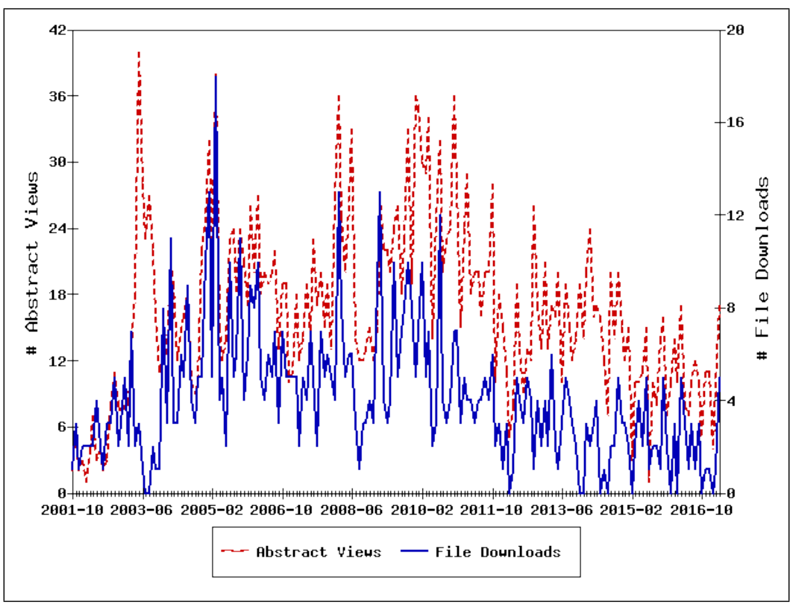
Problem Set #1

MACS 30200, Dr. Evans
Wanlin Ji

Data section

(a) The U.S. patent data used in this paper includes all 2,923,922 United States utility patents in the USPTO's TAF database granted between January 1963 and December 1999. These patent data comes from the The NBER U.S. Patent Citations Data File, a patent and citation database developed and curated by Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001) from National Bureau of Economic Research. These fully functional data is stored and can be accessed freely on the NBER website:http://nber.org/patents/. Due to that NBER requires every researcher using this free data to cite the working paper, we can take a glimpse of the trends of this patent databases by how frequently this working paper is downloaded on EconPapers.(2017)

Figure 1: Access statistics for working paper Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001)



(b) Citation of this working paper and research into this database remains active among recent economic research articles. Patents have long been recognized as a very rich and potentially fruitful source of data for the study of innovation and technical change, according to Hall B. H. et al.(2001). The idea of using patent data in a large scale for economic research goes back at least to Schmookler (1966), followed by Scherer (1982), and Griliches (1984), involved linking past patents based on industry and timing. The NBER U.S. Patent Citations Data File is one of the most important data sources on patents, and have large sclae of uses in recent years' research, including some newest research results. In the area of investment return associate with patents, Williams(2017) empirically used U.S. Patent Citations Data File as one of their data sources to construct three parameter for describing the private return and social contribution coordination mechanism of patent systems.

Another paper published on JPE by Zhen(2017) deny the ignorance hypothesis for patent examiners based on analysis of citations data from U.S. Patent Citations Data File and showed that US examiners tend to devote more search effort to weaker patents, implying that they can identify a substantial portion of the weak patents that they issue.

Balsmeier and Lee (2017) examines the effect of transition to independent boards on increasing citation and find the increase comes mainly from incremental patents, and showed no significant effect on riskier innovation strategy based on U.S. Patent Citations Data File. Their empirical research was published on JFE.

(c) U.S. Patent Citations Data File was developed from multiple different databases by Bronwyn H. Hall Adam B. Jaffe Manuel Trajtenberg. Detailed information was developed on almost 3 million U.S. patents granted between January 1963 and December 1999, mainly coming from numbering and reporting system to USPTO's The Office of Technology Assessment and Forecast database, which could dated to the 1870s. A reasonably broad match of patents to Compustat (the data set of all firms traded in the U. S. stock market) was also included in this NBER data file, but not used by this paper.

