Perspective Problem Set1

Alice Mee Seon Chung

(a) These data comprise detail information on almost 3 million U.S. patents granted from 1963 to 1999 and all citations made to these patents between 1975 and 1999. The data presents main trends in patenting over 30 years and made a broad match of patents to Compustat. The availability of patent data in these data format will boost researchers in economics to use the data and explore patent fields.
The patent data is stored in NBER US Patent Citation Database. NBER is working on a funded update and extension of this data so now a new release of files contained existing data up to date through 2006 are now available in Patent Data Project (https://sites.google.com/site/patentdataproject/Home). Original data from 1963 to 1999 are still available on NBER patent website(http://nber.org/patents/). All the data are freely available.
Primarily Bronwyn H. Hall, Adam B. Jaffe and Manuel Trajtenberg curate the data and they wrote the paper "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools." to describe these data in details and how users can use the data.

(b) Sonn and Storper(2008) used the NBER Patent Citation Data File to examine the importance of geographical proximity in knowledge production. The researchers used the information about patents, their inventors and the addresses of inventors and assignees to get the geographical data. They used the city and the state in the inventor's addresses to assign to a metropolitan area. [1]
Bessen and Hunt(2007) documented a increase in software patenting propensity and what aspect of the software patent can explain dramatic growth of software patent. The author used US Patent Office Database as primary data and uses the NBER Patent Citation Data File as their second dataset. The author extracted sample patents and matched the patents to firms and the assignees using NBER Patent Citation Data File. [2]
Balasubramanian and Sivadasan(2011) used the NBER Patent Citation Data File and U.S Census microdata to examine what happens when firms patent. The author wanted to build and use a new concordance between two datasets to provide new findings on the research question. To create the primary data set, the author linked NBER-Business Register Bridge to data five censuses. [3]
Thursby, Fuller and Thursby(2009) used National Research Council data and the NBER Patent Citation Data File to study the role of university faculty in university-industry technology transfer. The authors created a sample of faculty patents with the last names

---

[1] Jung Won Sonn and Michael Storper, 2008, "The increasing importance of geographical proximity in knowledge production: an analysis of US patent citations, 1975 -1997", Environment and Planning A 2008, volume 40, 1020-1039.

[2] James Bessen and Robert M. Hunt , 2007, "An Empirical Look at Software Patents", Journal of Economics & Management Strategy, Volume 16, Issue 1 Spring 2007, 157–189.

[3] Natarajan Balasubramanian and Jagadeesh Sivadasan, 2011, "What Happens When Firms Patent? New Evidence from U.S. Economic Census Data", The Review of Economics and Statistics, February 2011, Vol. 93, No. 1, 126-146.

using National Research Council data and compared faculty names with inventor names listed in the NBER Patent Citation Data File. [4]

(c) The patent data is provided by the Patent Office (USPTO's TAF database, United States Patent and Trademark Office). The data provided occupied about 60 magnetic tapes, and the inversion procedure would have necessitated computer resources because of computer technology at that time. However, computers and data availability improved by the time. The patent data were obtained from the Patent Office, except for the citations from patents granted in 1999 and it acquired from MicroPatent. After finished collecting, Jaffe and Trajtenberg did computations and add 10 new variables including "generality" and "originality".

(d) The 'Pat63_99' file includes 23 variables. Among 23 variables, 10 variables are USPTO original variables and 13 variables are new variables. Only 2 variables (country, postate) are character variable type and other 21 variables are numeric variables. A table that gives descriptive statistics for 10 key variables are below.

<Table1. Descriptive Statistics of 10 Variables>

Variables	Observations	Quartile 1	Median	Mean	Quartile 3	Std Dev.	Min	Max
Grant Year	2923922	1974	1985	1984	1993	10.98	1963	1999
Assignee Type	2923922	2	2	2.238	3	0.92	1	7
Number of Claims	1984055	5	10	12.1	16	10.27	1	868
Number Of Citations Made	2139314	3	6	7.7	9	9	0	770
Number of Citations Received	2923922	1	3	4.779	6	7.35	0	779
Percent of Citations Made to Patents	2923922	0.8	1	0.8	1	0.25	0	1
Generality	2240348	0	0.4	0.3	0.6	0.28	0	0.94
Originality	2042151	0	0.4	0.3	0.6	0.28	0	0.95

