Trademark Report Aug 30

Aniket Kesari

August 30, 2017

Overview

This report details the findings from plotting time series plots of the frequencies of trademark rebranding. The two datasets detailed here are the Frequency Table drawn from the full dataset (approximately 8 million observations), and a smaller dataset using only companies that appeared in the SEC file (approximately 50,000 observations).

I found two general findings that motivate further exploration. First, in both the full dataset and the sample data, there is evidence that the number of instances of prior registrations grows quickly throughout the Information Age. Second, this growth is not constant, and instead spikes suddenly around the year 1975. I suspsect there was a change in the way the USPTO measured or collected this data around that time, or some change in the trademarking legal regime that caused more people and corporations to register rebranded marks. That being said, the sudden spike in 1975 does not explain why the pace of growth continued to grow in the years since then, instead of flattening out.

One note about the SEC data: for some reason, filing dates and registration are missing after 1995 or so. However, records of company trademarks still appear, but do not have dates attached to them. This suggests that we need to match the trademark serial numbers to the dates that they were filed, and investigate the reasons as to why that information was missing from the USPTO's dataset in the first place. I don't think this is fatal - it suggests there was some systematic change in recording (possibly just for corporations?). I suspect that the bulk of these missing dates came after 1995 as the missing info far outnumbers the given info, which comports with our overall findings that rates of rebranding dramatically increased in the last 20-30 years.

Full Dataset

As mentioned earlier, the key finding from the data is that starting in 1975, there was a rapid acceleration in re-registering corporate trademarks. Here I present a few rows from the overall data to illustrate the underlying data structure, and then present graphs of the trend for one through ten prior registrations.

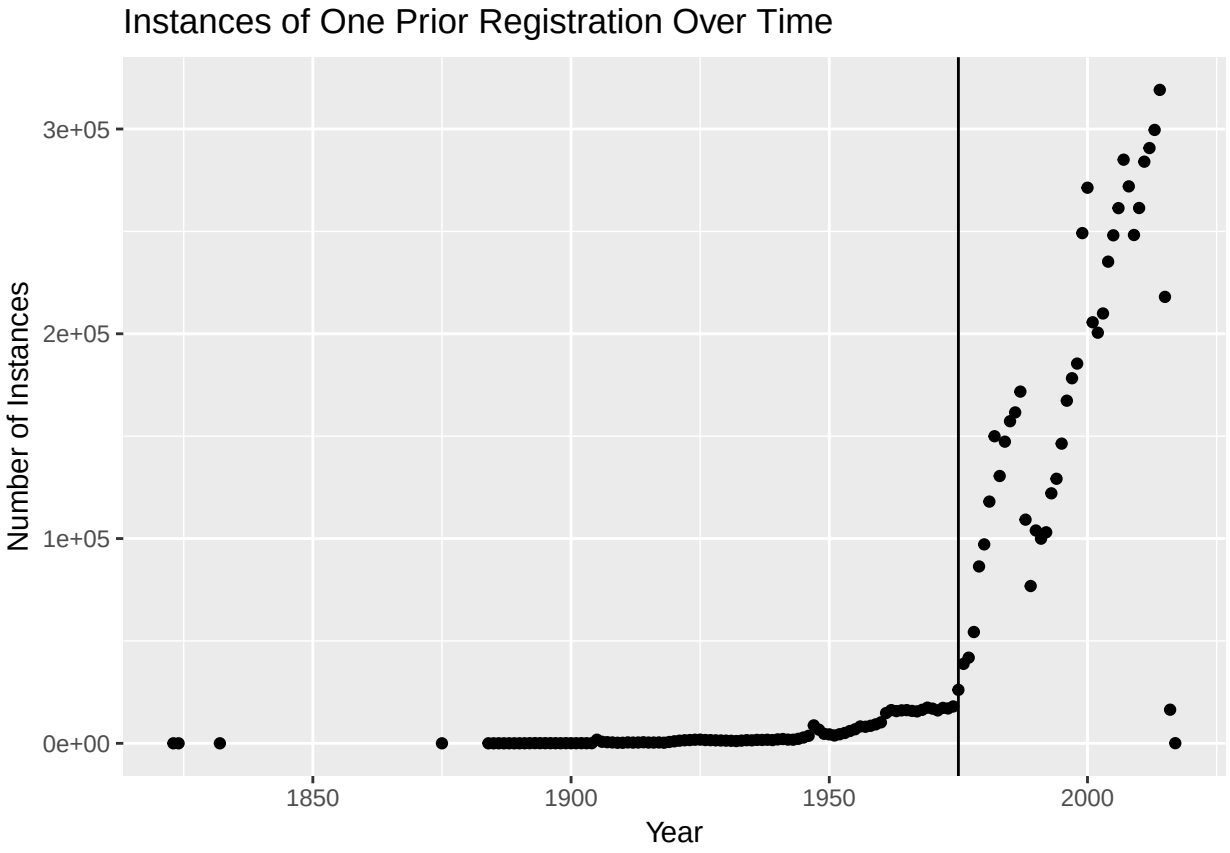
Data Preview

```
##         X Year prior_regs      n         prop
## 1228 1228 1999          5   4430 0.0145262095
## 1492 1492 2004         16    224 0.0007862463
## 1296 1296 2000         24    168 0.0005180133
## 1510 1510 2004         54     54 0.0001895415
## 778   778 1985          6   2628 0.0130318358
## 284   284 1936          7     21 0.0081332301
## 554   554 1952         13     78 0.0108198086
## 868   868 1988         29    464 0.0032151667
## 1240 1240 1999         17    272 0.0008919027
## 1686 1686 2015          1 218024 0.8879367924
```

