

Who Becomes an Inventor? The Importance of Exposure to Innovation

Codebook for Online Data Tables

November 2017

Table 2: Careers of Inventors

Innovation Rates by Current CZ/State, Gender, Year of Birth, and Age

Tables 2a and 2b report patenting outcomes for individuals aged 20 to 80 in years 1996-2012 by year of birth, gender, age and commuting zone (CZ) of residence (Table 2a) or state (Table 2b).

We report the fraction of individuals who file a patent application in a given year as well as the fraction of individuals who file a patent application in that year that is subsequently granted. We observe patent applications in years 2001 to 2012 and patent grants in years 1996 to 2014 (see Section II.B of the paper). All patent grants are public. For a fee, applicants can choose to have their filing kept secret; 15% of applicants choose to do so, and these patent applications do not appear in our data.

Since the grant data span 1996-2014, the data on patents subsequently granted are censored at the beginning and end of the sample frame; for instance, the grantee variable for 1996 includes only patent applications that were filed and granted in 1996, while the grantee variable for 2012 includes only patent applications filed in 2012 that were granted by 2014. This censoring leads to lower grantee rates at the beginning and end of our sample window.

We also report the average number of patent grants by patent category. We classify patents into technology categories using the classification developed in the NBER Patent Data Project (Hall et al. 2001). Note that an inventor may patent in more than one category in a given year.

The population counts reported for each cell are computed from a 10% sample of the taxpayer database.

There is one row in each of these tables for each CZ/state, year of birth, and age. Cells with fewer than 250 observations in the population (i.e., 100% sample) are omitted.

Users interested in correlating these measures with other CZ characteristics can download a set of CZ-level characteristics from Chetty et al. (2014, Online Data Table 8 [\[xls\]](#) [\[stata\]](#) [\[codebook\]](#)).

References:

Hall, B., A. Jae, and M. Trajtenberg. "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools." *NBER Working Paper No. 8498*, 2001.

Chetty, R., N. Hendren, P. Kline, and E. Saez. "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." *Quarterly Journal of Economics* 129(4): 1553-1623, 2014

Codebook for Table 2a – Innovation by Current CZ, Year of Birth, and Age

Variable	Description
cz	Current commuting zone of residence
czname	Commuting zone name
state	Current state Federal Information Processing Standard (FIPS) code; CZs that cross state borders are assigned to the state which contains the largest population in the CZ, based on the 2010 Census
stateabbrv	Two-letter state abbreviation
cohort	Year of birth
age	Age at which patenting outcomes are measured
year	Calendar year
count	Number of individuals in population
applicant	Fraction of individuals who apply for a patent in current calendar year
grantee	Fraction of individuals who apply for a patent in current year that is subsequently granted
num_grants	Average number of patents grants per individual, by application year
grantee_cat_[c]	Fraction of individuals granted a patent in technology category [c], by application year. Technology categories [c] are: 1 – Chemical 2 – Computers and Communications 3 – Drugs and Medical 4 – Electrical and Electronic 5 – Mechanical 6 – Others 7 – Design and Plant
[outcome]_g_m	Identical to variable [outcome], but restricting the sample to males.
[outcome]_g_f	Identical to variable [outcome], but restricting the sample to females.

Codebook for Table 2b – Innovation by Current State, Year of Birth, and Age

Table 2b contains the same variables as Table 2a with the exception of cz and czname, since statistics are reported at the state level. Variables in Table 2b are defined identically to variables in Table 2a, except that all statistics are computed directly at the state level in the microdata.