Software Patent Citations: A Consistent Weighted Ranking

Chaim Fershtman and Neil Gandal[1]

February 13, 2005

Version 5

Abstract

In recent years, economists have begun asking whether incentive schemes like the patent system provide proper incentives for firms to invest in path-breaking research and development. Such an analysis requires a good measure of the value of innovations and patents. This quantification is not an easy task especially in the digital economy, which is characterized by knowledge industries such as computer software. Yet, such quantification is especially important in the digital economy, where technology changes rapidly, the number of patents has grown exponentially and patenting has become an important strategy of firms.

Economists and other researchers have primarily used patent citation data as a proxy for the value of the underlying innovation and knowledge flows. The evidence, however, as to whether citations are a good measure of economic value is mixed. In the paper, we construct a refined measure of patent citations for software patents that weighs patent citations by the importance of the citing patent in such a way that the resulting weights are both endogenous and consistent. We then examine the difference between this consistent weighted ranking system and the traditional measure of patent counts. Our empirical work suggests that the CWR measure may be better in measuring patent value for these data than simply the number of citations.

Our empirical work also suggests that there may be strategic reasons for citing patents or omitting citations to certain patents. This may be an important strategy in the digital economy, where often the most important asset of the firm is the patent(s) that it holds.

We are especially grateful to Martin Peitz and also thank Josh Lerner and Mark Schankerman for helpful comments. We are also grateful to Michael Campbell-Kelly and Patrick Valduriez for providing us with their data. Financial support from NERA is gratefully acknowledged. Any opinions expressed are those of the authors.

[1] Fershtman: Tel Aviv University, fersht@post.tau.ac.il, Gandal: Tel Aviv University and CEPR, gandal@post.tau.ac.il.

1. Introduction

The importance of innovation in the digital economy to the U.S. and world economies probably cannot be overstated. As Scotchmer (2004) remarks, "Patents are the gold standard of intellectual property protection. With other forms of protection (like copyright), if a third party duplicates the protected innovation independently, he or she can use it. The absence of this independent-invention defense makes patent law uniquely powerful."[2] This is especially true in the case of software and other digital products, since reverse engineering is often feasible.[3]

In recent years, economists have begun asking whether incentive schemes like the patent system provide proper incentives for firms to invest in path-breaking research and development. Important questions include the optimal length of the patent, the patent scope, and whether software firms are excessively patenting.[4]

Such an analysis requires a good measure of the value of innovations and patents. This quantification is not an easy task especially in the digital economy, which is characterized by knowledge industries such as computer software. Yet, such quantification is especially important in the digital economy, where technology changes rapidly and the number of patents has grown exponentially.[5]

Economists and researchers have primarily used patents and patent citation data as a proxy for the value of the underlying innovation and knowledge flows. Intuitively the measure makes sense because in theory major innovations are important building blocks for subsequent innovations and hence would likely be highly cited.

[2] Scotchmer (2004), p.66.
[3] This does not imply that copyright protection is unimportant in protecting innovations. Most software firms obtain both copyright and patent protection. In the music industry, copyright protection is the essential form of intellectual property, since digital music can easily be reproduced at very low cost. As Peitz and Waelbroeck (2005, chapter 4 in this book) discuss in a paper that provides a detailed analysis of the digital music industry, Napster, a firm that allowed users to exchange music files was found guilty of copyright infringement and had to shut down. Gayer and Shy (2005, chapter 8 in this book) also examine copyright issues for products like digital music.
[4] For a summary of early empirical work, see Griliches (1990); for a summary of recent empirical work, see Jaffe and Trajtenberg (2002). For a summary of the theoretical work, see Scotchmer (1991, 2004).
[5] See Kortum and Lerner (1999). Patents are often essential for the rise of new business models, such as software platforms. See Evans, Hagiu, and Schmalensee (2005, chapter 3 in this book).

Nevertheless, evidence regarding whether patent citations are a good measure of the underlying value of the innovation is mixed.

