

**Read Me for**  
**“Megadeal” subsidies, local spillovers and corporate innovation”**

By Yoojin Lee, Shaphan Ng, and Aruhn Venkat

The replication package provides instructions and code to help replicate the data construction and empirical analyses.

***Data Cleaning***

*Relevant folders:*

*For codes – Replication Package > Sample Construction*

*For log files – Replication Package > Log > 01\_Data Clean*

1. Clean “Megadeal” subsidies, county, patent, Compustat, and college datasets.
  - a. Codes:
    - *01a\_Clean\_SubfirmsData*: Identify subsidized firms and counties.
    - *01b\_Clean\_CountyVariables*: Clean county variables from raw data files.
    - *01c\_Clean\_Patent*: Construct firm-county-year patent variables and firm-year patent variables for subsidy firms with GVKEY.
    - *01d\_Clean\_Compustat*: Create control variables.
    - *01h\_Clean\_UniCollege*: Clean and construct college variables.
  - b. Log: 01a.log, 01b.log, 01c.txt, 01d.log, and 01h.txt
2. Match subsidy firms with no GVKEY with patent assignee ID and create patent variables for these firms.
  - a. Codes:
    - *01e\_Clean\_SubfirmAssignee\_Match*: Match subsidy firms with no GVKEY with Patents View patent assignee ID.
    - *01f\_Clean\_SubPatentNoGVKEY*: Construct patent variables for subsidy firms with no GVKEY.
  - b. Log: 01e.log and 01f.txt
3. Clean and identify similar patents from Whalen et al. (2020) patent similarity dataset.
  - a. Code: *01g\_Clean\_LocalSimPat*
  - b. Log: 01g.txt
4. Clean corporate political contributions.
  - c. Code: *01i\_Clean\_PoliticalContributions*
  - d. Log: 01i.Rout
5. Clean local firm citations to Megadeal patents.
  - e. Code: *01j\_Clean\_LocalCites*
  - f. Log: 01j.Rout

Run codes sequentially from 01a – 01j.

***Sample construction***

*Relevant folders:*

*For code – Replication Package > Sample Construction*

*For log files – Replication Package > Log > 02\_Sample Construction*

1. Construct cohort-firm-county-year main sample used in Tables 3, 5, 6, 7, 8, 9, Panel B.
  - a. Code: *02a\_Data Construct\_Main*
  - b. Identifiers: *Identifiers\_Main*
  - c. Log: *02a.log*
2. Construct Megadeal sample for Table 2.
  - a. Code: *02b\_Data Construct\_Table2*
  - b. Log: *02b.log*
3. Construct sample used in Table 5, Panel C and create corporate political contributions variable.
  - a. Code: *02c\_Data Construct\_Table5C*
  - b. Log: *02c.Rout*
4. Construct sample used in Table 4 and create inventor movement variables.
  - a. Codes:
    - *03a\_Data Construct\_Table4A*: Construct Table 4 sample.
    - *03b\_Data Construct\_Table4B*: Obtain inventor location information for all patents.
    - *03c\_Data Construct\_Table4B*: Create inventor movement variables.
  - b. Identifiers: *Identifiers\_Table4*
  - c. Log: *03a.log, 03b.txt, 03c.log*
5. Create cross-sectional variables used in Table 6.
  - a. Codes:
    - *04a\_Data Construct\_Table6A*: Create subsidy firm innovativeness measure.
    - *04b\_Data Construct\_Table6B*: Create patent tech closeness measure.
    - *04c\_Data Construct\_Table6D*: Create college presence measure.
  - b. Log: *04a.log, 04b.log, 04c.log*
6. Construct patent similarity measures used in Table 7, Panels B and C.
  - a. Code: *05a\_Data Construct\_Table7B&C*
  - b. Log: *05a.log*
7. Construct cohort-firm-inventor-county-year sample used in Table 7, Panel D.
  - a. Code: *05b\_Data Construct\_Table7D*
  - b. Log: *05b.log*
8. Construct cohort-firm-county-year sample of local-firm citations to Megadeal patents used in Table 7, Panel A.
  - a. Code: *05c\_Data Construct\_Table7A*
  - b. Log: *05c.Rout*
9. Create cross-sectional variables used in Table 8.
  - a. Code: *06\_Data Construct\_Table8*
  - b. Log: *06.log*

10. Construct samples used in Table 9

- a. Code: *07\_Data Construct Table9*
- b. Identifiers: Identifiers\_Table9A (Panel A); Identifiers\_Table9B (Panel B)
- c. Log: 07.log

***Analysis***

*Relevant folders:*

*For codes – Replication Package > Analysis*

*For log files – Replication Package > Log > 03\_Output Tables*

1. Figure 1

- a. Code: *Fig1*
- b. Log: Fig1.log

Variable Dictionary

|                |                                               |
|----------------|-----------------------------------------------|
| tot_sub_splitm | <i>Total Megadeal Subsidy Amounts (\$mil)</i> |
| nsubs          | <i>Number of Megadeals</i>                    |