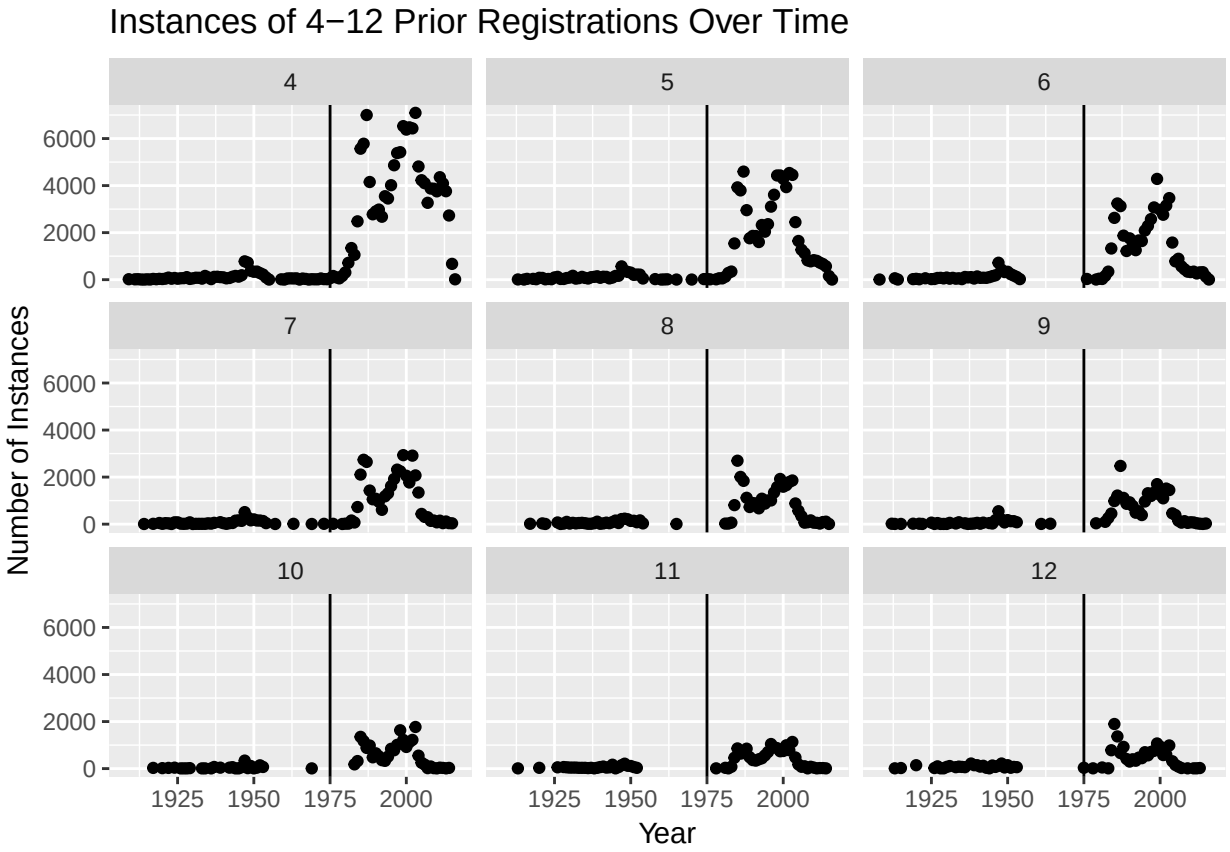
Here, the proper way to read the table (full table available in Dropbox and e-mail), is that each row represents a unique Year and Number of Prior Registrations combination, and "n" represents the count. So for instance, in the year 2015, there were 218,024 instances of trademarks that had one prior mark. Similarly, in 1999, there were 272 instances of trademarks that had 17 prior registrations.

