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

Constructed Variables of Source Data

| 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.  1 indicates Chemical  2 indicates Computers & Comm.  3 indicates Drugs and Medical  4 indicates Electrical & Electronic  5 indicates Mechanical  6 indicates All Others |
| 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 those 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. |

Table 3

Additional Variables Created for the Analysis

| Variable | Variable Type | Extended Name | Description |
| --- | --- | --- | --- |
| CAT02 | Numerical  Nominal | CAT 02 Indicator | Indicator variable  0 indicates patent not assigned to CAT 02  1 indicates patent assigned to CAT02 |
| CAT03 | Numerical  Nominal | CAT 03 Indicator | Indicator variable  0 indicates patent not assigned to CAT 03  1 indicates patent assigned to CAT03 |
| CAT04 | Numerical  Nominal | CAT 04 Indicator | Indicator variable  0 indicates patent not assigned to CAT 04  1 indicates patent assigned to CAT04 |
| CAT05 | Numerical  Nominal | CAT 05 Indicator | Indicator variable  0 indicates patent not assigned to CAT 05  1 indicates patent assigned to CAT05 |
| CAT06 | Numerical  Nominal | CAT 06 Indicator | Indicator variable  0 indicates patent not assigned to CAT 06  1 indicates patent assigned to CAT06 |
| CRECbinary | Numeric  Nominal | Number of Citations Received | 0 indicates 0 citations  1 indicates 1 or more  Takes on an integer value of 0 or 1. |
| CRECmdnSplt | Numeric  Nominal | Median Citations Received | 0 indicates less than or equal to median  1 indicates greater than median  Takes on an integer value of 0 or 1. |
| CRECordinal | Numeric  Ratio | Ordinal Level of Citations Received | The level of citations in other patents that reference the patent.  1 indicates 1 citation  2 indicates 2 citations  3 indicates 3 citations  4 indicates 4 citations  5 indicates 5 citations  6 indicates 6 citations  7 indicates 7 citations  8 indicates 8 citations  9 indicates 9 citations  10 indicates 10 citations  11 indicates 11 citations  12 indicates 12 citations  13 indicates 13 citations  14 indicates 14 citations  15 indicates 15 or more  Takes on integer values between 0 and 15. |
| CRECEIVEsqrt | Numerical  Ratio | Square Root of CRECEIVE | The square root of the value of the CRECEIVE variable |

Table 4

Correlation Matrix

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1. PATENT | 1.0000 |  |  |  |  |  |  |  |  |  |
| 2. GYEAR | 0.9856 | 1.0000 |  |  |  |  |  |  |  |  |
| 3. CRECEIVE | -0.1590 | -0.1495 | 1.0000 |  |  |  |  |  |  |  |
| 4. CAT | -0.0336 | -0.0268 | -0.0956 | 1.0000 |  |  |  |  |  |  |
| 5. CLAIMS | 0.0407 | 0.0402 | 0.1317 | -0.0172 | 1.0000 |  |  |  |  |  |
| 6. CMADE | 0.1023 | 0.0942 | 0.0632 | 0.0402 | 0.1656 | 1.0000 |  |  |  |  |
| 7. GENERAL | -0.1264 | -0.1189 | 0.4175 | -0.1055 | 0.1200 | 0.0882 | 1.0000 |  |  |  |
| 8. ORIGINAL | 0.0803 | 0.0809 | 0.0013 | -0.0607 | 0.0350 | 0.2531 | 0.2147 | 1.0000 |  |  |
| 9. FWDAPLAG | -0.1113 | 0.1052 | -0.1981 | 0.0149 | -0.0833 | -0.0734 | -0.2969 | -0.0092 | 1.0000 |  |
| 10. BCKGTLAG | -0.0113 | -0.0075 | -0.1357 | 0.1761 | -0.0714 | 0.0131 | -0.1064 | 0.2360 | 0.1309 | 1.0000 |