This patent data collects two main sets of variables for each patent case as the organization form, those that came from the Patent Office (original variables), and those created from them (constructed variables) by Bronwyn H. Hall Adam B. Jaffe Manuel Trajtenberg. Original Variables include:

1. Patent number
2. Grant year
3. Grant date
4. Application year (starting in 1967)
5. Country of first inventor
6. State of first inventor (if U. S.)
7. Assignee identifier, if the patent was assigned (starting in 1969)
8. Assignee type (i.e., individual, corporate, or government; foreign or domestic)
9. Main U.S. patent class
10. Number of claims (starting in 1975)

And Constructed variables include:

1. Technological category
2. Technological sub-category
3. Number of citations made
4. Number of citations received
5. Percent of citations made by this patent to patents granted since 19637
6. Measure of generality

(d) Among these patent variables, we have categorical variables as well as quantitative variables. So we make a brief descriptive statistics table of 9 major quantitative variables for a glimpse into these data. Here is a table of the mean and five number summary for these variables.

Table 1: Descriptive statistics

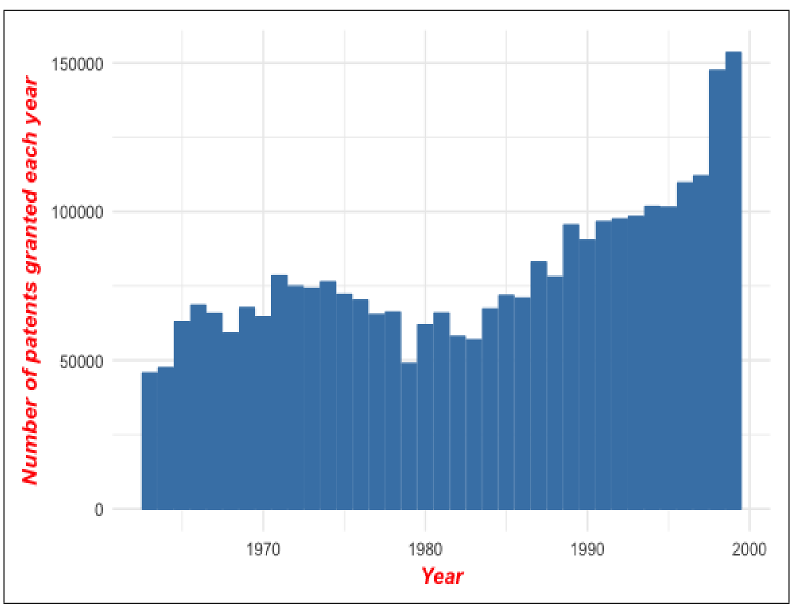
Variables	Mean	Min	1st Quartile	Median	3rd Quartile	Max
Grant Year	1984	1963	1974	1985	1993	1999
Application Year	1983	1901	1974	1985	1992	1999
# of Claims	12.1	1	5	10	16	868
# of Citations Made	7.7	0	3	6	9	770
# of Citations Received	4.8	0	1	3	6	779
Measure of Generality	0.3	0	0	0.4	0.6	0.9
Measure of Originality	0.3	0	0	0.4	0.6	1.0
Mean Forward Cita. Lag	8.3	0	4	7	11.5	96
Mean Backward Cita. Lag	14.1	0	6	10.5	18.2	154

* Calculated under R. All codes disclosed on the Github.

(e) Firstly, we want to pay attention to the number of patents each year as a overall signal of creativity. A quantitative frequency count would be an ideal signal to specify the tendency from 1963 to 1999. We decided to take the grant year for each case in the spreed sheet as our main variable to represent the creativity over the years, assuming the grant date for each patent to truly function.

The reason for not considering application year is determined by the fact the patent entered this database from 1963 to 1999 based on the grant date. And the application year could date back to 1901, where we have no sufficient information to present. The histogram of patents granted each year is below.