2. Table 2

- a. Code: *Table2*
- b. Log: Table 2.log

Variable Dictionary

|                  |                                                                                                                                                                                           |
|------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| tsub_splitm      | <i>Megadeals in \$mil</i>                                                                                                                                                                 |
| tjobs_split      | <i>Target number of jobs created</i>                                                                                                                                                      |
| ffi              | <i>Fama-French 12 Industry</i>                                                                                                                                                            |
| treat_year       | <i>Subsidy award year</i>                                                                                                                                                                 |
| tsub_split_`x`1  | <i>Megadeals in \$mil by investment type, where x can refer to plant (high-tech and non-high-tech), R&amp;D lab, headquarters office, data center, real estate, and other investments</i> |
| tjobs_split_`x`1 | <i>Target number of jobs created by investment type, where x is defined above</i>                                                                                                         |

3. Tables 3 and 5

- a. Code: *Table3&5*
- b. Log: Table 3 & 5.log

Variable Dictionary

|                    |                                                                                                                                                                   |
|--------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| log_numpat_`f`x`   | <i>Log(1+#Pat)<sub>t+`x`</sub></i>                                                                                                                                |
| numpat_`f`x`       | <i>#Pat<sub>t+`x`</sub></i>                                                                                                                                       |
| log_tsub_split_int | <i>LocalSubsidy</i>                                                                                                                                               |
| unemppercent       | <i>Unemploy</i>                                                                                                                                                   |
| log_incperscap     | <i>PerCapitaIncome</i>                                                                                                                                            |
| mlq_indemp         | <i>LQIndEmp</i>                                                                                                                                                   |
| log_amount         | <i>Poli_Contrib</i>                                                                                                                                               |
| foc_cfips          | <i>Cohort comprising individual treatment counties and their neighboring control counties; the cohort takes the value of the focal treatment county FIPS code</i> |
| sfips              | <i>State FIPS Code</i>                                                                                                                                            |
| cfips              | <i>County FIPS code</i>                                                                                                                                           |

4. Table 5, Panel C

- a. Code: *Table5C*

- b. Log: Table 5C.Rout

Variable Dictionary

log\_amount                      *Poli\_Contrib*  
Other variables are defined above

- 5. Figure 2
  - a. Code: *Fig2*
  - b. Log: Fig 2.Rout

Variable Dictionary

rr\_l'x', rr\_0, rr\_fx'              *Event-time indicators*  
Other variables are defined above

- 6. Figure 3
  - a. Code: *Fig3*
  - b. Log: Fig 3.Rout

Variable Dictionary

d\_quartile\_horizon=1              *Fast innovation cycle*  
d\_quartile\_horizon=0              *Slow innovation cycle*  
Other variables are defined above

- 7. Table 4
  - a. Code: *Table4*
  - b. Log: Table 4.log

Variable Dictionary

log\_ssub\_split\_int              *SubsidyFirm*  
log\_nchgfirmcflps1              *Log(1+#NewCtyInv)*  
Other variables are defined above

- 8. Table 6
  - a. Code: *Table6*
  - b. Log: Table 6.log

Variable Dictionary

hit                      *LocalSubsidy<sup>HighSubInno</sup>*  
lowt                      *LocalSubsidy<sup>LowSubInno</sup>*  
log\_tsub\_int\_hps1              *LocalSubsidy<sup>Close</sup>*  
log\_tsub\_int\_lps1              *LocalSubsidy<sup>NotClose</sup>*  
log\_tsub\_samesic3              *LocalSubsidy<sup>SameInd</sup>*  
log\_tsub\_diffsic3              *LocalSubsidy<sup>DiffInd</sup>*  
log\_tsub\_int\_huc              *LocalSubsidy<sup>HighCollege</sup>*  
log\_tsub\_int\_luc              *LocalSubsidy<sup>LowCollege</sup>*  
Other variables are defined above

- 9. Table 7, Panel A
  - a. Code: *Table7A*
  - b. Log: Table 7A.Rout

### Variable Dictionary

log\_cite\_fx'  $\text{Log}(1 + \#LocalCites)_{t+x}$

Other variables are defined above

### 10. Table 7, Panels B and C

a. Code: *Table7B&C*

b. Log: Table 7B & C.log

### Variable Dictionary

log\_nsimpatd\_fx'  $\text{Log}(1 + \#SimPat)_{t+x}$

msimscored\_fx'\_00  $PatSim_{t+x}$

Other variables are defined above

### 11. Table 7, Panel D

a. Code: *Table7D*

b. Log: Table 7D.log

### Variable Dictionary

prevsubfirm  $NewFromSub$

Other variables are defined above

### 12. Table 8

a. Code: *Table8*

b. Log: Table 8.log

### Variable Dictionary

log\_tsub\_newloc1  $LocalSubsidy^{Relocate}$

log\_tsub\_sameloc1  $LocalSubsidy^{Expand}$

log\_tsub\_split\_labs\_int  $LocalSubsidy^{Lab}$

log\_tsub\_split\_hqoffices\_int  $LocalSubsidy^{HQ}$

log\_thtsub\_split\_plants\_int  $LocalSubsidy^{TechPlant}$

log\_tltsub\_split\_plants\_int  $LocalSubsidy^{NonTechPlant}$

Other variables are defined above

### 13. Table 9

a. Code: *Table9*

b. Log: Table 9.log

### Variable Dictionary

log\_tsub\_split\_spill\_int  $LocalSubsidySpill$

Other variables are defined above