Paper Title

Paper Subtitle

Malcolm S. Townes

Saint Louis University

Abstract

[Enter text]

Keywords:

**Introduction**

[Enter text]

**Literature Review**

[Enter text]

**Research Questions**

[Enter text]

**Data and Methods**

**Data Sources**

This study uses patent data with constructed variables obtained from the National Bureau of Economic Research (NBER) website. The data file included all utility patents granted from January 1, 1963 to December 30, 1999 listed in the Technology Assessment and Forecast (TAF) database of the United States Patent and Trademark Office (USPTO). The data file contained data on 2,923,922 patents across 23 variables. Table 1 and Table 2 provide information about the variables.

Table 1

Original USPTO Patent Data Variables

| Variable | Variable Type | Extended Name | Description |
| --- | --- | --- | --- |
| PATENT | Numeric | Patent Number | The number assigned to the allowed patent by the USPTO. |
| GYEAR | Numeric | Grant Year | The year the USPTO allowed the patent. |
| GDATE | Numeric | Grant Date | The date the USPTO allowed the patent expressed in terms of the number of weeks elapsed since January 1, 1960. |
| APPYEAR | Numeric | Application Year | The year the patent application was submitted to the USPTO. |
| COUNTRY | Character | Country of First Inventor | The country of citizenship for the first inventor listed on the patent application. |
| POSTATE | Character | State of First Inventor (US) | The state of residency for the first inventor listed on the patent application if the country of citizenship is the United States of America. |
| ASSIGNEE | Numeric | Assignee Identifier | Unique identifier for the assignee of the patent. |
| ASSCODE | Numeric | Assignee Code | A one character code categorizing the type of assignee. |
| CLAIMS | Numeric | Number of Claims | Number of independent and dependent claims on the patent. |
| NCLASS | Numeric | Main Patent Class | A code that categorizes the patent into one of several broad classifications. |

Table 2

Constructed Patent Data Variables

| Variable | Variable Type | Extended Name | Description |
| --- | --- | --- | --- |
| CAT | Numeric | Technological Category | A higher-level classification of the Main Patent Class. |
| SUBCAT | Numeric | Technological Sub-category | The sub-category of the primary technological category to which the patent is assigned. |
| CMADE | Numeric | Number of Citations Made | The number of citations made by the patent. |
| CRECEIVE | Numeric | No. of Citations Received | The number of citations in other patents that reference the patent. |
| RATIOCIT | Numeric | Percent of Citations Made to Patents Granted Since 1963 | The ratio of the number of citations made by all patents granted since 1963 to the total number of citations made by the particular patent. |
| GENERAL | Numeric | Measure of Generality | A measure of how broad the influence of a patent spans across fields as determined by the number of different fields of all patents that cite the patent of interest.  Calculated as the following:  Generalityi = 1 - , where *sij* denotes the percentage of citations received by patent *i* that belong to patent class *j*, out of *ni* patent classes. |
| ORIGINAL | Numeric | Measure of Originality | A measure of the originality of a patent as determined by the number of different fields for all patents cited by the patent of interest.  Calculated as the following:  Originalityi = 1 - , where *sij* denotes the percentage of citations made by patent *i* that belong to patent class *j*, out of *ni* patent classes. |
| FWDAPLAG | Numeric | Mean Forward Citation Lag | The mean time difference between the application or grant date of the patent and that of the other patents citing this patent. |
| BCKGTLAG | Numeric | Mean Backward Citation Lag | The mean time difference between the application or grant date of the patent and that of the patents it cites. |
| SELFCTUB | Numeric | Share of Self-Citations Made – Upper Bound | The number of citations made by the patent to other patents with the same assignee divided by the total number of citations made by all patents with assignee codes. |
| SELFCTLB | Numeric | Share of Self-Citations Made – Lower Bound | The number of citations made by the patent to other patents with the same assignee divided by the total number of citations made by all patents. |
| SECUPBD | Numeric | Share of Self-Citations Received – Upper Bound | The number of citations received by the patent from other patents with the same assignee divided by the total number of citations received by all patents with assignee codes. |
| SECDLWBD | Numeric | Share of Self-Citations Received – Lower Bound | The number of citations received by the patent from other patents with the same assignee divided by the total number of citations received by all patents. |

**Data Selection and Modifications**

I used RStudio to create a subset of the data. After importing the data, I filtered the data for grant years between and including 1995 through 1999. I then filtered that data for patents that had at least 1 claim. I subsequently filtered that data for patents with at least 1 claim received. This generated a subset of 253,328 observations.

I inspected the final data sample using the miss\_var\_summary function to check for missing data by variables to ensure that there was no missing data in the CRECEIVE and CLAIMS variables. I then used the miss\_case\_summary function to check how many observations had missing data in the other variables. I used the get\_dupes function the verify that there were no duplicates in the PATENT variable thus ensuring that it could be used as a unique identifier. I then checked for duplicates across all variables to ensure that each observation was unique.

I used the sample function to select a random sample of 2,000 observations from the subset of 253,328 observations using the seed of 1972 for the function. I then saved this sample data as a CSV file.

**Analysis**

I analyzed the sample data using IBM SPSS Statistics 25. The SPSS 25 analysis output is shown in Exhibit A. I began by verifying that the correct data type and variable type was applied to each variable in the variable view tab. For this study, I assigned CRECEIVE as the dependent variable (DV). I prepared descriptive statistics for the dependent variable using the Analyze > Descriptive Statistics > Frequencies function. The specific statistics calculated for the DV included mean, standard error of mean, median, mode, minimum, maximum, and range, standard deviation, variance, skewness, and kurtosis. I created a histogram with the normal distribution curve superimposed to visually examine the data.

I then used the Graphs > Chart Builder function to create a scatter plot of the DV against the CLAIMS variable, which is one independent variable (IV) of interest. I later used the scatter plot to visually examine whether there was a potential relationship between the two variables.

Finally, I used the Analyze > Regression > Linear function to conduct a multiple regression analysis. The CRECEIVE variable remained the DV. The IVs included CAT, SUBCAT, NCLASS, CLAIMS, GENERAL, ORIGINAL, CMADE, RATIOCIT, BCKGTLAG, FWDAPLAG, SELFCTUB, SELFCTLB, SECDLWBD, and SECDUPBD.

**Discussion**

**Findings and Results**

[Enter text]

**Policy Implications**

[Enter text]

**Limitations of the Analysis**

[Enter text]

**Possible Extensions**

[Enter text]

**Conclusion**

[Enter text]

References

National Bureau of Economic Research. (2018). Patent data, including constructed variables [data file]. Retrieved from http://www.nber.org/patents/

Hall, B. H., Jaffe, A. B. and Trajtenberg, M. (2001). "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools." *NBER Working Paper 8498*. Retrieved from http://www.nber.org/patents/