A Binary Logistic Regression Analysis of Citations Received by Patents:   
Exploring Alternative Approaches to Measuring and Predicting Technology Transfer Outcomes

Malcolm S. Townes

Saint Louis University

Abstract

Technology transfer is the transition of technology or intellectual property from one person or entity to another person or entity. Improving transfer technology outcomes derived from federally-funded research and development (R&D) is a priority for the public policy of the United States of America. As such, understanding the drivers of technology transfer outcomes and improving methods for measuring those outcomes is an important topic for study. The potential benefits of developing models that describe the technology transfer process and facilitate better understanding of the factors associated with successful technology transfer include better management of technological innovation, more effective prioritization of high potential technologies for development, and more efficient resource allocation. This study used binary logistic regression analysis to explore an alternative approach to measuring and predicting technology transfer outcomes using citations received by patents as an indication of technology transfer.

Keywords: technology transfer, university technology transfer, technology commercialization, federally funded research and development, patents, patent citations, logistic regression

**Introduction**

This study continues the investigation of how success in technology transfer s can be defined and measured that I began on Assignments 01, 02, and 03 for SOC 6100 in the Fall 2018 semester. In this study, I conducted a binary logistic regression analysis to investigate the relationships between the number of citations received by patents and several patent citation data variables. The purpose of this study was to better understand the drivers of technology transfer outcomes and explore alternative approaches to measuring them. As with the previous analyses, I used patents issued by the United States Patent and Trademark Office (USPTO) as a proxy for units of technology and the number of citations a given U.S. patent receives from other U.S. patents as a measure of technology transfer.

**Literature Review**

[Enter text]

**Research Questions**

The purpose of this study is to investigate the issue of how technology transfer success can be defined and measured. While most other research conducted in this area conceptualize technology transfer as the transactional exchange of legally recognized intellectual property through a formal license, I consider an alternative conceptualization of technology transfer as a non-transactional exchange of knowledge. To explore this conceptualization, I use patents as a proxy for technology and citations of one patent by other patents as an indication of technology transfer. Specifically, I investigate the following questions:

1. Can patents that receive more than the median number of citations be effectively discriminated from patents that receive the median number of citations or less based on specific patent data related variables?
2. Is there a significant association between whether a patent receives more or less than the median number of citations received by all patents with the year a patent was granted, the number of claims made by a patent, the originality of the patent, and the generality of the patent?

**Data and Methods**

**Data Sources**

This study uses a subset of 2,000 observations taken patent data obtained from the National Bureau of Economic Research (NBER) website. The source data contains both original and constructed variables. The data file included all utility patents granted in the U.S. from January 1, 1963 to December 30, 1999 listed in the Technology Assessment and Forecast (TAF) database of the USPTO. The source file contained data on 2,923,922 patents across 23 variables.

Table 1 details the original USPTO variables of the source data and explanations of their meanings. PATENT indicates the number assigned by the USPTO to the allowed patent. GYEAR is the year the USPTO allowed the patent. APPYEAR is the year the patent application was submitted to the USPTO. GDATE is the number of weeks elapsed since January 1, 1960 to the date the USPTO allowed the patent. COUNTRY is the country of citizenship for the first inventor listed on the patent application. POSTATE is the state of residency for the first inventor listed on the patent application. ASSIGNEE indicates to whom the patent is assigned and is unique to each assignee. ASSCODE indicates the type of assignee. CLAIMS is the number of independent and dependent claims listed on the patent. NCLASS indicates the broad classification for the patent.

