Defining and Measuring University Technology Transfer:

The Relationship between Patent Citations Received and Various Patent Data Variables

Malcolm S. Townes

Saint Louis University

Malcom,

I truly enjoyed reading this paper which is nicely organized, following a journal format. I wrote some comments in the text below. For the second assignment, I suggest that you conduct a non-linear transformation and delete unnecessary IVs and outliners.

Is this going to be your dissertation topic? You might want to explore some possibilities in this course.

This is the best assignment 1 that I have reviewed for many years.

Excellent work! A

HM

Abstract

Technology transfer is the transition of technology or intellectual property from one person or entity to another person or entity. Improving the transfer of technology derived from federally funded research and development (R&D) to the private sector is a priority for the public policy of the United States of America (OMB, 2018). University technology transfer (UTT) is a subcategory of the broader technology transfer field. It focuses on the transfer of technology derived from research conducted at universities to the private sector. Identifying the drivers of successful UTT and improving methods to evaluate technology transfer efforts is an important topic for study. There are several potential benefits to developing predictive models describing UTT and understanding the factors associated with successful UTT. Such knowledge would be useful for managing technological innovation and efficiently identifying high potential technologies for further development (Choi, Jang, Jun & Park, 2015). This paper explores the question of how technology transfer success can be defined and how it can be measured. Insights from the broader field of technology transfer should be applicable to the narrower field of UTT. I propose that there is a positive linear relationship between the transfer of a technology as measured by the numbers of citations received by a patent and measures of the patent’s originality, generality, citation lag time, application year, and grant year.

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

**Introduction**

Technology transfer is the transition of technology or intellectual property from one person or entity to another person or entity. Improving the transfer of technology derived from federally funded research and development (R&D) to the private sector to achieve national objectives regarding economic growth and national security is a priority for the public policy of the United States of America (OMB, 2018). In fact, increasing the return on investment from federally funded R&D has been a top priority for the U.S. government going back to the Bush administration of the early 2000s (OMB, 2002) and interest in this topic can be traced as far back as the 1940s (Bush, 1949).

University technology transfer (UTT) is a subcategory of the broader technology transfer field. It focuses on the transfer of technology derived from research conducted at universities to the private sector. Identifying the drivers of successful UTT and improving methods to evaluate technology transfer efforts is an important topic for study. A significant portion of the federal R&D budget goes to American universities to conduct research of interest to the federal government.

In fiscal year 2018, the U.S. federal budget for total R&D was greater than US$142.9 billion (American Association for the Advancement of Science [AAAS], 2018a). In 2016, American universities received roughly US$32.7 billion from the federal government for research and development support (AAAS, 2018b). This amounts to nearly a quarter of the federal R&D budget. With total federal outlays of over US$3.9 trillion, the amount directed to U.S. universities for research is less than 1 percent of total government spending (Congressional Budget Office [CBO], 2018). One might consider this trivial in the grand scheme of things but the amount is significant in absolute terms given that it is greater than the gross domestic product (GDP) of over 112 countries (United Nations [UN], 2017). Moreover, there are other important problems of national interest to which the government could direct those dollars.

There are several potential benefits to developing predictive models describing UTT and understanding the factors associated with successful UTT. Such knowledge would be useful for managing technological innovation and efficiently identifying high potential technologies for further development (Choi, Jang, Jun & Park, 2015). However, there are important questions in the area of UTT that remain unanswered or underexplored including (1) how should success in UTT be defined, (2) how should UTT performance be measured, and (3) what are the drivers of successful UTT.

**Literature Review**

Several researchers have explored the issue of evaluating and predicting UTT performance. Most of these studies define UTT success in terms of the transfer of legally protected intellectual property through licenses. Anderson, Daim & Lavoie (2007) used an approach based on data envelopment analysis (DEA) to evaluate UTT productivity. Their study included patent applications and patents allowed as outputs. Ho, Liu, Lu & Huang (2014) investigated the efficiency of UTT in different stages of the technology transfer process. Their approach included patent applications and patents allowed among the intermediate input factors in a two-stage, networked-based DEA model. Choi, Jang, Jun & Park (2015) developed a predictive model based on patent analysis for evaluating the transfer potential of a technology and determining the relationship between technology transfer and a range of patent data variables. Their study included a multiple regression analysis on results of social network analysis (SNA) graphs. They proposed four models for predicting technology transfer. Three of the four models included patent citation variables as statistically significant contributors.