Seminal work by Hall, Jaffe, and Trajtenberg (2001) put data on all patents issued in the U.S. between 1963 and 1999 on the National Bureau of Economic Research (NBER) web site.[6] These data, which include all pair-wise patent citations between 1976 and 1999, are publicly available in a convenient format.

A small number of recent studies have examined whether patent citations are correlated with non-patent measures of value. Lanjouw and Schankerman (2004) find that a measure based on multiple factors including patent citations has statistically significant explanatory power in predicting whether a patent will be litigated. Shane (2002) finds that for M.I.T. patents there is a positive correlation between the number of patents citations and the probability that the patent will be licensed.

Hall, Jaffe and Trajtenberg (HJT 2000) find that "citation weighed patent stocks" are more highly correlated with firm market value than patent stocks themselves. Nevertheless, they also find that R&D stock is more highly correlated with firm market value than either patents or citations weighted patent stocks. In a study of University patents, Sampat and Ziedonis (2002) find that while citations are a good predictor of whether a patent will be licensed (a result similar to that of Shane, 2002), they are not a good predictor of revenues earned from licensing, that is, the number of patent citations may not be a good measure of the underlying value of the innovation.

Preliminary research by Campbell-Kelly and Valduriez (2004) suggests that the 50 most highly cited software patents are all incremental improvements in technology, rather than major innovations. This research is particularly interesting because the classifications (incremental innovation, dramatic innovation) are based on the authors' expertise and a detailed technical analysis of the patents themselves. Hence, the evidence as to whether citations are a good measure of economic value is mixed.

[6] See http://www.nber.org/patents.

In the paper, we examine a consistent measure of patent citations for the computer software industry. While research has shown that citations are a better measure of innovation than pure patent counts, it is probably important to "weigh" the citations as well. Consider the analogy to academic citations.[7] Citations that come from important papers may be more important in helping determine the value of the paper than a citation from a paper published in a less important journal. The same logic may be true for patents as well and thus citations should be weighed by the source of the citation. Is the citing patent itself an important or unimportant patent? If the citing patent has a lot of citations itself, its citation should be more heavily weighted than a citing patent that has very few citations.

We employ a measure which weighs patent citations by the importance of the citing patent – denoted by "Consistent Weighted Ranking" (CWR) scheme.[8] This measure is consistent in the sense that citation weights used in constructing the ranking are identical to the final ranking produced by our method. Our measure is quite different than counting the number of citations. Our empirical work suggests that the CWR measure may be better in measuring patent value for these data than simply the number of citations.

Our empirical work also suggests that there may be strategic reasons for citing patents or omitting citations to certain patents. This may be an important strategy in the digital economy, where often the most important asset of the firm is the patent(s) that it holds.[9]

In the following section, we explain the intuition behind the CWR. In section 3, we describe the formal methodology. In section 4 we construct a ranking of software patents using the CWR and compare these rankings to rankings based on the number of citations. In section 5 we examine the performance of the CWR and the number of citations using properties of the patents themselves. Section 6 provides brief conclusions.

[7] See Palacios-Huerta and Volig (2004) for an axiomatic approach of defining a consistent rating scheme for academic journals.
[8] See also Fershtman and Gandal (2004) for a consistent ranking method for sports teams.
[9] Other potentially important strategies in the digital economy include the preannouncement of products (Choi, Kristiansen, and Nahm (2005), chapter 7 in this book) and versioning (Belleflamme (2005), chapter 7 in this book).

2. Intuition for the CWR

To better understand the construction of CWR, consider the following example with six patents. The citations across patents are described by the table below.

Citations by each patent

Patent	1	2	3	4	5	6	Citations received	Initial weights	CWR first iteration	CWR final rating
1	0	1	1	1	0	0	**3**	**3**	**2**	**300**
2	0	0	0	0	1	1	**2**	**2**	**3**	**279**
3	0	0	0	0	0	0	**0**	**0**	**0**	**100**
4	0	0	0	0	0	0	**0**	**0**	**0**	**100**
5	0	0	0	1	0	0	**1**	**1**	**0**	**142**
6	1	0	0	0	1	0	**2**	**2**	**4**	**286**

The first row of the table shows the patents that cite patent #1, while the second column shows the patents cited by patent #2, etc. The total number of citations appears in column eight.