Graphs

The table view of the data suggest growth in the number of trademarks, and generally trademarks with many previous marks become more common over time. This because clearly visualized in the following graphs:




Instances of One, Two, and Three Prior Registrations Over Time
1
2
3
Number of Instances
3e+05
2e+05
1e+05
0e+00
1850
1900
1950
2000
Year



The clear trend that emerges from these graphs is that growth in trademark registration grows explosively around 1975 and continues into the present. I have added a vertical line to each graph to visualize the cut point.

SEC Data

Data Preview

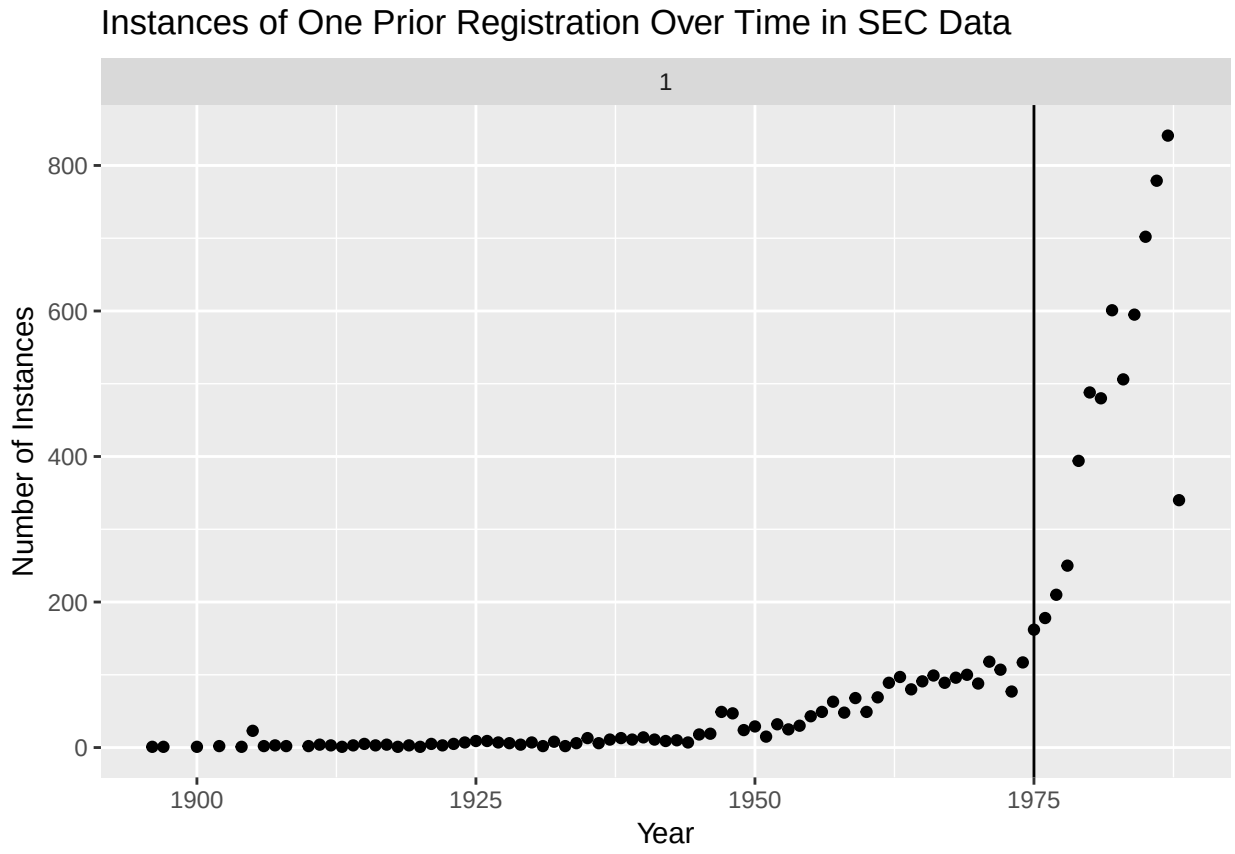
I matched the full dataset to the companies that appear in the SEC dataset. Overall, there were about 700 unique companies (out of about 8000 SEC companies) that appeared in the full dataset, keeping in mind that I only matched to company names that had a maximum Levenshtein distance of "1" (meaning only one deletion, addition, or substitition). After tidying the data into the same format as the frequency table, I found a similar trend as with the full data. However, as mentioned earlier, there is a substantial amount of missing data after the mid-1990s, suggesting there was a systematic change in the way USPTO recorded these trademarks that we need to investigate. Here is a sample of the data, with the larger dataframe in dropbox and attached by e-mail:

