Who Becomes an Inventor? The Importance of Exposure to Innovation

<u>Codebook for Online Data Tables</u>

November 2017

Table 1: Origins of Inventors
Innovation Rates by Childhood CZ/State, Gender and Parent Income

Tables 1a and 1b show patenting outcomes for children born in 1980-1984 by the commuting zone (Table 1a) or state (Table 1b) in which they grew up, gender, and parent income. We restrict the sample to U.S. citizens as of 2013 to exclude individuals who are likely to have immigrated to the U.S. as adults, for whom we cannot measure parent income.

We define a child as an inventor if he or she is listed on a patent application between 2001 and 2012 or grant between 1996 and 2014 (see Section II.B of the paper), and as a highly-cited inventor if he or she is among the 5% of inventors with the most patent citations by 2014 within his birth cohort. Each child is assigned a commuting zone (CZ) or state based on ZIP code from which his or her parents filed their 1040 tax return in the year the child was first claimed as a dependent (which is typically 1996, as our tax data begin in 1996). Parents are assigned percentile ranks by ranking them based on their mean household income from 1996 to 2000 relative to other parents with children in the same birth cohort. See Section II of the paper for further details on the sample construction and variable definitions.

We also report the share of inventors and highly-cited inventors broken down by patent category. We classify patents into technology categories using the classification developed in the NBER Patent Data Project (Hall et al. 2001). We assign each inventor to the category in which he or she patents most often in our sample frame, breaking ties randomly.

We provide statistics on the fraction of inventors by childhood CZ or state pooling all children, by gender, and by parent income quintile.

There is one row in each of these tables per CZ or state. Cells with less than 250 children 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 1a – Innovation Rates by Childhood CZ

Variable	Description
par_cz	Childhood commuting zone of residence
par_czname	Commuting zone name
par_state	Childhood 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
par_stateabbrv	Two-letter state abbreviation
kid_count	Number of children
inventor	Share of children who go on to become inventors
top5cit	Share of children with patent citations in top 5% of their birth cohort, using total number of citations

inventor_cat_[c]	Share of children who patent in technology category [c].
	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

top5cit_cat_[c]	Share of children who patent in technology category [c] and have total patent citations in top 5% of their birth cohort
[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.
[outcome]_pq_[quintile]	Identical to variable [outcome], but restricting the sample to children whose parental income is in quintile [quintile] of the parent income distribution of the children's birth cohort.

Codebook for Table 1b – Innovation Rates by Childhood State

Table 1b contains the same variables as Table 1a with the exception of par_cz and par_czname, since statistics are reported at the state level. Variables in Table 1b are defined identically to variables in Table 1a, except that all statistics are computed directly at the state level in the microdata.