**Research Questions**

The purpose of this study is to investigate the issue of how technology transfer success can be defined and how it can be measured. Insights from the broader field of technology transfer should be applicable to the narrower field of UTT. While most other research conducted in this area conceptualize technology transfer as the transfer of legally recognized intellectual property through licensing, I consider an alternative conceptualization of technology transfer as reflected by the transfer of knowledge as exhibited by the number of citations a patent receives. I explore an alternative approach to measuring this transfer of knowledge by using citations received as an indication of the importance of the citied technology (Hall, Jaffe & Trajtenberg, 2001). Specifically, I address the following questions:

1. Can patent citations data serve as a useful measure of technology transfer success?
2. Is technology transfer success as measured by patent citations received significantly related to other variables captured in the patent citation data?

I use patents as a proxy for units of technology. I propose that there is a positive linear relationship between the transfer of a technology as measured by the numbers of citations received by a patent and measures of the patent’s originality, generality, citation lag time, application year, and grant year. I theorize that certain independent variables (IVs) are associated with the dependent variable (DV) as follows:

CRECEIVE = β0 + β1(GRYEAR) + β2(APPYEAR) + β3(CLAIMS) + β4(GENERAL) + β5(ORIGINAL) + β6(CMADE) + β7(RATIOCIT) + β8(BCKGTLAG) + β9(FWDAPLAG) + β10(SELFCTUB) + β11(SELFCTLB) + β12 (SECDLWBD) + β13(SECDUPBD)+ ε

The meaning of each variable is explained in the Data and Methods section below. The null hypothesis is that the number of citations a patent receives is NOT related the IVs. That is, the regression coefficient for all IVs is equal to zero:

H0: β1 = β2 … = β13 = 0

The alternative hypothesis is that the number of citations a patent receives is related to at least one of the IVs. Stated another way, the regression coefficient for at least one IV is NOT equal to zero:

HA: β ≠ 0 for at least one independent variable

**Data and Methods**

**Data Sources**

This study uses patent data obtained from the National Bureau of Economic Research (NBER) website. The 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 file contained data on 2,923,922 patents across 23 variables. Table 1 and Table 2 provide information about the variables and explanations of their meanings.

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 raw data. After importing the raw data, I filtered it for grant years between and including 1995 through 1999. I then filtered that data for patents that had at least one (1) claim. I subsequently filtered that data for patents with at least one (1) citation received. This generated a subset of 253,328 observations.

I inspected this intermediate data set 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 determine how many observations had missing data in the other variables. I used the get\_dupes function to 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 set.seed function with a seed value of 1972. I then saved this sample data as a .csv file.

**Analysis**

I analyzed the sample data using IBM SPSS Statistics 25. 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 DV. I prepared descriptive statistics for the DV 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 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 GRYEAR, APPYEAR, CLAIMS, GENERAL, ORIGINAL, CMADE, RATIOCIT, BCKGTLAG, FWDAPLAG, SELFCTUB, SELFCTLB, SECDLWBD, and SECDUPBD. Although I suspect that the variables NCLASS, CAT, and SUBCAT all influence the number of citations a patent receives, I specifically excluded them from the analysis because they are nominal variables and would thus violate the assumptions of the linear regression. The options I selected for the linear regression statistics included model fit, R squared change, and part and partial correlations; estimates, a confidence level of 95 percent, and covariance matrix for the regression coefficients; and Durbin-Watson and casewise diagnotics for outliers beyond 3 standard deviations for the residuals.

**Results**

Table 3 shows descriptive statistics for the dependent variable CRECEIVE. The values of the DV ranged from 1 to 112 citations received. The mean value for the DV was 3.18 while the median value was 2.00 suggesting that the sample distribution is skewed. The standard deviation was 4.309, which seems relatively wide.

Table 3

Descriptive Statistics for CRECEIVE

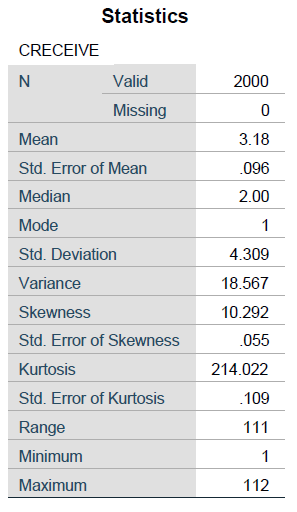


Figure 1 shows a histogram of the DV. Visual inspection of the data confirms that the sample distribution is in fact positively skewed to the right.