Both the second patent and the sixth patent are cited twice. However, if we weigh the citing patent by the number of citations it received, the weighted citations index in the tenth column shows that patent six is more important than patent two. This, of course, is just a single iteration of weights. The outcome of this iteration is another set of weights. In order to calculate the CWR, we require consistency which means that the weights used in calculating the weighted index will be identical to the resulting index itself. To perform this task we need to continue iterating until a fixed point is reached, or to use an algorithm that identifies such a fixed point given the matrix of citations.

Using such an algorithm, the final weights (and hence ratings) are shown in the final column of the above table and patent six indeed has a higher rating than patent two. Also notice that the rankings between patent one and patents two and six are much closer reflecting the fact that although patent one has 50% more citations, it is cited by relatively unimportant patents.

3. Our Formal Ranking Methodology

In this section, we present a system of indices for each patent that will capture not only the number of patents that cited it, but also the importance of the citing patents.

To achieve a ranking we will search for a vector of ratings $z = \{z_i\}_{i=1}^{N}$ which assigns each patent *i* with a respective rating z_i. All of the ratings (values of indices in z) will be interdependent. Hence we need to create a system of equations in which all ratings are determined simultaneously.

As we show below, in order to ensure existence of a solution to the system of equations, we must limit the range of possible z_i values, in particular $z_i \in [l, h]$ while *l* and *h* are exogenously determined. Moreover, to avoid the trivial fixed point for which $\forall_i . z_i = 0$ we require that *l* be positive.

3.1 The General Formula

We start with a simple equation:

$$z_i = l + b\sum_{j \neq i} z_j a_{i,j} , \tag{1}$$

where $a_{i,j}$ equals 1 if patent or article *i* is cited by patent *j* and *b* is a coefficient designed to ensure that $l \leq z_i \leq h$.[10]

We will let *b* be determined endogenously in a way that will ensure that the highest rating will be infinitesimally close to *h*. The lowest rating will inevitably be close to *l*. If a patent is not cited than it receives a rating of *l* regardless of the value of *b*.[11] The following condition must hold for the highest rated patent:

[10] In the example in section 2, *l*=100, *h*=300, and b=0.41.

[11] We determine *b* in that way, in order to achieve the maximum spread within the ratings' range, and insure that there is only one ranking possible for every given *l* and *h*. A simpler way of determining *b*, which relaxes these demands, and therefore requires a simpler computer algorithm is described in section 3.1 below.

$$
\begin{aligned}
&l + b \sum_{j \neq highest} z_j a_{highest,j} = h \\
(2) \qquad &\Rightarrow \\
&b\left(l, h, z, a_{highest}\right) = \frac{h - l}{\sum_{j \neq highest} z_j a_{highest,j}}
\end{aligned}
$$

where *highest* is the index of the highest rated patent.

After defining *b* we can define (1) as the condition for the index. Note that for every *i,j*, $a_{i,j}$ are given. The number of equations equals the number of patents. Hence we have a system of linear equations that can be solved and yield a fixed point.

After solving for the index z (as a function of *b*), we can update *b* so that (b, z) where $b\left(l, h, z, a_{highest}\right)$ fulfills condition (2) and $z(l, b, X)$[12] is determined by the system of equations resulting from condition (1).

We wish to stress that although *l* and *h* are parameters that are chosen to determine the spread of the ratings, they might influence the final ranking as well. As we choose a higher *l*, and a lower *h*, $(h - l)$ becomes lower and the differences in ratings decrease. In this case the ranking becomes similar to the old fashion ranking - merely counting the citations.

3.2 A Simpler Formula

If one is ready to relax the demand for the maximum spread possible within the ratings' range, and for the existence of only one possible ranking for every given *l* and *h*, a simpler formula can be employed. Condition (2) above can replaced with:

$$
\begin{aligned}
&l + b \sum_{j \neq i} h \cdot a_{highest,j} = h \\
(3) \qquad &\Rightarrow \\
&b\left(l, h, a_{highest}\right) = \frac{h - l}{h \cdot \sum_{j \neq i} a_{highest}}
\end{aligned}
$$

[12] X is a matrix of 1 and 0 and defines which patent cites another patent.