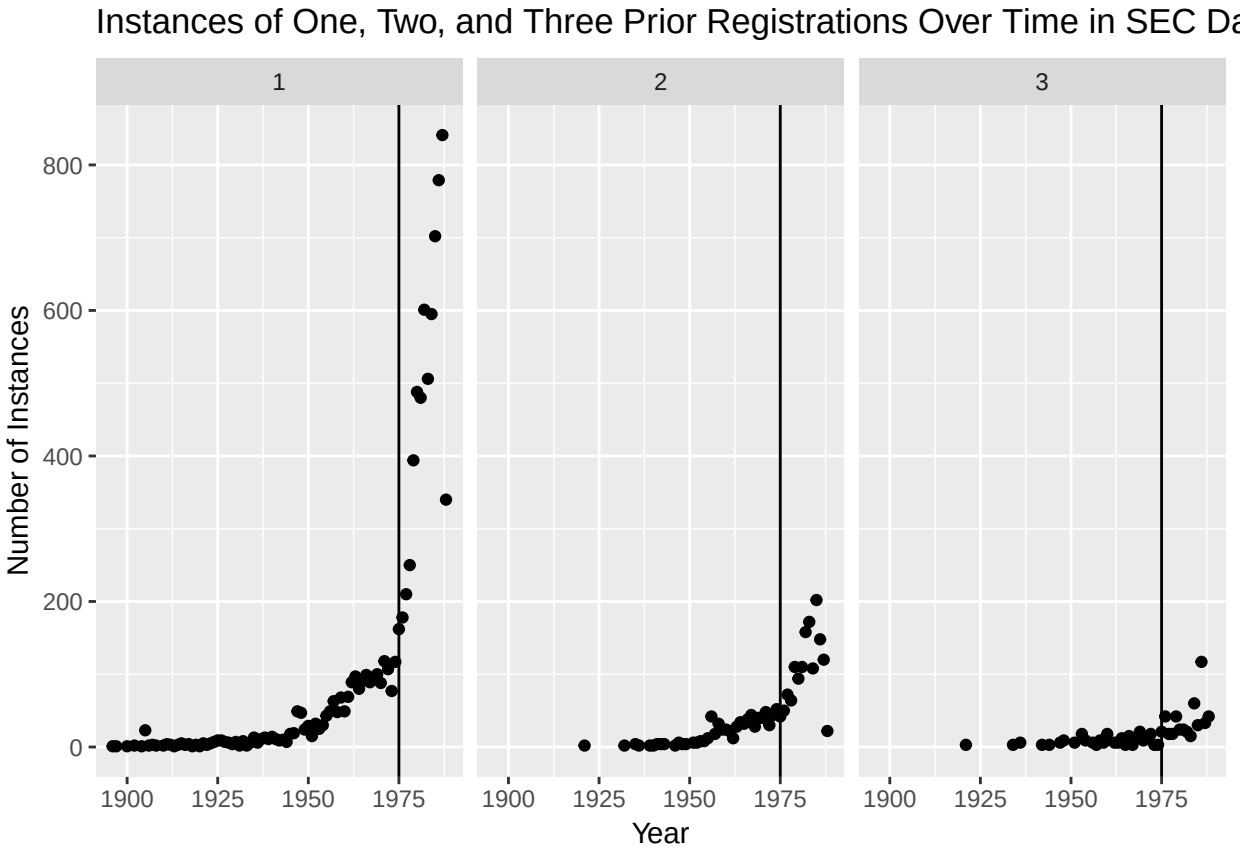
```
## # A tibble: 10 x 4
##     Year prior_regs     n         prop
##    <dbl>      <int> <int>        <dbl>
##  1  1978          1   250 0.0188054762
##  2  1985          8    24 0.0018053257
##  3  1981          1   480 0.0361065142
##  4  1985         10    10 0.0007522190
##  5  1956          2    42 0.0031593200
```