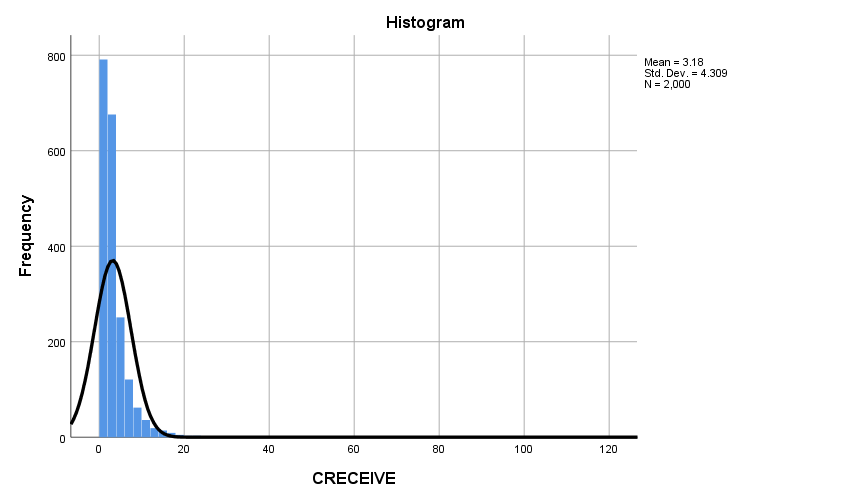


Figure 1. Histogram of CRECEIVE Values

Figure 2 shows a scatter plot of the DV against the independent variable CLAIMS. I chose this particular IV because conceptually one might expect a positive linear relationship between the number of citations a patent receives and the number of claims in the patent. More patent claims might be indicative of broad applicability of the technology or simply more opportunities for the patent to be cited. However, this does not appear to be the case. I discerned no obvious pattern upon visual inspection of the scatter plot.

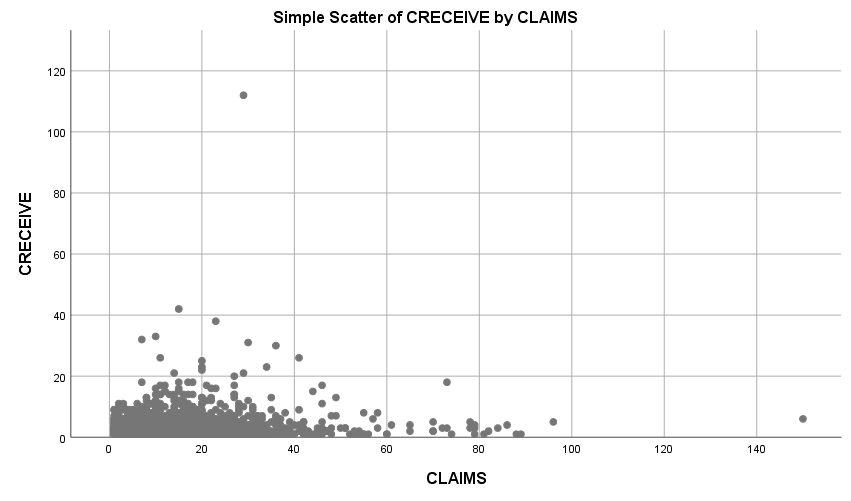


Figure 2. Scatter Plot of CRECEIVE Against CLAIMS

Table 4 shows the model summary for the regression. The value of the correlation coefficient, R, is 0.443 suggesting the amount of change in the DV (CRECEIVE) is only moderately determined by the IVs included in the model. The Adjusted R-Square value indicates that 19.1 percent of the variation in the number of citations a patent receives can be explained by the IVs included in the model.

Table 4

Regression Model Summary



Table 5 shows the analysis of variance (ANOVA) results. The significance value of the regression is 0.000, which is less than 0.05 to which the confidence level was set. This suggests that the probability that the variation in the DV variable explained by the IVs in the model is due to chance is low. The F statistic of 33.617 indicates that at least one variable in the regression was significant.

Table 5

ANOVA Results



Table 6 shows the results for the regression coefficients. The regression produced the following equation:

CRECEIVE = 1509.628 – 0.281(APPYEAR) – 0.064(BCKGTLAG) + 0.027(CLAIMS) – 0.034(FWDAPLAG) + 5.826(GENERAL) – 0.474(GYEAR) – 1.001(ORIGINAL) – 3.184(SECDLWBD) + 3.09(SECDUPBD) – 1.653(SELFCTLB) + 1.188(SELCTUB)

The GENERAL, SECDLWBD, and SECDUPBD variables have the greatest impact on the value of the DV. Somewhat surprisingly, the ORIGINAL variable, which is a measure of the originality of the patent, has a relatively small impact on the value of the DV. Moreover, it is inversely related to the value of the DV, which was not expected. The significance value for the variables FWDAPLG, SECDLWBD, SECDUPBD, SEFLCTLB, and SELFCTUB were all greater than 0.05, which indicates that these variables were not significant. While the significance value for the CLAIMS variable was less than 0.05, the impact of this IV on the DV was unexpectedly small. This seems to confirm what was observed in the scatter plot shown in Figure 2 above.