Note that in this new condition b is not a function of z.[13] This clearly simplifies the calculations. Condition (3) insures that no patent receives a rating of more than h, although it does not imply that any will reach h. From combining (1) and (3) we get

$$z_i = l + \left[\frac{h-l}{h \cdot \sum_{j \neq i} a_{highest}} \right] \sum_{j \neq i} z_j a_{i,j} \tag{4}$$

Hence we again have a system of linear equations that can be solved to yield the fixed point.

4. Data and Construction of CWR for Software Patents

Our data includes information on patents classified under International Patent Classification (IPC) G06F and granted between 1976 and 2000—a total of 76,920 patents. The data include information on "who" cites "who." This yields a matrix of 76,920 X 76,920 where each entry is either a zero or a one. This matrix is the input we use in our CWR calculations and it enables us to build the CWR at the level of the patent.[14]

We limited the data to IPC G06F in an effort to obtain an objective sample of software patents. The classification G06F refers to “electric digital data processing.”[15] The sub-classifications under G06F are shown in the Appendix.[16] Other definitions of software patents are, of course, possible. Indeed one can classify patent classes endogenously by other patents that they cite. Since we focus on the most highly cited software patents, it is likely that our results are robust to any reasonable classification scheme.

There are 76,290 software patents in the G06F class. Of these, 57,382 either cited at least one of the software patents in the G06F class or received a citation from at least

[13] *Highest* refers to the patent/article with the most citations.

[14] In the analysis in this section, we use the simpler formula described in section 3.1, with l=100, h=300.

[15] International Patent Classification, World International Property Organization website at http://www.wipo.int/classifications/en/index.html?wipo_content_frame=/classifications/en/ipc/index.html.

[16] For more detailed description of sub-classifications under G06F, see World International Property Organization website at <http://www.wipo.int/classifications/fulltext/new_ipc/index.htm>.

one of the software patents in the G06F class. We refer to this as the relevant "group."

On average, the total number of citations per patent is quite skewed. Excluding own citations (by the same firm), the mean number of citations per patent is 7.9, but the median is only 3. Further, 75% of the patents received ten citations or less.

In the case of citations from patents within the group, the number of citations per patent is even more skewed. Only 35,556 patents receive citations from other patents in our group. The mean number of citations per patent is 4.4, but the median is 1. 75% of the patents received five citations or less. Using the 57,382 patents, we compute the following measures:[17]

- All - All forward citations including citations from the firm that holds the patent.
- No Self (NS) – All forward citations from patents held by other firms.
- In Group (IG) – All forward citations from patents in the G0F6 class.
- CWR– Our ranking index.

We are primarily interested in the comparison between the "IG" and "CWR" rankings. We report the other results for the sake of completeness. When we consider the full group, we obtain the following correlations between the raw measures:

	All	No Self	In Group (IG)	CWR
All	1.00			
No Self	0.98	1.00		
In Group (IG)	0.84	0.83	1.00	
CWR	0.80	0.79	**0.95**	1.00

Table 1A: Correlations among measures: Full group (57,382 observations)

Table 1A shows that the correlation between the IG and CWR measures is quite high (0.95). This is primarily due to the fact that many of the patents do not receive even a

[17] Patents that did not receive any citations have a CWR ranking equal to the minimum value of the ranking index (100 in this case).

single citation. For all of these patents, IG equals zero and the CWR measure takes on the minimum possible value. Hence, it makes sense to restrict attention to patents that receive more than just a few forward citations from other patents in the group.

Table 1B shows the correlation between measures and patent ranks for all 6821 patents that received ten or more forward citations from other patents in the group. The correlation between IG and CWR is lower for these patents (0.89) than for the full group.

