Problem Set1 Part 1
MACS 30200, Dr. Evans
Mengchen Shi
April 15th, 2018

1 Data Source

The U.S. patent data used in this paper includes all 2,923,922 United States utility patents in the USPTO,Äôs 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/.

2 Key Papers Citation

The U.S. patent data has been used for a variety of papers. Hu and Jaffe(2001) study knowledge diffusion among Asian countries by studying the inventors patented in the US.

Hall, et al.(2005) used this dataset to study the relationship between a firm,Äôs patents and the market value of the firm. They confirm that the more important of patents granted to a firm, the more market value a firm has. This indicates that innovation can boost the market value of a firm. The same dataset is also used to study patent related issues for countries other than the US.

The paper published on JPE by Zhen(2017) denied 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.

3 Data Collection

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,Äôs 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.

4 Descriptive Statistics

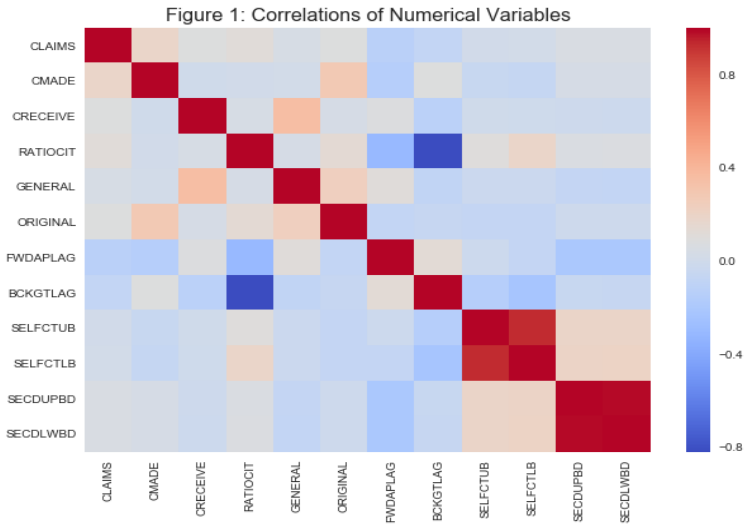
Tabelle 1: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
Number of Claims	1,984,055	12.083	10.268	1	868
Number of Citations Made	2,139,314	7.720	9.000	0	770
Number of Citations Received	2,923,922	4.779	7.346	0	779
Percent of Citations Made to Patents Granted Since 1963	2,088,795	0.843	0.249	0.000	1.000
Measure of Generality	2,240,348	0.321	0.285	0.000	0.940
Measure of Originality	2,042,151	0.349	0.281	0.000	0.951
Mean Forward Citation Lag	2,074,641	8.306	5.804	0.000	96.000
Mean Backward Citation Lag	2,088,785	14.100	11.769	0.000	154.000
Share of Self-Citations Made - Upper Bound	1,703,004	0.136	0.256	0.000	1.000
Share of Self-Citations Made - Lower Bound	1,703,004	0.110	0.218	0.000	1.000
Share of Self-Citations Received - Upper Bound	1,599,160	0.132	0.260	0.000	1.000
Share of Self-Citations Received - Lower Bound	1,599,160	0.125	0.250	0.000	1.000

Table 1 is the description statistics of numerical variables of the data. Number of observations, mean, median, standard deviation, minimum and maximum are provided for each varibale.

5 Key Visualization and Conditional Description

Let's look at the correlation between the numerical variables we described above. The heatmap (Figure 1) shows that there is a negative relationship between RATIO-



CIT (Percent of Citations Made to Patents Granted Since 1963) and BCKGTLAG

(Mean Backward Citation Lag). The correlation between RATIOCIT and FWDAPLAG (Mean Forward Citation Lag) is negative as well.