Table 6

Regression Coefficients



Table 7 provides collinearity statistics for the regression coefficients. The results suggest that the variables APPYEAR, FWDAPLAG, and GYEAR have a high degree of multicollinearity based on the collinearity statistic tolerance value. Likewise, the variables SECDLWBD and SECDUPBD have a high degree of multicollinearity. The same is true of the SELFCTLB and SELFCTUB variables. This suggests that these groups of variables are likely measuring the same characteristic.

Table 7

Collinearity Statistics

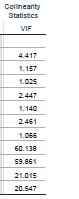


Table 8 shows the casewise diagnostics for the regression. There were 16 patents in which the DV had values beyond 3 standard deviations. Case number 230 was the most extreme having a value that was greater than 25 standard deviations from the mean. Ten other observations had values that were greater than 4 standard deviations from the mean. This is likely related to the skewness in the sample distribution.

Table 8

Casewise Diagnostics



**Discussion**

**Findings**

Based on the results of the analysis, I reject the null hypothesis because at least one of the regression coefficients was not equal to zero. This suggests that the fit of the model was improved by including at least one of the IVs compared with a model containing no IVs.

**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. Finally, 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 if 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. Given that the sample distribution for the DV is skewed, a transformation should be applied to it so that it approximates a normal distribution. Repeating the regression analysis with the removal of variables that exhibit high degrees of multicollinearity or regression coefficients near zero might also be worthwhile. Finally, comparing a baseline regression model to subsequent regression models with additional independent variables included could help reveal insights about the impact of those variables on the number of citations received by a patent as a measure of technology transfer.

**Conclusion**

In this paper, I have explored an alternative conceptualization of technology transfer and an approach to measuring technology transfer based on patent citations received. Using patent data, I conducted a multiple regression analysis using a variable measuring patent citations received as the dependent variable and measures of the patent’s originality, generality, citation lag time, application year, and grant year as independent variables.

The regression model generated by my analysis only explained 19.1 percent of the change in the DV. Several independent variables exhibited unexpected relationships with the DV, while others were insignificant or uncorrelated with 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.

References

American Association for the Advancement of Science [AAAS]. (2018a). *Defense, Nondefense, and Total R&D, 1976-2018* [Data file]. Retrieved from https://www.aaas.org/page/historical-trends-federal-rd

American Association for the Advancement of Science [AAAS]. (2018b). *Federal Support for Universities by Agency, 1990-2016 (Obligations)* [Data file]. Retrieved from https://www.aaas.org/page/rd-colleges-and-universities

Anderson, T. R., Daim, T. U., & Lavoie, F. F. (2007). Measuring the efficiency of university technology transfer. *Technovation*, 27(5), 306-318. doi:10.1016/j.technovation.2006.10.003

Bush, V. (1945). *Science, the endless frontier*. A report to the President: Washington, U.S. Government printing office, 1945.

Choi, J., Jang, D., Jun, S., & Park, S. (2015). A Predictive Model of Technology Transfer Using Patent Analysis. *Sustainability* (2071-1050), 7(12), 16175

Congressional Budget Office [CBO]. (2018). Historical Budget Data [Data file]. *The Budget and Economic Outlook: 2018 to 2028* Retrieved from https://www.cbo.gov/about/products/budget-economic-data#2

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/

Ho, M. H. C., Liu, J. S., Huang, C. C., & Lu, W. M. (2014). A new perspective to explore the technology transfer efficiencies in US universities. *Journal of Technology Transfer*, 39(2), 247-275. doi:10.1007/s10961-013-9298-7

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

Office of Management and Budget [OMB]. (2002). *The President's Management Agenda*. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a394421.pdf

Office of Management and Budget [OMB]. (2018). *The President's Management Agenda*. Retrieved from https://www.whitehouse.gov/wp-content/uploads/2018/03/Presidents-Management-Agenda.pdf

United Nations. (2017). *GDP and its breakdown at current prices in U.S. dollars* [Data file]. Retrieved from http://unstats.un.org/unsd/snaama/dnltransfer.asp?fID=2