	All	No Self	In Group (IG)	CWR
All	1.00			
No Self	0.97	1.00		
In Group (IG)	0.83	0.83	1.00	
CWR	0.75	0.75	**0.89**	1.00

Table 1B: Correlation among measures: 6821 patents with more than 10 forward citations within the group.

In Table 1C, we examine the correlation between measures and patent ranks for the 103 patents that received 70 or more forward citations from other patents in the class. Note that the correlation between IG and CWR is quite a bit lower for these highly cited patents (0.77) than for larger group of patents.[18]

	All	No Self	In Group (IG)	CWR
All	1.00			
No Self	0.98	1.00		
In Group (IG)	0.92	0.92	1.00	
CWR	0.75	0.72	**0.77**	1.00

Table 1C: Correlation among measures: 103 patents with 70 or more forward citations within the group.

The top 30 patents according to "In Group" citations are shown in Table 2. The table shows that with the exception of CWR, all other measures are virtually identical in the case of the top 10 patents. We are primarily interested in "In Group" and CWR rankings. Table 2 shows that these measures are quite different, even for the 30 most

[18] In the case of ranks, the correlation between IG and WCR is 0.57.

highly cited patents within the group. Table 3 shows the top 30 patents according to the CWR measure. This table includes the "In Group" rating as well.

The top patent in both the CWR and IG measures is a software management system patent owned by Xerox that automatically collects and recompiles component software objects over a computer network. "The component software objects are periodically updated, via a system editor, by various users at their personal computers and then stored in designated storage means."[19] This patent, which was issued in 1985, cites only one other patent.

The second highest rated patent according to the IG measure (#3 according to CWR) is a power manager inside of a laptop computer. It includes a software program that monitors and controls the distribution of power to the various units in the computer in order to conserve the battery. The patent issued in 1992 cites 17 other U.S. patents.

The second highest rated patent according to the CWR measure (#8 according to IG) is a multiprocessor system that interconnects two or more separate processors. The redundant multiprocessor system allows on-line maintenance of one of part of the system while the rest of the system is functional and includes a distributed power supply system that insures that each device controller has two separate power supplies and can function even if one of the power supplies shuts down. The patent issued in 1980 cites 13 other U.S. patents.

[19] For patent abstract for the patents discussed in this section, see www.uspto.gov.

			Forward Citation Measures				Rank			
Year	Assignee	Patent	All	No Self	In Group	CWR	IG	CWR	NS	All
1985	Xerox	4558413	267	263	252	3313	1	1	1	1
1992	Apple	5167024	252	247	226	1957	2	3	2	2
1993	Nexgen Microsystems	5226126	218	218	188	1440	3	20	3	3
1989	Hitachi, Ltd.	4858105	196	188	186	1435	4	22	8	8
1993	Eastman Kodak	5181162	198	198	171	1662	5	7	4	6
1991	Xerox	5008853	208	193	170	1638	6	8	7	5
1989	Cornell Univ.	4807115	197	197	167	1429	7	23	5	7
1980	Tandem Computers	4228496	218	194	164	2692	8	2	6	4
1993	NexGen Microsystems	5226130	176	176	161	1570	9	12	9	9
1992	AT&T	5093914	175	175	161	1251	9	30	10	10
1992	HP	5133075	166	164	150	1541	11	13	16	18
1992	Schlumberger Technology	5119475	157	157	143	1436	12	21	17	20
1989	Apollo Computer	4809170	146	145	140	1523	13	14	24	27
1989	Tektronix, Inc.	4821220	150	150	134	1587	14	10	20	24
1989	Tektronix, Inc.	4885717	148	148	134	1572	14	11	22	26
1990	HP	4953080	151	146	134	1455	14	16	23	23
1977	Siemens	4044338	134	134	127	946	17	90	33	41
1985	AT&T	4555775	172	170	120	1782	18	4	13	14
1989	Tektronix, Inc.	4853843	122	122	118	1475	19	15	42	51
1991	Intel	5075848	124	124	112	1246	20	32	40	48
1992	IBM	5151987	133	123	110	1133	21	43	41	43
1996	Sun Microsystems	5530852	136	125	109	881	22	126	38	39
1991	Xerox	5072412	137	120	108	1077	23	57	45	37
1978	Cray Research	4128880	141	112	105	1613	24	9	55	31
1987	Signetics	4669043	110	108	105	1445	24	19	64	72
1990	Fairchild Semiconductor	4928223	106	106	105	878	26	128	70	82
1982	Intel	4325120	121	115	104	1764	27	5	51	54
1987	Intel	4674089	123	112	102	1453	28	17	56	49
1981	Intel	4257095	107	107	102	1057	28	61	65	76
1992	Tektronix	5136705	109	109	98	1054	30	62	61	74

