Png LinkedIn Patent Inventor FIVE Data

by

Ivan Png

Distinguished Professor
NUS Business School and Department of Economics
National University of Singapore

©2016

Contents

Copyright © 2015 I.P.L. Png

1. Conditions of Use

The data sets described in this document are part of the FIVE Project on Firm and Industry Evolution and Entrepreneurship, organized by Constance Helfat (Tuck School of Business at Dartmouth). If you download or otherwise obtain any of the data sets described in this document, or obtain modified versions of these data sets, you agree to abide by the conditions of use set forth here.

The organizer of the FIVE Project, Constance Helfat, is not responsible for the content of any of the FIVE Project data sets. The FIVE Project organizers and the FIVE Project data set authors make no representations or warranties of any kind concerning any of the FIVE Project data, including with regard to the absence or presence of errors.

All FIVE Project data sets are protected by copyright. The copyright holders for each data set are indicated on the first page of the documentation for each data set. Permission is granted to reproduce FIVE Project data for non-profit educational and research use only. Requests to reproduce FIVE Project data for other uses should be addressed to: Professor Constance Helfat, Tuck School of Business at Dartmouth, Hanover, NH 03755, constance.helfat@dartmouth.edu.

This document describes the Png LinkedIn-Patent FIVE Data, which should be referred to by this name in any derivative works. In any written and published work, users of these data should cite this document, the author's original article (Ge, Huang, and Png 2016), and the FIVE Project: Data Overview (Helfat & Klepper, 2007).

You may create a new data set by modifying a FIVE Project data set, including but not limited to adding or removing data or merging data from different FIVE Project data sets. If you utilize any FIVE Project data to create new data sets, within six years from the date that you obtain the FIVE Project data, you must provide the FIVE Project with a copy of your modified data set by contacting: Constance Helfat, constance.helfat@dartmouth.edu.

Any written documents or statistical estimates that use FIVE Project data, including but not limited to working papers and publications, must be reported to: Constance Helfat, constance.helfat@dartmouth.edu.

2. Data Set Authors

Ge Chunmian
Assistant Professor
School of Business Administration
South China University of Technology
Guangzhou, China

Copyright © 2015 I.P.L. Png

Ke-wei Huang
Assistant Professor
School of Computing
National University of Singapore
Singapore

Ivan Png
Professor
NUS Business School
National University of Singapore
Singapore

3. Data Set Summary

The data cover 210,554 observations of the career histories of 14,293 inventors with patents and public LinkedIn profiles. The data include the first and last year in LinkedIn and patent records, field of technology, and change of employer computed by LinkedIn and patent records. Data File 1 contains detailed data on individual inventors, their patenting history, and their employment mobility. Data File 2 contains inventor IDs for each patent number in the data.

4. Author's Research Using This Data Set

Ge, Chunmian, Ke-wei Huang, and Ivan P.L. Png, "Engineer/Scientist Careers: Patents, Online Profiles, and Misclassification", *Strategic Management Journal*, Vol 37 No 1, January 2016, 232-253.

5. Additional References

Ganco M. 2013. Cutting the gordian knot: the effect of knowledge complexity on employee mobility and entrepreneurship. *Strategic Management Journal*, 34(6): 666–686.

Hall B H, Jaffe A B, Trajtenberg M. 2001. The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools. NBER working Paper 8498, National Bureau of Economic Research, Cambridge, MA.

Helfat, CE & Klepper, S. 2007. FIVE Project: Data Overview. http://papers.ssrn.com/paper=1028022.

Li G C, Lai R, D'Amour A, Doolin DM, Sun Y, Torvik VI, Yu AZ, Fleming L. 2014. Disambiguation and co-authorship networks of the U.S. patent inventor database (1975--2010). *Research Policy*, 43(6): 941--955.

Copyright © 2015 I.P.L. Png

Palomeras N, Melero E. 2010. Markets for inventors: learning-by-hiring as a driver of mobility. *Management Science*, 56(5): 881–895.

6. Data Set Sources and Construction

We collected public profiles in the English language from LinkedIn (www.linkedin.com). Members of LinkedIn can choose whether to open their profiles to public browsing, and if so, what information to make public. The minimum information required for a public profile is the individual's name, industry, location, and number of connections to other LinkedIn members. Between June-November 2013, we used Google to search for public profiles on LinkedIn that included one or more valid U.S. patent numbers.

