sd = sd(I_shared_office) if TREA
    > T == 0
    performance warning: -by- prefix may be slower than -by()-
1043bys permno12: gegen I_shared_office_m = mean(I_shared_office) if TRE
    > AT == 0
    performance warning: -by- prefix may be slower than -by()-
1044replace SAMPLE = 0 if I_shared_office_sd > 0 & I_shared_office_sd !=
    > .
    (16,203 real changes made)
1045replace SAMPLE = 0 if I_shared_office_m > 0 & I_shared_office_m != .
    >
    (2,443 real changes made)
1046* year indicators
1047g YEAR2000 = year == 2000
1048g YEAR2001 = year == 2001
1049g YEAR2002 = year == 2002
1050g YEAR2003 = year == 2003
1051g YEAR2004 = year == 2004
1052g YEAR2005 = year == 2005
1053g YEAR2006 = year == 2006
1054keep if SAMPLE == 1
    (19,251 observations deleted)
1055keep lcitation YEAR2000 YEAR2001 YEAR2002 YEAR2003 YEAR2004 YEAR2005
    > YEAR2006 POST TREAT $Control1 permno1 permno2 year

```

```
1056save "figureOA1.dta", replace
    file figureOA1.dta saved

1057
1058* cx new/old citations
1059* load KPSS data
1060clear all

1061capt frame drop KPSS

1062frame create KPSS

1063frame change KPSS

1064import delimited "KPSS_public.csv"
    (encoding automatically selected: ISO-8859-1)
    (7 vars, 3,053,011 obs)

1065drop xi_real xi_nominal cites

1066g filing_year = year(date(filing_date,"MDY"))
    (2 missing values generated)

1067g issue_year = year(date(issue_date,"MDY"))

1068drop filing_date issue_date

1069
1070* load patent citation data from patentsview
1071frame change default

1072use "citation_pair_public_firms_20231210.dta", clear

1073
1074* get info for citing patent
1075g patent_num = patent_id

1076frlink m:1 patent_num, frame(KPSS)
    (all observations in frame default matched)

1077frget filing_year, from(KPSS)
    (1 variable copied from linked frame)

1078drop year

1079rename filing_year year
```

```

1080drop KPSS

1081
1082* get info for cited patent
1083replace patent_num = citation_id
    (21,066,128 real changes made)

1084frlink m:1 patent_num, frame(KPSS)
    (all observations in frame default matched)

1085frget filing_year issue_year, from(KPSS)
    (2 variables copied from linked frame)

1086rename filing_year cited_filing_year

1087rename issue_year cited_issue_year

1088
1089* clean data
1090drop patent_num KPSS

1091drop if permno1 == . | permno2 == .
    (0 observations deleted)

1092sort year permno1 permno2

1093
1094* compute citation-lag
1095frame change default

1096* a patent is a new is citation lag is small
1097g novel = year - cited_issue_year <= 5

1098bys permno1 permno2 year: egen citations1 = count(patent_id) if nove
    > 1 == 1
    (10,926,099 missing values generated)

1099bys permno1 permno2 year: egen citations2 = count(patent_id) if nove
    > 1 -- 0
    (10,140,029 missing values generated)

1100bys permno1 permno2 year: egen citations1_m = mean(citations1)
    (3,634,146 missing values generated)

1101replace citations1 = citations1_m
    (7,291,953 real changes made)

```



```

1102drop citations1_m

1103bys permno1 permno2 year: egen citations2_m = mean(citations2)
    (1,975,866 missing values generated)

1104replace citations2 = citations2_m
    (8,164,163 real changes made)

1105drop citations2_m

1106
1107* save data
1108duplicates drop permno1 permno2 year , force

    Duplicates in terms of permno1 permno2 year
    (19,126,704 observations deleted)

1109keep permno1 permno2 year citations1 citations2

1110drop if year < 1980
    (110,993 observations deleted)

1111save "old_new.dta", replace
    file old_new.dta saved

1112
1113* cx - internal and external citation
1114clear all

1115capt frame drop KPSS

1116frame create KPSS

1117frame change KPSS

1118import delimited "KPSS_public.csv"
    (encoding automatically selected: ISO-8859-1)
    (7 vars, 3,053,011 obs)

1119drop xi_real xi_nominal cites

1120g filing_year = year(date(filing_date,"MDY"))
    (2 missing values generated)

1121g issue_year = year(date(issue_date,"MDY"))

```

```
1122drop filing_date issue_date
```

```
1123frame change default
```

```
1124
```

```
1125* self-citation
```

```
1126use "citation_pair_public_firms_20231210.dta", clear
```

```
1127g selfcite = permno1 == permno2
```

```
1128bys patent_id: egen selfcite_m = mean(selfcite)
```

```
1129duplicates drop patent_id, force
```

Duplicates in terms of **patent_id**

(18,959,114 observations deleted)

```
1130rename patent_id patent_num
```

```
1131keep patent_num selfcite_m year
```

```
1132sum selfcite_m
```

| Variable | Obs | Mean | Std. dev. | Min | Ma |
|------------|-----------|----------|-----------|-----|----|
| > x | | | | | |
| selfcite_m | 2,107,014 | .2521112 | .3280771 | 0 | |
| > 1 | | | | | |

```
1133save "self_citation.dta", replace
file self_citation.dta saved
```

```
1134
```

```
1135* internal citation rate of cited patent
```

```
1136frame change default
```

```
1137use "citation_pair_public_firms_20231210.dta", clear
```

```
1138keep if year>=2000 & year <=2018
```

(4,481,806 observations deleted)

```
1139g patent_num = citation_id
```

```
1140merge m:1 patent_num using "self_citation.dta"
```

(variable **patent_num** was **float**, now **double** to accommodate using data's values)

| Result | Number of obs | |
|-------------|---------------|--------------|
| Not matched | 2,207,328 | |
| from master | 1,347,485 | (_merge==1) |
| from using | 859,843 | (_merge==2) |
| Matched | 15,236,837 | (_merge==3) |

```

1141drop if _merge == 2
      (859,843 observations deleted)

1142drop _merge

1143
1144* a patent is internal if self citation rate is high
1145g internal = selfcite_m > 0.2

1146replace internal = 0 if selfcite_m == .
      (1,347,485 real changes made)

1147bys permno1 permno2 year: egen citations1 = count(year) if internal
      > == 1
      (10,081,320 missing values generated)

1148bys permno1 permno2 year: egen citations2 = count(year) if internal
      > == 0
      (6,503,002 missing values generated)

1149bys permno1 permno2 year: egen citations1_m = mean(citations1)
      (2,284,184 missing values generated)

1150replace citations1 = citations1_m
      (7,797,136 real changes made)

1151drop citations1_m

1152bys permno1 permno2 year: egen citations2_m = mean(citations2)
      (233,932 missing values generated)

1153replace citations2 = citations2_m
      (6,269,070 real changes made)

1154drop citations2_m

1155
1156* save data
1157duplicates drop permno1 permno2 year , force

      Duplicates in terms of permno1 permno2 year

      (15,396,408 observations deleted)

1158keep permno1 permno2 year citations1 citations2

```

```
1159drop if year <1980
      (0 observations deleted)

1160save "internal_external.dta", replace
      file internal_external.dta saved

1161
1162* firm-year controls
1163* for variable definitions, see He and Tian 2013
1164use "FUNDA.dta", clear
      (
      )

1165g SIZE = log(1+AT)
      (22,419 missing values generated)

1166g RDASSET = XRD/AT
      (176,902 missing values generated)

1167replace RDASSET = 0 if RDASSET == .
      (176,902 real changes made)

1168bys PERMNO: egen minyear = min(YEAR)
      (637 missing values generated)

1169g AGE = log(YEAR - minyear + 1)
      (1,802 missing values generated)

1170drop minyear

1171g ROA = OIBDP/AT
      (30,531 missing values generated)

1172g PPEASSET = PPENT / AT
      (29,106 missing values generated)

1173g LEVERAGE = (DLTT + DLC) / AT
      (26,621 missing values generated)

1174g CAPEXASSET = CAPX/AT
      (43,954 missing values generated)

1175g MVE = PRCC_C * CSHO
      (7,253 missing values generated)

1176g BVD = DLTT + DLC
      (26,594 missing values generated)
```



```
1177g BVE = CEQ
      (30,821 missing values generated)
```

```
1178g TOBINQ = (MVE+BVD)/(BVE+BVD)
      (36,683 missing values generated)
```

```
1179xtset PERMNO YEAR
```

```
Panel variable: PERMNO (unbalanced)
Time variable: YEAR, 1950 to 2020, but with gaps
Delta: 1 unit
```

```
1180g KZINDEX = -1.002*(DP+IB)/L.PPENT+0.283*TOBINQ + 3.139*LEVERAGE- 39
      > .36*(DVC+DVP)/L.PPENT - 1.315*CHE/L.PPENT
      (79,187 missing values generated)
```

```
1181bys YEAR SIC: egen s1 = sum(REVT)
```

```
1182g s2 = REVT/s1*100
      (23,987 missing values generated)
```

```
1183bys YEAR SIC: egen HHI = sum(s2^2)
```

```
1184replace HHI = int(HHI)/10000
      (309,892 real changes made)
```

```
1185g HHI2 = HHI*HHI
```

```
1186drop s1 s2
```

```
1187merge 1:1 PERMNO YEAR using "InstOwn_Perc.dta" /* download from wrds
      > -tf*/
      (variable YEAR was int, now float to accommodate using data's
      values)
```

| Result | Number of obs | |
|-------------|----------------|----------------------|
| Not matched | 139,678 | |
| from master | 122,594 | (_merge --1) |
| from using | 17,084 | (_merge ==2) |
| Matched | 189,628 | (_merge ==3) |

```
1188drop if _merge == 2
      (17,084 observations deleted)
```

```
1189drop _merge
```

```
1190rename InstOwn_Perc INSTIOWN
```

```
1191replace INSTIOWN = 0 if INSTIOWN == .
      (122,594 real changes made)
```

```
1192/* use crsp dsf file
      > use "dsf.sas7bdat.dta", clear
      > g ILLIQUIDITY = 1000000*abs(RET)/(abs(PRC)*VOL)
      > winsor2 ILLIQUIDITY, cuts(1 99) replace
      > g YEAR = year(DATE)
      > bys PERMNO YEAR: egen ILLIQUIDITY_m = mean(ILLIQUIDITY)
      > replace ILLIQUIDITY = ILLIQUIDITY_m
      > keep ILLIQUIDITY PERMNO YEAR
      > duplicates drop PERMNO YEAR, force
      > save "Illiquidity.dta"
      > */
```

```
1193merge 1:1 PERMNO YEAR using "Illiquidity.dta"
```

| Result | Number of obs | |
|-------------|----------------|--------------|
| Not matched | 112,325 | |
| from master | 3,184 | (_merge==1) |
| from using | 109,141 | (_merge==2) |
| Matched | 309,038 | (_merge==3) |

```
1194drop if _merge == 2
      (109,141 observations deleted)
```

```
1195drop _merge
```

```
1196g permno = PERMNO
```

```
1197g year = YEAR
      (1,802 missing values generated)
```

```
1198* auditor info set
```

```
1199merge m:1 permno year using "auditor_office_to_merge.dta"
      (variable year was float, now double to accommodate using data's
       values)
      (variable permno was float, now double to accommodate using data's
       values)
```

| Result | Number of obs | |
|-------------|----------------|--------------|
| Not matched | 227,364 | |
| from master | 225,691 | (_merge==1) |
| from using | 1,673 | (_merge==2) |
| Matched | 86,531 | (_merge==3) |

```

1200drop if _merge == 2
      (1,673 observations deleted)

1201drop _merge

1202drop if AUDITOR_CITY == "" | AUDITOR_STATE == ""
      (225,805 observations deleted)

1203quietly unique PERMNO, by(AUDITOR_NAME AUDITOR_STATE AUDITOR_CITY YE
    > AR) generate(F2)

1204bys AUDITOR_NAME AUDITOR_STATE AUDITOR_CITY YEAR: egen F2_m = mean(F
    > 2)

1205replace F2 = F2_m
      (71,347 real changes made)

1206quietly unique PERMNO, by(YEAR) generate(E)

1207bys YEAR: egen E_m = mean(E)

1208replace E = E_m
      (86,393 real changes made)

1209g AUINFO3 = (F2-1)/(E-1)

1210keep PERMNO YEAR SIZE RDASSET AGE ROA PPEASSET LEVERAGE CAPEXASSET T
    > OBINQ KZINDEX HHI HHI2 INSTIOWN ILLIQUIDITY AUINFO3 SICH AU

1211save "PERMNO YEAR 202311.dta", replace
      file PERMNO_YEAR_202311.dta saved

1212
1213* construct firm-year regression sample
1214clear all

1215* merge future 3 yrs innovation outcomes
1216forval I = 1/3{
    2. capt frame drop tmp`I'
    3. frame create tmp`I'
    4. frame change tmp`I'
    5. use "PERMNO YEAR 202311.dta", clear
    6. global Controls = "SIZE RDASSET AGE ROA PPEASSET LEVERAGE CAPEXAS
> SET TOBINQ KZINDEX HHI HHI2 INSTIOWN ILLIQUIDITY"
    7. winsor2 $Controls, cuts(1 99) by(YEAR) replace
    8. local end = 2018 - `I'
    9. keep if YEAR>=2000 & YEAR <=`end'
    10. replace YEAR = YEAR + `I'
    11. merge 1:1 PERMNO YEAR using "num_cites.dta"
    12. drop if _merge == 2
    13. drop _merge
    14. merge 1:1 PERMNO YEAR using "num_patents.dta"
    15. drop if _merge == 2
    16. drop _merge
    17. replace YEAR = YEAR - `I'
    18. replace PATENTS = 0 if PATENTS == .
    19. replace CITES = 0 if CITES == .
    20. g CITESperP = CITES/PATENTS

```



```

21. g LCITES = log(1+CITES/PATENTS)
22. replace LCITES = 0 if LCITES == .
23. g LPATENTS = log(1+PATENTS)
24. replace PATENTS = 0 if PATENTS == .
25. replace CITESperP = 0 if CITESperP == .
26. capt destring AU, replace
27. replace KZINDEX = KZINDEX/1000
28. g permno = PERMNO
29. g year = YEAR
30. foreach v of varlist LPATENTS AUINFO3 $Controls{
31.     drop if `v' == .
32. }
33. }
(
(13,256 observations deleted)
(73,161 real changes made)
(variable PERMNO was long, now double to accommodate using data's
values)

```

| Result | Number of obs | |
|-------------|---------------|--------------|
| Not matched | 52,750 | |
| from master | 45,659 | (_merge==1) |
| from using | 7,091 | (_merge==2) |
| Matched | 27,502 | (_merge==3) |

```
(7,091 observations deleted)
```

| Result | Number of obs | |
|-------------|----------------|--------------|
| Not matched | 132,310 | |
| from master | 56,626 | (_merge==1) |
| from using | 75,684 | (_merge==2) |
| Matched | 16,535 | (_merge==3) |

```

(75,684 observations deleted)
(73,161 real changes made)
(56,626 real changes made)
(45,659 real changes made)
(56,626 missing values generated)
(56,626 missing values generated)
(56,626 real changes made)
(0 real changes made)
(56,626 real changes made)
(64,064 real changes made)
(0 observations deleted)
(0 observations deleted)
(56 observations deleted)
(0 observations deleted)
(0 observations deleted)
(2,743 observations deleted)
(672 observations deleted)
(288 observations deleted)
(2,103 observations deleted)
(67 observations deleted)
(4,878 observations deleted)

```



```
(0 observations deleted)
(0 observations deleted)
(0 observations deleted)
(273 observations deleted)
(
(17,188 observations deleted)
(69,229 real changes made)
(variable PERMNO was long, now double to accommodate using data's
values)
```

| Result | Number of obs | |
|-------------|---------------|--------------|
| Not matched | 54,372 | |
| from master | 44,504 | (_merge==1) |
| from using | 9,868 | (_merge==2) |
| Matched | 24,725 | (_merge==3) |

```
(9,868 observations deleted)
```

| Result | Number of obs | |
|-------------|----------------|--------------|
| Not matched | 132,016 | |
| from master | 54,513 | (_merge==1) |
| from using | 77,503 | (_merge==2) |
| Matched | 14,716 | (_merge==3) |

```
(77,503 observations deleted)
(69,229 real changes made)
(54,513 real changes made)
(44,504 real changes made)
(54,513 missing values generated)
(54,513 missing values generated)
(54,513 real changes made)
(0 real changes made)
(54,513 real changes made)
(60,724 real changes made)
(0 observations deleted)
(0 observations deleted)
(55 observations deleted)
(0 observations deleted)
(0 observations deleted)
(2,565 observations deleted)
(632 observations deleted)
(273 observations deleted)
(2,085 observations deleted)
(54 observations deleted)
(4,540 observations deleted)
(0 observations deleted)
(0 observations deleted)
(0 observations deleted)
(254 observations deleted)
(
(21,143 observations deleted)
(65,274 real changes made)
(variable PERMNO was long, now double to accommodate using data's
values)
```

| Result | Number of obs | |
|-------------|---------------|--------------|
| Not matched | 55,661 | |
| from master | 43,171 | (_merge==1) |
| from using | 12,490 | (_merge==2) |
| Matched | 22,103 | (_merge==3) |

(12,490 observations deleted)

| Result | Number of obs | |
|-------------|----------------|--------------|
| Not matched | 131,375 | |
| from master | 52,215 | (_merge==1) |
| from using | 79,160 | (_merge==2) |
| Matched | 13,059 | (_merge==3) |

(79,160 observations deleted)
 (65,274 real changes made)
 (52,215 real changes made)
 (43,171 real changes made)
 (52,215 missing values generated)
 (52,215 missing values generated)
 (52,215 real changes made)
 (0 real changes made)
 (52,215 real changes made)
 (57,316 real changes made)
 (0 observations deleted)
 (0 observations deleted)
 (54 observations deleted)
 (0 observations deleted)
 (0 observations deleted)
 (2,381 observations deleted)
 (594 observations deleted)
 (259 observations deleted)
 (2,066 observations deleted)
 (44 observations deleted)
 (4,251 observations deleted)
 (0 observations deleted)
 (0 observations deleted)
 (0 observations deleted)
 (242 observations deleted)

1217frame copy tmp1 combine

```

1218frame change combine
1219rename LPATENTS LPATENTS1
1220rename LCITES LCITES1
1221frlink 1:1 PERMNO YEAR, frame(tmp2)
      (3,310 observations in frame combine unmatched)
1222frget LPATENTS LCITES, from(tmp2)
      (3,310 missing values generated)
      (3,310 missing values generated)
      (2 variables copied from linked frame)
1223rename LPATENTS LPATENTS2
1224rename LCITES LCITES2
1225frlink 1:1 PERMNO YEAR, frame(tmp3)
      (6,698 observations in frame combine unmatched)
1226frget LPATENTS LCITES, from(tmp3)
      (6,698 missing values generated)
      (6,698 missing values generated)
      (2 variables copied from linked frame)
1227rename LPATENTS LPATENTS3
1228rename LCITES LCITES3
1229drop SICH permno year
1230save "innovation.dta", replace
      file innovation.dta saved
1231use innovation.dta, clear
      (
      )
1232drop CITES- LCITES3
1233* see Dong et al. 2021 for innovation value vars
1234merge 1:1 PERMNO YEAR using innovation_value_to_merge.dta

```

| Result | Number of obs | |
|-------------|---------------|--------------|
| Not matched | 11,080 | |
| from master | 0 | (_merge==1) |
| from using | 11,080 | (_merge==2) |
| Matched | 62,081 | (_merge==3) |

```

1235drop if _merge == 2
      (11,080 observations deleted)

1236drop _merge

1237save "innovation value.dta", replace
      file innovation_value.dta saved

1238
1239* applicant - examiner
1240* load patentsview data
1241clear all

1242frame create patentview

1243frame change patentview

1244use "uspatentcitation.dta", clear

1245g examiner = category == "cited by examiner"

1246duplicates drop patent_id citation_id , force
      Duplicates in terms of patent_id citation_id
      (211,562 observations deleted)

1247frame change default

1248
1249* match examiner data
1250use "citation_pair_public_firms_20231210.dta", clear

1251capt drop patentview

1252frlink m:1 patent_id citation_id, frame(patentview)
      (all observations in frame default matched)

1253frget examiner, from(patentview)
      (1 variable copied from linked frame)

1254replace examiner = 0 if examiner == .
      (0 real changes made)

1255bys permno1 permno2 year: egen citations1 = count(year) if examiner
      > == 1
      (17,563,648 missing values generated)

```



```

1256bys permno1 permno2 year: egen citations2 = count(year) if examiner
    > == 0
    (3,502,480 missing values generated)

1257bys permno1 permno2 year: egen citations1_m = mean(citations1)
    (5,302,611 missing values generated)

1258replace citations1 = citations1_m
    (12,261,037 real changes made)

1259drop citations1_m

1260bys permno1 permno2 year: egen citations2_m = mean(citations2)
    (434,856 missing values generated)

1261replace citations2 = citations2_m
    (3,067,624 real changes made)

1262drop citations2_m

1263
1264* save data
1265gdrop duplicates drop permno1 permno2 year , force

    Duplicates in terms of permno1 permno2 year

    (19,126,458 observations deleted)

1266keep permno1 permno2 year citations1 citations2

1267save "applicant.dta", replace
    file applicant.dta saved

1268
1269* download TNIC data from Hoberg and Phillips (2016)
1270/* ccmxpf_linktable.dta is generated using the code:
    > import sas using "ccmxpf_linktable.sas7bdat", clear case(upper)
    > drop if LPERMNO == .
    > keep GVKEY LPERMNO LINKDT LINKENDDT
    > rename LPERMNO PERMNO
    > destring GVKEY, replace
    > save "ccmxpf_linktable.dta", replace
    > */
1271forval year=2000/2018{
    2.      import delimited "TNIC\tnic_all_data\tnicall`year'.txt", clea
    > r
    3.      rename year YEAR
    4.      rename gvkey1 GVKEY
    5.      joinby GVKEY using "ccmxpf_linktable.dta"
    6.      drop if YEAR < year(LINKDT) | ( YEAR > year(LINKENDDT) &
    > LINKENDDT != . )
    7.      drop if PERMNO == .
    8.      drop GVKEY
    9.      rename PERMNO PERMNO1
    10.     drop LINKDT LINKENDDT
    11.     rename gvkey2 GVKEY
    12.     joinby GVKEY using "ccmxpf_linktable.dta"
    13.     drop if YEAR < year(LINKDT) | ( YEAR > year(LINKENDDT) &

```

```

> LINKENDDT != . )
14.      drop if PERMNO == .
15.      drop GVKEY
16.      rename PERMNO PERMNO2
17.      drop LINKDT LINKENDDT
18.      drop if score == .
19.      duplicates drop YEAR PERMNO1 PERMNO2, force
20.      drop ball
21.      save "TNIC\tnic_all_data\tnicall`year`.txt.dta", replace
22. }

```

(encoding automatically selected: ISO-8859-9)

(5 vars, 45,283,778 obs)

(10,180,328 observations deleted)

(0 observations deleted)

(10,379,346 observations deleted)

(0 observations deleted)

(7,246 observations deleted)

Duplicates in terms of **YEAR PERMNO1 PERMNO2**

(1,099,159 observations deleted)

file **TNIC\tnic_all_data\tnicall2000.txt.dta** saved

(encoding automatically selected: ISO-8859-9)

(5 vars, 37,817,819 obs)

(8,602,251 observations deleted)

(0 observations deleted)

(8,771,267 observations deleted)

(0 observations deleted)

(6,596 observations deleted)

Duplicates in terms of **YEAR PERMNO1 PERMNO2**

(809,005 observations deleted)

file **TNIC\tnic_all_data\tnicall2001.txt.dta** saved

(encoding automatically selected: ISO-8859-9)

(5 vars, 32,261,316 obs)

(7,477,975 observations deleted)

(0 observations deleted)

(7,636,877 observations deleted)

(0 observations deleted)

(6,187 observations deleted)

Duplicates in terms of **YEAR PERMNO1 PERMNO2**

(827,179 observations deleted)

file **TNIC\tnic_all_data\tnicall2002.txt.dta** saved

(encoding automatically selected: ISO-8859-9)

(5 vars, 28,177,916 obs)

(6,663,761 observations deleted)

(0 observations deleted)

(6,794,111 observations deleted)

(0 observations deleted)

(5,747 observations deleted)

Duplicates in terms of **YEAR PERMNO1 PERMNO2**

