Downloads vs. Citations: Relationships, Contributing Factors and Beyond[1]

Heting Chu[a] & Thomas Krichel[b]
Palmer School of Library & Information Science, Long Island University,
720 Northern Blvd., Brookville, NY 11548, USA
[a]hchu@liu.edu
[b]krichel@openlib.org

Abstract

Citations to 200 top downloaded papers at RePEc, a digital library in economics, were obtained from SSCI and Google Scholar respectively to address questions relating to downloads and their corresponding citations. This study finds that top downloaded documents are used in various degrees when citation is regarded as an indicator of usage. The results also show that a single downloaded paper selected for this study on average receives twice as many citations from Google Scholar as that from SSCI although the latter has been established much earlier in time. According to the coefficients computed, downloads appear having a moderate relationship with citations. However, other measures such as the download-citation ratio indicate a stronger connection between the two. While an author's reputation positively affects both download and citation frequencies, other factors (e.g., targeted readers and subject content) seem in play differently for the documents that are repeatedly downloaded or cited. The study suggests that an infrastructure which encourages downloading at digital libraries would eventually lead to higher usage of their resources.

Keywords

Digital libraries; Downloads; Citations; Usage analysis; Google Scholar

[1] This research was partially funded by the authors' institution. The authors also wish to thank Ms. Nai-yu Liu for her assistance in data collection.

Introduction

Along with the development of information technology, particularly the advent of the Internet, digital libraries of various kinds have been established to better serve people in all walks of life. Among all the digital libraries created so far, some are devoted to research communities in specific subject areas and meanwhile belong to the open access (OA) movement. arXiv (http://arxiv.org), CiteSeer (http://citeseer.ist.psu.edu) and RePEc (http://repec.org) are examples of such digital libraries.

Documents collected in digital libraries of this kind can often be freely downloaded for future reading and use. On the other hand, people may wonder whether documents downloaded from those OA digital libraries are actually used. In other words, are those documents in fact read after being downloaded? It appears difficult to find a method which allows us to directly address the afore-mentioned question. Most approaches have a feasibility problem. For example, it appears impractical to survey people who downloaded documents from digital libraries in order to solicit answers to the question about their usage of the downloaded documents even if we are able to trace them for survey purpose. Likewise, how can we observe others regarding their usage of downloaded documents without intrusion to their privacy even if we are permitted to do so? The quandary, however, can be circumvented by the citation approach – a time tested method although it is not controversy free.

Simply put, our approach is to gather citation counts to documents downloaded from digital libraries to find out if the documents are actually used after being downloaded. The rationale behind this methodology is that, in general, one must read a document before citing it. Based on this rationale, we intend to gather citations to the top 200 documents downloaded from RePEc (Research Papers in Economics) in the hope to answer the above question as well as to explore the following questions:

- Do any relationships exist between download frequency and citation count for the top 200 downloaded documents at RePEc?
- What factors affect downloads and citations?
- What implications will there be for RePEc given the answer to the preceding question?

Background

OA digital libraries are on the rise. The same also holds true for OA citation products and services. RePEc has been selected for this study because one of the present writers is its founder and remains actively involved in its operations and management. The reason for choosing citation tools for this study is given shortly.

The RePEc Digital Library

RePEc has been specially created to facilitate the dissemination of working papers, journal articles and software objects in the field of economics. Based on a collaborative effort of hundreds of volunteers in 54 countries, RePEc holds over 367,000 items of interest, over 266,000 of which are available online. The founders of RePEc have provided detailed descriptions about the digital library in their writings (e.g., Barrueco Cruz & Krichel, 2000). As most citation indexes traditionally cover journal articles only, documents selected for this study from RePEc are thus all journals articles even though there are actually other types of items in the collection.

Citations Products and Services

Besides the renowned citation indexes (e.g., Social Sciences Citation Index - SSCI) published by the Institute for Scientific Information (ISI, now known as Thomson Scientific), some other citation products

have emerged under different auspices in recent years (Roth, 2005). Examples of such citation products include Google Scholar (scholar.google.com), Scopus (www.scopus.com), and CiteSeer. Both Google Scholar and CiteSeer fall into the OA category while Scopus is a fee-based service. Since CiteSeer focuses primarily on the literature in computer and information science, it has little relevance to economics – the subject covered by RePEc. On the other hand, Scopus has not reached the momentum that ISI enjoys with its citation indexes. ISI's citation indexes and Google Scholar are, therefore, chosen as the sources of citations for the current study. More specifically, ISI's SSCI is the target citation index for this study as economics is one discipline in the social sciences.

Literature Review

An extensive literature search for publications on downloads versus citations turned out few titles that are right on target for this study. The most frequently discussed topic along this line appears to be comparisons among various citation products or services. For example, Google Scholar and ISI citation indexes are contrasted by Charbonneau (2006) and Noruzi (2005). Jasco (2005), on the other hand, examined Scopus as well as Google Scholar and ISI citation indexes in his comparative study. Bauer and Bakkalbasi (2005) compared citation counts from WoS (Web of Science - the Web version of ISI citation indexes), Scopus, and Google Scholar for articles published in JASIS&T (Journal of the American Society for Information Science & Technology) in 1985 and 2000 respectively.