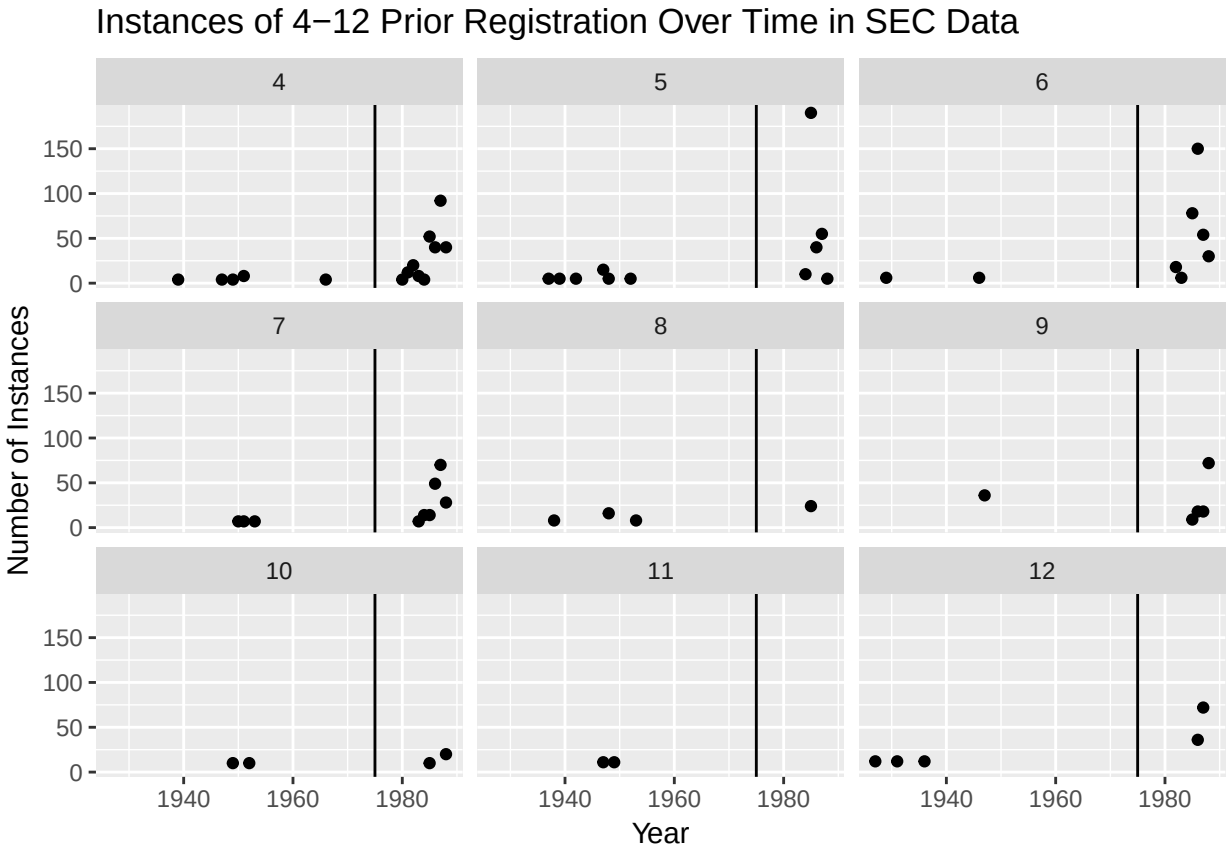
```
##  6  1934          1     6 0.0004513314
##  7  1948          1    47 0.0035354295
##  8  1960          3    18 0.0013539943
##  9  1978          3    18 0.0013539943
## 10  1988          9    72 0.0054159771
```