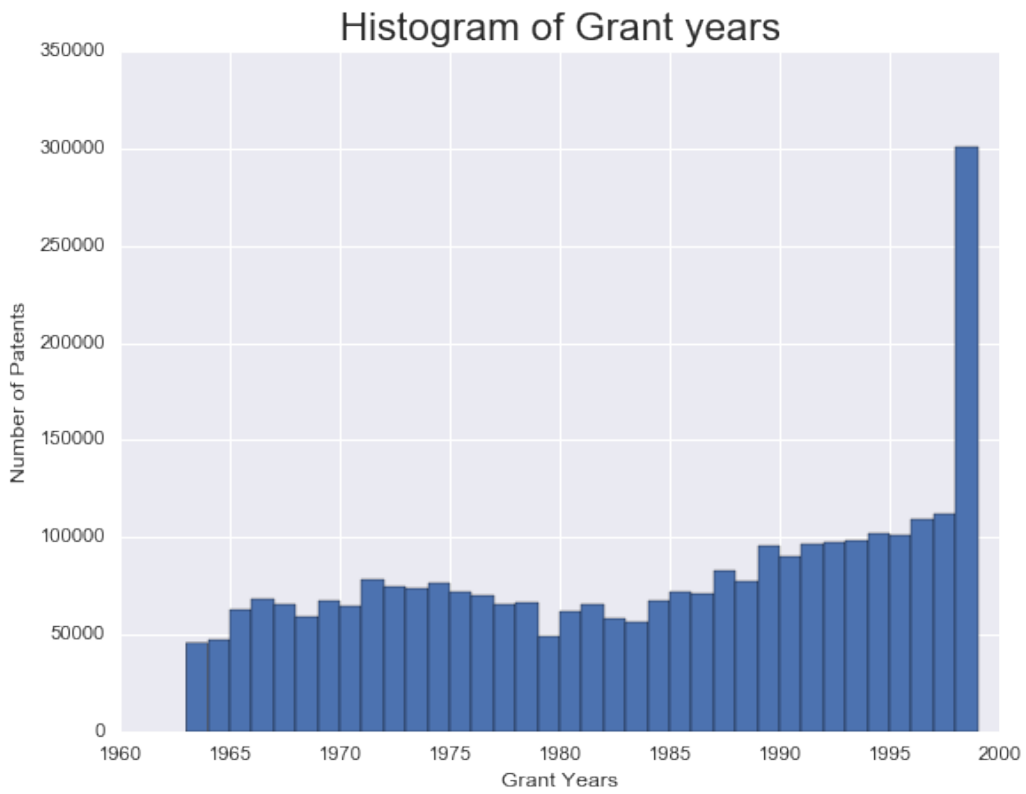
[4] Jerry Thursby, Anne W. Fuller, Marie Thursby, 2009, "US faculty patenting: Inside and outside the university", Research Policy 38, 14–25.

Mean Forward Citation Lag	2074641	4	7	8.3	11.5	5.8	0	96
Mean Backward Citation Lag	2088785	6	10.5	14.1	18.2	11.77	0	154

Descriptive statistics table tells basic statistics information about the dataset. Interestingly, when we see number of citation made and received, the means are 7.7 and 4.7, but the maximum citations are 770 and 779. It means that the distributions have high standard deviation and in fact, it is 9 and 7.35.

(e)