Table 5

Binomial Logistic Regression for CRECbinary

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Coefficient Estimate | Std. Error | | Z value | | p value | | Sig. | Exp (Coef) | | | C.I. Lower Exp(Coef) | | C.I. Upper Exp(Coef) |
| Intercept | 18.62 | 4.195(104) | | 0.000 | | 1.000 | |  | 1.222(108) | | | 0.0000 | | Inf |
| GYEAR | 0.31 | 0.880 | | 0.355 | | 0.723 | |  | 1.366 | | | 0.3095 | | 22.7335 |
| CAT02 | 51.87 | 2.259(105) | | 0.000 | | 1.000 | |  | 3.365(1022) | | | 0.0000 | | Inf |
| CAT03 | 57.59 | 3.690(104) | | 0.002 | | 0.999 | |  | 1.023(1025) | | | 0.0000 | | Inf |
| CAT04 | 76.29 | 3.578(104) | | 0.002 | | 0.998 | |  | 1.359(1033) | | | 0.0000 | | Inf |
| CAT05 | 18.54 | 4.344(104) | | 0.000 | | 1.000 | |  | 1.130(108) | | | 0.0000 | | Inf |
| CAT06 | 61.74 | 3.574(104) | | 0.002 | | 0.999 | |  | 6.530(1026) | | | 0.0000 | | Inf |
| CMADE | -0.02 | 0.188 | | -0.105 | | 0.916 | |  | 0.980 | | | 0.5223 | | 1.3160 |
| CLAIMS | -0.02 | 0.174 | | -0.136 | | 0.892 | |  | 0.977 | | | 0.5599 | | 1.2531 |
| ORIGINAL | 2.19 | 4.565 | | 0.480 | | 0.631 | |  | 8.956 | | | 0.0017 | | 2.66(106) |
| GENERAL | 70.98 | 3.623(104) | | 0.002 | | 0.998 | |  | 6.686(1030) | | | 0.0000 | | Inf |
| FWDAPAG | -68.76 | 2.461(103) | | -0.028 | | 0.978 | |  | 1.376(10-30) | | | 0.0000 | | 0.0000 |
| BCKGTLAG | 0.0028 | 0.054 | | 0.052 | | 0.959 | |  | 1.003 | | | 0.8447 | | 1.1269 |
| \*\*\* 0.001 \*\* 0.01 \* 0.05 | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | |
| -2 log likelihood null | | | 1,744.459 | |  | |  | | |  |  | |  | |
| -2 log likelihood residual | | | 9.167 | |  | |  | | |  |  | |  | |
|  | | |  | |  | |  | | |  |  | |  | |
| McFadden R2 | | | 0.9948 | |  | |  | | |  |  | |  | |
|  | | |  | |  | |  | | |  |  | |  | |

Table 6

Binomial Logistic Regression for CRECmdnSplt

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Coefficient Estimate | Std. Error | | Z value | | p value | | Sig. | Exp (Coef) | | | C.I. Lower Exp(Coef) | | C.I. Upper Exp(Coef) |
| Intercept | 530.049 | 71.396 | | 7.424 | | 0.0000 | | \*\*\* | 1.575(10230) | | | 6.937(10169) | | 2.817(10291) |
| GYEAR | -0.266 | 0.036 | | -7.435 | | 0.0000 | | \*\*\* | 0.766 | | | 0.7138 | | 0.8215 |
| CAT02 | 1.002 | 0.223 | | 4.496 | | 0.0000 | | \*\*\* | 2.722 | | | 1.7645 | | 4.2277 |
| CAT03 | 0.725 | 0.218 | | 3.323 | | 0.0008 | | \*\*\* | 2.065 | | | 1.3482 | | 3.1728 |
| CAT04 | 0.401 | 0.184 | | 2.178 | | 0.0294 | | \* | 1.494 | | | 1.0417 | | 2.1464 |
| CAT05 | 0.115 | 0.182 | | 0.634 | | 0.5262 | |  | 1.122 | | | 0.7860 | | 1.6035 |
| CAT06 | 0.350 | 0.186 | | 1.878 | | 0.0603 | |  | 1.418 | | | 0.9856 | | 2.0449 |
| CMADE | 0.024 | 0.007 | | 3.226 | | 0.0013 | | \*\* | 1.024 | | | 1.0100 | | 1.0402 |
| CLAIMS | 0.015 | 0.006 | | 2.441 | | 0.0146 | | \* | 1.015 | | | 1.0030 | | 1.0275 |
| ORIGINAL | -1.101 | 0.221 | | -4.978 | | 0.0000 | | \*\*\* | 0.333 | | | 0.2148 | | 0.5114 |
| GENERAL | 4.277 | 0.231 | | 18.487 | | 0.0000 | | \*\*\* | 72.030 | | | 46.0593 | | 114.1223 |
| FWDAPAG | -0.199 | 0.027 | | -7.276 | | 0.0000 | | \*\*\* | 0.819 | | | 0.7755 | | 0.8635 |
| BCKGTLAG | -0.015 | 0.005 | | -2.999 | | 0.0027 | | \*\* | 0.986 | | | 0.9761 | | 0.9948 |
| \*\*\* 0.001 \*\* 0.01 \* 0.05 | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | |
| -2 log likelihood null | | | 2,734.8 | |  | |  | | |  |  | |  | |
| -2 log likelihood residual | | | 1,927.5 | |  | |  | | |  |  | |  | |
|  | | |  | |  | |  | | |  |  | |  | |
| McFadden R2 | | | 0.2952 | |  | |  | | |  |  | |  | |
|  | | |  | |  | |  | | |  |  | |  | |