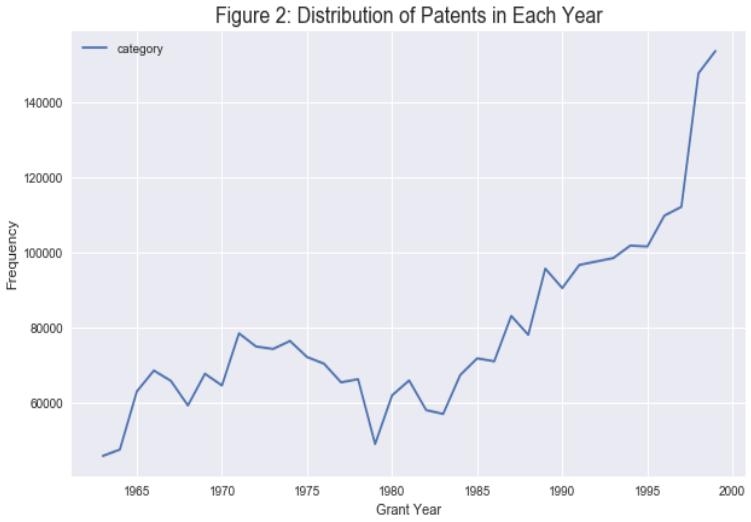


Figure 2 shows that the general trend of number of patents from 1963 to 1999 is upward. However, there is a significant decrease in the number of patents from 1075 to 1979, followed by another decrease in 1983.

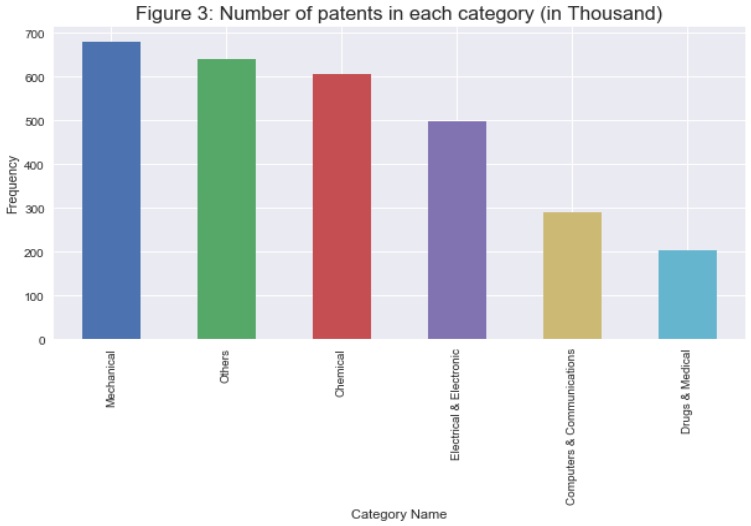
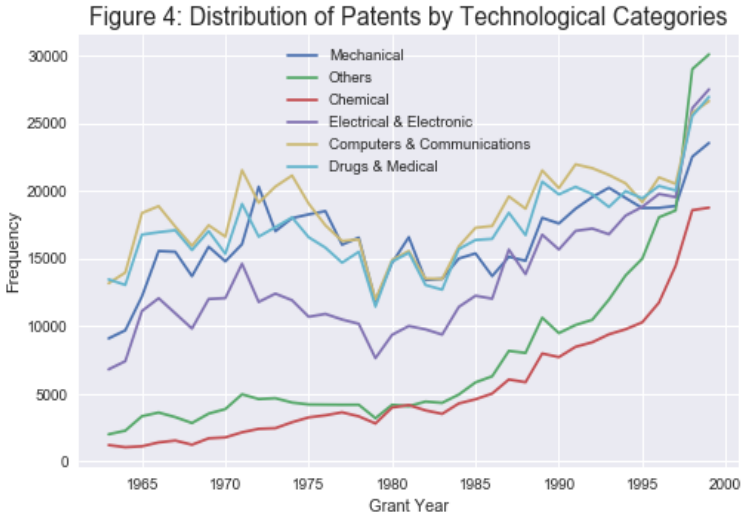


Figure 3 shows that the number of patents in mechanical is the greatest from 1963 to 1999, and the number of patents in drugs medical is the lowest.

We can tell from figure 4 that the patents in chemical and other categories are much lower than other four categories from 1963 to 1990. After 1990, the number of patents in these two categories increase so dramatically that the gap becames smaller and smaller. The number of patents in öthers"became the greatest among all categories in 1998.



6 References

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

Bronwyn H. Hall Adam B. Jaffe Manuel Trajtenberg, 2001. "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools", NBER.

Bronwyn H., Adam Jaffe, and Manuel Trajtenberg. "Market value and patent citations."RAND Journal of economics (2005): 16-38.

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.