```
(716,241 observations deleted)
file TNIC\tnic_all_data\tnicall2003.txt.dta saved
(encoding automatically selected: ISO-8859-9)
(5 vars, 26,776,441 obs)
(6,245,107 observations deleted)
(0 observations deleted)
(6,397,615 observations deleted)
(0 observations deleted)
(5,642 observations deleted)
```

Duplicates in terms of **YEAR PERMNO1 PERMNO2**

```
(722,375 observations deleted)
file TNIC\tnic_all_data\tnicall2004.txt.dta saved
(encoding automatically selected: ISO-8859-9)
(5 vars, 25,698,479 obs)
(5,995,404 observations deleted)
(0 observations deleted)
(6,144,988 observations deleted)
(0 observations deleted)
(5,530 observations deleted)
```

Duplicates in terms of **YEAR PERMNO1 PERMNO2**

```
(720,475 observations deleted)
file TNIC\tnic_all_data\tnicall2005.txt.dta saved
(encoding automatically selected: ISO-8859-9)
(5 vars, 25,027,695 obs)
(5,860,471 observations deleted)
(0 observations deleted)
(5,995,916 observations deleted)
(0 observations deleted)
(5,419 observations deleted)
```

Duplicates in terms of **YEAR PERMNO1 PERMNO2**

```
(669,018 observations deleted)
file TNIC\tnic_all_data\tnicall2006.txt.dta saved
(encoding automatically selected: ISO-8859-9)
(5 vars, 24,239,275 obs)
(5,428,024 observations deleted)
(0 observations deleted)
(5,564,329 observations deleted)
(0 observations deleted)
(5,347 observations deleted)
```

Duplicates in terms of **YEAR PERMNO1 PERMNO2**

```
(679,723 observations deleted)
file TNIC\tnic_all_data\tnicall2007.txt.dta saved
(encoding automatically selected: ISO-8859-9)
(5 vars, 21,799,661 obs)
(4,873,729 observations deleted)
(0 observations deleted)
(4,998,451 observations deleted)
(0 observations deleted)
(5,081 observations deleted)
```


Duplicates in terms of **YEAR PERMNO1 PERMNO2**

(681,329 observations deleted)
file **TNIC\tnic_all_data\tnicall2008.txt.dta** saved
(encoding automatically selected: ISO-8859-9)
(5 vars, 19,640,813 obs)
(4,437,787 observations deleted)
(0 observations deleted)
(4,553,352 observations deleted)
(0 observations deleted)
(4,827 observations deleted)

Duplicates in terms of **YEAR PERMNO1 PERMNO2**

(602,264 observations deleted)
file **TNIC\tnic_all_data\tnicall2009.txt.dta** saved
(encoding automatically selected: ISO-8859-9)
(5 vars, 18,368,732 obs)
(4,071,486 observations deleted)
(0 observations deleted)
(4,182,936 observations deleted)
(0 observations deleted)
(4,696 observations deleted)

Duplicates in terms of **YEAR PERMNO1 PERMNO2**

(608,521 observations deleted)
file **TNIC\tnic_all_data\tnicall2010.txt.dta** saved
(encoding automatically selected: ISO-8859-9)
(5 vars, 17,300,575 obs)
(3,838,916 observations deleted)
(0 observations deleted)
(3,938,839 observations deleted)
(0 observations deleted)
(4,556 observations deleted)

Duplicates in terms of **YEAR PERMNO1 PERMNO2**

(589,752 observations deleted)
file **TNIC\tnic_all_data\tnicall2011.txt.dta** saved
(encoding automatically selected: ISO-8859-9)
(5 vars, 16,459,786 obs)
(3,576,485 observations deleted)
(0 observations deleted)
(3,673,558 observations deleted)
(0 observations deleted)
(4,439 observations deleted)

Duplicates in terms of **YEAR PERMNO1 PERMNO2**

(582,767 observations deleted)
file **TNIC\tnic_all_data\tnicall2012.txt.dta** saved
(encoding automatically selected: ISO-8859-9)
(5 vars, 16,979,127 obs)
(3,526,138 observations deleted)
(0 observations deleted)
(3,634,654 observations deleted)
(0 observations deleted)
(4,551 observations deleted)

Duplicates in terms of **YEAR PERMNO1 PERMNO2**

(650,705 observations deleted)
file **TNIC\tnic_all_data\tnicall2013.txt.dta** saved
(encoding automatically selected: ISO-8859-9)
(5 vars, 18,167,775 obs)
(3,640,954 observations deleted)
(0 observations deleted)
(3,754,403 observations deleted)
(0 observations deleted)
(4,727 observations deleted)

Duplicates in terms of **YEAR PERMNO1 PERMNO2**

(691,582 observations deleted)
file **TNIC\tnic_all_data\tnicall2014.txt.dta** saved
(encoding automatically selected: ISO-8859-9)
(5 vars, 17,891,766 obs)
(3,598,233 observations deleted)
(0 observations deleted)
(3,715,561 observations deleted)
(0 observations deleted)
(4,701 observations deleted)

Duplicates in terms of **YEAR PERMNO1 PERMNO2**

(687,309 observations deleted)
file **TNIC\tnic_all_data\tnicall2015.txt.dta** saved
(encoding automatically selected: ISO-8859-9)
(5 vars, 16,703,336 obs)
(3,271,099 observations deleted)
(0 observations deleted)
(3,385,677 observations deleted)
(0 observations deleted)
(4,592 observations deleted)

Duplicates in terms of **YEAR PERMNO1 PERMNO2**

(673,096 observations deleted)
file **TNIC\tnic_all_data\tnicall2016.txt.dta** saved
(encoding automatically selected: ISO-8859-9)
(5 vars, 16,364,522 obs)
(3,177,850 observations deleted)
(0 observations deleted)
(3,294,910 observations deleted)
(0 observations deleted)
(4,544 observations deleted)

Duplicates in terms of **YEAR PERMNO1 PERMNO2**

(724,357 observations deleted)
 file **TNIC\tnic_all_data\tnicall2017.txt.dta** saved
 (encoding automatically selected: ISO-8859-9)
 (5 vars, 16,257,070 obs)
 (3,047,057 observations deleted)
 (0 observations deleted)
 (3,144,636 observations deleted)
 (0 observations deleted)
 (4,466 observations deleted)

Duplicates in terms of **YEAR PERMNO1 PERMNO2**

(612,895 observations deleted)
 file **TNIC\tnic_all_data\tnicall2018.txt.dta** saved

1272

end of do-file

1273* step4: empirical analyses and outputs

1274do ./4_empirical_analysis.do

1275*****

1276* PART 4. statistical analysis & regression outputs

1277*****

1278clear all

1279capt cd data

1280global Control1 = "CommonAnalyst CommonAlliance CommonInventor Commo
 > nOwnership TechSimilarity RelativeAnalyst RelativeSize RelativePaten
 > tStock RelativeCitation"

1281global Control2 = "c.CommonAnalyst c.CommonAlliance c.CommonInventor
 > c.CommonOwnership c.TechSimilarity c.RelativeAnalyst c.RelativeSize
 > c.RelativePatentStock c.RelativeCitation"

1282

1283* Table 1

1284* summarize in 2_construct_main.do

1285

1286* Table 2

1287use "main.dta", clear

1288* Panel A

```
1289g I_shared_other = (I_shared_auditor == 1 & I_shared_office == 0 )
```

```
1290fsum citations_raw I_shared_office I_shared_other I_shared_auditor,
> s(N mean sd p1 p25 p50 p75 p99) format(%12.3f)
```

| Variable | N | Mean | SD | P1 |
|--------------------------|---------|-------|--------|-------|
| P25 Median P75 P99 | | | | |
| citations_raw | 1568167 | 3.323 | 46.128 | 0.000 |
| 0.000 0.000 1.000 51.000 | | | | |
| I_shared_office | 1568167 | 0.017 | 0.129 | 0.000 |
| 0.000 0.000 0.000 1.000 | | | | |
| I_shared_other | 1568167 | 0.221 | 0.415 | 0.000 |
| 0.000 0.000 0.000 1.000 | | | | |
| I_shared_auditor | 1568167 | 0.238 | 0.426 | 0.000 |
| 0.000 0.000 0.000 1.000 | | | | |

```
1291fsum $Controll1, s(N mean sd p1 p25 p50 p75 p99) format(%12.3f)
```

| Variable | N | Mean | SD | P1 |
|--------------------------|---------|-------|-------|-------|
| P25 Median P75 P99 | | | | |
| CommonAnalyst | 1568167 | 0.597 | 2.268 | 0.000 |
| 0.000 0.000 0.000 12.000 | | | | |
| CommonAlliance | 1568167 | 0.000 | 0.018 | 0.000 |
| 0.000 0.000 0.000 0.000 | | | | |
| CommonInventor | 1568167 | 0.097 | 3.003 | 0.000 |
| 0.000 0.000 0.000 2.000 | | | | |
| CommonOwnership | 1568167 | 0.142 | 0.934 | 0.000 |
| 0.000 0.000 0.000 1.182 | | | | |
| TechSimilarity | 1568167 | 0.190 | 0.252 | 0.000 |
| 0.000 0.074 0.283 0.954 | | | | |
| RelativeAnalyst | 1568167 | 0.003 | 0.006 | 0.000 |
| 0.001 0.001 0.003 0.030 | | | | |
| RelativeSize | 1568167 | 0.119 | 0.501 | 0.000 |
| 0.000 0.001 0.016 3.472 | | | | |
| RelativePatentStock | 1568167 | 0.147 | 0.655 | 0.000 |
| 0.000 0.001 0.017 3.870 | | | | |
| RelativeCitation | 1568167 | 1.116 | 4.754 | 0.000 |
| 0.000 0.002 0.055 26.969 | | | | |

>

```

1292//asdoc fsum lcitation citations citations_raw lcitation I_shared au
> ditor I_shared office I_shared other, s(N mean sd p1 p25 p50 p75 p99
> ) format(%12.3f) f(%12.3f) save(output\t2_a.rtf) replace
1293//asdoc fsum $Control1 , s(N mean sd p1 p25 p50 p75 p99) format(%12.
> 3f) f(%12.3f) save(output\t2_a.rtf) append
1294
1295* Panel B
1296pwcrr citations_raw lcitation I_shared_office $Control1 , star(0.0
> 5)

```

| > ommon~p | citati~w lcitat~n I_shar~e Common~t Common~e Common~r C | | | | | | |
|--------------|---|----------|----------|----------|----------|----------|--|
| citations ~w | 1.0000 | | | | | | |
| lcitation | 0.1203* | 1.0000 | | | | | |
| I_shared_o~e | 0.0091* | 0.0294* | 1.0000 | | | | |
| CommonAnal~t | 0.1323* | 0.1650* | 0.0658* | 1.0000 | | | |
| CommonAlli~e | 0.0292* | 0.0148* | 0.0019* | 0.0264* | 1.0000 | | |
| CommonInve~r | 0.1666* | 0.0428* | 0.0083* | 0.0456* | 0.0098* | 1.0000 | |
| CommonOwne~p | 0.0248* | 0.0379* | 0.0003 | 0.0314* | 0.0030* | 0.0105* | |
| > 1.0000 | | | | | | | |
| TechSimila~y | 0.0962* | 0.2321* | 0.0636* | 0.2900* | 0.0219* | 0.0409* | |
| > 0.0065* | | | | | | | |
| RelativeAn~t | -0.0170* | -0.0583* | -0.0159* | -0.0751* | -0.0046* | -0.0079* | |
| > -0.0418* | | | | | | | |
| RelativeSize | -0.0095* | -0.0329* | -0.0217* | -0.0548* | -0.0027* | -0.0055* | |
| > -0.0303* | | | | | | | |
| RelativePa~k | -0.0131* | -0.0442* | -0.0121* | -0.0366* | 0.0022* | -0.0017* | |
| > -0.0012 | | | | | | | |
| RelativeCi~n | -0.0054* | -0.0089* | -0.0105* | -0.0368* | -0.0011 | -0.0063* | |
| > -0.0050* | | | | | | | |
| | TechSi~y Relati~t Relati~e Relati~k Relati~n | | | | | | |
| TechSimila~y | 1.0000 | | | | | | |
| RelativeAn~t | -0.1065* | 1.0000 | | | | | |
| RelativeSize | -0.0714* | 0.4884* | 1.0000 | | | | |
| RelativePa~k | -0.0641* | 0.2175* | 0.3476* | 1.0000 | | | |
| RelativeCi~n | -0.1589* | 0.2831* | 0.3113* | 0.3636* | 1.0000 | | |

```

1297//asdoc pwcrr citations_raw lcitation I_shared_office $Control1 ,
> star(0.05) dec(2) save(output\t2_b.rtf) replace
1298

```


1299* Table 3

1300use "main.dta", clear

1301* i) without controls

1302reghdfe lcitation I_shared office, absorb(permno1##year permno2##yea

> r) cluster(permno1 permno2) version(5)

(running historical version of reghdfe: 5)(dropped 2911 singleton observations)

(MWFE estimator converged in 10 iterations)

Warning: VCV matrix was non-positive semi-definite; adjustment from Ca

> meron, Gelbach & Miller applied.

HDFE Linear regression

Number of obs = 1

> ,565,256

Absorbing 2 HDFE groups

F(1, 2689) =

> 52.94

Statistics robust to heteroskedasticity

Prob > F =

> 0.0000

R-squared =

> 0.2465

Adj R-squared =

> 0.2211

Number of clusters (permno1) = 2,690

Within R-sq. =

> 0.0011

Number of clusters (permno2) = 3,167

Root MSE =

> 0.6783

(Std. err. adjusted for 2,690 clusters in permn

> o1 permno2)

| | Coefficient | Robust std. err. | t | P> t | [95% conf |
|-----------------|-------------|---------------------|--------|-------|-----------|
| lcitation | | | | | |
| > . interval] | | | | | |
| I_shared office | .1814392 | .0249367 | 7.28 | 0.000 | .1325423 |
| > .2303362 | | | | | |
| cons | .3712212 | .0003854 | 963.28 | 0.000 | .3704655 |
| > .3719769 | | | | | |

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs |
|--------------|------------|-------------|--------------|
| permno1#year | 19106 | 19106 | 0 * |
| permno2#year | 31879 | 31879 | 0 * |

* = FE nested within cluster; treated as redundant for DoF computation

1303est store reg1

1304* ii) without fixed effects

1305reghdfe lcitation I_shared_office \$Control1, cluster(permno1 permno2
>)

(MWFE estimator converged in 1 iterations)

HDFE Linear regression

Number of obs = 1

> ,568,167

Absorbing 1 HDFE group

F(10, 2811) =

> 99.10

Statistics robust to heteroskedasticity

Prob > F =

> 0.0000

R-squared =

> 0.1205

Adj R-squared =

> 0.1205

Number of clusters (permno1) = 2,812

Within R-sq. =

> 0.1205

Number of clusters (permno2) = 3,465

Root MSE =

> 0.7203

(Std. err. adjusted for 2,812 clusters in per

> mno1 permno2)

| | lcitation | Coefficient | Robust std. err. | t | P> t | [95% co |
|----------------|-------------------|-------------|---------------------|-------|-------|----------|
| > n | | | | | | |
| > f. interval] | | | | | | |
| > 1 | I_shared_office | .0705647 | .0331775 | 2.13 | 0.034 | .0055 |
| > 8 | CommonAnalyst | .0591381 | .0034018 | 17.38 | 0.000 | .052467 |
| > 1 | CommonAlliance | .5748485 | .1236497 | 4.65 | 0.000 | .332395 |
| > 8 | CommonInventor | .0171369 | .005992 | 2.86 | 0.004 | .005387 |
| > 5 | CommonOwnership | .0388173 | .0239431 | 1.62 | 0.105 | -.008130 |
| > 9 | TechSimilarity | .6964939 | .0488329 | 14.26 | 0.000 | .600741 |
| > 9 | RelativeAnalyst | -4.587906 | .9459231 | -4.85 | 0.000 | -6.44267 |
| > 8 | RelativeSize | .0034858 | .0070689 | 0.49 | 0.622 | -.010374 |
| > 4 | RelativePatents~k | -.0515257 | .0141871 | -3.63 | 0.000 | -.07934 |

```

>      -.0237075
RelativeCitation |      .0084889      .0009307      9.12      0.000      .00666
> 4
>      .0103138
      _cons |      .2087482      .0128804      16.21      0.000      .183492
> 2
>      .2340042

```

1306est store reg2

1307* iii) comprehensive reg

```

1308reghdfe lcitation I_shared_office $Controll, absorb(permnol##year pe
> rmno2##year) cluster(permnol permno2)
(dropped 2911 singleton observations)
(MWFE estimator converged in 11 iterations)
Warning: VCV matrix was non-positive semi-definite; adjustment from Ca
> meron, Gelbach & Miller applied.

```

```

HDFE Linear regression                                Number of obs   =   1
> ,565,256                                           F(   10,   2689) =
Absorbing 2 HDFE groups                             Prob > F         =
>   111.04                                           R-squared       =
Statistics robust to heteroskedasticity              Adj R-squared   =
>   0.0000                                           Within R-sq.    =
>   0.3692                                           Root MSE       =
>   0.3480
Number of clusters (permnol) =           2,690
>   0.1638
Number of clusters (permno2) =           3,167
>   0.6206

```

```

> mnol permno2) (Std. err. adjusted for 2,690 clusters in per

```

| | lcitation | Coefficient | Robust std. err. | t | P> t | [95% co |
|-----------------|-----------|-------------|---------------------|-------|----------|---------|
| > n | | | | | | |
| > f. interval] | | | | | | |
| I_shared_office | .0661216 | .0166121 | 3.98 | 0.000 | .033547 | |
| > 8 | | | | | | |
| > .0986955 | | | | | | |
| CommonAnalyst | .0578618 | .0023999 | 24.11 | 0.000 | .05315 | |
| > 6 | | | | | | |
| > .0625676 | | | | | | |
| CommonAlliance | .298337 | .0688833 | 4.33 | 0.000 | .163267 | |
| > 3 | | | | | | |
| > .4334066 | | | | | | |
| CommonInventor | .0096318 | .003631 | 2.65 | 0.008 | .00251 | |
| > 2 | | | | | | |
| > .0167517 | | | | | | |
| CommonOwnership | .0103757 | .0064942 | 1.60 | 0.110 | -.002358 | |


```

> 4
>      .0231098
> TechSimilarity |      1.000857      .0540921      18.50      0.000      .894790
> 9
>      1.106924
> RelativeAnalyst |     -9.825906      1.231278      -7.98      0.000     -12.2402
> 5
>      -7.411558
> RelativeSize |      -.046855      .0081052      -5.78      0.000     -.062748
> 1
>      -.0309619
> RelativePatentS~k |    -.0597317      .0078403      -7.62      0.000     -.075105
> 3
>      -.044358
> RelativeCitation |    -.0154495      .001272     -12.15      0.000     -.017943
> 7
>      -.0129553
>      _cons |      .2055674      .0082829      24.82      0.000      .18932
> 6
>      .2218088

```

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs | |
|--------------|------------|-------------|--------------|---|
| permno1#year | 19106 | 19106 | 0 | * |
| permno2#year | 31879 | 31879 | 0 | * |

* = FE nested within cluster; treated as redundant for DoF computation

1309est store reg3

1310* output regression results

```

1311esttab reg1 reg2 reg3 using "output\t3.rtf", cells(b(star fmt(3)) se
> (par fmt(3))) stats(r2_a N_full, fmt(%9.3f %9.0g)) starlevels(* 0.1
> ** 0.05 *** 0.01) order(I_shared_office $Control1) modelwidth(12) re
> place /* report N_full as FEs varies */
(output written to output\t3.rtf)

```

1312

1313* Table 4

1314use "table4.dta", clear

```

1315reghdfe lcitation POST##TREAT $Control1, absorb(permno1##permno2 per
> mno1##year permno2##year) cluster(permno1 permno2)
(dropped 1212 singleton observations)
(MWFE estimator converged in 10 iterations)
note: 1bn.POST is probably collinear with the fixed effects (all parti
> alled-out values are close to zero; tol = 1.0e-09)
note: 1bn.TREAT is probably collinear with the fixed effects (all part
> ialled-out values are close to zero; tol = 1.0e-09)
Warning: VCV matrix was non-positive semi-definite; adjustment from Ca
> meron, Gelbach & Miller applied.

```



```

HDFE Linear regression
> 427,344
Absorbing 3 HDFE groups
> 14.12
Statistics robust to heteroskedasticity
> 0.0000
> 0.7676
> 0.6920
Number of clusters (permno1) = 1,171
> 0.0020
Number of clusters (permno2) = 1,884
> 0.3947

```

Number of obs =

F(10, 1170) =

Prob > F =

R-squared =

Adj R-squared =

Within R-sq. =

Root MSE =

(Std. err. adjusted for 1,171 clusters in per

> mno1 permno2)

| | lcitation | Coefficient | Robust std. err. | t | P> t | [95% co |
|----------------|-------------------|-------------|---------------------|-------|-------|----------|
| > n | | | | | | |
| > f. interval] | | | | | | |
| | 1.POST | 0 | (omitted) | | | |
| | 1.TREAT | 0 | (omitted) | | | |
| | POST#TREAT | | | | | |
| | 1 1 | .0601659 | .0224071 | 2.69 | 0.007 | .016203 |
| > 3 | | | | | | |
| > | .1041285 | | | | | |
| | CommonAnalyst | .0117007 | .0023884 | 4.90 | 0.000 | .007014 |
| > 6 | | | | | | |
| > | .0163868 | | | | | |
| | CommonAlliance | -.0162472 | .0456366 | -0.36 | 0.722 | -.10578 |
| > 6 | | | | | | |
| > | .0732916 | | | | | |
| | CommonInventor | .01111 | .0050232 | 2.21 | 0.027 | .001254 |
| > 5 | | | | | | |
| > | .0209654 | | | | | |
| | CommonOwnership | .03125 | .0136945 | 2.28 | 0.023 | .004381 |
| > 5 | | | | | | |
| > | .0581184 | | | | | |
| | TechSimilarity | .1456977 | .0164653 | 8.85 | 0.000 | .113392 |
| > 8 | | | | | | |
| > | .1780025 | | | | | |
| | RelativeAnalyst | -.6060514 | .564956 | -1.07 | 0.284 | -1.71449 |
| > 2 | | | | | | |
| > | .5023887 | | | | | |
| | RelativeSize | -.0146676 | .0093887 | -1.56 | 0.118 | -.033088 |
| > 2 | | | | | | |
| > | .003753 | | | | | |
| | RelativePatentS~k | -.0075339 | .0064093 | -1.18 | 0.240 | -.020108 |
| > 9 | | | | | | |
| > | .005041 | | | | | |
| | RelativeCitation | -.002591 | .0006743 | -3.84 | 0.000 | -.003913 |

```

> 9
>      -.0012681
>      _cons |      .306424      .0041851      73.22      0.000      .298212
> 8
>      .3146353

```

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs |
|-----------------|------------|-------------|--------------|
| permno1#permno2 | 90930 | 90930 | 0 * |
| permno1#year | 4908 | 4908 | 0 * |
| permno2#year | 9066 | 9066 | 0 * |

* = FE nested within cluster; treated as redundant for DoF computation

1316est store reg1

```

1317reghdfe lcitation (YEAR2000 YEAR2001 YEAR2003 YEAR2004 YEAR2002)##TR
> EAT $Control1, absorb(permno1##permno2 permno1##year permno2##year)
> cluster(permno1 permno2)
(dropped 1212 singleton observations)
(MWFE estimator converged in 10 iterations)
note: 1bn.YEAR2000 is probably collinear with the fixed effects (all p
> artialled-out values are close to zero; tol = 1.0e-09)
note: 1bn.YEAR2001 is probably collinear with the fixed effects (all p
> artialled-out values are close to zero; tol = 1.0e-09)
note: 1bn.YEAR2003 is probably collinear with the fixed effects (all p
> artialled-out values are close to zero; tol = 1.0e-09)
note: 1bn.YEAR2004 is probably collinear with the fixed effects (all p
> artialled-out values are close to zero; tol = 1.0e-09)
note: 1bn.YEAR2002 is probably collinear with the fixed effects (all p
> artialled-out values are close to zero; tol = 1.0e-09)
note: 1bn.TREAT is probably collinear with the fixed effects (all part
> ialled-out values are close to zero; tol = 1.0e-09)
note: 0b.YEAR2002#0b.TREAT omitted because of collinearity

```

```

HDFE Linear regression
> 427,344
Absorbing 3 HDFE groups
> 11.41
Statistics robust to heteroskedasticity
> 0.0000