We compiled career histories in the following way. If the LinkedIn profile for an individual showed two consecutive jobs with different employers, we deemed that the person changed employers in the ending year of the first job. If the profile showed a gap of more than one year between employers, we deemed the person to have entered and left self-employment during the gap, and to have changed employers at the beginning and end of the gap.

Next, we used the Harvard Patent Inventor Database (Li et al., *Research Policy*, 2014) to compile career histories. Following the majority of previous research, we inferred a change of employer as occurring at the midpoint of two applications for patents with different assignees. To screen out false positives due to contract or collaborative R&D, merger or acquisition, or organizational policy or name change, we inferred a change of employer only if the difference in assignee met three conditions. One was that there be no patent assigned to the previous assignee up to 360 days after the second patent application, and the second was that there be no patent assigned to the new assignee up to 360 days before the previous patent application. The third was to explicitly exclude differences in assignees due to merger or acquisition as identified in the NBER Patent Database (Hall, Jaffe, and Trajtenberg, NBER Working Paper, 2001).

We consolidated the data from LinkedIn and patent records on an annual basis, so that the measure of mobility is whether the individual changed employers at least once during the year. We matched each inventor’s LinkedIn profile with their patent records by the inventor's first or last name, and a U.S. patent number, subject to the patent record and online profile overlapping by at least one year. (For patents filed by multiple co-inventors, the patent number alone does not identify an inventor unambiguously.)

Excluding inventors whose public LinkedIn profiles do not report their employment history and also excluding inventors with only one patent, the matched sample of patent inventors with online profiles comprises 14,293 individuals.

Copyright © 2015 I.P.L. Png

7. Variable List and Definitions: File 1: PngLinkedIninventorFIVEdata

Observations: 210,554
Variables: 59
Size: 42,110,800

Note: Data items that are blank in the file are missing data.

Variable name	Description	Detailed construction	Source
Patent			
lower	Inventor identifier (lower)	The Harvard Patent Inventor Database creates two identifiers for each inventor – lower and upper bounds of matching. We use the "lower."	Harvard Patent Inventor Database
zipcode	Zip code of inventor's address		Harvard Patent Inventor Database
patlife	Lifetime patents	Total number of patents applied for by the inventor	Harvard Patent Inventor Database
patfirstyr	Application year of the first patent		Harvard Patent Inventor Database
patlastyr	Application year of the last patent		Harvard Patent Inventor Database
patfirstno	USPTO number of first patent		Harvard Patent Inventor Database
echg360yr	Employer change during the year (360)	Indicator for whether the inventor changed jobs based on the employment history from patent records. Inventor is assumed to move at the date of the earlier patent application among two consecutive patents. Dummy variable, 1 if	

Copyright © 2015 I.P.L. Png

		employer change occurred during the year, 0 otherwise. The “360” refers to 360-day forward and backward check for same assignee.	
patcareer	Patent career	Number of years between first patent application year and last patent application year	
tenure	Tenure	Current year minus the application year of the first patent	
patyr	Patent rate	Number of patent applications in the year	
cites_avg	Citations	Following Palomeras and Melero (2010), stipulated as the average of standardized citations, i.e., the number of citations to each patent divided by the mean number of citations received by the population of patents granted in the same year and category of technology (Hall et al. 2001), with the average taken over all patents that the inventor ever applied for until the current year.	
pat_breadth	Patent breadth	Average number of main patent classes, with the average taken over all patents that inventor applied for in the year.	
cplxty	Complexity	Following Ganco (2013), calculated as the average of the ratio of interdependencies to the number of components, with the average taken over all patents that the inventor applied for in the year. To ensure functional equivalence of subclasses, interdependencies are	

Copyright © 2015 I.P.L. Png

		calculated within each technology category (Hall et al. 2001).	
co_inventor	Co-inventors	Average number of co-inventors, over all patents that the inventor applied for in the year.	
sv	Silicon Valley	Silicon Valley zip codes from AnnaLee Saxenian, "Silicon Valley's New Immigrant Entrepreneurs", Public Policy Institute of California, 1999, pp. 79-80, plus zip code for Santa Clara.	
self	Ever apply for a patent without assignee	Dummy variable, 1 if yes, 0 otherwise	
invtr_pct	Quintile of average patents per year	For each inventor; quintiles by average patents per year among all inventors, so, for example, the first quintile comprises inventors with the fewest patents and the fifth (last) comprises inventors with the most patents.	Harvard Patent Inventor Database
invtrq1	Average patent per year quintile 1	Dummy variable, 1 if inventor belongs to first quintile, 0 otherwise.	
invtrq2	Average patent per year quintile 2	Dummy variable, 1 if inventor belongs to second quintile, 0 otherwise.	
invtrq3	Average patent per year quintile 3	Dummy variable, 1 if inventor belongs to third quintile, 0 otherwise.	
invtrq4	Average patent per year quintile 4	Dummy variable, 1 if inventor belongs to fourth quintile, 0 otherwise.	
invtrq5	Average patent per year quintile 5	Dummy variable, 1 if inventor belongs to fifth quintile, 0 otherwise.	
tech	Tech subclass	Match the class information	NBER Patent