Table 2: Patents with the most in group citations in the G06F classification.

Year	Assignee	patent	IG rank	CWR rank
1985	Xerox Corporation	4558413	1	1
1980	Tandem Computers	4228496	8	2
1992	Apple Computer, Inc.	5167024	2	3
1985	AT&T Bell Laboratories	4555775	18	4
1982	Intel Corporation	4325120	27	5
1978	Codex Corporation	4096571	33	6
1993	Eastman Kodak Company	5181162	5	7
1991	Xerox Corporation	5008853	6	8
1978	Cray Research, Inc.	4128880	24	9
1989	Tektronix, Inc.	4885717	14	10
1990	Hewlett-Packard Company	4953080	14	11
1992	AT&T Bell Laboratories	5093914	9	12
1992	Hewlett-Packard Company	5133075	11	13
1989	Apollo Computer, Inc.	4809170	13	14
1989	Tektronix, Inc.	4853843	19	15
1989	Tektronix, Inc.	4821220	14	16
1981	Intel Corporation	4257095	29	17
1987	Measurex Corporation	4635189	55	18
1987	Signetics Corporation	4669043	25	19
1993	Nexgen Microsystems	5226126	3	20
1992	Schlumberger Technology	5119475	12	21
1989	Hitachi, Ltd.	4858105	4	22
1989	Cornell Research	4807115	7	23
1985	Texas Instruments	4562535	36	24
1978	Bolt Beranek and Newman	4130865	128	25
1986	IBM	4594655	69	26
1984	IBM	4442487	84	27
1980	IBM	4200927	117	28
1978	Bunker Ramo Corporation	4075691	39	29
1993	NexGen Microsystems	5226130	9	30

Table 3: Patents with the highest CWR measure in the G06F classification.

5. A Formal Analysis Using the CWR and IG Measures

The difference between the "In Group" citation and CWR rankings raises the question of whether one of the measures better captures the value of a patent. In this section, we examine whether observable characteristics of the patents can explain the number of citations and the CWR measure. We employ characteristics from the NBER patent database (such as the number of claims and the year in which the patent was granted) as well as characteristics from work by Campbell-Kelly and Valduriez (2004). These characteristics -- scope and technical depth -- are especially interesting because they are based on a scientific examination of the patents by researchers familiar with the technologies described by these patents.[20] Campbell-Kelly and Valduriez (2004) determined these characteristics for the 50 most highly cited patents.[21]

Our analysis in this section employs the following variables:

- Claims – the number of claims made by the patent.
- Scope – A dummy variable that takes on the value one if the scope of the patent is broad and zero if the scope is narrow.
- Depth – A dummy variable that takes on the value one if the technical depth is high and zero if the technical depth is medium or low.[22]
- Year1976 – The difference between the year in which the patent was granted and 1976, the first year for which the patent data are available.

Descriptive statistics are available in table 4.

Variable	Mean	Minimum	Maximum	Std. Dev.
CWR rating	868.26	143.06	3313.50	611.61
In group citations	77.91	1	252	52.50
CWR ranking	1296.98	1	12956	2529.31
In group ranking	972.47	1	12926	2383.53
Claims	28.35	6	85	18.54
Scope	0.37	0	1	0.49
Depth	0.63	0	1	0.49
Year1976	15.16	5	20	3.90

Table 4: Descriptive Statistics (49 observations)

[20] We are grateful to Campbell-Kelly and Valduriez for providing us with these data.