Table 7

Ordinal Logistic Regression

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Coefficient Estimate | Std. Error | | t value | | p value | | Sig. | Exp (Coef) | | | C.I. Lower Exp(Coef) | | C.I. Upper Exp(Coef) |
| Variables: |  |  | |  | |  | |  |  | | |  | |  |
| GYEAR | -0.3659 | 0.00010 | | -3,750.527 | | 0.0000 | | \*\*\* | 0.694 | | | 0.6935 | | 0.6936 |
| CAT02 | 0.9424 | 0.11150 | | 8.453 | | 0.0000 | | \*\*\* | 2.566 | | | 1.8637 | | 3.5377 |
| CAT03 | 0.9285 | 0.10950 | | 8.480 | | 0.0000 | | \*\*\* | 2.531 | | | 1.8336 | | 3.4959 |
| CAT04 | 0.4930 | 0.09072 | | 5.435 | | 0.0000 | | \*\*\* | 1.637 | | | 1.2525 | | 2.1411 |
| CAT05 | 0.0345 | 0.08924 | | 0.386 | | 0.6995 | |  | 1.035 | | | 0.7950 | | 1.3477 |
| CAT06 | 0.2973 | 0.09029 | | 3.292 | | 0.0010 | | \*\*\* | 1.346 | | | 1.0329 | | 1.7550 |
| CMADE | 0.0242 | 0.00501 | | 4.839 | | 0.0000 | | \*\*\* | 1.025 | | | 1.0143 | | 1.0350 |
| CLAIMS | 0.0176 | 0.00455 | | 3.865 | | 0.0001 | | \*\*\* | 1.018 | | | 1.0087 | | 1.0269 |
| ORIGINAL | -0.8521 | 0.12358 | | -6.895 | | 0.0000 | | \*\*\* | 0.427 | | | 0.3132 | | 0.5804 |
| GENERAL | 4.6444 | 0.06795 | | 68.351 | | 0.0000 | | \*\*\* | 104.002 | | | 72.4010 | | 150.0701 |
| FWDAPAG | -0.4885 | 0.02004 | | -24.372 | | 0.0000 | | \*\*\* | 0.614 | | | 0.5878 | | 0.6397 |
| BCKGTLAG | -0.0067 | 0.00324 | | -2.060 | | 0.0395 | | \*\* | 0.993 | | | 0.9870 | | 0.9997 |
|  |  |  | |  | |  | |  |  | | |  | |  |
| Intercepts: |  |  | |  | |  | |  |  | | |  | |  |
| 0|1 | -732.991 | 0.00198 | | -369,549.5 | | 0.0000 | | \*\*\* | 0.000 | | | 0.000 | | 0.000 |
| 1|2 | -730.923 | 0.10504 | | -6,958.3 | | 0.0000 | | \*\*\* | 0.000 | | | 0.000 | | 0.000 |
| 2|3 | -729.860 | 0.11145 | | -6,549.0 | | 0.0000 | | \*\*\* | 0.000 | | | 0.000 | | 0.000 |
| 3|4 | -729.079 | 0.11439 | | -6,373.8 | | 0.0000 | | \*\*\* | 0.000 | | | 0.000 | | 0.000 |
| 4|5 | -728.525 | 0.11660 | | -6,248.5 | | 0.0000 | | \*\*\* | 0.000 | | | 0.000 | | 0.000 |
| 5|6 | -728.049 | 0.11891 | | -6,122.7 | | 0.0000 | | \*\*\* | 0.000 | | | 0.000 | | 0.000 |
| 6|7 | -727.700 | 0.12091 | | -6,018.7 | | 0.0000 | | \*\*\* | 0.000 | | | 0.000 | | 0.000 |
| 7|8 | -727.412 | 0.12283 | | -5,922.2 | | 0.0000 | | \*\*\* | 0.000 | | | 0.000 | | 0.000 |
| 8|9 | -727.149 | 0.12487 | | -5,823.4 | | 0.0000 | | \*\*\* | 0.000 | | | 0.000 | | 0.000 |
| 9|10 | -726.933 | 0.12652 | | -5,745.7 | | 0.0000 | | \*\*\* | 0.000 | | | 0.000 | | 0.000 |
| 10|11 | -726.699 | 0.12874 | | -5,644.7 | | 0.0000 | | \*\*\* | 0.000 | | | 0.000 | | 0.000 |
| 11|12 | -726.499 | 0.13029 | | -5,575.9 | | 0.0000 | | \*\*\* | 0.000 | | | 0.000 | | 0.000 |
| 12|13 | -726.323 | 0.13116 | | -5,537.6 | | 0.0000 | | \*\*\* | 0.000 | | | 0.000 | | 0.000 |
| 13|14 | -726.188 | 0.13145 | | -5,524.3 | | 0.0000 | | \*\*\* | 0.000 | | | 0.000 | | 0.000 |
| 14|15 | -726.030 | 0.13180 | | -5,508.8 | | 0.0000 | | \*\*\* | 0.000 | | | 0.000 | | 0.000 |
| \*\*\* 0.001 \*\* 0.01 \* 0.05 | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | |
| -2 log likelihood null | | | 9,744.1 | |  | |  | | |  |  | |  | |
| -2 log likelihood residual | | | 7,904.2 | |  | |  | | |  |  | |  | |
|  | | |  | |  | |  | | |  |  | |  | |
| McFadden R2 | | | 0.1888 | |  | |  | | |  |  | |  | |
|  | | |  | |  | |  | | |  |  | |  | |