Table 1

Original USPTO Variables of Source Data

| Variable | Variable Type | Extended Name | Description |
| --- | --- | --- | --- |
| PATENT | Numeric  Nominal | Patent Number | The number assigned to the allowed patent by the USPTO.  Takes on values integer values between 3070801 and 6009554. |
| GYEAR | Numeric  Interval | Grant Year | The year the USPTO allowed the patent.  Takes on integer values between 1963 – 1999. |
| GDATE | Numeric  Interval | Grant Date | The date the USPTO allowed the patent expressed in terms of the number of weeks elapsed since  January 1, 1960.  Takes on integer values between 156 and 2,028. |
| APPYEAR | Numeric  Interval | Application Year | The year the patent application was submitted to the USPTO.  Takes on integer values between 1963 – 1999. |
| COUNTRY | Character  Nominal | Country of First Inventor | The country of citizenship for the first inventor listed on the patent application.  Takes on values of two character string data. |
| POSTATE | Character  Nominal | 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.  Takes on values of two character string data. |
| ASSIGNEE | Numeric  Nominal | Assignee Identifier | Unique identifier for the assignee of the patent.  Takes on values from 10950 to 99550. |
| ASSCODE | Numeric  Nominal | Assignee Code | A one character code categorizing the type of assignee.  Takes on values from 1 to 7. |
| CLAIMS | Numeric  Interval | Number of Claims | Number of independent and dependent claims on the patent.  Takes on integer values from 1 to . |
| NCLASS | Numeric  Nominal | Main Patent Class | A code that categorizes the patent into one of several broad classifications.  Takes on integer values from 1 to 800. |

Table 2 provides information about the source data constructed variables and explanations of their meanings. CAT is a higher-level classification of the main patent class. SUBCAT is a sub-category of the main patent class. CMADE indicates the number of citations made by the patent. CRECEIVE indicates the number of citations in other patents that reference the patent. RATIOCIT is the ratio of the number of citations made by all patents granted since 1963 to the total number of citations made by the patent. GENERAL is a measure of how broad the influence of a patent spans across fields. ORIGINAL is a measure of the originality of the patent. FWDAPLAG measures forward citations lag. BCKGTLAG measures backward citations lag. SELFCTUB is the upper bound for the share of citations the patent makes to other patents assigned to the same assignee (i.e., self-citations made). SELFCTLB is the lower bound for the share of citations the patent makes to other patents assigned to the same assignee. SECUPBD is the upper bound for the share of citations the patent receives from other patents assigned to the same assignee (i.e., self-citations received). SECDLWBD is the lower bound for the share of citations the patent receives from other patents assigned to the same assignee.

Table 2

Source Data Constructed Variables

| Variable | Variable Type | Extended Name | Description |
| --- | --- | --- | --- |
| CAT | Numeric  Nominal | Technological Category | A higher-level classification of the Main Patent Class.  Takes on integer values from 1 to 6. |
| SUBCAT | Numeric  Nominal | Technological Sub-category | The sub-category of the primary technological category to which the patent is assigned.  Takes on integer values from 1 to 69. |
| CMADE | Numeric  Interval | Number of Citations Made | The number of citations made by the patent.  Takes on integer values from 1 to . |
| CRECEIVE | Numeric  Interval | No. of Citations Received | The number of citations in other patents that reference the patent.  Takes on integer values from 1 to . |
| RATIOCIT | Numeric  Ratio | 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.  Takes on continuous values between 0 and 1. |
| GENERAL | Numeric  Ratio | 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.  Takes on continuous values between 0 and 1. |
| ORIGINAL | Numeric  Ratio | 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.  Takes on continuous values between 0 and 1. |
| FWDAPLAG | Numeric  Ratio | 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.  Takes on continuous values between 0 and 1. |
| BCKGTLAG | Numeric  Ratio | Mean Backward Citation Lag | The mean time difference between the application or grant date of the patent and that of the patents it cites.  Takes on continuous values between 0 and 1. |
| SELFCTUB | Numeric  Ratio | 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.  Takes on continuous values between 0 and 1. |
| SELFCTLB | Numeric  Ratio | 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.  Takes on continuous values between 0 and 1. |
| SECUPBD | Numeric  Ratio | 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.  Takes on continuous values between 0 and 1. |
| SECDLWBD | Numeric  Ratio | 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.  Takes on continuous values between 0 and 1. |