Graphs

As with the full data, I created graphs using just the SEC data. As with the larger dataset, the SEC companies had the same tendency to suddenly start registering more trademarks after 1975. The missing data makes it impossible to tell if the trend continue past the 1990s, but up until that point, it certainly moved in that direction. The trend becomes discintly less noticeable once we start looking at more than 4 prior registrations, as it does with the full data. But the steep slope is encouraging nonetheless, and I suspect after we figure out the missing data problem, it will look a lot clearer.




Instances of One, Two, and Three Prior Registrations Over Time in SEC Da
1
2
3
Number of Instances
800
600
400
200
0
1900
1925
1950
1975
Year



Summaries by Company

Finally, I constructed a dataset that gives summaries for each company. This process necessarily loses some information in the translation process, but may be helpful for further work distinguishing companies by type, age, etc. Here I provide measures of each company's average number of trademark registrations from 1900-2017, and average number of trademark registrations from 1975 - 2017. I also suggest that we come up with a measure for how often companies rebrand more than once.

```
## # A tibble: 10 x 3
##                          own_name avgperyear avgpost1975
##                             <chr>      <dbl>       <dbl>
##  1                      PPM, INC.   0.500000           1
##  2           TIBCO SOFTWARE INC.    0.500000           1
##  3            ROCKY BRANDS, INC.   1.153846           1
##  4        TOWNSQUARE MEDIA, INC.   0.500000           1
##  5                     IBID, INC.   0.500000           1
##  6                      BRL, INC.   0.500000           1
##  7 EDUCATIONAL DEVELOPMENT CORP.   0.500000          NA
##  8                      JCM, INC.   0.500000           1
##  9             PSB HOLDINGS INC.    1.500000           3
## 10                     XTRA, Inc.   1.500000           3
```

Conclusion + Next Steps

Overall, the most striking finding was the rapid pace of growth in trademark rebrand grequency post-1975. This trend is most clear for trademarks with fewer than four prior marks, but there is also a clear growth in trademarks with several dozen prior marks. In terms of next steps, here are my suggestions:

1. Investigate the reason for missing dates for corporate trademarks.
2. Come up with a plausible story (possible a change in legal regime) that explains the post-1975 change in registration rates.
3. Join the company summary data to Christine's data about firm characteristics. This information should give us covariates for model specification. Depending on the reason for the post-1975 growth, I think a regression discontinuity design model will be appropriate.

Once these three steps are taken, I think we will be on track to have interesting findings, and learn a bit about what characteristics are associated with rebranding.