Table 8

Adjusted R2 for Regression Subsets

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Adjusted R2 | 0.23 | X | X | X | X | X |  |  |  | X | X | X | X |  |
| 0.23 | X | X | X | X | X |  |  |  | X |  | X | X | X |
| 0.23 | X | X | X | X | X |  |  |  | X |  | X | X |  |
| 0.23 | X | X | X | X |  |  |  |  | X | X | X | X |  |
| 0.23 | X | X | X | X |  |  |  |  | X |  | X | X |  |
| 0.23 | X | X | X | X | X |  |  |  |  |  | X | X |  |
| 0.22 | X | X | X | X |  |  |  |  |  |  | X | X |  |
| 0.22 | X | X | X | X |  |  |  |  | X |  | X |  |  |
| 0.21 | X | X | X | X |  |  |  |  |  |  | X |  |  |
| 0.21 | X |  | X | X |  |  |  |  |  | X | X |  |  |
| 0.20 | X |  | X | X |  |  |  |  |  |  | X |  |  |
| 0.20 | X | X |  | X |  |  |  |  |  |  | X |  |  |
| 0.19 | X |  |  | X |  |  |  |  |  |  | X |  |  |
| 0.18 | X |  | X |  |  |  |  |  |  |  | X |  |  |
| 0.17 | X |  |  |  |  |  |  |  |  |  | X |  |  |
|  | | Intercept | GYEAR | CAT02 | CAT03 | CAT04 | CAT05 | CAT06 | CMADE | CLAIMS | ORIGINAL | GENERAL | FWDAPLAG | BCKGTLAG |

Table 9

Multiple Regression Results

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | CRECEIVE | | | | CRECEIVEsqrt | | | |
|  | Coefficient Estimate | p-value | Sig. | VIF | Coefficient Estimate | p-value | Sig. | VIF |
| Intercept | 998.65051 | 0.000 | \*\*\* |  | 169.336058 | 0.000 | \*\*\* |  |
| GYEAR | -0.49997 | 0.000 | \*\*\* | 1.0589 | -0.084001 | 0.000 | \*\*\* | 1.0458 |
| CAT02 | 3.05106 | 0.000 | \*\*\* | 1.0768 | 0.135559 | 0.004 | \*\* | 1.0617 |
| CAT03 | 3.45773 | 0.000 | \*\*\* | 1.0554 | 0.141919 | 0.002 | \*\* | 1.0434 |
| CAT04 | 1.37359 | 0.000 | \*\*\* | 1.0755 | 0.072884 | 0.034 | \* | 1.0630 |
| CLAIMS | 0.05989 | 0.000 | \*\*\* | 1.0216 | 0.004406 | 0.003 | \*\* | 1.0135 |
| ORIGINAL | -1.55882 | 0.002 | \*\* | 1.0728 | -0.249616 | 0.000 | \*\*\* | 1.0601 |
| GENERAL | 9.57013 | 0.000 | \*\*\* | 1.2192 | 1.473686 | 0.000 | \*\*\* | 1.1878 |
| FWDAPAG | -0.21028 | 0.000 | \*\*\* | 1.1375 | -0.158777 | 0.000 | \*\*\* | 1.1428 |

\*\*\* 0.001 \*\* 0.01 \* 0.05

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Residual SE | DF | Multiple R2 | Adjusted R2 | F-statistic | DF | p-value |
| CRECEIVE | 6.357 | 1,989 | 0.2373 | 0.2342 | 77.4 | 8 & 1,989 | 0.000\*\*\* |
| CRECEIVEsqrt | 0.544 | 1,754 | 0.6417 | 0.6401 | 392.7 | 8 & 1,754 | 0.000\*\*\* |

\*\*\* 0.001 \*\* 0.01 \* 0.05