Number of obs =
F( 13, 1170) =
Prob > F =
R-squared =
Adj R-squared =
Number of clusters (permno1) = 1,171
Within R-sq. =
Number of clusters (permno2) = 1,884
Root MSE =
> 0.3947

```

(Std. err. adjusted for 1,171 clusters in per

> mno1 permno2)

| lcitation | Coefficient | Robust std. err. | t | P> t | [95% co |
|-----------------------|-------------|---------------------|-------|-------|----------|
| > n | | | | | |
| > f. interval] | | | | | |
| 1.YEAR2000 | 0 | (omitted) | | | |
| 1.YEAR2001 | 0 | (omitted) | | | |
| 1.YEAR2003 | 0 | (omitted) | | | |
| 1.YEAR2004 | 0 | (omitted) | | | |
| 1.YEAR2002 | 0 | (omitted) | | | |
| 1.TREAT | 0 | (omitted) | | | |
| YEAR2000#TREAT 1 1 | -.0393893 | .0332237 | -1.19 | 0.236 | -.104574 |
| > 1 | | | | | |
| > .0257955 | | | | | |
| YEAR2001#TREAT 1 1 | .017907 | .027322 | 0.66 | 0.512 | -.035698 |
| > 5 | | | | | |
| > .0715126 | | | | | |
| YEAR2003#TREAT 1 1 | .0231942 | .0325818 | 0.71 | 0.477 | -.040731 |
| > 1 | | | | | |
| > .0871196 | | | | | |
| YEAR2004#TREAT 1 1 | .0855116 | .032846 | 2.60 | 0.009 | .021067 |
| > 9 | | | | | |
| > .1499553 | | | | | |
| YEAR2002#TREAT 1 1 | 0 | (omitted) | | | |
| CommonAnalyst | .0116848 | .0023868 | 4.90 | 0.000 | .007001 |
| > 8 | | | | | |
| > .0163677 | | | | | |
| CommonAlliance | -.0162808 | .045625 | -0.36 | 0.721 | -.105796 |
| > 8 | | | | | |
| > .0732351 | | | | | |
| CommonInventor | .0111066 | .0050221 | 2.21 | 0.027 | .001253 |
| > 3 | | | | | |
| > .02096 | | | | | |
| CommonOwnership | .0312216 | .0136958 | 2.28 | 0.023 | .004350 |
| > 5 | | | | | |
| > .0580928 | | | | | |
| TechSimilarity | .1457711 | .0164647 | 8.85 | 0.000 | .113467 |
| > 5 | | | | | |
| > .1780747 | | | | | |
| RelativeAnalyst | -.6089887 | .5649478 | -1.08 | 0.281 | -1.71741 |
| > 3 | | | | | |
| > .4994352 | | | | | |


```

> RelativeSize | -.0146645 .0093901 -1.56 0.119 -.033087
> 9
> .0037589
RelativePatentS~k | -.007559 .0064083 -1.18 0.238 -.02013
> 2
> .005014
RelativeCitation | -.0025917 .0006745 -3.84 0.000 -.003915
> 1
> -.0012683
_cons | .3064596 .0041759 73.39 0.000 .298266
> 5
> .3146527

```

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs |
|-----------------|------------|-------------|--------------|
| permno1#permno2 | 90930 | 90930 | 0 * |
| permno1#year | 4908 | 4908 | 0 * |
| permno2#year | 9066 | 9066 | 0 * |

* = FE nested within cluster; treated as redundant for DoF computation

1318est store reg2

1319* output regression results

```

1320esttab reg1 reg2 using "output\t4.rtf", cells(b(star fmt(3)) se(par
> fmt(3))) stats(r2_a N_full, fmt(%9.3f %9.0g)) starlevels(* 0.1 ** 0.
> 05 *** 0.01) order(I_shared_office $Control1) modelwidth(12) replace
>
(output written to output\t4.rtf)

```

1321

1322

1323* Table 5

1324* Col. 1-2 Innovation

1325use "main.dta", clear

1326capt drop I

1327g I = (I_AM2AT_50_1_citing == 1 & I_AM2AT_50_1_cited == 1)

```

1328reghdfe lcitation I_shared_office $Control1 if I==0 , absorb(permno1#year permno2#year) cluster(permno1 permno2)
(dropped 3505 singleton observations)
(MWFE estimator converged in 11 iterations)

```

Warning: VCV matrix was non-positive semi-definite; adjustment from Cameron, Gelbach & Miller applied.


```

HDFE Linear regression                                Number of obs   =   1
> ,155,733                                           F(   10,   2641) =
Absorbing 2 HDFE groups                               Prob > F         =
>   114.70                                           R-squared        =
Statistics robust to heteroskedasticity               Adj R-squared    =
>   0.0000                                           Within R-sq.     =
>   0.3718                                           Root MSE         =
>   0.3438
Number of clusters (permno1) =           2,642
>   0.1573
Number of clusters (permno2) =           3,135
>   0.6099

```

(Std. err. adjusted for 2,642 clusters in permno1 permno2)

```

> mnol permno2)

```

| | l citation | Coefficient | Robust std. err. | t | P> t | [95% co f. interval] |
|-----|-------------------|-------------|---------------------|--------|-------|-------------------------|
| > n | I_shared_office | .0559438 | .0154122 | 3.63 | 0.000 | .025722 |
| > 7 | | .0861649 | | | | |
| > | CommonAnalyst | .0558726 | .002541 | 21.99 | 0.000 | .050889 |
| > 9 | | .0608552 | | | | |
| > | CommonAlliance | .2742592 | .0796004 | 3.45 | 0.001 | .118173 |
| > 6 | | .4303447 | | | | |
| > | CommonInventor | .0077854 | .0027755 | 2.81 | 0.005 | .002343 |
| > 1 | | .0132277 | | | | |
| > | CommonOwnership | .0154029 | .007864 | 1.96 | 0.050 | -.000017 |
| > 2 | | .0308231 | | | | |
| > | TechSimilarity | .9765192 | .058251 | 16.76 | 0.000 | .862297 |
| > 1 | | 1.090741 | | | | |
| > | RelativeAnalyst | -9.05041 | 1.204839 | -7.51 | 0.000 | -11.4129 |
| > 3 | | -6.687885 | | | | |
| > | RelativeSize | -.045979 | .0083485 | -5.51 | 0.000 | -.062349 |
| > 2 | | -.0296088 | | | | |
| > | RelativePatentS~k | -.0551142 | .0081588 | -6.76 | 0.000 | -.071112 |
| > 5 | | -.0391159 | | | | |
| > | RelativeCitation | -.0151913 | .0012231 | -12.42 | 0.000 | -.017589 |
| > 7 | | -.0127929 | | | | |
| > | _cons | .2038159 | .0085732 | 23.77 | 0.000 | .187005 |
| > 1 | | .2206267 | | | | |

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs | |
|--------------|------------|-------------|--------------|---|
| permno1#year | 18838 | 18838 | 0 | * |
| permno2#year | 30455 | 30455 | 0 | * |

* = FE nested within cluster; treated as redundant for DoF computation

1329est store reg1

1330reghdfe lcitation I_shared_office \$Controll if I--1 , absorb(permno1#year permno2#year) cluster(permno1 permno2)
(dropped 1455 singleton observations)
(MWFE estimator converged in 11 iterations)
Warning: VCV matrix was non-positive semi-definite; adjustment from Cameron, Gelbach & Miller applied.

| | | |
|---|---------------|---|
| HDFE Linear regression | Number of obs | = |
| > 407,474 | F(10, 1548) | = |
| Absorbing 2 HDFE groups | Prob > F | = |
| > 88.67 | R-squared | = |
| Statistics robust to heteroskedasticity | Adj R-squared | = |
| > 0.0000 | Within R-sq. | = |
| > 0.3935 | Root MSE | = |
| > 0.3612 | | |
| Number of clusters (permno1) = | 1,549 | |
| > 0.1875 | | |
| Number of clusters (permno2) = | 1,942 | |
| > 0.6486 | | |

(Std. err. adjusted for 1,549 clusters in permno1 permno2)

> mnol permno2)

| lcitation | Coefficient | Robust std. err. | t | P> t | [95% co |
|-----------------|-------------|---------------------|-------|-------|---------|
| > n | | | | | |
| > f. interval] | | | | | |
| I_shared_office | .0859006 | .0230089 | 3.73 | 0.000 | .040768 |
| > 7 | | | | | |
| > .1310325 | | | | | |
| CommonAnalyst | .0604261 | .0029153 | 20.73 | 0.000 | .054707 |
| > 8 | | | | | |
| > .0661444 | | | | | |
| CommonAlliance | .3344926 | .0977867 | 3.42 | 0.001 | .142684 |
| > 2 | | | | | |
| > .526301 | | | | | |
| CommonInventor | .0561661 | .0184833 | 3.04 | 0.002 | .019911 |
| > 2 | | | | | |
| > .092421 | | | | | |

```

CommonOwnership | .0055465 .0042806 1.30 0.195 -.002849
> 9
> .0139429
TechSimilarity | 1.071903 .0602362 17.79 0.000 .953749
> 4
> 1.190056
RelativeAnalyst | -12.60955 1.911286 -6.60 0.000 -16.3585
> 3
> -8.860567
RelativeSize | -.0557885 .0086131 -6.48 0.000 -.07268
> 3
> -.0388939
RelativePatentStock | -.0787025 .0115308 -6.83 0.000 -.101320
> 2
> -.0560848
RelativeCitation | -.0154843 .0016642 -9.30 0.000 -.018748
> 6
> -.01222
_cons | .2050549 .0109542 18.72 0.000 .183568
> 3
> .2265414

```

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs |
|--------------|------------|-------------|--------------|
| permno1#year | 8343 | 8343 | 0 * |
| permno2#year | 12250 | 12250 | 0 * |

* = FE nested within cluster; treated as redundant for DoF computation

1331est store reg2

1332* test coef diff

```

1333reghdfe lcitation i.I##(i.I_shared_office c.CommonAnalyst c.CommonAl
> liance c.CommonInventor c.CommonOwnership c.TechSimilarity c.Relativ
> eAnalyst c.RelativeSize c.RelativePatentStock c.RelativeCitation ) ,
> absorb(I##permno1##year I##permno2##year) cluster(permno1 permno2)
(dropped 4960 singleton observations)
(MWFE estimator converged in 11 iterations)
note: 1bn.I is probably collinear with the fixed effects (all partiall
> ed-out values are close to zero; tol = 1.0e-09)
Warning: VCV matrix was non-positive semi-definite; adjustment from Ca
> meron, Gelbach & Miller applied.
warning: missing F statistic; dropped variables due to collinearity or
> too few clusters

```



```

HDFE Linear regression                                Number of obs   =   1
> ,563,207                                           F(   20,   2675) =
Absorbing 2 HDFE groups                               Prob > F         =
> .                                                  R-squared        =
Statistics robust to heteroskedasticity              Adj R-squared    =
> .                                                  Within R-sq.     =
> 0.3786                                           Root MSE         =
> 0.3495
Number of clusters (permno1) =          2,676
> 0.1661
Number of clusters (permno2) =          3,157
> 0.6201

```

(Std. err. adjusted for 2,676 clusters in permno1)

```
> mnol permno2)
```

| | lcitation | Coefficient | Robust std. err. | t | P> t | [95% co |
|-------------------|-----------|-------------|---------------------|--------|-------|----------|
| > n | | | | | | |
| > f. interval] | | | | | | |
| | 1.I | 0 (omitted) | | | | |
| 1.I_shared_office | | .0559438 | .0154229 | 3.63 | 0.000 | .025701 |
| > 8 | | | | | | |
| > .0861858 | | | | | | |
| CommonAnalyst | | .0558726 | .002541 | 21.99 | 0.000 | .0508 |
| > 9 | | | | | | |
| > .0608552 | | | | | | |
| CommonAlliance | | .2742592 | .0796008 | 3.45 | 0.001 | .118173 |
| > 8 | | | | | | |
| > .4303445 | | | | | | |
| CommonInventor | | .0077854 | .0027755 | 2.81 | 0.005 | .00234 |
| > 3 | | | | | | |
| > .0132278 | | | | | | |
| CommonOwnership | | .0154029 | .0078656 | 1.96 | 0.050 | -.000020 |
| > 3 | | | | | | |
| > .0308261 | | | | | | |
| TechSimilarity | | .9765192 | .0582509 | 16.76 | 0.000 | .862297 |
| > 8 | | | | | | |
| > 1.090741 | | | | | | |
| RelativeAnalyst | | -9.05041 | 1.204846 | -7.51 | 0.000 | -11.4129 |
| > 3 | | | | | | |
| > -6.687885 | | | | | | |
| RelativeSize | | -.045979 | .0083493 | -5.51 | 0.000 | -.062350 |
| > 8 | | | | | | |
| > -.0296072 | | | | | | |
| RelativePatents~k | | -.0551142 | .0081588 | -6.76 | 0.000 | -.071112 |
| > 5 | | | | | | |
| > -.0391159 | | | | | | |
| RelativeCitation | | -.0151913 | .0012231 | -12.42 | 0.000 | -.017589 |
| > 7 | | | | | | |
| > -.0127929 | | | | | | |
| I#I_shared_office | | | | | | |

| | | | | | | |
|-------------------|-----------|-----------|----------|-------|-------|----------|
| | 1 1 | .0299568 | .0165095 | 1.81 | 0.070 | -.002415 |
| > 8 | | | | | | |
| > | .0623294 | | | | | |
| I#c.CommonAnalyst | 1 | .0045535 | .0028833 | 1.58 | 0.114 | -.001100 |
| > 2 | | | | | | |
| > | .0102073 | | | | | |
| I# | | | | | | |
| c.CommonAlliance | 1 | .0602334 | .1110065 | 0.54 | 0.587 | -.157433 |
| > 7 | | | | | | |
| > | .2779006 | | | | | |
| I# | | | | | | |
| c.CommonInventor | 1 | .0483807 | .0185188 | 2.61 | 0.009 | .01206 |
| > 8 | | | | | | |
| > | .0846934 | | | | | |
| I# | | | | | | |
| c.CommonOwnership | 1 | -.0098564 | .0038702 | -2.55 | 0.011 | -.017445 |
| > 3 | | | | | | |
| > | -.0022675 | | | | | |
| I# | | | | | | |
| c.TechSimilarity | 1 | .0953834 | .0539443 | 1.77 | 0.077 | -.010393 |
| > 3 | | | | | | |
| > | .20116 | | | | | |
| I# | | | | | | |
| c.RelativeAnalyst | 1 | -3.559139 | 1.580409 | -2.25 | 0.024 | -6.65808 |
| > 5 | | | | | | |
| > | -.460193 | | | | | |
| I#c.RelativeSize | 1 | -.0098095 | .0083562 | -1.17 | 0.241 | -.026194 |
| > 8 | | | | | | |
| > | .0065758 | | | | | |
| I# | | | | | | |
| c. | | | | | | |
| RelativePatentS~k | 1 | -.0235883 | .0100644 | -2.34 | 0.019 | -.04332 |
| > 3 | | | | | | |
| > | -.0038535 | | | | | |
| I# | | | | | | |
| c. | | | | | | |
| RelativeCitation | 1 | -.000293 | .0014755 | -0.20 | 0.843 | -.003186 |
| > 2 | | | | | | |
| > | .0026003 | | | | | |

```

      _cons |      .2041388      .008239      24.78      0.000      .187983
> 4
>      .2202942

```

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs |
|----------------|------------|-------------|--------------|
| I#permno1#year | 27181 | 27181 | 0 * |
| I#permno2#year | 42705 | 42705 | 0 * |

* = FE nested within cluster; treated as redundant for DoF computation

```
1334test 1.I#1.I_shared_office=0
```

```
( 1) 1.I#1.I_shared_office = 0
```

```

      F( 1, 2675) =      3.29
      Prob > F =      0.0697

```

```
1335est store reg3
```

```
1336* output regression results
```

```

1337esttab reg1 reg2 reg3 using "output\t5.rtf", cells(b(star fmt(3)) se
> (par fmt(3))) stats(r2_a N_full, fmt(%9.3f %9.0g)) starlevels(* 0.1
> ** 0.05 *** 0.01) order(I_shared_office $Control1) modelwidth(12) re
> place
(output written to output\t5.rtf)

```

```
1338
```

```
1339* Table 6
```

```
1340use "main.dta", clear
```

```
1341
```

```
1342* new v.s. old
```

```
1343merge 1:1 permno1 permno2 year using "old_new.dta"
```

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 2,568,680 | |
| from master | 1,154,208 | (_merge==1) |
| from using | 1,414,472 | (_merge==2) |
| Matched | 413,959 | (_merge==3) |

```

1344drop if _merge == 2
      (1,414,472 observations deleted)

1345drop _merge

1346replace citations1 = 0 if citations1 == . //novel
      (1,318,853 real changes made)

1347replace citations2 = 0 if citations2 == . //non-novel
      (1,307,636 real changes made)

1348winsor2 citations1 citations2 , cuts(1 99) replace by(YEAR)

1349g lcitations1 = log(1+citations1)

1350g lcitations2 = log(1+citations2)

1351reghdfe lcitations1 I_shared_office $Control1 , absorb(permno1##year
> permno2##year) cluster(permno1 permno2)
      (dropped 2911 singleton observations)
      (MWFE estimator converged in 11 iterations)
Warning: VCV matrix was non-positive semi-definite; adjustment from Ca
> meron, Gelbach & Miller applied.

HDFE Linear regression                                Number of obs   =   1
> ,565,256                                           F(   10,   2689) =
Absorbing 2 HDFE groups                             Prob > F         =
>   77.62                                           R-squared        =
Statistics robust to heteroskedasticity              Adj R-squared    =
>   0.0000                                           Within R-sq.     =
>   0.3374                                           Root MSE         =
>   0.3151
Number of clusters (permno1) =           2,690
>   0.1536
Number of clusters (permno2) =           3,167
>   0.4783

                                (Std. err. adjusted for 2,690 clusters in per
> mno1 permno2)

```

| | Coefficient | Robust std. err. | t | P> t | [95% co |
|-----------------|-------------|---------------------|-------|-------|---------|
| lcitations1 | | | | | |
| I_shared_office | .0559129 | .0148618 | 3.76 | 0.000 | .026771 |
| CommonAnalyst | .0462839 | .0022422 | 20.64 | 0.000 | .041887 |
| CommonAlliance | .340232 | .0707182 | 4.81 | 0.000 | .201564 |

```

CommonInventor | .0103942 .0037344 2.78 0.005 .003071
> 7
> .0177167
CommonOwnership | .0069896 .0042092 1.66 0.097 -.00126
> 4
> .0152432
TechSimilarity | .6907975 .0463732 14.90 0.000 .599866
> 8
> .7817281
RelativeAnalyst | -9.315736 1.145308 -8.13 0.000 -11.5615
> 1
> -7.069962
RelativeSize | -.0327762 .0072133 -4.54 0.000 -.046920
> 5
> -.018632
RelativePatentS~k | -.0375606 .0066476 -5.65 0.000 -.050595
> 5
> -.0245257
RelativeCitation | -.0124392 .000912 -13.64 0.000 -.014227
> 5
> -.0106509
_cons | .0991331 .0066252 14.96 0.000 .08614
> 2
> .1121242

```

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs |
|--------------|------------|-------------|--------------|
| permno1#year | 19106 | 19106 | 0 * |
| permno2#year | 31879 | 31879 | 0 * |

* = FE nested within cluster; treated as redundant for DoF computation

1352est store reg1

```

1353reghdfe lcitations2 I_shared_office $Controll , absorb(permno1##yea
> r permno2##year) cluster(permno1 permno2)
(dropped 2911 singleton observations)
(MWFE estimator converged in 11 iterations)
Warning: VCV matrix was non-positive semi-definite; adjustment from Ca
> meron, Gelbach & Miller applied.

```

```

HDFE Linear regression
> ,565,256
Absorbing 2 HDFE groups
> 53.84
Statistics robust to heteroskedasticity
> 0.0000
> 0.3241
> 0.3013
Number of clusters (permno1) = 2,690
> 0.1170
Number of clusters (permno2) = 3,167

```

Number of obs = 1
F(10, 2689) =
Prob > F =
R-squared =
Adj R-squared =
Within R-sq. =
Root MSE =

> 0.5229

(Std. err. adjusted for 2,690 clusters in per

> mno1 permno2)

| | Coefficient | Robust std. err. | t | P> t | [95% co |
|-------------------|-------------|---------------------|-------|-------|----------|
| lcitations2 | | | | | |
| > n | | | | | |
| > f. interval] | | | | | |
| I_shared_office | .0365364 | .0136125 | 2.68 | 0.007 | .009844 |
| > 3 | | | | | |
| .0632285 | | | | | |
| CommonAnalyst | .0432141 | .0026415 | 16.36 | 0.000 | .038034 |
| > 5 | | | | | |
| .0483936 | | | | | |
| CommonAlliance | .203254 | .0570142 | 3.56 | 0.000 | .091457 |
| > 9 | | | | | |
| .3150501 | | | | | |
| CommonInventor | .00915 | .0030855 | 2.97 | 0.003 | .003099 |
| > 9 | | | | | |
| .0152001 | | | | | |
| CommonOwnership | .0096746 | .0060109 | 1.61 | 0.108 | -.00211 |
| > 2 | | | | | |
| .0214611 | | | | | |
| TechSimilarity | .6574835 | .0468528 | 14.03 | 0.000 | .565612 |
| > 3 | | | | | |
| .7493546 | | | | | |
| RelativeAnalyst | -4.450957 | .8993737 | -4.95 | 0.000 | -6.2144 |
| > 9 | | | | | |
| -2.687423 | | | | | |
| RelativeSize | -.0328167 | .0064927 | -5.05 | 0.000 | -.045547 |
| > 9 | | | | | |
| -.0200855 | | | | | |
| RelativePatentS~k | -.049177 | .0092107 | -5.34 | 0.000 | -.067237 |
| > 7 | | | | | |
| -.0311163 | | | | | |
| RelativeCitation | -.010694 | .001236 | -8.65 | 0.000 | -.013117 |
| > 6 | | | | | |
| -.0082705 | | | | | |
| _cons | .1145387 | .0075027 | 15.27 | 0.000 | .099827 |
| > 1 | | | | | |
| .1292504 | | | | | |

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs |
|--------------|------------|-------------|--------------|
| permno1#year | 19106 | 19106 | 0 * |
| permno2#year | 31879 | 31879 | 0 * |

* = FE nested within cluster; treated as redundant for DoF computation

```

1354est store reg2

1355
1356* test coef diff
1357save tmp1.dta, replace
    (file tmp1.dta not found)
    file tmp1.dta saved

1358g n = 0

1359append using tmp1.dta

1360replace n = 1 if n == .
    (1,568,167 real changes made)

1361rm tmp1.dta

1362g lcitations = lcitations1

1363replace lcitations = lcitations2 if n == 1
    (398,374 real changes made)

1364reghdfe lcitations i.n#(c.I_shared_office c.CommonAnalyst c.CommonAl
    > liance c.CommonInventor c.RelativeAnalyst c.RelativeSize c.RelativeP
    > atentStock c.RelativeCitation c.TechSimilarity c.CommonOwnership) ,
    > absorb(n##permno1##year n##permno2##year) cluster(permno1 permno2 )
    (dropped 5822 singleton observations)
    (MWFE estimator converged in 11 iterations)
Warning: VCV matrix was non-positive semi-definite; adjustment from Ca
> meron, Gelbach & Miller applied.