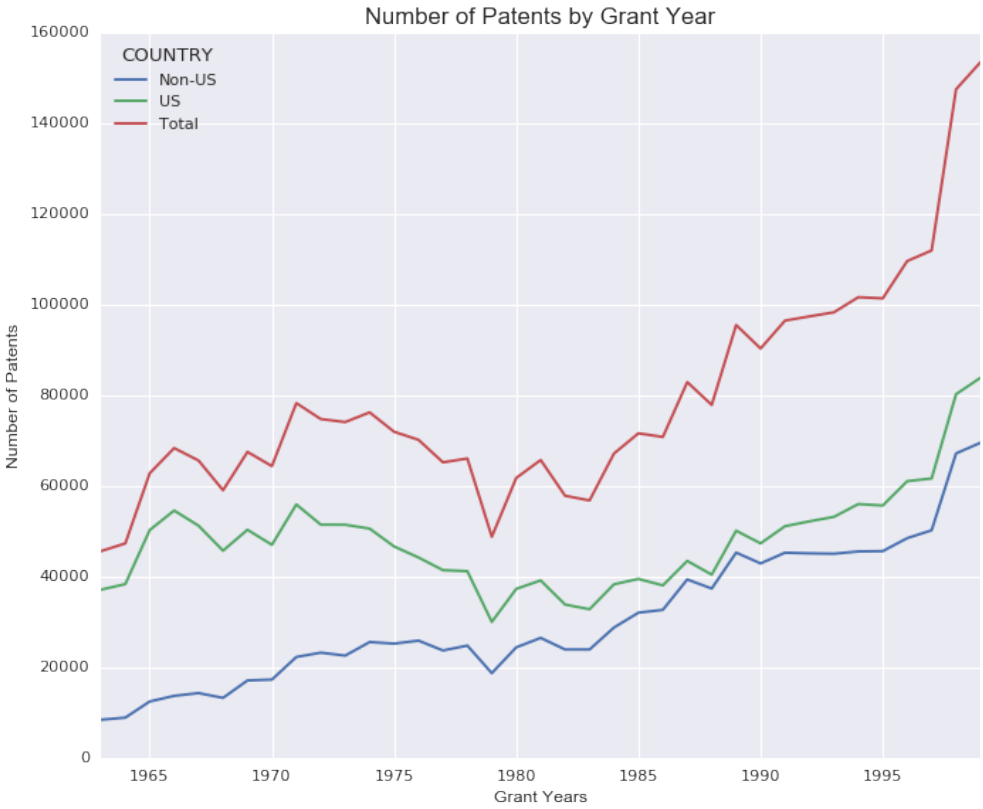
Fig1. Histogram



From above histogram, we can say the number of patents is increasing over 37 years overall. From 1963 to 1983, the trend fluctuated in small range and in fact, in 1979, the the number of patens granted decreased. After 1997, the number of patents started to grow and between 1998 to 1999, the number of patents increased rapidly comparing to the past 35 years.

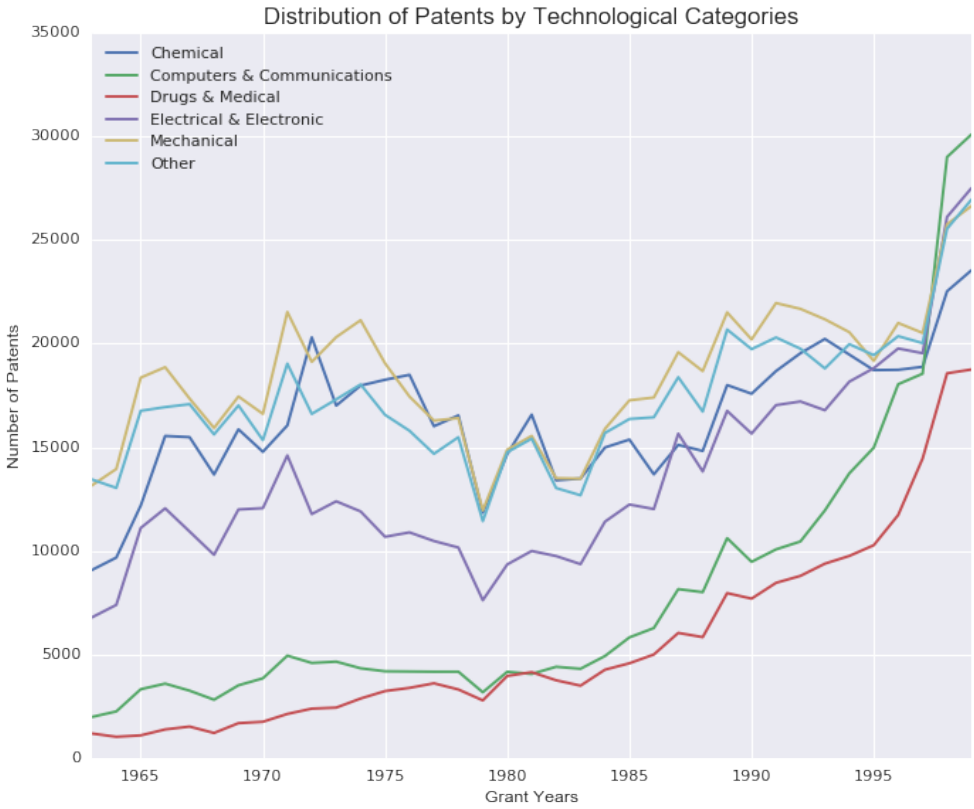
(f)

Fig2. Plot



As we saw in (e), the number of patents is increasing over 37 years and the graph shows the trend more explicitly then histogram. Overall, the number of patents granted in US is higher than in Non-US countries. In 1979, the number of patents rapidly decreased in both US and non-US countries but in 1997 the number of patents rapid increased in both US and non-US countries.

Fig3. Plot



The graph of distribution of patents by technological categories over 37 years shows that the number of patents granted in Mechanical field is generally larger than other fields. Interestingly, the number of patents in the computers $ communication field has rapidly increased than any other fields over 37 years and it has been in the first place from 1997 to 1999. In 1979, the number of patents rapidly decreased in all fields but in 1997, the number of patents rapidly increased in all fields.

References

Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001). "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools." NBER Working Paper 8498.

James Bessen and Robert M. Hunt, 2007, “An Empirical Look at Software Patents”, Journal of Economics & Management Strategy, Volume 16, Issue 1 Spring 2007, 157–189.

Jerry Thursby, Anne W. Fuller, Marie Thursby, 2009, “US faculty patenting: Inside and outside the university”, Research Policy 38, 14–25.

Jung Won Sonn and Michael Storper, 2008, “The increasing importance of geographical proximity in knowledge production: an analysis of US patent citations, 1975 -1997”, Environment and Planning A 2008, volume 40, 1020-1039.

Natarajan Balasubramanian and Jagadeesh Sivadasan, 2011, “What Happens When Firms Patent? New Evidence from U.S. Economic Census Data“, The Review of Economics and Statistics, February 2011, Vol. 93, No. 1, 126-146.