[21] In particular, the 50 patents include the 41 most highly cited patents (not including citations from the same firm) and 9 patents granted since 1990 with the highest number of forward citations within three years from the year the patent was granted. They chose the sample in this manner in order that several more recent patents would be in their data set. See Campbell-Kelly and Valduriez (2004) for details. One of the nine patents granted since 1990 has no forward citations within the group. Thus, the CWR and IG rankings are not defined and there are 49 observations.

[22] Campbell-Kelly and Valduriez (2004) also determine whether the disclosure level is high or low. This variable was not significant in any of the regressions.

Table 5 shows the correlation matrix between the dependent and independent variables. The CWR and In Group are quite highly correlated. Scope and Claims have a correlation coefficient of 0.47, while Claims and Depth have a much smaller correlation coefficient (0.25); Scope and Depth are virtually uncorrelated (-0.12).

	CWR	In Group	Claims	Depth	Scope
CWR	1.00				
In Group	0.96	1.00			
Claims	-0.33	-0.33	1.00		
Depth	0.25	0.23	0.25	1.00	
Scope	-0.35	-0.32	0.47	-0.12	1.00
Year	-0.37	-0.22	0.08	-0.11	0.09

Table 5: Correlations between dependent and independent variables (49 observations)

The first column in Table 6 below shows a regression of the CWR measure (not the ranking) on the four variables described above. The second column in the table shows a regression of "in group" citations on the same variables. The third and fourth regressions in the table repeat the analysis using the natural logarithm of the dependent variables from the regressions in columns one and two respectively and the natural logarithm of claims.

The regressions in Table 6 suggest that more claims lead to lower CWR rankings and fewer "in group" citations. This may reflect a strategic incentive on behalf of the citing firm. If a patent with many claims is cited, the citing firm may be less likely to receive a patent or it may take a longer time for a patent to be issued. This effect is statistically significant in both of the first two regressions in Table 6. This strategic incentive may be especially important in the digital economy, since often a patent is the most important asset that a firm holds. To the best of our knowledge, there is no theoretical or empirical work on strategic citations.

The regressions in Table 6 also suggest that patents with broader scope have lower CWR rankings and fewer "in group" citations. This effect seems similar to the effect described above. If a very broad patent is cited, the citing firm may be less likely to receive a patent or it may take a longer time for a patent to be issued. This effect is not statistically significant, however, in any of the four regressions in Table 6. Nevertheless, it suggests a strategic incentive that may be important for firms competing in the digital economy.

Finally, the regressions in Table 6 suggest that patents described in greater technical depth receive higher CWR ratings and more citations. This might be because other things being equal, these patents are clear and relatively easy to understand. This effect is statistically significant all four of the regressions in table 6.

The adjusted R-squared values are higher in the CWR regressions, than in the corresponding "in group" citations regression. If the variables in the regression are truly characteristics that explain patent value, this result suggests that the CWR measure is better in capturing patent value for these data than the number of citations. Of course, this analysis is only suggestive at best since it based on such a small number of observations.[23]

	Dependent Variable			
	CWR Rating	In Group	log(CWR)	log(In Group)
Independent Variables				
Constant	1716.59 (5.11)	120.19 (3.93)	7.24 (17.35)	4.62 (6.64)
Claims/ log(claims)	-10.10 (-2.05)	-0.93 (-2.07)	-0.22 (-1.22)	-0.23 (-1.04)
Depth	343.39 (2.05)	30.46 (1.99)	0.47 (2.20)	0.58 (2.28)
Scope	-183.17 (-1.01)	-13.11 (-0.79)	-0.23 (-0.98)	-0.16 (-0.57)
Year1976	-46.94 (-2.39)	-2.00 (-1.12)	-0.042 (-1.68)	-0.0086 (-0.29)
Adjusted R^2	0.27	0.18	0.19	0.08
Number of obs.	49	49	49	49

Table 6: Regression results

[23] The number of observations is limited by the Campbell-Kelly and Valduriez (2004) study.