HDFE Linear regression                                Number of obs   =   3
> ,130,512                                           F(   20,   2689) =
Absorbing 2 HDFE groups                               Prob > F        =
>    61.17                                           R-squared       =
Statistics robust to heteroskedasticity               Adj R-squared   =
>    0.0000                                           Within R-sq.    =
>    0.3304                                           Root MSE        =
>    0.3079
Number of clusters (permno1) =          2,690
>    0.1341
Number of clusters (permno2) =          3,167
>    0.5011

                                (Std. err. adjusted for 2,690 clusters in per
> mno1 permno2)

```

| | Coefficient | Robust std. err. | t | P> t | [95% co |
|-------------------|-------------|---------------------|---|------|---------|
| lcitations | | | | | |
| > n | | | | | |
| > f. interval] | | | | | |
| n# | | | | | |
| c.I_shared_office | | | | | |

| | | | | | | |
|-------------------|------------------|-----------|----------|-------|-------|----------|
| | 0 | .0559129 | .0148618 | 3.76 | 0.000 | .026771 |
| > 2 | | | | | | |
| > | .0850546 | | | | | |
| | 1 | .0365364 | .0136125 | 2.68 | 0.007 | .009844 |
| > 3 | | | | | | |
| > | .0632285 | | | | | |
| n#c.CommonAnalyst | | | | | | |
| | 0 | .0462839 | .0022422 | 20.64 | 0.000 | .041887 |
| > 2 | | | | | | |
| > | .0506806 | | | | | |
| | 1 | .0432141 | .0026415 | 16.36 | 0.000 | .038034 |
| > 5 | | | | | | |
| > | .0483936 | | | | | |
| | n# | | | | | |
| c.CommonAlliance | | | | | | |
| | 0 | .340232 | .0707182 | 4.81 | 0.000 | .201564 |
| > 4 | | | | | | |
| > | .4788996 | | | | | |
| | 1 | .203254 | .0570142 | 3.56 | 0.000 | .091457 |
| > 9 | | | | | | |
| > | .3150501 | | | | | |
| | n# | | | | | |
| c.CommonInventor | | | | | | |
| | 0 | .0103942 | .0037344 | 2.78 | 0.005 | .003071 |
| > 7 | | | | | | |
| > | .0177167 | | | | | |
| | 1 | .00915 | .0030855 | 2.97 | 0.003 | .003099 |
| > 9 | | | | | | |
| > | .0152001 | | | | | |
| | n# | | | | | |
| c.RelativeAnalyst | | | | | | |
| | 0 | -9.315736 | 1.145308 | -8.13 | 0.000 | -11.5615 |
| > 1 | | | | | | |
| > | -7.069962 | | | | | |
| | 1 | -4.450957 | .8993737 | -4.95 | 0.000 | -6.2144 |
| > 9 | | | | | | |
| > | -2.687423 | | | | | |
| | n#c.RelativeSize | | | | | |
| | 0 | -.0327762 | .0072133 | -4.54 | 0.000 | -.046920 |
| > 5 | | | | | | |
| > | -.018632 | | | | | |
| | 1 | -.0328167 | .0064927 | -5.05 | 0.000 | -.045547 |
| > 9 | | | | | | |
| > | -.0200855 | | | | | |
| | n# | | | | | |
| c. | | | | | | |
| RelativePatentS~k | | | | | | |
| | 0 | -.0375606 | .0066476 | -5.65 | 0.000 | -.050595 |
| > 5 | | | | | | |
| > | -.0245257 | | | | | |
| | 1 | -.049177 | .0092107 | -5.34 | 0.000 | -.067237 |
| > 7 | | | | | | |

```

>      -.0311163
      n#
      c.
RelativeCitation
0      -.0124392      .000912      -13.64      0.000      -.014227
> 5
>      -.0106509
      1      -.010694      .001236      -8.65      0.000      -.013117
> 6
>      -.0082705
      n#
      c.TechSimilarity
0      .6907975      .0463732      14.90      0.000      .599866
> 8
>      .7817281
      1      .6574835      .0468528      14.03      0.000      .565612
> 3
>      .7493546
      n#
      c.CommonOwnership
0      .0069896      .0042092      1.66      0.097      -.00126
> 4
>      .0152432
      1      .0096746      .0060109      1.61      0.108      -.00211
> 2
>      .0214611
      _cons      .1068359      .0067187      15.90      0.000      .093661
> 6
>      .1200102

```

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | - Num. Coefs |
|----------------|------------|-------------|--------------|
| n#permno1#year | 38212 | 38212 | 0 * |
| n#permno2#year | 63758 | 63758 | 0 * |

* = FE nested within cluster; treated as redundant for DoF computation

1365test 0bn.n#c.I_shared_office=1bn.n#c.I_shared_office

(1) 0b.n#c.I_shared_office - 1.n#c.I_shared_office = 0

F(1, 2689) = 4.48
 Prob > F = 0.0345

1366est store reg3

1367* output regression results

1368esttab reg1 reg2 reg3 using "output\t6_a.rtf", cells(b(star fmt(3))
 > se(par fmt(3))) stats(r2_a N_full, fmt(%9.3f %9.0g)) starlevels(* 0.
 > 1 ** 0.05 *** 0.01) order(I_shared_office \$Control1) modelwidth(12)
 > replace
 (output written to output\t6_a.rtf)

1369

1370* internal-external

1371use "main.dta", clear

1372merge 1:1 permno1 permno2 year using "internal_external.dta"

| Result | Number of obs | |
|-------------|------------------|--------------|
| Not matched | 1,928,243 | |
| from master | 1,154,248 | (_merge==1) |
| from using | 773,995 | (_merge==2) |
| Matched | 413,919 | (_merge==3) |

1373drop if _merge == 2
 (773,995 observations deleted)

1374drop _merge

1375replace citations1 = 0 if citations1 == . //internal
 (1,407,653 real changes made)

1376replace citations2 = 0 if citations2 == . //external
 (1,201,585 real changes made)

1377winsor2 citations1 citations2 , cuts(1 99) replace by(YEAR)

1378g lcitations1 = log(1+citations1)

1379g lcitations2 = log(1+citations2)

1380reghdfe lcitations1 I_shared_office \$Control1 , absorb(permno1##year
 > permno2##year) cluster(permno1 permno2)
 (dropped 2911 singleton observations)
 (MWFE estimator converged in 11 iterations)
 Warning: VCV matrix was non-positive semi-definite; adjustment from Ca
 > meron, Gelbach & Miller applied.

```

HDFE Linear regression                                Number of obs   =   1
> ,565,256                                           F(   10,   2689) =
Absorbing 2 HDFE groups                             Prob > F         =
>   43.81                                           R-squared        =
Statistics robust to heteroskedasticity              Adj R-squared    =
>   0.0000                                           Within R-sq.     =
>   0.2765                                           Root MSE         =
>   0.2521
Number of clusters (permno1) =           2,690
>   0.1229
Number of clusters (permno2) =           3,167
>   0.4148

```

(Std. err. adjusted for 2,690 clusters in permno1 permno2)

```

> mnno1 permno2)

```

| | Coefficient | Robust std. err. | t | P> t | [95% co f. interval] |
|-------------------|-------------|---------------------|--------|-------|-------------------------|
| I_shared_office | .0296774 | .0136112 | 2.18 | 0.029 | .00298 |
| CommonAnalyst | .0361304 | .0024092 | 15.00 | 0.000 | .031406 |
| CommonAlliance | .2674223 | .0666639 | 4.01 | 0.000 | .136704 |
| CommonInventor | .0089478 | .0030853 | 2.90 | 0.004 | .00289 |
| CommonOwnership | .0073768 | .0044043 | 1.67 | 0.094 | -.001259 |
| TechSimilarity | .5199474 | .0423429 | 12.28 | 0.000 | .436919 |
| RelativeAnalyst | -4.651072 | .8024824 | -5.80 | 0.000 | -6.22461 |
| RelativeSize | -.0208279 | .0050156 | -4.15 | 0.000 | -.030662 |
| RelativePatentS~k | -.0354068 | .0084132 | -4.21 | 0.000 | -.051903 |
| RelativeCitation | -.0092764 | .0009047 | -10.25 | 0.000 | -.011050 |
| _cons | .0476261 | .0066125 | 7.20 | 0.000 | .034660 |

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs | |
|--------------|------------|-------------|--------------|---|
| permno1#year | 19106 | 19106 | 0 | * |
| permno2#year | 31879 | 31879 | 0 | * |

* = FE nested within cluster; treated as redundant for DoF computation

1381est store reg1

```
1382reghdfe lcitations2 I_shared_office $Controll , absorb(permno1##yea
> r permno2##year) cluster(permno1 permno2)
(dropped 2911 singleton observations)
(MWFE estimator converged in 11 iterations)
Warning: VCV matrix was non-positive semi-definite; adjustment from Ca
> meron, Gelbach & Miller applied.
```

```
HDFE Linear regression                                Number of obs   =   1
> ,565,256                                           F(   10,   2689) =
Absorbing 2 HDFE groups                               Prob > F         =
>   100.47                                           R-squared        =
Statistics robust to heteroskedasticity              Adj R-squared    =
>   0.0000                                           Within R-sq.     =
>   0.3543                                           Root MSE         =
>   0.3326
Number of clusters (permno1) =           2,690
>   0.1597
Number of clusters (permno2) =           3,167
>   0.5759
```

```
(Std. err. adjusted for 2,690 clusters in per
> mno1 permno2)
```

| | Coefficient | Robust std. err. | t | P> t | [95% co |
|-----------------|-------------|---------------------|-------|-------|---------|
| lcitations2 | | | | | |
| > n | | | | | |
| > f. interval] | | | | | |
| I_shared_office | .061101 | .0158476 | 3.86 | 0.000 | .030026 |
| > 2 | | | | | |
| > .0921757 | | | | | |
| CommonAnalyst | .0555587 | .0024293 | 22.87 | 0.000 | .050795 |
| > 1 | | | | | |
| > .0603222 | | | | | |
| CommonAlliance | .3167795 | .0696329 | 4.55 | 0.000 | .1802 |
| > 4 | | | | | |
| > .453319 | | | | | |
| CommonInventor | .0107944 | .0038682 | 2.79 | 0.005 | .003209 |
| > 5 | | | | | |
| > .0183794 | | | | | |

```

CommonOwnership | .0095931 .0060961 1.57 0.116 -.002360
> 4
> .0215466
TechSimilarity | .8796664 .0515278 17.07 0.000 .778628
> 3
> .9807046
RelativeAnalyst | -9.382767 1.223499 -7.67 0.000 -11.7818
> 6
> -6.983674
RelativeSize | -.0456005 .00753 -6.06 0.000 -.060365
> 8
> -.0308352
RelativePatentS~k | -.0527451 .0074562 -7.07 0.000 -.067365
> 6
> -.0381245
RelativeCitation | -.0145185 .0011523 -12.60 0.000 -.016777
> 9
> -.0122591
_cons | .1704097 .0077883 21.88 0.000 .15513
> 8
> .1856813

```

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs | |
|--------------|------------|-------------|--------------|---|
| permno1#year | 19106 | 19106 | 0 | * |
| permno2#year | 31879 | 31879 | 0 | * |

* = FE nested within cluster; treated as redundant for DoF computation

1383est store reg2

1384

1385* test coef diff

1386save tmp1.dta, replace

(file **tmp1.dta** not found)

file **tmp1.dta** saved

1387g n = 0

1388append using tmp1.dta

1389replace n = 1 if n == .

(1,568,167 real changes made)


```
1390rm tmp1.dta
```

```
1391g lcitations = lcitations1
```

```
1392replace lcitations = lcitations2 if n == 1
      (394,627 real changes made)
```

```
1393reghdfe lcitations i.n#(c.I_shared_office c.CommonAnalyst c.CommonAl
> liance c.CommonInventor c.RelativeAnalyst c.RelativeSize c.RelativeP
> atentStock c.RelativeCitation c.TechSimilarity c.CommonOwnership) ,
> absorb(n##permno1##year n##permno2##year) cluster(permno1 permno2 )
      (dropped 5822 singleton observations)
      (MWFE estimator converged in 11 iterations)
Warning: VCV matrix was non-positive semi-definite; adjustment from Ca
> meron, Gelbach & Miller applied.
warning: missing F statistic; dropped variables due to collinearity or
> too few clusters
```

```
HDFE Linear regression                                Number of obs   =   3
> ,130,512                                           F(   20,   2689) =
Absorbing 2 HDFE groups                             Prob > F        =
> .                                                  R-squared       =
Statistics robust to heteroskedasticity              Adj R-squared   =
> .                                                  Within R-sq.    =
> 0.3441                                           Root MSE        =
> 0.3220
Number of clusters (permno1) =           2,690
> 0.1475
Number of clusters (permno2) =           3,167
> 0.5018
```

```
(Std. err. adjusted for 2,690 clusters in per
> mno1 permno2)
```

| | lcitations | Coefficient | Robust std. err. | t | P> t | [95% co |
|-------------------|------------|-------------|---------------------|-------|-------|---------|
| > n | | | | | | |
| > f. interval] | | | | | | |
| | n# | | | | | |
| c.I_shared_office | 0 | .0296774 | .013704 | 2.17 | 0.030 | .00280 |
| > 6 | | | | | | |
| > .0565488 | 1 | .061101 | .0158476 | 3.86 | 0.000 | .030026 |
| > 2 | | | | | | |
| > .0921757 | | | | | | |
| n#c.CommonAnalyst | 0 | .0361304 | .0024093 | 15.00 | 0.000 | .031406 |
| > 3 | | | | | | |
| > .0408546 | 1 | .0555587 | .0024294 | 22.87 | 0.000 | .050794 |
| > 9 | | | | | | |

```

> .0603224
      n#
c.CommonAlliance
0 | .2674223 .0675029 3.96 0.000 .135059
> 5
> .3997852
      1 | .3167795 .0704832 4.49 0.000 .178572
> 8
> .4549862
      n#
c.CommonInventor
0 | .0089478 .003086 2.90 0.004 .002896
> 6
> .014999
      1 | .0107944 .0038691 2.79 0.005 .003207
> 7
> .0183811
      n#
c.RelativeAnalyst
0 | -4.651072 .8027985 -5.79 0.000 -6.22523
> 7
> -3.076908
      1 | -9.382767 1.223544 -7.67 0.000 -11.7819
> 5
> -6.983585
      n#c.RelativeSize
0 | -.0208279 .0050158 -4.15 0.000 -.030663
> 1
> -.0109927
      1 | -.0456005 .0075301 -6.06 0.000 -.060365
> 9
> -.0308351
      n#
c.
RelativePatentS~k
0 | -.0354068 .0084178 -4.21 0.000 -.051912
> 9
> -.0189008
      1 | -.0527451 .0074579 -7.07 0.000 -.067368
> 8
> -.0381213
      n#
c.
RelativeCitation
0 | -.0092764 .0009054 -10.25 0.000 -.011051
> 8
> -.0075011
      1 | -.0145185 .0011524 -12.60 0.000 -.016778
> 1
> -.0122589
      n#

```

| | | | | | | |
|-------------------|--|----------|----------|-------|-------|----------|
| c.TechSimilarity | | | | | | |
| 0 | | .5199474 | .0423431 | 12.28 | 0.000 | .43691 |
| > 9 | | | | | | |
| > .6029757 | | | | | | |
| 1 | | .8796664 | .0515278 | 17.07 | 0.000 | .778628 |
| > 3 | | | | | | |
| > .9807046 | | | | | | |
| | | | | | | |
| n# | | | | | | |
| c.CommonOwnership | | | | | | |
| 0 | | .0073768 | .0044044 | 1.67 | 0.094 | -.001259 |
| > 5 | | | | | | |
| > .016013 | | | | | | |
| 1 | | .0095931 | .0060961 | 1.57 | 0.116 | -.002360 |
| > 5 | | | | | | |
| > .0215467 | | | | | | |
| | | | | | | |
| _cons | | .1090179 | .0069854 | 15.61 | 0.000 | .095320 |
| > 6 | | | | | | |
| > .1227152 | | | | | | |

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs |
|----------------|------------|-------------|--------------|
| n#permno1#year | 38212 | 38212 | 0 * |
| n#permno2#year | 63758 | 63758 | 0 * |

* = FE nested within cluster; treated as redundant for DoF computation

1394test 0bn.n#c.I_shared_office=1bn.n#c.I_shared_office

(1) 0b.n#c.I_shared_office - 1.n#c.I_shared_office = 0

F(1, 2689) = 12.59
Prob > F = 0.0004

1395est store reg3

1396* output regression results

1397esttab reg1 reg2 reg3 using "output\t6_b.rtf", cells(b(star fmt(3))
> se(par fmt(3))) stats(r2_a N_full, fmt(%9.3f %9.0g)) starlevels(* 0.
> 1 ** 0.05 *** 0.01) order(I_shared_office \$Control1) modelwidth(12)
> replace
(output written to output\t6_b.rtf)

```

1398
1399* table 7
1400use "main.dta", clear

1401* col 1
1402reghdfe lcitation I_shared_auditor $Control1, absorb(permno1##year p
> ermno2##year) cluster(permno1 permno2)
(dropped 2911 singleton observations)
(MWFE estimator converged in 11 iterations)
Warning: VCV matrix was non-positive semi-definite; adjustment from Ca
> meron, Gelbach & Miller applied.

```

```

HDFE Linear regression                                Number of obs   =   1
> ,565,256                                           F(   10,   2689) =
Absorbing 2 HDFE groups                             Prob > F         =
> 114.14                                             R-squared        =
Statistics robust to heteroskedasticity              Adj R-squared    =
> 0.0000                                           Within R-sq.     =
> 0.3691                                           Root MSE         =
> 0.3479
Number of clusters (permno1) =           2,690
> 0.1637
Number of clusters (permno2) =           3,167
> 0.6207

```

(Std. err. adjusted for 2,690 clusters in per

```
> mno1 permno2)
```

| | lcitation | Coefficient | Robust std. err. | t | P> t | [95% co |
|------------------|-----------|-------------|---------------------|-------|-------|----------|
| > n | | | | | | |
| > f. interval] | | | | | | |
| I_shared_auditor | | .0073796 | .0044634 | 1.65 | 0.098 | -.001372 |
| > 4 | | | | | | |
| > .0161315 | | | | | | |
| CommonAnalyst | | .0580018 | .0024025 | 24.14 | 0.000 | .05329 |
| > 1 | | | | | | |
| > .0627127 | | | | | | |
| CommonAlliance | | .2984282 | .0688403 | 4.34 | 0.000 | .163442 |
| > 8 | | | | | | |
| > .4334135 | | | | | | |
| CommonInventor | | .0096438 | .0036363 | 2.65 | 0.008 | .002513 |
| > 5 | | | | | | |
| > .0167741 | | | | | | |
| CommonOwnership | | .0103819 | .0064987 | 1.60 | 0.110 | -.00236 |
| > 1 | | | | | | |
| > .0231248 | | | | | | |
| TechSimilarity | | 1.001936 | .0541278 | 18.51 | 0.000 | .895799 |
| > 4 | | | | | | |
| > 1.108072 | | | | | | |
| RelativeAnalyst | | -9.83534 | 1.23342 | -7.97 | 0.000 | -12.2538 |
| > 9 | | | | | | |
| > -7.416792 | | | | | | |


```

      RelativeSize |  -.0469185   .0081179   -5.78   0.000   -.062836
> 5
      -.0310005
RelativePatentS~k |  -.0597024   .0078497   -7.61   0.000   -.075094
> 6
      -.0443103
RelativeCitation |  -.0154537   .0012726  -12.14   0.000   -.01794
> 9
      -.0129583
      _cons |      .204674   .0083769   24.43   0.000   .188248
> 2
      .2210998

```

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs |
|--------------|------------|-------------|--------------|
| permno1#year | 19106 | 19106 | 0 * |
| permno2#year | 31879 | 31879 | 0 * |

* = FE nested within cluster; treated as redundant for DoF computation

1403est store regl

1404* col 2.

1405use "main_1980_2018.dta", clear /* change the sample period in Part
> I */

1406reghdfe lcitation I_shared_auditor \$Control1, absorb(permno1##year p
> ermno2##year) cluster(permno1 permno2)
(dropped 5912 singleton observations)
(MWFE estimator converged in 11 iterations)
Warning: VCV matrix was non-positive semi-definite; adjustment from Ca
> meron, Gelbach & Miller applied.

HDFE Linear regression

> ,560,510

Absorbing 2 HDFE groups

> 119.83

Statistics robust to heteroskedasticity

> 0.0000

> 0.3591

> 0.3385

Number of clusters (permno1) = 5,015

> 0.1460

Number of clusters (permno2) = 5,424

> 0.5102

Number of obs = 3

F(10, 5014) =

Prob > F =

R-squared =

Adj R-squared =

Within R-sq. =

Root MSE =

(Std. err. adjusted for 5,015 clusters in per

> mno1 permno2)

| | Coefficient | Robust std. err. | t | P> t | [95% co |
|-------------------|-------------|---------------------|--------|-------|----------|
| lcitation | | | | | |
| > n | | | | | |
| > f. interval] | | | | | |
| I_shared_auditor | .0061546 | .0029821 | 2.06 | 0.039 | .000308 |
| > 4 | | | | | |
| > .0120009 | | | | | |
| CommonAnalyst | .0425362 | .0018525 | 22.96 | 0.000 | .038904 |
| > 5 | | | | | |
| > .046168 | | | | | |
| CommonAlliance | .300671 | .041263 | 7.29 | 0.000 | .219777 |
| > 4 | | | | | |
| > .3815646 | | | | | |
| CommonInventor | .0119014 | .0043905 | 2.71 | 0.007 | .003294 |
| > 2 | | | | | |
| > .0205086 | | | | | |
| CommonOwnership | .0152698 | .0094979 | 1.61 | 0.108 | -.003350 |
| > 3 | | | | | |
| > .0338899 | | | | | |
| TechSimilarity | .8197161 | .04259 | 19.25 | 0.000 | .736221 |
| > 2 | | | | | |
| > .903211 | | | | | |
| RelativeAnalyst | -5.019194 | .5852473 | -8.58 | 0.000 | -6.16653 |
| > 5 | | | | | |
| > -3.871853 | | | | | |
| RelativeSize | -.0410486 | .005521 | -7.44 | 0.000 | -.051872 |
| > 1 | | | | | |
| > -.0302251 | | | | | |
| RelativePatentS~k | -.0388139 | .0042207 | -9.20 | 0.000 | -.047088 |
| > 2 | | | | | |
| > -.0305395 | | | | | |
| RelativeCitation | -.0173235 | .0013801 | -12.55 | 0.000 | -.020029 |
| > 1 | | | | | |
| > -.0146179 | | | | | |
| _cons | .1291754 | .0064592 | 20.00 | 0.000 | .116512 |
| > 4 | | | | | |
| > .1418383 | | | | | |

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs |
|--------------|------------|-------------|--------------|
| permno1#year | 40304 | 40304 | 0 * |
| permno2#year | 70530 | 70530 | 0 * |

* = FE nested within cluster; treated as redundant for DoF computation

```

1407est store reg2

1408* col 3
1409use "main.dta", clear

1410g I_shared_other = (I_shared_auditor == 1 & I_shared_office == 0 )

1411reghdfe lcitation I_shared_office I_shared_other $Control1 , absorb(
> permno1##year permno2##year) cluster(permno1 permno2)
(dropped 2911 singleton observations)
(MWFE estimator converged in 11 iterations)
Warning: VCV matrix was non-positive semi-definite; adjustment from Ca
> meron, Gelbach & Miller applied.

```

```

HDFE Linear regression                                Number of obs   =   1
> ,565,256                                           F(   11,   2689) =
Absorbing 2 HDFE groups                               Prob > F         =
> 103.64                                             R-squared        =
Statistics robust to heteroskedasticity              Adj R-squared    =
> 0.0000                                           Within R-sq.     =
> 0.3692                                           Root MSE         =
> 0.3480
Number of clusters (permno1) =           2,690
> 0.1638
Number of clusters (permno2) =           3,167
> 0.6206

```

```

                                (Std. err. adjusted for 2,690 clusters in per
> mno1 permno2)

```

| | lcitation | Coefficient | Robust std. err. | t | P> t | [95% co |
|----------------|-----------------|-------------|---------------------|-------|-------|----------|
| > n | | | | | | |
| > f. interval] | | | | | | |
| > 9 | I_shared_office | .0668974 | .0164834 | 4.06 | 0.000 | .034575 |
| > 8 | I_shared_other | .0029657 | .0047604 | 0.62 | 0.533 | -.006368 |
| > 2 | CommonAnalyst | .0578629 | .0023993 | 24.12 | 0.000 | .053158 |
| > 1 | CommonAlliance | .2984343 | .0688959 | 4.33 | 0.000 | .163340 |
| > 8 | CommonInventor | .0096329 | .0036306 | 2.65 | 0.008 | .002513 |
| > 5 | CommonOwnership | .0103738 | .0064917 | 1.60 | 0.110 | -.002355 |
| > | TechSimilarity | 1.000856 | .0540909 | 18.50 | 0.000 | .894791 |

```

> 7
>      1.10692
RelativeAnalyst | -9.824012   1.231356   -7.98   0.000   -12.2385
> 1
>     -7.409511
RelativeSize | -.0468504   .0081115   -5.78   0.000   -.062755
> 9
>     -.0309449
RelativePatentS~k | -.0597147   .0078396   -7.62   0.000   -.075086
> 9
>     -.0443425
RelativeCitation | -.0154493   .0012718  -12.15   0.000   -.017943
> 1
>     -.0129554
      _cons |   .2048888   .0083801   24.45   0.000   .188456
> 6
>     .221321

```

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs |
|--------------|------------|-------------|--------------|
| permno1#year | 19106 | 19106 | 0 * |
| permno2#year | 31879 | 31879 | 0 * |

* = FE nested within cluster; treated as redundant for DoF computation

```
1412test I_shared_other == I_shared_office
```

```
( 1) - I_shared_office + I_shared_other = 0
```

```

F( 1, 2689) = 13.41
Prob > F = 0.0003

```

```
1413est store reg3
```

```
1414* output regression results
```

```

1415esttab reg1 reg2 reg3 using "output\t7.rtf", cells(b(star fmt(3)) se
> (par fmt(3))) stats(r2_a N_full, fmt(%9.3f %9.0g)) starlevels(* 0.1
> ** 0.05 *** 0.01) order(I_shared_office $Control1) modelwidth(12) re
> place
(output written to output\t7.rtf)

```

