A Path Analysis Model to Explore the Feasibility of Using

Citations Received by Patents to Assess Technology Transfer Outcomes

Malcolm S. Townes

Saint Louis University

**Abstract**

**Introduction**

This study continues the investigation of how technology transfer success can be defined and measured that I began on Assignments 01 and 02 for SOC 6100 in the Fall 2018 semester. Specifically, I conducted a path analysis to evaluate direct and indirect dependencies among various patent citation data variables. For the dependent variable (DV), I used the natural logarithm of the number of citations a given U.S. patent receives from other U.S. patents (CRECEIVEln) as a measure of technology transfer. The various independent variables

**Data and Methods**

**Data Sources**

This study uses a subset of 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 United States Patent and Trademark Office (USPTO). The source file contained data on 2,923,922 patents across 23 variables.

Table 1 and Table 2 provide information about the source data variables and explanations of their meanings.

Table 1

Source Data Original USPTO 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

Source Data Constructed 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 Modification**

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 with other variables: APPYEAR, BCKGTLAG, FWDAPLAG, SELFCTLB, and SECDLWBD. Based on a scatter plot of the CRECEIVE variable against the CLAIMS variable, I removed observations with CLAIMS greater than 90 claims and CRECEIVE greater than 40 citations received as outliers. This resulted in 42 outlier observations being removed from the analysis for a final sample count of 1,568 observations. Additionally, I created a new variable (CRECEIVEln) using the Transform > Compute Variable function of SPSS Statistics 25. The CRECEIVEln variable is the natural logarithm transformation of the CRECEIVE variable. I decided to perform a transformation on the CRECEIVE variable because the data was skewed to the right (i.e., positively skewed) based on a visual inspection of a histogram. I chose a natural logarithm transformation because it appeared to bring out potential linear relationships between the CRECEIVE variable and primary independent variable (IV) of interest, which is the CLAIMS variable, better than other transformations that I tested. The transformed variable (CRECEIVEln) is what I used as the dependent variable (DV) in the analysis.

**Theoretical Model**

Figure 1 shows the theoretical path model that I developed from logical consideration of the relationships among the variables.



Figure 1. Logically Derived Theoretical Path Model

The theoretical path model uses the CRECEIVEln variable as the final DV of interest. Sub-model 1 uses GENERAL as the DV and ORIGINAL as the IV. Sub-model 2 uses CLAIMS as the DV and ORIGINAL, GENERAL, GYEAR, and RATIOCIT as the IVs. Sub-model 3 uses CRECEIVEln as the DV and ORIGINAL, GENERAL, CLAIMS, GYEAR, and RATIOCIT as the IVs.

**Analysis**

I used IBM SPSS Statistics 25 to analyze the theoretical path model. I used the Analyze > Regression > Linear function to prepare regression analyses for each sub-model. The options I selected included model fit, R square change, part and partial correlations, and collinearity diagnostics for the regression statistics; estimates, a confidence level of 95 percent, and covariance matrix for the regression coefficients; and Durbin-Watson, casewise diagnostics for outliers beyond 3 standard deviations for the residuals. For each model I used the enter method. The complete SPSS Statistics 25 output file for the analysis is shown in Appendix A.

**Results and Findings**

Figure 2 shows theoretical path model with standardized coefficients and p-values from the various regression analyses.



Figure 2. Path Model with Standardized Coefficients and P-Values

Figure 3 shows the final path model with standardized coefficients and p-values. I removed the RATIOCIT variable from the final model because it was not significant at least at the 0.05 level. However, the p-value was 0.055 which is just above the threshold for significance.



Figure 3. Final Path Model with Standardized Coefficients and P-Values

Table 1 summarizes the direct, indirect, and total effects for the final path model.

Table 1

Path Model Direct and Indirect Effects



**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 identify possible factors that should be considered when forming public policy regarding technology transfer. As such, this study may help point policymakers in the right direction.

**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 with at least 10 years of subsequent data 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 subset the data by category and subcategory of the patent since these variables are nominal and could not be directly included in the regression. Finally, comparing an analysis of various baseline path analysis models could result in a more optimized final path analysis model.

**Conclusion**

In this paper, I have continued the exploration of 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, and grant year and citations ratio as independent variables. The path analysis model that I developed explained 43.2 percent of the change in the DV.

Finally, I identify potential policy implications for this study. It provides information to help policymakers identify factors that possibly should be considered when forming public policy regarding technology transfer. As such, this study may help point policymakers in the right direction when forming public policy in this area.

Appendix A. IBM Statistics SPSS 25 Output