Figure 2: Tendency of patents granted each year



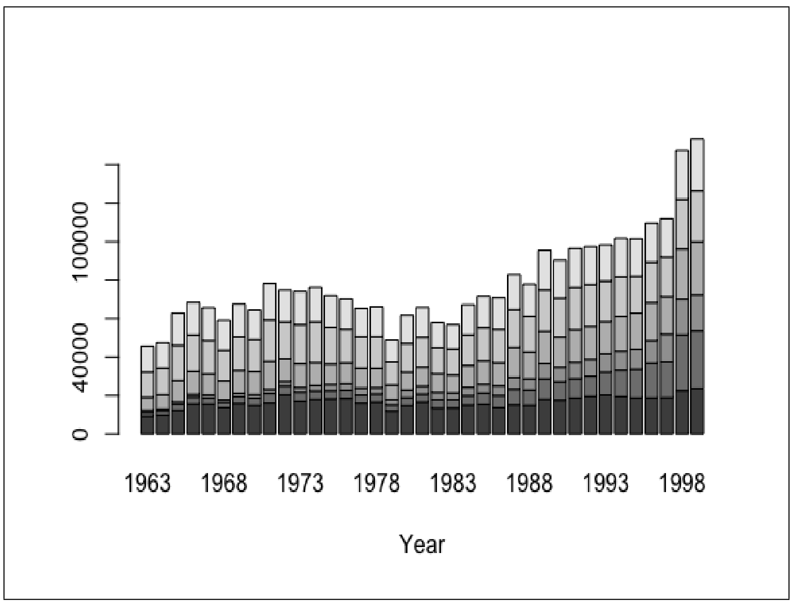
From the graph, we can find that the number of patents remains the overall increasing pattern. However, there is some striking deviations deserving our attention. Firstly, the patents granted number remained largely same or even decreased from

1973 to 1983, with the largest decrease in 1979. Considering macroeconomic environment at that time, the 1973 oil crisis maybe the cause for lack of creativity.

Another pattern deserves attention is that number of patents received each year dramatically increased from 1997 to 1999. We assume the motivation maybe the developing of Internet industry (and mostly we can find from the data, the creativity level under computer and communication) at the end of 20th century, right before the bubble burst. These two are our major assumptions we establish from the graphs.

(f) Since we have made some assumptions about the reason behind the changes of patents over the years, now we want to take a closer look into the composition of these patents to test our hypothesis. Let's begin by taking a look at the proportion of patents.

Figure 3: Proportion of Patents Each Year



The colors ranges from lower to higher is corresponding to categorical code 1 through 6, representing Chemical, Computers Communications, Drugs Medical, Electrical Electronic, Mechanical and Others. We can find that 1, 5 and 6 remains roughly the same over these year, indicating their changes are quite stable across the history.

Let's take a look at our second assumption about Internet development by taking a look at creativity level of Internet industry the 1997-1999 period of our graph. In the last part of our analysis, we made an assumption that the increase from 1997 to 1999 is partly due to Internet. If we found that the number of patents that belongs to category 2 has increased significantly, then we cannot deny our assumption that Internet contributed to the total numbers. From the graph, we can find that the patents under Computers Communications did increase at a higher rate of around 30%. Although this part of patent increased gradually from 1983, this sudden increase did shows some significance comparing to history.

Then we take a look at our first assumption. Due to our database do not shows any macroeconomic indicator, it is impossible to fully testify this assumption. However, we found that the average number of patterns has decreased for almost every category during 1973 and 1983, which is corresponding to the overall depression across these industries.

Reference:

[1] Bronwyn H. Hall Adam B. Jaffe Manuel Trajtenberg, 2001. "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools," NBER Working Papers 8498, National Bureau of Economic Research, Inc.

[2] Access Statistics for the working paper "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools", IDEAS, 2017

[3] Schmookler, J. Invention and Economic Growth. Cambridge: Harvard University Press, 1966.

[4] Scherer, F.M. Inter-Industry Technology Flows and Productivity Growth, Review of Economics and Statistics, 64, November 1982.

[5] Griliches, Zvi, (ed.) RD, Patents, and Productivity, NBER Conference Proceedings. University of Chicago Press, 1984

[6] Heidi L. Williams, 2017. "How Do Patents Affect Research Investments?", NBER Working Papers 23088, National Bureau of Economic Research, Inc.

[7] Balsmeier, Benjamin Fleming, Lee Manso, Gustavo, 2017. "Independent boards and innovation", Journal of Financial Economics, Elsevier, vol. 123(3), pages 536-557.

[8] Lei, Zhen Wright, Brian D., 2017. "Why weak patents? Testing the examiner ignorance hypothesis," Journal of Public Economics, Elsevier, vol. 148(C), pages 43-56.