```
1416
```



```

1417* table 8
1418use "innovation.dta", clear
(
)

1419global Controls = "SIZE RDASSET AGE ROA PPEASSET LEVERAGE CAPEXASSET
> TOBINQ KZINDEX HHI HHI2 INSTIOWN ILLIQUIDITY"

1420* Panel A. summary stat.
1421qui: reghdfe LPATENTS1 AUINFO3 $Controls if AU != ., cluster(PERMNO
> )

1422fsum PATENTS CITESperP AUINFO3 $Controls if e(sample) == 1, s(N mean
> sd p1 p25 p50 p75 p99)

```

| Variable | | N | Mean | SD | P1 | P25 | Median |
|------------|--------|-------|--------|-------|-------|-------|--------|
| P75 | P99 | | | | | | |
| PATENTS | 58058 | 14.60 | 140.59 | 0.00 | 0.00 | 0.00 | |
| 1.00 | 308.00 | | | | | | |
| CITESperP | 58058 | 1.36 | 12.86 | 0.00 | 0.00 | 0.00 | |
| 0.00 | 23.71 | | | | | | |
| AUINFO3 | 58058 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | |
| 0.01 | 0.02 | | | | | | |
| SIZE | 58058 | 6.45 | 2.02 | 2.33 | 4.99 | 6.42 | |
| 7.79 | 11.42 | | | | | | |
| RDASSET | 58058 | 0.05 | 0.12 | 0.00 | 0.00 | 0.00 | |
| 0.05 | 0.62 | | | | | | |
| AGE | 58058 | 2.55 | 0.87 | 0.69 | 1.95 | 2.64 | |
| 3.22 | 4.11 | | | | | | |
| ROA | 58058 | 0.05 | 0.22 | -0.93 | 0.02 | 0.09 | |
| 0.15 | 0.40 | | | | | | |
| PPEASSET | 58058 | 0.23 | 0.24 | 0.00 | 0.04 | 0.13 | |
| 0.34 | 0.88 | | | | | | |
| LEVERAGE | 58058 | 0.22 | 0.22 | 0.00 | 0.02 | 0.16 | |
| 0.34 | 0.94 | | | | | | |
| CAPEXASSET | 58058 | 0.04 | 0.05 | 0.00 | 0.01 | 0.03 | |
| 0.05 | 0.28 | | | | | | |
| TOBINQ | 58058 | 2.44 | 2.53 | 0.39 | 1.12 | 1.62 | |
| 2.70 | 15.14 | | | | | | |
| KZINDEX | 58058 | -0.02 | 0.06 | -0.29 | -0.01 | -0.00 | |
| 0.00 | 0.01 | | | | | | |
| HHI | 58058 | 0.25 | 0.21 | 0.04 | 0.10 | 0.19 | |
| 0.33 | 1.00 | | | | | | |

```

>      HHI2 |      58058      0.11      0.18      0.00      0.01      0.04
>      0.11 |      1.00
>
>      INSTIOWN |      58058      0.55      0.33      0.00      0.27      0.60
>      0.83 |      1.12
>
>      ILLIQUIDITY |      58058      0.89      3.03      0.00      0.00      0.01
>      0.14 |      17.07
>

```

1423* Panel B. regression

```

1424foreach v of varlist LPATENTS1 LCITES1 LPATENTS2 LCITES2 LPATENTS3 L
> CITES3 {
> 2. reghdfe `v' AUINFO3 $Controls if AU!=. , absorb(PERMNO YEAR AU) c
> luster(PERMNO)
> 3. est store reg_`v'
> 4. }
(dropped 1045 singleton observations)
(MWFE estimator converged in 21 iterations)

```

```

HDFE Linear regression                                Number of obs   =
> 57,013                                              F( 14, 6772) =
Absorbing 3 HDFE groups                               Prob > F       =
> 18.98                                              R-squared      =
Statistics robust to heteroskedasticity              Adj R-squared  =
> 0.0000                                           Within R-sq.   =
> 0.8871                                           Root MSE       =
> 0.8717
> 0.0213
Number of clusters (PERMNO) = 6,773
> 0.4610

```

(Std. err. adjusted for 6,773 clusters in
> PERMNO)

| | Coefficient | Robust std. err. | t | P> t | [95% conf. i nterval] |
|-----------|-------------|---------------------|-------|-------|--------------------------|
| LPATENTS1 | | | | | |
| AUINFO3 | 3.687324 | 1.720559 | 2.14 | 0.032 | .3144887 |
| SIZE | .1497623 | .0118207 | 12.67 | 0.000 | .12659 |
| RDASSET | .2269104 | .0935653 | 2.43 | 0.015 | .043493 |
| AGE | -.0371552 | .0210027 | -1.77 | 0.077 | -.0783272 |
| ROA | .0093076 | .0308153 | 0.30 | 0.763 | -.0511001 |
| PPEASSET | .2070763 | .0569422 | 3.64 | 0.000 | .0954516 |
| LEVERAGE | -.1930521 | .0354919 | -5.44 | 0.000 | -.2626274 |

| | | | | | |
|-------------|-----------|----------|-------|-------|-----------|
| CAPEXASSET | .1203124 | .0803698 | 1.50 | 0.134 | -.0372378 |
| > .2778625 | | | | | |
| TOBINQ | .0115508 | .0019011 | 6.08 | 0.000 | .0078241 |
| > .0152775 | | | | | |
| KZINDEX | .0499787 | .0747193 | 0.67 | 0.504 | -.0964947 |
| > .1964521 | | | | | |
| HHI | -.038364 | .1317005 | -0.29 | 0.771 | -.2965383 |
| > .2198102 | | | | | |
| HHI2 | .0672359 | .1304827 | 0.52 | 0.606 | -.1885511 |
| > .3230229 | | | | | |
| INSTIOWN | .0361861 | .0255609 | 1.42 | 0.157 | -.0139212 |
| > .0862934 | | | | | |
| ILLIQUIDITY | .0005322 | .0008328 | 0.64 | 0.523 | -.0011004 |
| > .0021649 | | | | | |
| _cons | -.3463098 | .0990683 | -3.50 | 0.000 | -.5405148 |
| > .1521049 | | | | | |

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs |
|-------------|------------|-------------|--------------|
| PERMNO | 6773 | 6773 | 0 * |
| YEAR | 18 | 1 | 17 |
| AU | 21 | 1 | 20 |

* = FE nested within cluster; treated as redundant for DoF computation
(dropped 1045 singleton observations)
(MWFE estimator converged in 21 iterations)

HDFE Linear regression
> 57,013
Absorbing 3 HDFE groups
> 3.70
Statistics robust to heteroskedasticity
> 0.0000

> 0.5964
> 0.5415
> 0.0022
Number of clusters (PERMNO) = 6,773
> 0.4709

Number of obs =
F(14, 6772) =
Prob > F =
R-squared =
Adj R-squared =
Within R-sq. =
Root MSE =

(Std. err. adjusted for 6,773 clusters in
> PERMNO)

| | Coefficient | Robust std. err. | t | P> t | [95% conf. i |
|------------|-------------|---------------------|------|-------|--------------|
| > nterval] | | | | | |
| AUINFO3 | 5.774719 | 1.568366 | 3.68 | 0.000 | 2.700228 |
| > 8.84921 | | | | | |
| SIZE | .0348103 | .0080615 | 4.32 | 0.000 | .0190072 |


```

> .0506133
  RDASSET | .0048937 .0707522 0.07 0.945 -.1338028
> .1435903
  AGE | .0399261 .0142118 2.81 0.005 .0120666
> .0677856
  ROA | -.0260771 .0288296 -0.90 0.366 -.0825922
> .030438
  PPEASSET | .0142206 .0465178 0.31 0.760 -.0769689
> .1054101
  LEVERAGE | -.0480708 .0270429 -1.78 0.076 -.1010834
> .0049417
  CAPEXASSET | -.0878419 .0705293 -1.25 0.213 -.2261016
> .0504178
  TOBINQ | .0052104 .0018402 2.83 0.005 .0016031
> .0088177
  KZINDEX | -.0442062 .0611232 -0.72 0.470 -.164027
> .0756145
  HHI | .0310454 .0947545 0.33 0.743 -.1547033
> .216794
  HHI2 | -.0380562 .0885815 -0.43 0.667 -.2117038
> .1355913
  INSTIOWN | -.0002367 .0226513 -0.01 0.992 -.0446403
> .0441669
  ILLIQUIDITY | .0008284 .0007876 1.05 0.293 -.0007156
> .0023723
  cons | -.0815279 .0658705 -1.24 0.216 -.2106547
> .047599

```

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs |
|-------------|------------|-------------|--------------|
| PERMNO | 6773 | 6773 | 0 * |
| YEAR | 18 | 1 | 17 |
| AU | 21 | 1 | 20 |

* - FE nested within cluster; treated as redundant for DoF computation
 (dropped 1065 singleton observations)
 (MWFE estimator converged in 21 iterations)

HDFE Linear regression

> 53,853

Absorbing 3 HDFE groups

> 14.06

Statistics robust to heteroskedasticity

> 0.0000

> 0.8815

> 0.8648

> 0.0156

Number of clusters (PERMNO) = 6,588

> 0.4681

Number of obs =

F(14, 6587) =

Prob > F =

R-squared =

Adj R-squared =

Within R-sq. =

Root MSE =

(Std. err. adjusted for **6,588** clusters in> **PERMNO)**

| | Coefficient | Robust std. err. | t | P> t | [95% conf. i |
|-------------------------|------------------|---------------------|--------------|--------------|------------------|
| LPATENTS2 > nterval] | | | | | |
| AUINFO3 | 4.174357 | 1.815192 | 2.30 | 0.021 | .6159927 |
| > 7.732722 | | | | | |
| SIZE | .1242279 | .0120173 | 10.34 | 0.000 | .1006701 |
| > .1477858 | | | | | |
| RDASSET | .0611529 | .0973007 | 0.63 | 0.530 | -.129588 |
| > .2518937 | | | | | |
| AGE | -.0344692 | .0228734 | -1.51 | 0.132 | -.0793086 |
| > .0103702 | | | | | |
| ROA | .0012988 | .0316706 | 0.04 | 0.967 | -.0607858 |
| > .0633834 | | | | | |
| PPEASSET | .1694742 | .0613928 | 2.76 | 0.006 | .0491244 |
| > .2898239 | | | | | |
| LEVERAGE | -.1926774 | .0375052 | -5.14 | 0.000 | -.2661998 |
| > -.119155 | | | | | |
| CAPEXASSET | .0587051 | .0792853 | 0.74 | 0.459 | -.0967199 |
| > .2141301 | | | | | |
| TOBINQ | .0123731 | .001956 | 6.33 | 0.000 | .0085387 |
| > .0162075 | | | | | |
| KZINDEX | .070507 | .0691862 | 1.02 | 0.308 | -.0651205 |
| > .2061344 | | | | | |
| HHI | -.1083613 | .1408498 | -0.77 | 0.442 | -.3844726 |
| > .1677499 | | | | | |
| HHI2 | .1199639 | .1391584 | 0.86 | 0.389 | -.1528317 |
| > .3927595 | | | | | |
| INSTIOWN | .043069 | .0295073 | 1.46 | 0.144 | -.0147748 |
| > .1009128 | | | | | |
| ILLIQUIDITY | .0012436 | .0008095 | 1.54 | 0.125 | -.0003433 |
| > .0028305 | | | | | |
| cons | -.1882811 | .1033095 | -1.82 | 0.068 | -.3908013 |
| > .014239 | | | | | |

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs |
|-------------|------------|-------------|--------------|
| PERMNO | 6588 | 6588 | 0 * |
| YEAR | 17 | 1 | 16 |
| AU | 21 | 1 | 20 |

* = FE nested within cluster; treated as redundant for DoF computation
 (dropped 1065 singleton observations)
 (MWFE estimator converged in 21 iterations)

```

HDFE Linear regression
> 53,853
Absorbing 3 HDFE groups
> 5.19
Statistics robust to heteroskedasticity
> 0.0000
> 0.5793
> 0.5202
> 0.0026
Number of clusters (PERMNO) = 6,588
> 0.4737

Number of obs =
F( 14, 6587) =
Prob > F =
R-squared =
Adj R-squared =
Within R-sq. =
Root MSE =

(Std. err. adjusted for 6,588 clusters in
> PERMNO)

```

| | Coefficient | Robust std. err. | t | P> t | [95% conf. i |
|---------------------|-------------|---------------------|-------|-------|--------------|
| LCITES2 nterval] | | | | | |
| AUINFO3 | 3.957015 | 1.616857 | 2.45 | 0.014 | .7874519 |
| > 7.126578 | | | | | |
| SIZE | .0500824 | .0088768 | 5.64 | 0.000 | .0326811 |
| > .0674837 | | | | | |
| RDASSET | .061924 | .0766349 | 0.81 | 0.419 | -.0883053 |
| > .2121533 | | | | | |
| AGE | .0166655 | .0150531 | 1.11 | 0.268 | -.0128434 |
| > .0461744 | | | | | |
| ROA | -.0281459 | .0290591 | -0.97 | 0.333 | -.0851111 |
| > .0288193 | | | | | |
| PPEASSET | .0227175 | .0497493 | 0.46 | 0.648 | -.0748073 |
| > .1202424 | | | | | |
| LEVERAGE | -.0559199 | .0281602 | -1.99 | 0.047 | -.111123 |
| > .0007168 | | | | | |
| CAPEXASSET | .1427467 | .0720889 | 1.98 | 0.048 | .0014291 |
| > .2840644 | | | | | |
| TOBINQ | .0029342 | .0017845 | 1.64 | 0.100 | -.0005639 |
| > .0064323 | | | | | |
| KZINDEX | -.0870671 | .088425 | -0.98 | 0.325 | -.2604087 |
| > .0862745 | | | | | |
| HHI | -.0069775 | .097361 | -0.07 | 0.943 | -.1978367 |
| > .1838816 | | | | | |
| HHI2 | -.0030277 | .0911591 | -0.03 | 0.974 | -.1817292 |
| > .1756738 | | | | | |
| INSTIOWN | -.0094254 | .0229015 | -0.41 | 0.681 | -.0543198 |
| > .035469 | | | | | |
| ILLIQUIDITY | -.0026874 | .0008362 | -3.21 | 0.001 | -.0043266 |
| > .0010481 | | | | | |
| cons | -.1181618 | .0701986 | -1.68 | 0.092 | -.2557737 |
| > .0194502 | | | | | |

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs |
|-------------|------------|-------------|--------------|
| PERMNO | 6588 | 6588 | 0 * |
| YEAR | 17 | 1 | 16 |
| AU | 21 | 1 | 20 |

* = FE nested within cluster; treated as redundant for DoF computation
(dropped 1134 singleton observations)
(MWFE estimator converged in 21 iterations)

```

HDFE Linear regression
> 50,582
Absorbing 3 HDFE groups
> 9.36
Statistics robust to heteroskedasticity
> 0.0000
> 0.8813
> 0.8642
> 0.0100
Number of clusters (PERMNO) = 6,312
> 0.4640

Number of obs =
F( 14, 6311) =
Prob > F =
R-squared =
Adj R-squared =
Within R-sq. =
Root MSE =

(STD. err. adjusted for 6,312 clusters in
> PERMNO)

```

| LPATENTS3 > nterval] | Coefficient | Robust std. err. | t | P> t | [95% conf. i |
|-------------------------|-------------|---------------------|-------|-------|--------------|
| AUINFO3 | 3.457878 | 1.929639 | 1.79 | 0.073 | -.3248705 |
| > 7.240627 | | | | | |
| SIZE | .0929431 | .0123105 | 7.55 | 0.000 | .0688104 |
| > .1170758 | | | | | |
| RDASSET | -.0117643 | .1105722 | -0.11 | 0.915 | -.2285233 |
| > .2049948 | | | | | |
| AGE | -.0205582 | .0244337 | -0.84 | 0.400 | -.0684566 |
| > .0273403 | | | | | |
| ROA | -.0406466 | .0320697 | -1.27 | 0.205 | -.1035141 |
| > .0222208 | | | | | |
| PPEASSET | .1708848 | .0659793 | 2.59 | 0.010 | .041543 |
| > .3002267 | | | | | |
| LEVERAGE | -.1641744 | .0402137 | -4.08 | 0.000 | -.243007 |
| > .0853418 | | | | | |
| CAPEXASSET | .0713397 | .0790245 | 0.90 | 0.367 | -.0835751 |
| > .2262545 | | | | | |
| TOBINQ | .010324 | .0020606 | 5.01 | 0.000 | .0062846 |
| > .0143634 | | | | | |
| KZINDEX | .0886686 | .0702085 | 1.26 | 0.207 | -.0489638 |
| > .2263011 | | | | | |
| HHI | -.1131462 | .1473015 | -0.77 | 0.442 | -.4019072 |
| > .1756148 | | | | | |


```

      HHI2 |      .1128016   .1457376   0.77   0.439   -.1728936
> .3984967
      INSTIOWN |      .064953   .0323345   2.01   0.045   .0015664
> .1283396
      ILLIQUIDITY |      .0002513   .0007271   0.35   0.730   -.0011741
> .0016767
      _cons |     -.0483431   .1082403   -0.45   0.655   -.2605309
> .1638446

```

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs |
|-------------|------------|-------------|--------------|
| PERMNO | 6312 | 6312 | 0 * |
| YEAR | 16 | 1 | 15 |
| AU | 21 | 1 | 20 |

* = FE nested within cluster; treated as redundant for DoF computation
(dropped 1134 singleton observations)
(MWFE estimator converged in 21 iterations)

```

HDFE Linear regression                                Number of obs   =
> 50,582
Absorbing 3 HDFE groups                              F( 14, 6311) =
> 6.93
Statistics robust to heteroskedasticity              Prob > F        =
> 0.0000
                                                    R-squared       =
> 0.5667
                                                    Adj R-squared   =
> 0.5044
                                                    Within R-sq.    =
> 0.0032
Number of clusters (PERMNO) = 6,312                Root MSE       =
> 0.4717

                                                    (Std. err. adjusted for 6,312 clusters in
> PERMNO)

```

| | Coefficient | Robust std. err. | t | P> t | [95% conf. i |
|-----------------------|-------------|---------------------|-------|-------|--------------|
| LCITES3 > nterval] | | | | | |
| AUINFO3 > 6.711463 | 3.481369 | 1.647721 | 2.11 | 0.035 | .2512749 |
| SIZE > .0645776 | .046772 | .0090829 | 5.15 | 0.000 | .0289664 |
| RDASSET > .185983 | .0011986 | .0942614 | 0.01 | 0.990 | -.1835858 |
| AGE > .0317205 | -.0003624 | .016366 | -0.02 | 0.982 | -.0324452 |
| ROA > .0503463 | -.0073999 | .0294572 | -0.25 | 0.802 | -.0651461 |
| PPEASSET | .0496396 | .0537682 | 0.92 | 0.356 | -.0557643 |


```

> .1550435
  LEVERAGE | -.0972439 .0316208 -3.08 0.002 -.1592315 -
> .0352562
  CAPEXASSET | .1323956 .0754063 1.76 0.079 -.0154265
> .2802176
  TOBINQ | .0025272 .0019881 1.27 0.204 -.0013702
> .0064247
  KZINDEX | -.0642248 .0764711 -0.84 0.401 -.2141341
> .0856845
  HHI | -.0660154 .1055693 -0.63 0.532 -.2729671
> .1409363
  HHI2 | .05276 .0978689 0.54 0.590 -.1390963
> .2446163
  INSTIOWN | .0080901 .0221257 0.37 0.715 -.0352837
> .0514639
  ILLIQUIDITY | -.0047477 .0009158 -5.18 0.000 -.0065429 -
> .0029524
  cons | -.0547589 .0745181 -0.73 0.462 -.2008397
> .0913218

```

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs |
|-------------|------------|-------------|--------------|
| PERMNO | 6312 | 6312 | 0 * |
| YEAR | 16 | 1 | 15 |
| AU | 21 | 1 | 20 |

* = FE nested within cluster; treated as redundant for DoF computation

1425* output regression results

```

1426esttab reg_LPATENTS1 reg_LCITES1 reg_LPATENTS2 reg_LCITES2 reg_LPATE
> NTS3 reg_LCITES3 using "output\t8.rtf", cells(b(star fmt(3)) se(par
> fmt(3))) stats(r2_a N full, fmt(%9.3f %9.0g)) starlevels(* 0.1 ** 0.
> 05 *** 0.01) order(AUINFO3 $Control) modelwidth(12) replace
(output written to output\t8.rtf)

```

1427* Panel C. innovation value

```

1428use "innovation_value.dta", clear
( )

```

1429* summary

```

1430qui: reghdfe lSCOPE1 AUINFO3 $Controls if AU != ., cluster(PERMNO)

```

```
1431fsum SCOPE1 ORIGINALITY1 NOVELTY1 if e(sample) == 1, s(N mean sd p1
> p25 p50 p75 p99) format(%12.3f)
```

| Variable | | N | Mean | SD | P1 |
|--------------|--------|-------|-------|-------|-------|
| P25 | Median | P75 | P99 | | |
| SCOPE1 | 0.000 | 58058 | 0.042 | 0.132 | 0.000 |
| 0.000 | 0.000 | 0.000 | 0.625 | | |
| ORIGINALITY1 | 0.000 | 58058 | 0.044 | 0.158 | 0.000 |
| 0.000 | 0.000 | 0.000 | 0.752 | | |
| NOVELTY1 | 0.000 | 58058 | 0.308 | 0.959 | 0.000 |
| 0.000 | 0.000 | 0.000 | 5.287 | | |

```
1432* regression col (1) -(3)
1433foreach v of varlist lSCOPE1 lORIGINALITY1 lNOVELTY1{
  2. reghdfe `v' AUINFO3 $Controls, absorb(PERMNO YEAR AU) cluster(PER
  > MNO)
  3. est store reg_`v'
  4. }
(dropped 1045 singleton observations)
(MWFE estimator converged in 21 iterations)
```

| | | |
|---|---------------|-------|
| HDFE Linear regression | Number of obs | = |
| > 57,013 | | |
| Absorbing 3 HDFE groups | F(14, 6772) | = |
| > 6.61 | | |
| Statistics robust to heteroskedasticity | Prob > F | = |
| > 0.0000 | | |
| | R-squared | = |
| > 0.4904 | | |
| | Adj R-squared | = |
| > 0.4211 | | |
| | Within R-sq. | = |
| > 0.0054 | | |
| Number of clusters (PERMNO) | = | 6,773 |
| > 0.0817 | Root MSE | = |