**Data Selection and Modification**

For this study, I only used the GYEAR, CLAIMS, CRECEIVE, GENERAL, and ORIGINAL variables in the path analysis. Based on the results of my previous analyses, I made several modifications to the data and incorporated several previous observations into the initial theoretical path analysis model. I removed the following variables because of high multicollinearity: APPYEAR, BCKGTLAG, FWDAPLAG, SELFCTLB, and SECDLWBD. I created a new variable called CRECBINARY using the Transform > Recode into Different Variables function of SPSS Statistics 25. The CRECBINARY variable is a dichotomous variable calculated from the CRECEIVE variable using a median split of the data. The median of the CRECEIVE data was 2.0 citations received. I coded cases with values of the median or less as 1.0 and cases with 3.0 or more citations received as 2.0. The CRECBINARY variable is what I used as the dependent variable (DV) of interest in the analysis. Table 3 lists the final variables that I used in the analysis.

Table 3

Variables Used in Analysis

| Variable | Variable Type | Extended Name | Description |
| --- | --- | --- | --- |
| GYEAR | Numeric  Interval | Grant Year | The year the USPTO allowed the patent.  Takes on integer values between 1963 – 1999. |
| CLAIMS | Numeric  Interval | Number of Claims | Number of independent and dependent claims on the patent.  Takes on integer values from 1 to . |
| CRECBINARY | Numeric  Nominal | No. of Citations Received | 1 indicates 0-2 citations  2 indicates 3 or more  Takes on an integer value of 1 or 2. |
| GENERAL | Numeric  Ratio | 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.  Takes on continuous values between 0 and 1. |
| ORIGINAL | Numeric  Ratio | 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.  Takes on continuous values between 0 and 1. |

**Analysis and Results**

I used IBM SPSS Statistics 25 to analyze the theoretical path model. I used the Analyze > Regression > Binary Logistic function to perform a binary logistic regression analyses using CRECBINARY and the DV and CLAIMS, GYEAR, GENERAL, and ORIGINAL as the covariates or independent variables (IVs). I used the Enter method for the regression. For the Logistic Regression Options I selected Correlations of estimates, Hosmer-Lemeshow goodness-of-fit, Iteration history, and Confidence interval for exp(B) of 95 percent. I left the Classification cutoff at the 0.5 default, set the maximum iterations to 30, and included the constant in the model. The complete SPSS Statistics 25 output file for the analysis is shown in Appendix A.

**Discussion**

**Policy Implications**

There are several possible policy implications of this study. The analysis provides insight into a topic that is of considerable interest to policymakers. It provides information to help policymakers better understand the drivers of the technology transfer outcomes and identify possible factors that should be considered when forming public policy regarding technology transfer. As such, this study may help policymakers formulate public policy that leads to greater levels of transfer of federally-funded research to the private sector.

**Limitations of the Analysis**

As with any research project or study, this analysis has limitations. Since this analysis was focused on patent data for a five year period from 1995 to 1999, findings based on the data may not be relevant to time frames before or after this period. Additionally, there is a truncation effect in the data. Patents issued in the earliest part of the study period have the potential of receiving citations from patents over a longer period than patents issued in the latter part of the study period. This could potentially be contributing to the skewness observed in the sample distribution.

**Future Study**

There are several opportunities to improve upon and extend the analysis presented in this paper. To begin, it might prove useful to secure more recent data and to examine a subset of data buffered by at least 5 years of data on both sides of the period of study to minimize truncation effects. By merging the data with data containing information about patent assignees, it should be possible to further subset the patent data specifically for university technologies. It might also be useful to introduce the category and subcategory of patents into the analysis to determine if the type of technology is associated with technology transfer outcomes. Finally, comparing an analysis of various baseline path analysis models could help identify a more optimized path analysis model.

**Conclusion**

In this paper, I have continued to explore an alternative conceptualization of technology transfer and an approach to measuring technology transfer based on patent citations received. Using patent data, I conducted a path analysis using a variable measuring patent citations received as the dependent variable and measures of the patent’s originality, generality, claims, the year the patent was granted, and citations ratio as independent variables. The path analysis model that I developed indicated that the generality of a patent was the overwhelming contributor of effects on the number of citations the patent received. Moreover, the study revealed an inverse relationship between the year a patent was granted and the number of citations the patent received.

References

[Enter text]