Copyright © 2015 I.P.L. Png

	(HJT)	in Harvard Patent Inventor Database with NBER Patent Database to get the technology class	Database (Hall et al. 2001) patent_HJTclass.dta
tech1	Chemicals	Dummy variable, 1 if technology class is chemical, 0 otherwise	
tech2	Computers & communications	Dummy variable, 1 if technology class is computer & communications , 0 otherwise	
tech3	Medical (other than drugs)	Dummy variable, 1 if technology class is medical (other than drugs), 0 otherwise	
tech31	Drugs	Dummy variable, 1 if technology class is drugs, 0 otherwise	
tech4	Electrical & electronic (excluding semiconductors)	Dummy variable, 1 if technology class is electrical and electronic (excluding semiconductors), 0 otherwise	
tech46	Semiconductors	Dummy variable, 1 if technology class is semiconductors, 0 otherwise	
tech5	Mechanical	Dummy variable, 1 if technology class is mechanical, 0 otherwise	
tech6	Other technologies	Dummy variable, 1 if technology class is 6 (other technologies), 0 otherwise	
tech99	Not in tech1-tech6	Dummy variable, 1 if not in tech1-tech6, 0 otherwise	
LinkedIn			
lkn_id	Inventor LinkedIn identifier	Constructed by authors	
url	Inventor's LinkedIn profile		LinkedIn.com

Copyright © 2015 I.P.L. Png

	weblink		
year_begin	First job start year		LinkedIn.com
year_end	2013 or end year of last job if prior to 2013		LinkedIn.com
lkncareer	LinkedIn career	Number of years between year_end and year_begin	
echgyr	Job change in year	Indicator for whether the inventor changed jobs, equal to 1 if current year employer is not the same as previous year, equals 0 otherwise. Job history is based on LinkedIn record.	LinkedIn.com
Patent & LinkedIn			
lower_id	Inventor patent identifier	Identifier created by authors from Harvard identifier lower (Stata: egen lower_id = group(lower))	
LinkedIn_Flag	Indicator for whether the inventor is in the LinkedIn database	Dummy variable (1 if in LinkedIn database, 0 otherwise)	
fp	False positive	Equals 0 if echg360yr = 0 and echgyr = 0 Equals 1 if echg360yr = 1 & echgyr = 0	
fn	False negative	Equals 0 if echg360yr = 1 and echgyr = 1 Equals 1 if echg360yr = 0 & echgyr = 1	
yr45_80	Year 1945-1980 indicator	Dummy variable (1 if year=1945-1980, 0 otherwise)	
yr81_85	Year 1981-1985 indicator	Dummy variable (1if year=1981-1985, 0 otherwise)	

Copyright © 2015 I.P.L. Png

yr86_90	Year 1986-1990 indicator	Dummy variable (1if year=1986-1990, 0 otherwise)	
yr91_95	Year 1991-1995 indicator	Dummy variable (1if year=1991-1995, 0 otherwise)	
yr96_00	Year 1996-2000 indicator	Dummy variable (1if year=1981-1985, 0 otherwise)	
yr01_05	Year 2001-2005 indicator	Dummy variable (1if year=2001-2005, 0 otherwise)	
yr06_14	Year 2006-2014 indicator	Dummy variable (1if year=2006-2014, 0 otherwise)	

8. Variable List and Definitions: File 2: PngLinkedInpatentFIVEdata

Observations: 96,104
Variables: 4
Size: 7,976,632

Variable name	Description	Detailed construction	Source
lkn_id	Inventor's LinkedIn ID		Assigned by authors
lower	Inventor identifier (lower)		Harvard Patent Inventor Database
number	Patent number		Inventor's LinkedIn profile
pat_num_final	Patent number	Removed characters from *number*	

Copyright © 2015 I.P.L. Png