```
(Std. err. adjusted for 6,773 clusters in
> PERMNO)
```

| lSCOPE1 | Coefficient | Robust std. err. | t | P> t | [95% conf. i |
|------------|-------------|---------------------|-------|-------|--------------|
| > nterval] | | | | | |
| AUINFO3 | 1.67628 | .3075419 | 5.45 | 0.000 | 1.073402 |
| > 2.279159 | | | | | |
| SIZE | .0040429 | .0019374 | 2.09 | 0.037 | .000245 |
| > .0078407 | | | | | |
| RDASSET | -.0249452 | .0143176 | -1.74 | 0.082 | -.0530122 |
| > .0031219 | | | | | |

| | | | | | | | |
|---|-------------|-----------|----------|-------|-------|-----------|---|
| | AGE | .0117019 | .0037291 | 3.14 | 0.002 | .0043917 | |
| > | .0190122 | | | | | | |
| | ROA | -.0231744 | .0055944 | -4.14 | 0.000 | -.0341412 | - |
| > | .0122077 | | | | | | |
| | PPEASSET | .0466119 | .0106545 | 4.37 | 0.000 | .0257258 | |
| > | .067498 | | | | | | |
| | LEVERAGE | -.0176823 | .0056501 | -3.13 | 0.002 | -.0287583 | - |
| > | .0066063 | | | | | | |
| | CAPEXASSET | .0045299 | .0156509 | 0.29 | 0.772 | -.0261508 | |
| > | .0352106 | | | | | | |
| | TOBINQ | .0006899 | .000336 | 2.05 | 0.040 | .0000312 | |
| > | .0013487 | | | | | | |
| | KZINDEX | -.0042072 | .0077137 | -0.55 | 0.585 | -.0193284 | |
| > | .010914 | | | | | | |
| | HHI | -.0246889 | .0224185 | -1.10 | 0.271 | -.0686362 | |
| > | .0192584 | | | | | | |
| | HHI2 | .0183372 | .0227648 | 0.81 | 0.421 | -.026289 | |
| > | .0629634 | | | | | | |
| | INSTIOWN | .0024236 | .004101 | 0.59 | 0.555 | -.0056157 | |
| > | .0104629 | | | | | | |
| | ILLIQUIDITY | -.0000613 | .0002313 | -0.26 | 0.791 | -.0005146 | |
| > | .0003921 | | | | | | |
| | cons | -.032104 | .0162272 | -1.98 | 0.048 | -.0639144 | - |
| > | .0002937 | | | | | | |

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs | |
|-------------|------------|-------------|--------------|---|
| PERMNO | 6773 | 6773 | 0 | * |
| YEAR | 18 | 1 | 17 | |
| AU | 21 | 1 | 20 | |

* = FE nested within cluster; treated as redundant for DoF computation
(dropped 1045 singleton observations)
(MWFE estimator converged in 21 iterations)

HDFE Linear regression

> 57,013

Absorbing 3 HDFE groups

> 6.91

Statistics robust to heteroskedasticity

> 0.0000

> 0.4725

> 0.4008

> 0.0070

Number of clusters (PERMNO) = 6,773

> 0.0950

Number of obs =

F(14, 6772) =

Prob > F =

R-squared =

Adj R-squared =

Within R-sq. =

Root MSE =

(Std. err. adjusted for 6,773 clusters in

> PERMNO)

| | Coefficient | Robust std. err. | t | P> t | [95% conf. i |
|----------------------------|-------------|---------------------|-------|-------|--------------|
| lORIGINALI~1 > nterval] | | | | | |
| AUINFO3 | 2.096555 | .3753624 | 5.59 | 0.000 | 1.360726 |
| > 2.832383 | | | | | |
| SIZE | .005427 | .0023036 | 2.36 | 0.019 | .0009113 |
| > .0099427 | | | | | |
| RDASSET | -.0074389 | .0164841 | -0.45 | 0.652 | -.0397529 |
| > .0248752 | | | | | |
| AGE | .0192951 | .0045834 | 4.21 | 0.000 | .0103102 |
| > .02828 | | | | | |
| ROA | -.0145534 | .0062199 | -2.34 | 0.019 | -.0267464 |
| > .0023605 | | | | | |
| PPEASSET | .0637814 | .0125516 | 5.08 | 0.000 | .0391764 |
| > .0883865 | | | | | |
| LEVERAGE | -.0202606 | .006704 | -3.02 | 0.003 | -.0334026 |
| > .0071187 | | | | | |
| CAPEXASSET | -.0038724 | .016915 | -0.23 | 0.819 | -.0370311 |
| > .0292864 | | | | | |
| TOBINQ | .0005682 | .0003648 | 1.56 | 0.119 | -.000147 |
| > .0012833 | | | | | |
| KZINDEX | -.0091113 | .0085669 | -1.06 | 0.288 | -.0259051 |
| > .0076826 | | | | | |
| HHI | -.0201248 | .0277703 | -0.72 | 0.469 | -.0745632 |
| > .0343137 | | | | | |
| HHI2 | .0102187 | .0282635 | 0.36 | 0.718 | -.0451867 |
| > .065624 | | | | | |
| INSTIOWN | .0088037 | .0050472 | 1.74 | 0.081 | -.0010903 |
| > .0186978 | | | | | |
| ILLIQUIDITY | -.0001361 | .0002345 | -0.58 | 0.562 | -.0005958 |
| > .0003236 | | | | | |
| cons | -.0708976 | .0196364 | -3.61 | 0.000 | -.1093911 |
| > .0324042 | | | | | |

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs |
|-------------|------------|-------------|--------------|
| PERMNO | 6773 | 6773 | 0 * |
| YEAR | 18 | 1 | 17 |
| AU | 21 | 1 | 20 |

* = FE nested within cluster; treated as redundant for DoF computation
 (dropped 1045 singleton observations)
 (MWFE estimator converged in 21 iterations)


```

HDFE Linear regression
> 57,013
Absorbing 3 HDFE groups
> 11.59
Statistics robust to heteroskedasticity
> 0.0000
> 0.5441
> 0.4822
> 0.0059
Number of clusters (PERMNO) = 6,773
> 0.2741

```

```

Number of obs =
F( 14, 6772) =
Prob > F =
R-squared =
Adj R-squared =
Within R-sq. =
Root MSE =

```

```

> PERMNO) (Std. err. adjusted for 6,773 clusters in

```

| lnNOVELTY1 > nterval] | Coefficient | Robust std. err. | t | P> t | [95% conf. i |
|--------------------------|-------------|---------------------|-------|-------|--------------|
| AUINFO3 | 5.242248 | .8488926 | 6.18 | 0.000 | 3.578152 |
| > 6.906345 | | | | | |
| SIZE | .0206701 | .004995 | 4.14 | 0.000 | .0108782 |
| > .0304619 | | | | | |
| RDASSET | .0491425 | .0450479 | 1.09 | 0.275 | -.0391655 |
| > .1374505 | | | | | |
| AGE | .014625 | .0098861 | 1.48 | 0.139 | -.0047548 |
| > .0340048 | | | | | |
| ROA | -.0313356 | .0158553 | -1.98 | 0.048 | -.062417 - |
| > .0002542 | | | | | |
| PPEASSET | .1496541 | .0275321 | 5.44 | 0.000 | .0956825 |
| > .2036257 | | | | | |
| LEVERAGE | -.0907009 | .0165343 | -5.49 | 0.000 | -.1231134 - |
| > .0582885 | | | | | |
| CAPEXASSET | .0264807 | .0435597 | 0.61 | 0.543 | -.05891 |
| > .1118715 | | | | | |
| TOBINQ | .0036908 | .000996 | 3.71 | 0.000 | .0017385 |
| > .0056432 | | | | | |
| KZINDEX | -.0212524 | .0297709 | -0.71 | 0.475 | -.0796128 |
| > .0371079 | | | | | |
| HHI | -.1167301 | .0586106 | -1.99 | 0.046 | -.2316253 - |
| > .0018348 | | | | | |
| HHI2 | .1034407 | .0596996 | 1.73 | 0.083 | -.0135893 |
| > .2204706 | | | | | |
| INSTIOWN | .016544 | .0117678 | 1.41 | 0.160 | -.0065246 |
| > .0396125 | | | | | |
| ILLIQUIDITY | -.0022464 | .0004881 | -4.60 | 0.000 | -.0032031 - |
| > .0012896 | | | | | |
| cons | -.0450784 | .0436629 | -1.03 | 0.302 | -.1306714 |
| > .0405146 | | | | | |

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs | |
|-------------|------------|-------------|--------------|---|
| PERMNO | 6773 | 6773 | 0 | * |
| YEAR | 18 | 1 | 17 | |
| AU | 21 | 1 | 20 | |

* = FE nested within cluster; treated as redundant for DoF computation

```

1434* regression col (4) -(6)
1435foreach v of varlist lSCOPE2 lORIGINALITY2 lNOVELTY2{
    2. reghdfe `v' AUINFO3 $Controls, absorb(PERMNO YEAR AU) cluster(PER
    > MNO)
    3. est store reg_`v'
    4. }
(dropped 1065 singleton observations)
(MWFE estimator converged in 21 iterations)

```

```

HDFE Linear regression                                Number of obs    =
> 53,853                                              F( 14, 6587) =
Absorbing 3 HDFE groups                               Prob > F         =
> 7.49                                              R-squared        =
Statistics robust to heteroskedasticity              Adj R-squared    =
> 0.0000                                           Within R-sq.     =
> 0.4513                                           Root MSE         =
> 0.3742
> 0.0062
Number of clusters (PERMNO) = 6,588
> 0.0780

```

(Std. err. adjusted for 6,588 clusters in
> PERMNO)

| | lSCOPE2 | Coefficient | Robust std. err. | t | P> t | [95% conf. i | |
|------------|------------|-------------|---------------------|-------|-------|--------------|---|
| > nterval] | | | | | | | |
| | AUINFO3 | 1.813883 | .3006903 | 6.03 | 0.000 | 1.224433 | |
| > 2.403334 | SIZE | .0036034 | .0018945 | 1.90 | 0.057 | -.0001104 | |
| > .0073172 | RDASSET | -.0501276 | .014163 | -3.54 | 0.000 | -.0778917 | - |
| > .0223636 | AGE | .0104228 | .0036875 | 2.83 | 0.005 | .0031942 | |
| > .0176514 | ROA | -.0237656 | .0054514 | -4.36 | 0.000 | -.0344522 | - |
| > .0130791 | PPEASSET | .0448034 | .0105828 | 4.23 | 0.000 | .0240577 | |
| > .0655491 | LEVERAGE | -.0127528 | .0057493 | -2.22 | 0.027 | -.0240233 | - |
| > .0014823 | CAPEXASSET | .0078218 | .0145701 | 0.54 | 0.591 | -.0207403 | |

```

> .0363838
   TOBINQ | .0013447 .0003425 3.93 0.000 .0006732
> .0020162
   KZINDEX | -.0060836 .00775 -0.78 0.432 -.0212761
> .0091089
   HHI | -.0059065 .0217251 -0.27 0.786 -.0484948
> .0366818
   HHI2 | -.0052954 .0222323 -0.24 0.812 -.0488779
> .0382871
   INSTIOWN | .0004052 .0039217 0.10 0.918 -.0072826
> .0080929
   ILLIQUIDITY | -.0000997 .0002206 -0.45 0.651 -.0005322
> .0003328
   cons | -.0341593 .0158687 -2.15 0.031 -.0652671 -
> .0030515

```

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs |
|-------------|------------|-------------|--------------|
| PERMNO | 6588 | 6588 | 0 * |
| YEAR | 17 | 1 | 16 |
| AU | 21 | 1 | 20 |

* = FE nested within cluster; treated as redundant for DoF computation
 (dropped 1065 singleton observations)
 (MWFE estimator converged in 21 iterations)

```

HDFE Linear regression
> 53,853
Absorbing 3 HDFE groups
> 7.17
Statistics robust to heteroskedasticity
> 0.0000
> 0.4412
> 0.3627
> 0.0077
Number of clusters (PERMNO) = 6,588
> 0.0916

```

Number of obs =

F(14, 6587) =

Prob > F =

R-squared =

Adj R-squared =

Within R-sq. =

Root MSE =

(Std. err. adjusted for 6,588 clusters in

> PERMNO)

| | Coefficient | Robust std. err. | t | P> t | [95% conf. i |
|----------------------------|-------------|---------------------|------|-------|--------------|
| 1ORIGINALI~2 > nterval] | | | | | |
| AUINFO3 | 2.208815 | .3899468 | 5.66 | 0.000 | 1.444392 |
| > 2.973237 | | | | | |
| SIZE | .0053531 | .0023017 | 2.33 | 0.020 | .000841 |
| > .0098652 | | | | | |


```

      RDASSET | -.0333421 .016986 -1.96 0.050 -.0666401 -
> .0000441
      AGE | .0193925 .0046213 4.20 0.000 .0103333
> .0284517
      ROA | -.0191768 .0060597 -3.16 0.002 -.0310559 -
> .0072978
      PPEASSET | .0638252 .0129222 4.94 0.000 .0384935
> .0891568
      LEVERAGE | -.0155308 .0066829 -2.32 0.020 -.0286315 -
> .0024301
      CAPEXASSET | .0026686 .0165853 0.16 0.872 -.0298439
> .0351811
      TOBINQ | .001545 .0003833 4.03 0.000 .0007936
> .0022963
      KZINDEX | -.0097468 .0095925 -1.02 0.310 -.0285512
> .0090577
      HHI | -.0007464 .0274828 -0.03 0.978 -.0546215
> .0531288
      HHI2 | -.0141711 .0284709 -0.50 0.619 -.0699834
> .0416411
      INSTIOWN | .0027368 .0048945 0.56 0.576 -.0068581
> .0123317
      ILLIQUIDITY | .0000145 .0002437 0.06 0.953 -.0004633
> .0004922
      cons | -.0767904 .0197256 -3.89 0.000 -.1154589 -
> .0381218

```

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs |
|-------------|------------|-------------|--------------|
| PERMNO | 6588 | 6588 | 0 * |
| YEAR | 17 | 1 | 16 |
| AU | 21 | 1 | 20 |

* = FE nested within cluster; treated as redundant for DoF computation
 (dropped 1065 singleton observations)
 (MWFE estimator converged in 21 iterations)

```

HDFE Linear regression
> 53,853
Absorbing 3 HDFE groups
> 7.34
Statistics robust to heteroskedasticity
> 0.0000

> 0.5365
> 0.4713
> 0.0044
Number of clusters (PERMNO) = 6,588
> 0.2638

```

Number of obs =
 F(14, 6587) =
 Prob > F =
 R-squared =
 Adj R-squared =
 Within R-sq. =
 Root MSE =

(Std. err. adjusted for 6,588 clusters in

> PERMNO)

| | Coefficient | Robust std. err. | t | P> t | [95% conf. i |
|------------------------------------|-------------|---------------------|-------|-------|--------------|
| lnNOVELTY2 [95% conf. interval] | | | | | |
| AUINFO3 | 3.395862 | .84047 | 4.04 | 0.000 | 1.748268 |
| > 5.043456 | | | | | |
| SIZE | .0170673 | .0051317 | 3.33 | 0.001 | .0070074 |
| > .0271271 | | | | | |
| RDASSET | -.0302095 | .0469303 | -0.64 | 0.520 | -.1222081 |
| > .061789 | | | | | |
| AGE | .0116399 | .0104014 | 1.12 | 0.263 | -.0087503 |
| > .0320301 | | | | | |
| ROA | -.0373567 | .0170994 | -2.18 | 0.029 | -.070877 |
| > .0038364 | | | | | |
| PPEASSET | .1333908 | .0288171 | 4.63 | 0.000 | .0769 |
| > .1898816 | | | | | |
| LEVERAGE | -.0854873 | .0171571 | -4.98 | 0.000 | -.1191208 |
| > .0518537 | | | | | |
| CAPEXASSET | -.0112278 | .0420994 | -0.27 | 0.790 | -.0937563 |
| > .0713007 | | | | | |
| TOBINQ | .0051537 | .0010597 | 4.86 | 0.000 | .0030763 |
| > .007231 | | | | | |
| KZINDEX | .0146321 | .0305125 | 0.48 | 0.632 | -.0451823 |
| > .0744465 | | | | | |
| HHI | -.0999318 | .0614019 | -1.63 | 0.104 | -.2202995 |
| > .0204358 | | | | | |
| HHI2 | .0838127 | .0615205 | 1.36 | 0.173 | -.0367875 |
| > .2044129 | | | | | |
| INSTIOWN | .0090031 | .0127625 | 0.71 | 0.481 | -.0160156 |
| > .0340218 | | | | | |
| ILLIQUIDITY | -.0010287 | .0004684 | -2.20 | 0.028 | -.0019469 |
| > .0001106 | | | | | |
| cons | -.012786 | .0446851 | -0.29 | 0.775 | -.1003833 |
| > .0748112 | | | | | |

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs |
|-------------|------------|-------------|--------------|
| PERMNO | 6588 | 6588 | 0 * |
| YEAR | 17 | 1 | 16 |
| AU | 21 | 1 | 20 |

* = FE nested within cluster; treated as redundant for DoF computation

```

1436* regression col (7) -(9)
1437foreach v of varlist lSCOPE3 lORIGINALITY3 lNOVELTY3{
  2. reghdfe `v' AUINFO3 $Controls, absorb(PERMNO YEAR AU) cluster(PER
  > MNO)
  3. est store reg_`v'
  4. }
(dropped 1134 singleton observations)
(MWFE estimator converged in 21 iterations)

```

```

HDFE Linear regression
> 50,582
Absorbing 3 HDFE groups
> 6.97
Statistics robust to heteroskedasticity
> 0.0000
> 0.4181
> 0.3344
> 0.0052
Number of clusters (PERMNO) = 6,312
> 0.0721

Number of obs =
F( 14, 6311) =
Prob > F =
R-squared =
Adj R-squared =
Within R-sq. =
Root MSE =

(Std. err. adjusted for 6,312 clusters in
> PERMNO)

```

| | | Coefficient | Robust std. err. | t | P> t | [95% conf. i |
|---|-----------------------|-------------|---------------------|-------|-------|--------------|
| | lSCOPE3 > nterval] | | | | | |
| | AUINFO3 | 1.325955 | .289169 | 4.59 | 0.000 | .7590851 |
| > | 1.892824 | | | | | |
| | SIZE | .0018919 | .0018277 | 1.04 | 0.301 | -.0016909 |
| > | .0054748 | | | | | |
| | RDASSET | -.0493414 | .0139974 | -3.53 | 0.000 | -.0767811 |
| > | .0219017 | | | | | |
| | AGE | .0088139 | .0036259 | 2.43 | 0.015 | .001706 |
| > | .0159218 | | | | | |
| | ROA | -.0266948 | .005042 | -5.29 | 0.000 | -.0365788 |
| > | .0168108 | | | | | |
| | PPEASSET | .0356229 | .0105336 | 3.38 | 0.001 | .0149736 |
| > | .0562723 | | | | | |
| | LEVERAGE | -.0027337 | .0056801 | -0.48 | 0.630 | -.0138685 |
| > | .0084011 | | | | | |
| | CAPEXASSET | .014341 | .0137851 | 1.04 | 0.298 | -.0126825 |
| > | .0413644 | | | | | |
| | TOBINQ | .0009432 | .0003333 | 2.83 | 0.005 | .0002898 |
| > | .0015967 | | | | | |
| | KZINDEX | -.017014 | .0078828 | -2.16 | 0.031 | -.032467 |
| > | .0015611 | | | | | |
| | HHI | .0357429 | .02159 | 1.66 | 0.098 | -.0065807 |
| > | .0780666 | | | | | |
| | HHI2 | -.0436307 | .0222108 | -1.96 | 0.050 | -.0871714 |
| > | -.00009 | | | | | |
| | INSTIOWN | .0006988 | .003624 | 0.19 | 0.847 | -.0064055 |

```

> .0078031
ILLIQUIDITY | -.0005495 .0001845 -2.98 0.003 -.0009111 -
> .0001878
cons | -.0280667 .015623 -1.80 0.072 -.058693
> .0025597

```

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs |
|-------------|------------|-------------|--------------|
| PERMNO | 6312 | 6312 | 0 * |
| YEAR | 16 | 1 | 15 |
| AU | 21 | 1 | 20 |

* = FE nested within cluster; treated as redundant for DoF computation
(dropped 1134 singleton observations)
(MWFE estimator converged in 21 iterations)

```

HDFE Linear regression
> 50,582
Absorbing 3 HDFE groups
> 7.31
Statistics robust to heteroskedasticity
> 0.0000
> 0.4044
> 0.3187
> 0.0071
Number of clusters (PERMNO) = 6,312
> 0.0869

```

Number of obs =

F(14, 6311) =

Prob > F =

R-squared =

Adj R-squared =

Within R-sq. =

Root MSE =

(Std. err. adjusted for 6,312 clusters in
> PERMNO)

| lORIGINALI~3 > nterval] | Coefficient | Robust std. err. | t | P> t | [95% conf. i |
|----------------------------|-------------|---------------------|-------|-------|--------------|
| AUINFO3 | 1.772288 | .4017128 | 4.41 | 0.000 | .9847939 |
| > 2.559781 | | | | | |
| SIZE | .0037035 | .0022474 | 1.65 | 0.099 | -.0007022 |
| > .0081091 | | | | | |
| RDASSET | -.0328841 | .0170222 | -1.93 | 0.053 | -.0662535 |
| > .0004852 | | | | | |
| AGE | .017245 | .0045839 | 3.76 | 0.000 | .008259 |
| > .0262309 | | | | | |
| ROA | -.0275097 | .0058977 | -4.66 | 0.000 | -.0390711 - |
| > .0159482 | | | | | |
| PPEASSET | .0578668 | .0126946 | 4.56 | 0.000 | .0329812 |
| > .0827525 | | | | | |
| LEVERAGE | -.0145807 | .0066501 | -2.19 | 0.028 | -.0276171 - |
| > .0015444 | | | | | |

| | | | | | | |
|-------------|-----------|----------|-------|-------|-----------|---|
| CAPEXASSET | .0194038 | .0159632 | 1.22 | 0.224 | -.0118896 | |
| > .0506971 | | | | | | |
| TOBINQ | .0013346 | .0003721 | 3.59 | 0.000 | .0006051 | |
| > .0020641 | | | | | | |
| KZINDEX | -.0203456 | .0097497 | -2.09 | 0.037 | -.0394584 | - |
| > .0012328 | | | | | | |
| HHI | .048374 | .0268276 | 1.80 | 0.071 | -.0042172 | |
| > .1009652 | | | | | | |
| HHI2 | -.0576627 | .0279601 | -2.06 | 0.039 | -.112474 | - |
| > .0028515 | | | | | | |
| INSTIOWN | .003434 | .004532 | 0.76 | 0.449 | -.0054502 | |
| > .0123183 | | | | | | |
| ILLIQUIDITY | -.000581 | .0002037 | -2.85 | 0.004 | -.0009803 | - |
| > .0001818 | | | | | | |
| cons | -.0699031 | .0195118 | -3.58 | 0.000 | -.1081528 | - |
| > .0316535 | | | | | | |

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs | |
|-------------|------------|-------------|--------------|---|
| PERMNO | 6312 | 6312 | 0 | * |
| YEAR | 16 | 1 | 15 | |
| AU | 21 | 1 | 20 | |

* = FE nested within cluster; treated as redundant for DoF computation
(dropped 1134 singleton observations)
(MWFE estimator converged in 21 iterations)

HDFE Linear regression
> 50,582
Absorbing 3 HDFE groups
> 4.37
Statistics robust to heteroskedasticity
> 0.0000

> 0.5373
> 0.4708
> 0.0026
Number of clusters (PERMNO) = 6,312
> 0.2492

Number of obs =

F(14, 6311) =

Prob > F =

R-squared =

Adj R-squared =

Within R-sq. =

Root MSE =

(Std. err. adjusted for 6,312 clusters in
> PERMNO)

| | Coefficient | Robust std. err. | t | P> t | [95% conf. i |
|------------|-------------|---------------------|------|-------|--------------|
| lnNOVELTY3 | | | | | |
| > nterval] | | | | | |
| AUINFO3 | 3.252617 | .8378344 | 3.88 | 0.000 | 1.610177 |
| > 4.895057 | | | | | |
| SIZE | .0115541 | .0052395 | 2.21 | 0.027 | .001283 |


```

> .0218253
RDASSET | -.0225307 .0465915 -0.48 0.629 -.1138658
> .0688045
AGE | .0171376 .0106397 1.61 0.107 -.0037198
> .0379949
ROA | -.0472452 .0164091 -2.88 0.004 -.0794127 -
> .0150777
PPEASSET | .098201 .0305886 3.21 0.001 .0382369
> .1581651
LEVERAGE | -.055569 .0177072 -3.14 0.002 -.0902812 -
> .0208569
CAPEXASSET | -.0008623 .0401741 -0.02 0.983 -.0796172
> .0778926
TOBINQ | .0021601 .0011136 1.94 0.052 -.000023
> .0043431
KZINDEX | -.045182 .0346083 -1.31 0.192 -.113026
> .0226619
HHI | -.0571961 .0613846 -0.93 0.351 -.1775308
> .0631386
HHI2 | .0371702 .0616833 0.60 0.547 -.08375
> .1580903
INSTIOWN | .0060937 .0132232 0.46 0.645 -.0198283
> .0320156
ILLIQUIDITY | -.0009957 .0004416 -2.25 0.024 -.0018615
> -.00013
cons | -.0008321 .0456972 -0.02 0.985 -.0904141
> .0887499

```

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs | |
|-------------|------------|-------------|--------------|---|
| PERMNO | 6312 | 6312 | 0 | * |
| YEAR | 16 | 1 | 15 | |
| AU | 21 | 1 | 20 | |

* - FE nested within cluster; treated as redundant for DoF computation

1438* output regression results

```

1439esttab reg lSCOPE1 reg lORIGINALITY1 reg lNOVELTY1 using "output\t8_
> a.rtf", cells(b(star fmt(3)) se(par fmt(3))) stats(r2_a N full, fmt(
> %9.3f %9.0g)) starlevels(* 0.1 ** 0.05 *** 0.01) order(AUINFO3 $Cont
> rol) modelwidth(12) replace
(output written to output\t8_a.rtf)

```

```

1440esttab reg lSCOPE2 reg lORIGINALITY2 reg lNOVELTY2 using "output\t8_
> b.rtf", cells(b(star fmt(3)) se(par fmt(3))) stats(r2_a N full, fmt(
> %9.3f %9.0g)) starlevels(* 0.1 ** 0.05 *** 0.01) order(AUINFO3 $Cont
> rol) modelwidth(12) replace
(output written to output\t8_b.rtf)