6. Further Discussion

Our paper constructed a new metric of patent valuation for software patents. If the independent variables in the regressions in Table 6 are truly characteristics that explain patent valuation, our results suggest that the CWR measure may be better in measuring patent value for these data than simply the number of citations. Additional research is, of course, necessary to examine this issue in greater detail.

Our analysis also suggests that a citation (or an "omitted" citation) may be in part due to strategic reasons, since firms in oligopolistic industries may have incentives to "under" or over "cite" patents of competitors. This may be especially true in the digital economy, where often the sole asset of a firm is the patent(s) that it holds. To the best of our knowledge no empirical work has been done on the strategic patenting issue.[24]

[24] Other research suggests that patent examiners and patent lawyers also play a non-trivial role in determining which patents are cited. See Cockburn, Kortum, and Stern (2002).

References:

Belleflamme, P., 2005, "Versioning Information Goods," in ??? M. Peitz editor, MIT Press.

Campbell-Kelly, M., and P. Valduriez, 2004, "A Technical Critique of Fifty Software Patents," mimeo.

Choi, J., Kristiansen, E., and J. Nahm, 2005, "Preannouncing Information Goods," in ??? M. Peitz editor, MIT Press.

Cockburn, I., Kortum, S., and S. Stern, 2002, "Are All Patent Examiners Equal? The Impact of Examiner Characteristics On Patent Statistics and Litigation Outcomes," NBER Working Paper #8980.

Evans, D., Hagiu, A., and R. Schmalansee, 2005, "Software Platforms," in ??? M. Peitz editor, MIT Press.

Fershtman, C, and N. Gandal, "A Consistent Weighted Ranking Scheme: An Application to College Football Rankings," Tel Aviv University mimeo.

Gayer, M., and O.Shy, 2005, "Copyright of Information Goods," in ??? M. Peitz editor, MIT Press.

Griliches, Z., 1990, "Patent Statistics as Economic Indicators: A Survey" *Journal of Economic Literature*, XXVIII: 1661-1707.

Hall, B., Jaffe, A., and M. Trajtenberg, 2000, "Market Value and Patent Citations: A First Look," NBER Working Paper W7741.

Hall, B., Jaffe, A., and M. Tratjenberg, 2001, "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools," NBER Working Paper 8498.

Jaffe, A., and M. Trajtenberg, Patents, Citations and Innovations: A Window on the Knowledge Economy, MIT Press, Cambridge, MA.

Lanjouw, J. and M. Schankerman, 2004, "Patent Quality and Research Productivity: Measuring Innovation with Multiple Indicators," forthcoming *The Economic Journal*.

Kortum, S., and J. Lerner, 1999, "Stronger Protection or Technological Revolution: What is Behind the Recent Surge in Patenting?" *Research Policy* 28:1-22.

Merton, R., "The Matthew Effect in Science II," 1988, Isis, 79: 606-623.

Palacios-Huerta, I., and O. Volij, 2004 "The Measure of Intellectual Influence," *Econometrica*, 72: 963-977.

Peitz, M., and M. Waelbroeck, 2005 "Digital Music," in ??? M. Peitz editor, MIT Press.

Shane, S. “Selling university technology: Patterns from MIT” Management Science, 48(1), 122-137, 2002.

Sampat, B., and A. Ziedonis, 2002, “Cite Seeing: Patent Citations and the Economic Value of Patents,” mimeo.

Scotchmer, S., 1991, “Standing on the Shoulders of Giants: Cumulative Research and the Patent Law", Journal of Economic Perspectives 5:29-41.

Scotchmer, S., 2004, Innovations and Incentives, MIT Press.

Appendix: Sub-classifications under G06F

3/ Input arrangements for transferring data to be processed into a form capable of being handled by the computer…

5/ Methods or arrangements for data conversion without changing the order or content of the data handled…

7/ Methods or arrangements for processing data by operating upon the order or content of the data handled…

9/ Arrangements for programme control…

11/ Error detection; Error correction; Monitoring…

12/ Accessing, addressing or allocating within memory systems or architectures…

13/ Interconnection of, or transfer of information or other signals between, memories, input/output devices or central processing units…

15/ Digital computers in general…