1441esttab reg lSCOPE3 reg lORIGINALITY3 reg lNOVELTY3 using "output\t8_
> c.rtf", cells(b(star fmt(3)) se(par fmt(3))) stats(r2_a N full, fmt(
> %9.3f %9.0g)) starlevels(* 0.1 ** 0.05 *** 0.01) order(AUINFO3 $Cont
> rol) modelwidth(12) replace
(output written to output\t8_c.rtf)

1442
1443* figure oal
1444use "figureOAl.dta"

1445qui: reghdfe lcitation (YEAR2000 YEAR2001 YEAR2003 YEAR2004 YEAR2005
> YEAR2006 YEAR2002)##TREAT $Control2, absorb(permnol##permno2 permno
> 1##year permno2##year) cluster(permnol permno2)

1446est store reg1

1447coefplot reg1 , keep(1.YEAR2000#1.TREAT 1.YEAR2001#1.TREAT 1.YEAR200
> 3#1.TREAT 1.YEAR2004#1.TREAT 1.YEAR2005#1.TREAT 1.YEAR2006#1.TREAT 1
> .YEAR2002#1.TREAT) ///
> order(1.YEAR2000#1.TREAT 1.YEAR2001#1.TREAT 1.YEAR2002#1.TREAT 1.YEA
> R2003#1.TREAT 1.YEAR2004#1.TREAT 1.YEAR2005#1.TREAT 1.YEAR2006#1.TRE
> AT ) vertical coeflabels(1.YEAR2000#1.TREAT = "2000" 1.YEAR2001#1.TR
> EAT ="2001" 1.YEAR2002#1.TREAT= "2002" 1.YEAR2003#1.TREAT ="2003"
> 1.YEAR2004#1.TREAT ="2004" 1.YEAR2005#1.TREAT ="2005" 1.YEAR2006#1
> .TREAT ="2006") levels(90) baselevels yline(0, lwidth (medthin) lcol
> or(cranberry)) xline(3, lpattern(dash) lcolor(black)) omitted xtitl
> e("Year") legend(order(2 "Point Estimate" 1 "90%CI") pos(6) rows(1))
> mcolor(cranberry) ciopts(recast(rcap) lcol( "0 91 150") ) mlabel(s
> tring(@b, "%9.3f")) mlabposition(3) mlabgap(5pt) mlabcolor("0 91 150
> ") grid(none) ylabel(-0.1(0.05)0.2) yscale(range(-0.1(0.05)0.2))

1448graph export "output\oal.png", replace
file output\oal.png saved as PNG format

1449
1450* OA 1. summary by industry
1451use "main.dta", clear

1452g sic2_1 = int(siccd1/100)

```

```
1453g sic2_2 = int(siccd2/100)
```

```
1454* panel A
```

```
1455groups sic2_1, order(h) select(30)
```

| sic2_1 | Freq. | Percent | %<= |
|--------|--------|---------|-------|
| 36 | 332376 | 21.20 | 21.20 |
| 35 | 252258 | 16.09 | 37.28 |
| 38 | 190677 | 12.16 | 49.44 |
| 28 | 169476 | 10.81 | 60.25 |
| 73 | 161311 | 10.29 | 70.53 |
| 37 | 111730 | 7.12 | 77.66 |
| 48 | 49253 | 3.14 | 80.80 |
| 26 | 24226 | 1.54 | 82.34 |
| 87 | 23336 | 1.49 | 83.83 |
| 34 | 22670 | 1.45 | 85.28 |
| 30 | 18391 | 1.17 | 86.45 |
| 39 | 17355 | 1.11 | 87.56 |
| 99 | 15971 | 1.02 | 88.58 |
| 13 | 15757 | 1.00 | 89.58 |
| 20 | 15635 | 1.00 | 90.58 |
| 33 | 13721 | 0.87 | 91.45 |
| 29 | 13631 | 0.87 | 92.32 |
| 25 | 11011 | 0.70 | 93.02 |
| 60 | 9940 | 0.63 | 93.66 |
| 50 | 9299 | 0.59 | 94.25 |
| 63 | 8318 | 0.53 | 94.78 |
| 27 | 6709 | 0.43 | 95.21 |
| 78 | 6407 | 0.41 | 95.62 |
| 61 | 5763 | 0.37 | 95.99 |
| 79 | 5503 | 0.35 | 96.34 |
| 49 | 5486 | 0.35 | 96.69 |
| 53 | 5388 | 0.34 | 97.03 |
| 51 | 5362 | 0.34 | 97.37 |
| 62 | 4831 | 0.31 | 97.68 |
| 32 | 4619 | 0.29 | 97.97 |

```
1456//asdoc groups sic2_1, order(h) select(30) sep(0) save(output\oal_a.
> rtf) replace
```



```
1457* panel B
1458groups sic2_2, order(h) select(30)
```

| sic2_2 | Freq. | Percent | %<= |
|--------|--------|---------|-------|
| 36 | 325686 | 20.77 | 20.77 |
| 35 | 239887 | 15.30 | 36.07 |
| 28 | 188756 | 12.04 | 48.10 |
| 38 | 185210 | 11.81 | 59.91 |
| 73 | 166729 | 10.63 | 70.55 |
| 37 | 103622 | 6.61 | 77.15 |
| 48 | 40301 | 2.57 | 79.72 |
| 34 | 25162 | 1.60 | 81.33 |
| 87 | 24486 | 1.56 | 82.89 |
| 26 | 22798 | 1.45 | 84.34 |
| 29 | 20680 | 1.32 | 85.66 |
| 20 | 19787 | 1.26 | 86.92 |
| 13 | 19330 | 1.23 | 88.16 |
| 30 | 17951 | 1.14 | 89.30 |
| 33 | 17636 | 1.12 | 90.43 |
| 39 | 13318 | 0.85 | 91.27 |
| 50 | 12692 | 0.81 | 92.08 |
| 63 | 11281 | 0.72 | 92.80 |
| 49 | 9422 | 0.60 | 93.40 |
| 25 | 9257 | 0.59 | 93.99 |
| 27 | 8168 | 0.52 | 94.52 |
| 99 | 6041 | 0.39 | 94.90 |
| 32 | 5857 | 0.37 | 95.27 |
| 60 | 5723 | 0.36 | 95.64 |
| 51 | 5161 | 0.33 | 95.97 |
| 78 | 4649 | 0.30 | 96.26 |
| 62 | 4549 | 0.29 | 96.55 |
| 67 | 4483 | 0.29 | 96.84 |
| 61 | 3675 | 0.23 | 97.07 |
| 53 | 3542 | 0.23 | 97.30 |

```
1459//asdoc groups sic2_2, order(h) select(30) save(output\oal_b.rtf) re
> place
1460* panel C
```



```
1461 groups sic2_1 sic2_2, order(h) select(30)
```

| sic2_1 | sic2_2 | Freq. | Percent |
|--------|--------|--------|---------|
| 36 | 36 | 120862 | 7.71 |
| 28 | 28 | 64182 | 4.09 |
| 36 | 35 | 53416 | 3.41 |
| 35 | 36 | 52817 | 3.37 |
| 35 | 35 | 46834 | 2.99 |
| 73 | 73 | 45847 | 2.92 |
| 38 | 38 | 43727 | 2.79 |
| 36 | 38 | 34294 | 2.19 |
| 73 | 36 | 34169 | 2.18 |
| 36 | 73 | 33182 | 2.12 |
| 38 | 36 | 31520 | 2.01 |
| 35 | 73 | 27844 | 1.78 |
| 38 | 35 | 27820 | 1.77 |
| 35 | 38 | 27773 | 1.77 |
| 38 | 28 | 27228 | 1.74 |
| 73 | 35 | 25282 | 1.61 |
| 37 | 36 | 21621 | 1.38 |
| 28 | 38 | 21423 | 1.37 |
| 36 | 37 | 21080 | 1.34 |
| 35 | 28 | 19863 | 1.27 |
| 37 | 35 | 18760 | 1.20 |
| 35 | 37 | 18350 | 1.17 |
| 36 | 28 | 16436 | 1.05 |
| 28 | 35 | 16408 | 1.05 |
| 38 | 37 | 12975 | 0.83 |
| 37 | 38 | 12903 | 0.82 |
| 48 | 36 | 12824 | 0.82 |
| 28 | 36 | 12757 | 0.81 |
| 37 | 37 | 12295 | 0.78 |
| 73 | 38 | 11516 | 0.73 |

```
1462 //asdoc groups sic2_1 sic2_2, order(h) select(30) save(output\oa1_c.
    > rtf) replace
1463
1464 * OA 2. citing * cited fe
```

```
1465use "main.dta", clear
```

```
1466* col (1)
```

```
1467reghdfe lcitation I_shared_office , absorb(permnol##year permno2##ye
> ar permno1##permno2) cluster(permnol permno2)
(dropped 16968 singleton observations)
(MWFE estimator converged in 11 iterations)
```

```
HDFE Linear regression
```

```
> ,551,199
```

```
Absorbing 3 HDFE groups
```

```
> 5.24
```

```
Statistics robust to heteroskedasticity
```

```
> 0.0221
```

```
> 0.7085
```

```
> 0.6658
```

```
Number of clusters (permno1) = 2,230
```

```
> 0.0000
```

```
Number of clusters (permno2) = 2,912
```

```
> 0.4452
```

```
Number of obs = 1
```

```
F( 1, 2229) =
```

```
Prob > F =
```

```
R-squared =
```

```
Adj R-squared =
```

```
Within R-sq. =
```

```
Root MSE =
```

```
(Std. err. adjusted for 2,230 clusters in permn
```

```
> o1 permno2)
```

| | Coefficient | Robust std. err. | t | P> t | [95% conf |
|-----------------|-------------|---------------------|---------|-------|-----------|
| lcitation | | | | | |
| > . interval] | | | | | |
| I_shared_office | .0186705 | .0081531 | 2.29 | 0.022 | .0026821 |
| > .034659 | | | | | |
| cons | .3748595 | .0001389 | 2699.08 | 0.000 | .3745872 |
| > .3751319 | | | | | |

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs |
|-----------------|------------|-------------|--------------|
| permno1#year | 18646 | 18646 | 0 * |
| permno2#year | 31622 | 31622 | 0 * |
| permno1#permno2 | 148005 | 148005 | 0 * |

* = FE nested within cluster; treated as redundant for DoF computation

1468est store reg1

1469* col (2)

1470reghdfe lcitation I_shared_office \$Control1 , absorb(permnol##year p
 > ermno2##year permno1##permno2) cluster(permnol permno2)
 (dropped 16968 singleton observations)
 (MWFE estimator converged in 12 iterations)
 Warning: VCV matrix was non-positive semi-definite; adjustment from Ca
 > meron, Gelbach & Miller applied.

| | | | |
|---|---------------|---|---|
| HDFE Linear regression | Number of obs | = | 1 |
| > ,551,199 | F(10, 2229) | = | |
| Absorbing 3 HDFE groups | Prob > F | = | |
| > 40.01 | R-squared | = | |
| Statistics robust to heteroskedasticity | Adj R-squared | = | |
| > 0.0000 | Within R-sq. | = | |
| > 0.7101 | Root MSE | = | |
| > 0.6676 | | | |
| Number of clusters (permno1) = | 2,230 | | |
| > 0.0053 | | | |
| Number of clusters (permno2) = | 2,912 | | |
| > 0.4440 | | | |

(Std. err. adjusted for 2,230 clusters in per

> mno1 permno2)

| | Coefficient | Robust std. err. | t | P> t | [95% co |
|-----------------|-------------|---------------------|-------|-------|----------|
| lcitation | | | | | |
| > n | | | | | |
| > f. interval] | | | | | |
| I_shared_office | .0144521 | .0078734 | 1.84 | 0.067 | -.000987 |
| > 9 | | | | | |
| > .029892 | | | | | |
| CommonAnalyst | .0157885 | .0015168 | 10.41 | 0.000 | .01281 |
| > 4 | | | | | |
| > .018763 | | | | | |
| CommonAlliance | -.0730927 | .0296622 | -2.46 | 0.014 | -.131261 |
| > 1 | | | | | |
| > -.0149242 | | | | | |
| CommonInventor | .0023873 | .0011256 | 2.12 | 0.034 | .000180 |
| > 1 | | | | | |
| > .0045946 | | | | | |
| CommonOwnership | .0003827 | .0006986 | 0.55 | 0.584 | -.000987 |
| > 4 | | | | | |
| > .0017527 | | | | | |
| TechSimilarity | .247305 | .018292 | 13.52 | 0.000 | .211433 |
| > 9 | | | | | |
| > .2831762 | | | | | |
| RelativeAnalyst | -1.366393 | .5819255 | -2.35 | 0.019 | -2.50756 |
| > 6 | | | | | |
| > -.2252203 | | | | | |
| RelativeSize | -.0230502 | .0069853 | -3.30 | 0.001 | -.036748 |
| > 7 | | | | | |

```

>      -.0093518
RelativePatentS~k | -.0270349   .008757   -3.09   0.002   -.044207
> 5
>      -.0098622
RelativeCitation | -.005649   .0007782   -7.26   0.000   -.007175
> 1
>      -.0041229
      _cons | .3348339   .0037107   90.23   0.000   .32755
> 7
>      .3421108

```

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs |
|-----------------|------------|-------------|--------------|
| permno1#year | 18646 | 18646 | 0 * |
| permno2#year | 31622 | 31622 | 0 * |
| permno1#permno2 | 148005 | 148005 | 0 * |

* = FE nested within cluster; treated as redundant for DoF computation

1471est store reg2

1472* output regression results

```

1473esttab reg1 reg2 using "output\oa2.rtf", cells(b(star fmt(3)) se(par
>   fmt(3))) stats(r2_a N_full, fmt(%9.3f %9.0g)) starlevels(* 0.1 ** 0
>   .05 *** 0.01) order(I_shared_office $Control1) modelwidth(12) replac
>   e
(output written to output\oa2.rtf)

```

1474

1475* OA 3. control product similarity

1476use "main.dta", clear

1477* merge with TNIC data

1478capt drop score

1479g PERMNO1 = permno1

1480g PERMNO2 = permno2

1481g _score = .

(1,568,167 missing values generated)


```

1482forval yr = 2000/2018{
    2. di "****`yr`****"
    3. * process year-over-year to avoid memory overflow
1483merge 1:1 PERMNO1 PERMNO2 YEAR using "TNIC\tnic_all_data\tnicall`yr'
    > .txt.dta"
    4. drop if _merge == 2
    5. replace _score = score if _merge == 3
    6. drop _merge
    7. drop _score
    8. }
***2000***
(variable PERMNO1 was float, now double to accommodate using data's
values)
(variable PERMNO2 was float, now double to accommodate using data's
values)

```

| Result | Number of obs | |
|-------------|-------------------|--------------|
| Not matched | 47,298,606 | |
| from master | 1,451,037 | (_merge==1) |
| from using | 45,847,569 | (_merge==2) |
| Matched | 117,130 | (_merge==3) |

(45,847,569 observations deleted)

(117,130 real changes made)

2001

| Result | Number of obs | |
|-------------|-------------------|--------------|
| Not matched | 39,845,204 | |
| from master | 1,454,670 | (_merge==1) |
| from using | 38,390,534 | (_merge==2) |
| Matched | 113,497 | (_merge==3) |

(38,390,534 observations deleted)

(113,497 real changes made)

2002

| Result | Number of obs | |
|-------------|-------------------|--------------|
| Not matched | 34,170,294 | |
| from master | 1,462,558 | (_merge==1) |
| from using | 32,707,736 | (_merge==2) |
| Matched | 105,609 | (_merge==3) |

(32,707,736 observations deleted)

(105,609 real changes made)

2003

| Result | Number of obs | |
|-------------|-------------------|--------------|
| Not matched | 29,926,922 | |
| from master | 1,466,916 | (_merge==1) |
| from using | 28,460,006 | (_merge==2) |

| | | |
|---------|----------------|--------------|
| Matched | 101,251 | (_merge==3) |
|---------|----------------|--------------|

(28,460,006 observations deleted)
 (101,251 real changes made)
 2004

| | | |
|-------------|-------------------|--------------|
| Result | Number of obs | |
| Not matched | 28,735,500 | |
| from master | 1,467,787 | (_merge==1) |
| from using | 27,267,713 | (_merge==2) |
| Matched | 100,380 | (_merge==3) |

(27,267,713 observations deleted)
 (100,380 real changes made)
 2005

| | | |
|-------------|-------------------|--------------|
| Result | Number of obs | |
| Not matched | 27,640,518 | |
| from master | 1,470,636 | (_merge==1) |
| from using | 26,169,882 | (_merge==2) |
| Matched | 97,531 | (_merge==3) |

(26,169,882 observations deleted)
 (97,531 real changes made)
 2006

| | | |
|-------------|-------------------|--------------|
| Result | Number of obs | |
| Not matched | 26,907,869 | |
| from master | 1,476,496 | (_merge==1) |
| from using | 25,431,373 | (_merge==2) |
| Matched | 91,671 | (_merge==3) |

(25,431,373 observations deleted)
 (91,671 real changes made)
 2007

| | | |
|-------------|-------------------|--------------|
| Result | Number of obs | |
| Not matched | 26,179,844 | |
| from master | 1,480,433 | (_merge==1) |
| from using | 24,699,411 | (_merge==2) |
| Matched | 87,734 | (_merge==3) |

(24,699,411 observations deleted)
 (87,734 real changes made)
 2008

| Result | Number of obs | |
|-------------|-------------------|--------------|
| Not matched | 23,646,200 | |
| from master | 1,485,172 | (_merge==1) |
| from using | 22,161,028 | (_merge==2) |
| Matched | 82,995 | (_merge==3) |

(22,161,028 observations deleted)
(82,995 real changes made)
2009

| Result | Number of obs | |
|-------------|-------------------|--------------|
| Not matched | 21,488,131 | |
| from master | 1,492,324 | (_merge==1) |
| from using | 19,995,807 | (_merge==2) |
| Matched | 75,843 | (_merge==3) |

(19,995,807 observations deleted)
(75,843 real changes made)
2010

| Result | Number of obs | |
|-------------|-------------------|--------------|
| Not matched | 20,198,028 | |
| from master | 1,494,617 | (_merge==1) |
| from using | 18,703,411 | (_merge==2) |
| Matched | 73,550 | (_merge==3) |

(18,703,411 observations deleted)
(73,550 real changes made)
2011

| Result | Number of obs | |
|-------------|-------------------|--------------|
| Not matched | 19,047,583 | |
| from master | 1,497,185 | (_merge==1) |
| from using | 17,550,398 | (_merge==2) |
| Matched | 70,982 | (_merge==3) |

(17,550,398 observations deleted)
(70,982 real changes made)
2012

| Result | Number of obs | |
|-------------|-------------------|--------------|
| Not matched | 18,212,788 | |
| from master | 1,500,020 | (_merge==1) |
| from using | 16,712,768 | (_merge==2) |

| | | |
|---------|---------------|--------------|
| Matched | 68,147 | (_merge==3) |
|---------|---------------|--------------|

(16,712,768 observations deleted)
(68,147 real changes made)
2013

| | | |
|-------------|-------------------|--------------|
| Result | Number of obs | |
| Not matched | 18,822,416 | |
| from master | 1,502,217 | (_merge==1) |
| from using | 17,320,199 | (_merge==2) |
| Matched | 65,950 | (_merge==3) |

(17,320,199 observations deleted)
(65,950 real changes made)
2014

| | | |
|-------------|-------------------|--------------|
| Result | Number of obs | |
| Not matched | 20,068,825 | |
| from master | 1,506,374 | (_merge==1) |
| from using | 18,562,451 | (_merge==2) |
| Matched | 61,793 | (_merge==3) |

(18,562,451 observations deleted)
(61,793 real changes made)
2015

| | | |
|-------------|-------------------|--------------|
| Result | Number of obs | |
| Not matched | 19,842,054 | |
| from master | 1,511,861 | (_merge==1) |
| from using | 18,330,193 | (_merge==2) |
| Matched | 56,306 | (_merge==3) |

(18,330,193 observations deleted)
(56,306 real changes made)
2016

| | | |
|-------------|-------------------|--------------|
| Result | Number of obs | |
| Not matched | 18,683,751 | |
| from master | 1,518,161 | (_merge==1) |
| from using | 17,165,590 | (_merge==2) |
| Matched | 50,006 | (_merge==3) |

(17,165,590 observations deleted)
(50,006 real changes made)
2017

| Result | Number of obs | |
|-------------|-------------------|--------------|
| Not matched | 18,338,740 | |
| from master | 1,522,603 | (_merge==1) |
| from using | 16,816,137 | (_merge==2) |
| Matched | 45,564 | (_merge==3) |

(16,816,137 observations deleted)
(45,564 real changes made)
2018

| Result | Number of obs | |
|-------------|-------------------|--------------|
| Not matched | 18,180,988 | |
| from master | 1,526,463 | (_merge==1) |
| from using | 16,654,525 | (_merge==2) |
| Matched | 41,704 | (_merge==3) |

(16,654,525 observations deleted)
(41,704 real changes made)

1484g score = _score
(60,524 missing values generated)

1485* col (1)
1486* replace control var: tech sim -> prod sim
1487reghdfe lcitation I_shared office CommonAnalyst CommonAlliance Commo
> nInventor CommonOwnership RelativeAnalyst RelativeSize RelativePaten
> tStock RelativeCitation score , absorb(permnol##year permno2##year)
> cluster(permnol permno2)
(dropped 2937 singleton observations)
(MWFE estimator converged in 11 iterations)
Warning: VCV matrix was non-positive semi-definite; adjustment from Ca
> meron, Gelbach & Miller applied.

| | | | |
|---|---------------|--------------|---|
| HDFE Linear regression | Number of obs | = | 1 |
| > ,504,706 | | | |
| Absorbing 2 HDFE groups | F(10, 2677) | = | |
| > 95.53 | | | |
| Statistics robust to heteroskedasticity | Prob > F | = | |
| > 0.0000 | | | |
| | R-squared | = | |
| > 0.3409 | | | |
| | Adj R-squared | = | |
| > 0.3180 | | | |
| Number of clusters (permno1) = | 2,678 | Within R-sq. | = |
| > 0.1291 | | | |
| Number of clusters (permno2) = | 3,152 | Root MSE | = |
| > 0.6331 | | | |

(Std. err. adjusted for 2,678 clusters in per

> mno1 permno2)

| | lcitation | Coefficient | Robust std. err. | t | P> t | [95% co |
|----------------|-------------------|-------------|---------------------|--------|-------|----------|
| > n | | | | | | |
| > f. interval] | | | | | | |
| | I_shared_office | .0472267 | .0176129 | 2.68 | 0.007 | .012690 |
| > 5 | | | | | | |
| > | .0817629 | | | | | |
| | CommonAnalyst | .0570999 | .0027389 | 20.85 | 0.000 | .051729 |
| > 2 | | | | | | |
| > | .0624706 | | | | | |
| | CommonAlliance | .3276191 | .0853478 | 3.84 | 0.000 | .160264 |
| > 8 | | | | | | |
| > | .4949735 | | | | | |
| | CommonInventor | .0105311 | .0037292 | 2.82 | 0.005 | .003218 |
| > 7 | | | | | | |
| > | .0178434 | | | | | |
| | CommonOwnership | .0092598 | .0056007 | 1.65 | 0.098 | -.001722 |
| > 4 | | | | | | |
| > | .0202419 | | | | | |
| | RelativeAnalyst | -10.67967 | 1.3361 | -7.99 | 0.000 | -13.2995 |
| > 6 | | | | | | |
| > | -8.059774 | | | | | |
| | RelativeSize | -.0310323 | .0085347 | -3.64 | 0.000 | -.047767 |
| > 5 | | | | | | |
| > | -.014297 | | | | | |
| | RelativePatents~k | -.0627932 | .0073032 | -8.60 | 0.000 | -.077113 |
| > 7 | | | | | | |
| > | -.0484727 | | | | | |
| | RelativeCitation | -.0145323 | .0011311 | -12.85 | 0.000 | -.016750 |
| > 1 | | | | | | |
| > | -.0123144 | | | | | |
| | score | 3.679936 | .1942242 | 18.95 | 0.000 | 3.29909 |
| > 1 | | | | | | |
| > | 4.06078 | | | | | |
| | _cons | .2109425 | .0085444 | 24.69 | 0.000 | .194188 |
| > 3 | | | | | | |
| > | .2276967 | | | | | |

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs |
|--------------|------------|-------------|--------------|
| permno1#year | 18915 | 18915 | 0 * |
| permno2#year | 31498 | 31498 | 0 * |

* = FE nested within cluster; treated as redundant for DoF computation

1488est store reg1

1489* col (2)

1490reghdfe lcitation I shared office \$Controll score, absorb(permnol##y
 > ear permno2##year) cluster(permnol permno2)
 (dropped 2937 singleton observations)
 (MWFE estimator converged in 11 iterations)
 Warning: VCV matrix was non-positive semi-definite; adjustment from Ca
 > meron, Gelbach & Miller applied.

| | | | |
|---|---------------|---|---|
| HDFE Linear regression | Number of obs | = | 1 |
| > ,504,706 | F(11, 2677) | = | |
| Absorbing 2 HDFE groups | Prob > F | = | |
| > 102.98 | R-squared | = | |
| Statistics robust to heteroskedasticity | Adj R-squared | = | |
| > 0.0000 | Within R-sq. | = | |
| > 0.3755 | Root MSE | = | |
| > 0.3539 | | | |
| Number of clusters (permno1) = | 2,678 | | |
| > 0.1749 | | | |
| Number of clusters (permno2) = | 3,152 | | |
| > 0.6162 | | | |

(Std. err. adjusted for 2,678 clusters in per

> mnol permno2)

| | lcitation | Coefficient | Robust std. err. | t | P> t | [95% co |
|-----------------|-----------|-------------|---------------------|-------|-------|----------|
| > n | | | | | | |
| > f. interval] | | | | | | |
| I_shared_office | | .0366107 | .015945 | 2.30 | 0.022 | .005344 |
| > 9 | | | | | | |
| > .0678765 | | | | | | |
| CommonAnalyst | | .045721 | .0023237 | 19.68 | 0.000 | .041164 |
| > 5 | | | | | | |
| > .0502775 | | | | | | |
| CommonAlliance | | .2612753 | .0720301 | 3.63 | 0.000 | .120035 |
| > 1 | | | | | | |
| > .4025155 | | | | | | |
| CommonInventor | | .009249 | .0034213 | 2.70 | 0.007 | .002540 |
| > 4 | | | | | | |
| > .0159576 | | | | | | |
| CommonOwnership | | .0093073 | .0057044 | 1.63 | 0.103 | -.001878 |
| > 2 | | | | | | |
| > .0204928 | | | | | | |
| TechSimilarity | | .8555021 | .0514924 | 16.61 | 0.000 | .754533 |
| > 2 | | | | | | |
| > .956471 | | | | | | |
| RelativeAnalyst | | -10.08009 | 1.205375 | -8.36 | 0.000 | -12.4436 |
| > 5 | | | | | | |
| > -7.716526 | | | | | | |
| RelativeSize | | -.0338099 | .0081 | -4.17 | 0.000 | -.049692 |
| > 8 | | | | | | |


```

>      -.0179271
RelativePatentS~k | -.0556317   .0077652   -7.16   0.000   -.07085
> 8
>      -.0404053
RelativeCitation | -.0157406   .0012401  -12.69   0.000   -.018172
> 3
>      -.0133088
          score |   2.139432   .1323559   16.16   0.000   1.87990
> 2
>      2.398962
          _cons |   .1326003   .0105538   12.56   0.000   .111905
> 9
>      .1532946

```

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs |
|--------------|------------|-------------|--------------|
| permno1#year | 18915 | 18915 | 0 * |
| permno2#year | 31498 | 31498 | 0 * |

* = FE nested within cluster; treated as redundant for DoF computation

1491est store reg2

1492* output regression results

```

1493esttab reg1 reg2 using "output\oa3.rtf", cells(b(star fmt(3)) se(par
>   fmt(3))) stats(r2_a N_full, fmt(%9.3f %9.0g)) starlevels(* 0.1 ** 0
>   .05 *** 0.01) order(I_shared_office $Control1) modelwidth(12) replac
>   e
(output written to output\oa3.rtf)

```

1494

1495* OA 4. alternative cluster levels

1496use "main.dta", clear

1497* col (1)

```

1498reghdfe lcitation I_shared_office $Control1, absorb(permno1##year pe
>   rmno2##year) cluster(permno1)
(dropped 2911 singleton observations)
(MWFE estimator converged in 11 iterations)

```

HDFE Linear regression

> ,565,256

Absorbing 2 HDFE groups

> 303.66

Statistics robust to heteroskedasticity

> 0.0000

> 0.3692

> 0.3480

> 0.1638

Number of clusters (permno1) = 2,690

> 0.6206

Number of obs = 1

F(10, 2689) =

Prob > F =

R-squared =

Adj R-squared =

Within R-sq. =

Root MSE =

(Std. err. adjusted for 2,690 cluster
> s in permno1)

| | lcitation | Coefficient | Robust std. err. | t | P> t | [95% co |
|----------------|-------------------|-------------|---------------------|--------|-------|----------|
| > n | | | | | | |
| > f. interval] | | | | | | |
| > 2 | I_shared_office | .0661216 | .0148316 | 4.46 | 0.000 | .037039 |
| > 3 | .0952041 | | | | | |
| > 1 | CommonAnalyst | .0578618 | .0019142 | 30.23 | 0.000 | .054108 |
| > 5 | .0616153 | | | | | |
| > 6 | CommonAlliance | .298337 | .0628243 | 4.75 | 0.000 | .175148 |
| > 9 | .4215258 | | | | | |
| > 8 | CommonInventor | .0096318 | .0038416 | 2.51 | 0.012 | .00209 |
| > 7 | .0171647 | | | | | |
| > 4 | CommonOwnership | .0103757 | .005765 | 1.80 | 0.072 | -.000928 |
| > 2 | .0216799 | | | | | |
| > 1 | TechSimilarity | 1.000857 | .0342166 | 29.25 | 0.000 | .933763 |
| > 3 | 1.067951 | | | | | |
| > 5 | RelativeAnalyst | -9.825906 | .7819006 | -12.57 | 0.000 | -11.3590 |
| > 6 | -8.292719 | | | | | |
| > 4 | RelativeSize | -.046855 | .0075965 | -6.17 | 0.000 | -.061750 |
| > 2 | -.0319595 | | | | | |
| > 1 | RelativePatentS~k | -.0597317 | .0071849 | -8.31 | 0.000 | -.073820 |
| > 3 | -.0456431 | | | | | |
| > 5 | RelativeCitation | -.0154495 | .0009786 | -15.79 | 0.000 | -.017368 |
| > 8 | -.0135306 | | | | | |
| > 4 | _cons | .2055674 | .0059528 | 34.53 | 0.000 | .193894 |
| > 2 | .2172399 | | | | | |

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs | |
|--------------|------------|-------------|--------------|---|
| permno1#year | 19106 | 19106 | 0 | * |
| permno2#year | 31879 | 1 | 31878 | |

* = FE nested within cluster; treated as redundant for DoF computation

1499est store regl

1500* col (2)

1501capt drop siccd1_2

1502g siccd1_2 = int(siccd1/100) //SIC2 of citing company

1503reghdfe lcitation I_shared_office \$Controll, absorb(permnol##year pe
> rmno2##year) cluster(siccd1_2)
(dropped 2911 singleton observations)
(MWFE estimator converged in 11 iterations)

HDFE Linear regression

> ,565,256

Absorbing 2 HDFE groups

> 1229.82

Statistics robust to heteroskedasticity

> 0.0000

> 0.3692

> 0.3480

> 0.1638

Number of clusters (siccd1_2) = 66

> 0.6206

Number of obs = 1

F(10, 65) =

Prob > F =

R-squared =

Adj R-squared =

Within R-sq. =

Root MSE =

(Std. err. adjusted for 66 clusters

> in siccd1_2)

| | lcitation | Coefficient | Robust std. err. | t | P> t | [95% co |
|----------------|-----------------|-------------|---------------------|-------|-------|----------|
| > n | | | | | | |
| > f. interval] | | | | | | |
| > 5 | I_shared_office | .0661216 | .0218899 | 3.02 | 0.004 | .022404 |
| > 7 | CommonAnalyst | .0578618 | .0039447 | 14.67 | 0.000 | .049983 |
| > 2 | CommonAlliance | .298337 | .0618664 | 4.82 | 0.000 | .174781 |
| > 3 | CommonInventor | .0096318 | .0023927 | 4.03 | 0.000 | .004853 |
| > 8 | CommonOwnership | .0103757 | .004978 | 2.08 | 0.041 | .000433 |
| > 4 | TechSimilarity | 1.000857 | .0303654 | 32.96 | 0.000 | .940213 |
| > 8 | RelativeAnalyst | -9.825906 | 1.014788 | -9.68 | 0.000 | -11.8525 |
| > | | -7.799234 | | | | |

```

      RelativeSize |   -.046855   .010828   -4.33   0.000   -.0684
> 8
      -.02523
RelativePatentS~k |   -.0597317   .0089382   -6.68   0.000   -.077582
> 5
      -.0418808
RelativeCitation |   -.0154495   .0010799   -14.31   0.000   -.017606
> 3
      -.0132928
      _cons |   .2055674   .0063642   32.30   0.000   .192857
> 3
      .2182775

```

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs |
|--------------|------------|-------------|--------------|
| permno1#year | 19106 | 19106 | 0 * |
| permno2#year | 31879 | 1 | 31878 |

* = FE nested within cluster; treated as redundant for DoF computation

1504est store reg2

1505* output regression results

```

1506esttab reg1 reg2 using "output\oa4.rtf", cells(b(star fmt(3)) se(par
>   fmt(3))) stats(r2_a N_full, fmt(%9.3f %9.0g)) starlevels(* 0.1 ** 0
>   .05 *** 0.01) order(I_shared_office $Controll) modelwidth(12) replac
>   e
(output written to output\oa4.rtf)

```

1507

1508* OA 5. other robustness checks

1509use "main.dta", clear

1510* col (1)

```

1511reghdfe lcitation I_shared_office $Controll if lcitation>0, absorb(p
>   ermno1##year permno2##year) cluster(permno1 permno2)
(dropped 7103 singleton observations)
(MWFE estimator converged in 14 iterations)
Warning: VCV matrix was non-positive semi-definite; adjustment from Ca
>   meron, Gelbach & Miller applied.

```

HDFE Linear regression

> 406,816

Absorbing 2 HDFE groups

> 191.59

Statistics robust to heteroskedasticity

> 0.0000

> 0.4567

> 0.4029

Number of clusters (permno1) = 2,454

> 0.1972

Number of clusters (permno2) = 2,599

Number of obs =

F(10, 2453) =

Prob > F =

R-squared =

Adj R-squared =

Within R-sq. =

Root MSE =

> 0.6752

(Std. err. adjusted for 2,454 clusters in per

> mno1 permno2)

| | lcitation | Coefficient | Robust std. err. | t | P> t | [95% co |
|----------------|-------------------|-------------|---------------------|-------|-------|----------|
| > n | | | | | | |
| > f. interval] | | | | | | |
| > 1 | I_shared_office | .0781028 | .0194814 | 4.01 | 0.000 | .03990 |
| > 1 | .1163045 | | | | | |
| > 7 | CommonAnalyst | .0393003 | .0020075 | 19.58 | 0.000 | .035363 |
| > 7 | .0432369 | | | | | |
| > 7 | CommonAlliance | .1623487 | .047353 | 3.43 | 0.001 | .069492 |
| > 7 | .2552047 | | | | | |
| > 5 | CommonInventor | .0041752 | .0016695 | 2.50 | 0.012 | .000901 |
| > 5 | .007449 | | | | | |
| > 4 | CommonOwnership | .0100823 | .0045733 | 2.20 | 0.028 | .001114 |
| > 4 | .0190502 | | | | | |
| > 9 | TechSimilarity | 1.206725 | .0488627 | 24.70 | 0.000 | 1.11090 |
| > 9 | 1.302541 | | | | | |
| > 1 | RelativeAnalyst | -11.40654 | 1.917438 | -5.95 | 0.000 | -15.1665 |
| > 1 | -7.646579 | | | | | |
| > 1 | RelativeSize | -.0642025 | .0122827 | -5.23 | 0.000 | -.088288 |
| > 1 | -.040117 | | | | | |
| > 4 | RelativePatents~k | -.0641005 | .0075009 | -8.55 | 0.000 | -.078809 |
| > 4 | -.0493917 | | | | | |
| > 4 | RelativeCitation | -.0116 | .001398 | -8.30 | 0.000 | -.014341 |
| > 4 | -.0088585 | | | | | |
| > 7 | _cons | 1.070196 | .0125135 | 85.52 | 0.000 | 1.04565 |
| > 7 | 1.094734 | | | | | |

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs |
|--------------|------------|-------------|--------------|
| permno1#year | 16160 | 16160 | 0 * |
| permno2#year | 20456 | 20456 | 0 * |

* = FE nested within cluster; treated as redundant for DoF computation

1512est store reg1

1513* col (2)

1514reghdfe Icitation I_shared_office \$Control1, absorb(permno1##year pe
> rmno2##year) cluster(permno1 permno2)
(dropped 2911 singleton observations)
(MWFE estimator converged in 11 iterations)
Warning: VCV matrix was non-positive semi-definite; adjustment from Ca
> meron, Gelbach & Miller applied.

| | | | |
|---|---------------|---|---|
| HDFE Linear regression | Number of obs | = | 1 |
| > ,565,256 | F(10, 2689) | = | |
| Absorbing 2 HDFE groups | Prob > F | = | |
| > 314.62 | R-squared | = | |
| Statistics robust to heteroskedasticity | Adj R-squared | = | |
| > 0.0000 | Within R-sq. | = | |
| > 0.2643 | Root MSE | = | |
| > 0.2396 | | | |
| Number of clusters (permno1) = | 2,690 | | |
| > 0.0781 | | | |
| Number of clusters (permno2) = | 3,167 | | |
| > 0.3845 | | | |

(Std. err. adjusted for 2,690 clusters in per

> mno1 permno2)

| | Icitation | Coefficient | Robust std. err. | t | P> t | [95% co |
|-----------------|-----------|-------------|---------------------|-------|-------|----------|
| > n | | | | | | |
| > f. interval] | | | | | | |
| I_shared_office | | .0248925 | .0056138 | 4.43 | 0.000 | .013884 |
| > 7 | | | | | | |
| > .0359002 | | | | | | |
| CommonAnalyst | | .0186652 | .0007722 | 24.17 | 0.000 | .017150 |
| > 9 | | | | | | |
| > .0201795 | | | | | | |
| CommonAlliance | | .0565359 | .021577 | 2.62 | 0.009 | .014226 |
| > 7 | | | | | | |
| > .0988452 | | | | | | |
| CommonInventor | | .0011148 | .0005961 | 1.87 | 0.062 | -.000054 |
| > 2 | | | | | | |
| > .0022837 | | | | | | |
| CommonOwnership | | .0027482 | .0019646 | 1.40 | 0.162 | -.001104 |
| > 1 | | | | | | |
| > .0066005 | | | | | | |
| TechSimilarity | | .4608068 | .0148793 | 30.97 | 0.000 | .431630 |
| > 8 | | | | | | |
| > .4899828 | | | | | | |
| RelativeAnalyst | | -4.031819 | .4539646 | -8.88 | 0.000 | -4.92197 |
| > 4 | | | | | | |
| > -3.141664 | | | | | | |
| RelativeSize | | -.0179228 | .0034551 | -5.19 | 0.000 | -.024697 |
| > 6 | | | | | | |

```

>      -.0111479
RelativePatentS~k | -.0255974   .0044131   -5.80   0.000   -.034250
> 8
>      -.016944
RelativeCitation | -.0049376   .0005343   -9.24   0.000   -.005985
> 2
>      -.00389
      _cons |   .1874176   .002559   73.24   0.000   .182399
> 8
>      .1924355

```

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs | |
|--------------|------------|-------------|--------------|---|
| permno1#year | 19106 | 19106 | 0 | * |
| permno2#year | 31879 | 31879 | 0 | * |

* = FE nested within cluster; treated as redundant for DoF computation

1515est store reg2

1516* col (3)

1517ppmlhdfe citations I_shared_office \$Controll, absorb(permno1##year p
> ermno2##year) cluster(permno1 permno2)

(dropped 68879 observations that are either singletons or separated by
> a fixed effect)

**(ReLU separation check: maximum number of iterations reached; aborting
>)**

Iteration 1: deviance = 4.9281e+06 eps = . iters = 7 tol
> = 1.0e-04

> min(eta) = -6.25 P
Iteration 2: deviance = 4.1508e+06 eps = 1.87e-01 iters = 6 tol
> = 1.0e-04

> min(eta) = -7.73
Iteration 3: deviance = 4.0725e+06 eps = 1.92e-02 iters = 5 tol
> = 1.0e-04

> min(eta) = -9.35
Iteration 4: deviance = 4.0670e+06 eps = 1.35e-03 iters = 5 tol
> = 1.0e-04

> min(eta) = -10.51
Iteration 5: deviance = 4.0666e+06 eps = 1.18e-04 iters = 5 tol
> = 1.0e-04

> min(eta) = -11.25
Iteration 6: deviance = 4.0665e+06 eps = 1.04e-05 iters = 5 tol
> = 1.0e-04

> min(eta) = -11.84
Iteration 7: deviance = 4.0665e+06 eps = 4.76e-07 iters = 5 tol
> = 1.0e-05

> min(eta) = -11.99
Iteration 8: deviance = 4.0665e+06 eps = 3.09e-09 iters = 6 tol
> = 1.0e-06

> min(eta) = -12.00 S
Iteration 9: deviance = 4.0665e+06 eps = 5.81e-13 iters = 7 tol
> = 1.0e-08

> min(eta) = -12.00 S

```
Iteration 10: deviance = 4.0665e+06 eps = 0.00e+00 iters = 5 tol
> = 1.0e-09
> min(eta) = -12.00 S O
```

```
(legend: p: exact partial-out s: exact solver h: step-halving o:
> epsilon below tolerance)
Converged in 10 iterations and 56 HDFE sub-iterations (tol = 1.0e-08)
Warning: VCV matrix was non-positive semi-definite; adjustment from Ca
> meron, Gelbach & Miller applied.
```

```
HDFE PPML regression No. of obs =
> 1499288
Absorbing 2 HDFE groups Residual df =
> 2,608
Statistics robust to heteroskedasticity Wald chi2(10) =
> 1576.53
Deviance = 4066517.44 Prob > chi2 =
> 0.0000
Log pseudolikelihood = -2639324.111 Pseudo R2 =
> 0.5108
```

```
Number of clusters (permno1)= 2,609
Number of clusters (permno2)= 3,012
(Std. err. adjusted for 2,609 clusters in per
> mnol permno2)
```

| | Coefficient | Robust std. err. | z | P> z | [95% co |
|-----------------|-------------|---------------------|-------|-------|----------|
| citations | | | | | |
| I_shared_office | .1659432 | .0654604 | 2.54 | 0.011 | .037643 |
| 1 | .2942434 | | | | |
| CommonAnalyst | .0268247 | .0035301 | 7.60 | 0.000 | .019905 |
| 7 | .0337436 | | | | |
| CommonAlliance | -.2704349 | .1399704 | -1.93 | 0.053 | -.544771 |
| 9 | .003902 | | | | |
| CommonInventor | -.0028932 | .0025935 | -1.12 | 0.265 | -.007976 |
| 4 | .00219 | | | | |
| CommonOwnership | .002707 | .0009054 | 2.99 | 0.003 | .000932 |
| 4 | .0044817 | | | | |
| TechSimilarity | 2.861571 | .0864764 | 33.09 | 0.000 | 2.69208 |
| 1 | 3.031062 | | | | |
| RelativeAnalyst | -11.31204 | 6.396823 | -1.77 | 0.077 | -23.8495 |
| 9 | 1.225501 | | | | |
| RelativeSize | -.1188311 | .0414213 | -2.87 | 0.004 | -.200015 |
| 4 | | | | | |
| | -.0376468 | | | | |


```

RelativePatentS~k | -.1120228 .1025003 -1.09 0.274 -.312919
> 8
> .0888742
RelativeCitation | -.0007631 .0027235 -0.28 0.779 -.006101
> 1
> .0045749
      _cons | .4407408 .0356207 12.37 0.000 .370925
> 6
> .5105561

```

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | - Num. Coefs |
|--------------|------------|-------------|--------------|
| permno1#year | 17933 | 17933 | 0 * |
| permno2#year | 25330 | 25330 | 0 * |

* = FE nested within cluster; treated as redundant for DoF computation

1518est store reg3

1519* col (4)

1520use "main.dta", clear

1521keep if YEAR>=2002 & YEAR <=2018 /*appl-examiner data is not availa
> ble until 2002 */
(239,642 observations deleted)

1522merge 1:1 permno1 permno2 year using "applicant.dta"

| Result | Number of obs | |
|-------------|---------------|--------------|
| Not matched | 2,552,455 | |
| from master | 970,655 | (_merge==1) |
| from using | 1,581,800 | (_merge==2) |
| Matched | 357,870 | (_merge==3) |

1523drop if _merge == 2
(1,581,800 observations deleted)

1524drop _merge


```
1525replace citations2 = 0 if citations2 == . /* applicant citations */
    (1,047,557 real changes made)
```

```
1526winsor2 citations2 , cuts(1 99) replace by(YEAR)
```

```
1527g lcitations2 = log(1+citations2)
```

```
1528reghdfe lcitations2 I_shared_office $Control1 , absorb(permno1##yea
> r permno2##year) cluster(permno1 permno2)
(dropped 2601 singleton observations)
(MWFE estimator converged in 11 iterations)
Warning: VCV matrix was non-positive semi-definite; adjustment from Ca
> meron, Gelbach & Miller applied.
```

```
HDFE Linear regression                                Number of obs   =   1
> ,325,924                                           F(   10,   2361) =
Absorbing 2 HDFE groups                               Prob > F         =
>   82.91                                           R-squared        =
Statistics robust to heteroskedasticity              Adj R-squared    =
>   0.0000                                           Within R-sq.     =
>   0.3490                                           Root MSE         =
>   0.3269
Number of clusters (permno1) =           2,362
>   0.1459
Number of clusters (permno2) =           2,763
>   0.6114
```

```
(Std. err. adjusted for 2,362 clusters in per
> mno1 permno2)
```

| | lcitations2 | Coefficient | Robust std. err. | t | P> t | [95% co |
|-----------------|-------------|-------------|---------------------|-------|-------|----------|
| > n | | | | | | |
| > f. interval] | | | | | | |
| I_shared_office | | .0544342 | .0166538 | 3.27 | 0.001 | .021776 |
| > 6 | | | | | | |
| > .0870918 | | | | | | |
| CommonAnalyst | | .0530428 | .0026201 | 20.24 | 0.000 | .04790 |
| > 5 | | | | | | |
| > .0581807 | | | | | | |
| CommonAlliance | | .3188004 | .0781941 | 4.08 | 0.000 | .165464 |
| > 2 | | | | | | |
| > .4721366 | | | | | | |
| CommonInventor | | .0106107 | .0038249 | 2.77 | 0.006 | .003110 |
| > 1 | | | | | | |
| > .0181113 | | | | | | |
| CommonOwnership | | .0087064 | .0058392 | 1.49 | 0.136 | -.002744 |
| > 1 | | | | | | |
| > .020157 | | | | | | |
| TechSimilarity | | .895927 | .0552878 | 16.20 | 0.000 | .787509 |
| > 4 | | | | | | |
| > 1.004345 | | | | | | |
| RelativeAnalyst | | -8.579965 | 1.427728 | -6.01 | 0.000 | -11.379 |

```

> 7
>      -5.780235
RelativeSize | -.0403153   .0071465   -5.64   0.000   -.054329
> 4
>      -.0263012
RelativePatentS~k | -.0712645   .0147013   -4.85   0.000   -.100093
> 2
>      -.0424358
RelativeCitation | -.0145363   .001325   -10.97   0.000   -.017134
> 6
>      -.011938
      _cons |   .1629946   .0086298   18.89   0.000   .146071
> 9
>      .1799173

```

Absorbed degrees of freedom:

| Absorbed FE | Categories | - Redundant | = Num. Coefs |
|--------------|------------|-------------|--------------|
| permno1#year | 16354 | 16354 | 0 * |
| permno2#year | 27183 | 27183 | 0 * |

* = FE nested within cluster; treated as redundant for DoF computation

1529est store reg4

1530* output regression results

```

1531esttab reg1 reg2 reg3 reg4 using "output\oa5.rtf", cells(b(star fmt(
> 3)) se(par fmt(3))) stats(r2_a N_full, fmt(%9.3f %9.0g)) starlevels(
> * 0.1 ** 0.05 *** 0.01) order(I_shared_office $Control1) modelwidth(
> 12) replace
(output written to output\oa5.rtf)

```

1532

1533* generate permno identifiers for JAR

1534use "main.dta", clear

1535keep permno1 permno2

1536gduplicates drop

Duplicates in terms of all variables

(1,405,684 observations deleted)

```

1537export delimited "identifier_permnos.csv", replace
file identifier_permnos.csv saved

```

1538

1539

end of do-file

1540/* End of code */

1541

1542capt log close