The most related study to the present research is an editorial of International Journal of Cardiology (IJC), in which Coats (2005) compared two sets of 10 IJC papers, 10 most cited and 10 top downloaded, taken from the same one year time period in terms of their subject content and document type (e.g., review or research report). The study found no overlaps in these two sets of papers. On the other hand, Coats did not explore whether there exists any relationship between top downloaded papers and citations to them - a theme for the current study. By contrast, Bollen, Van de Sompel, Smith, and Luce (2005) did make a comparison of download and citation data. Their argument, however, was that downloading data should also be used for computing impact factor (IF) which traditionally is calculated with citation data alone.

Increasingly, scholars realize that both citations and downloads need to be combined to more completely assess the impact of journals (Kaplan & Nelson, 2002). The same notion is echoed by Darmoni, Roussel, Benichou, Faure, Thirion, and Pinhas (2000). After comparing the IF for a medical collection with a reading factor which consists of the ratio of a particular journal's download frequency to the total downloads of all journals as recorded in the system, the authors claim that the reading factor seems a credible alternative to IF.

In analyzing access data of an online archive of Physical Review and citations from Physical Review Letters to other Physical Review articles, Fosmire (2004) concluded that each article was downloaded ten times every year while both usage (measured in half-life which is again based on citations) and citation rates showed exponential decay rates with different intrinsic time scales. While Fosmire's study no doubt includes both download and citation data, it does not associate the citations with download documents. Rather, each type of data is considered independently of the other.

As one of the several major OA digital libraries, arXiv has been the subject of research on citations as well as downloads. For example, according to Brown (2001), about two-thirds of all articles submitted to arXiv since 1991 were ultimately cited somewhere between 1998 and 1999. Taking a different approach, Davis and Fromerth (2006) found that papers in the arXiv received 35% more citations on average than non-deposited articles based on an analysis of 2765 articles published in four math journals from 1997 to 2005. However, arXiv-deposited articles received 23% fewer downloads from the publisher's website (about 10 fewer downloads per article) in all but the most recent two years after publication. That is to say, an OA digital library such as arXiv enjoy a citation advantage but suffers in downloads at the publisher's

website. Although the findings are interesting, the focus of both studies is placed upon whether OA has any impact on downloads and/or citations.

Relationship between downloads and citations for one single electronic journal is explored in Moed (2005) from a bibliometric perspective. A synchronous approach applied in the research revealed that downloads and citations show different patterns of obsolescence of the used materials while a diachronous approach showed that, as a cohort of documents grows older, its download distribution becomes more and more skewed, and more statistically similar to its citation distribution. In addition, Moed examined the effect of downloads and citations on each other statistically.

As shown, none of the studies reviewed above aims to investigate if top downloaded documents are also highly cited and thus frequently used by fellow researchers when citation is treated as an indicator of usage. We hence set to accomplish this research objective by analyzing citations to the 200 top downloaded journal papers from RePEc. The citation data, as described earlier, is obtained from SSCI and Google Scholar respectively.

Methodology

Our data collection was completed in about two weeks of time in early 2006. Download data for the top 200 documents, including the name of first author (as SSCI only indexes the first cited author), article title, journal name, publication year, volume/issue/page number, and download frequency, was automatically extracted from LogEc at http://logec.repec.org/ using a purpose-written script. Since LogEc updates and ranks its downloading data regularly, the download data for this study was saved for future reference after being extracted.[2]

Citations to the selected 200 documents were gathered from SSCI and Google Scholar respectively. While citation collection from Google Scholar, again automated but followed by an extensive manual verification, was mainly based on the article title of each top download document, citation counts from SSCI were searched in Dialog with the CR (cited reference) command (e.g., CR=AKERLOF GA, 1970, V84, P488?). Citation frequencies from both SSCI and Google Scholar were gathered in about one week each to minimize the result variation due to the time difference in data collection.

The three sets of data (i.e., downloads as well as citations from SSCI and Google Scholar respectively) were then analyzed quantitatively and qualitatively to address the research questions of this study.

Results and Discussion

Downloads vs. Citations: A Descriptive Overview

Download counts and citation frequencies were first analyzed to present a general picture of the data collected for this study. Table 1 shows the summary measures (e.g., mean and standard deviation) for each set of the data. A visual display of the download and citation frequencies in quartile is provided in Figure 1. Although what is displayed in Table 1 and Figure 1 appears self-explanatory, several points deserve further elaboration.

[2] A list of the 200 top downloaded papers, including their bibliographic information and download frequency, is available upon request.

Table 1. Overview of Downloads vs. Citations

Summary Measure	Download Count	Citation Frequency	
		SSCI	Google Scholar
Mean	1468[3]	344	713
Maximum	8739	3436	4488
Minimum	836	0	0
Standard Deviation	1023	508	762

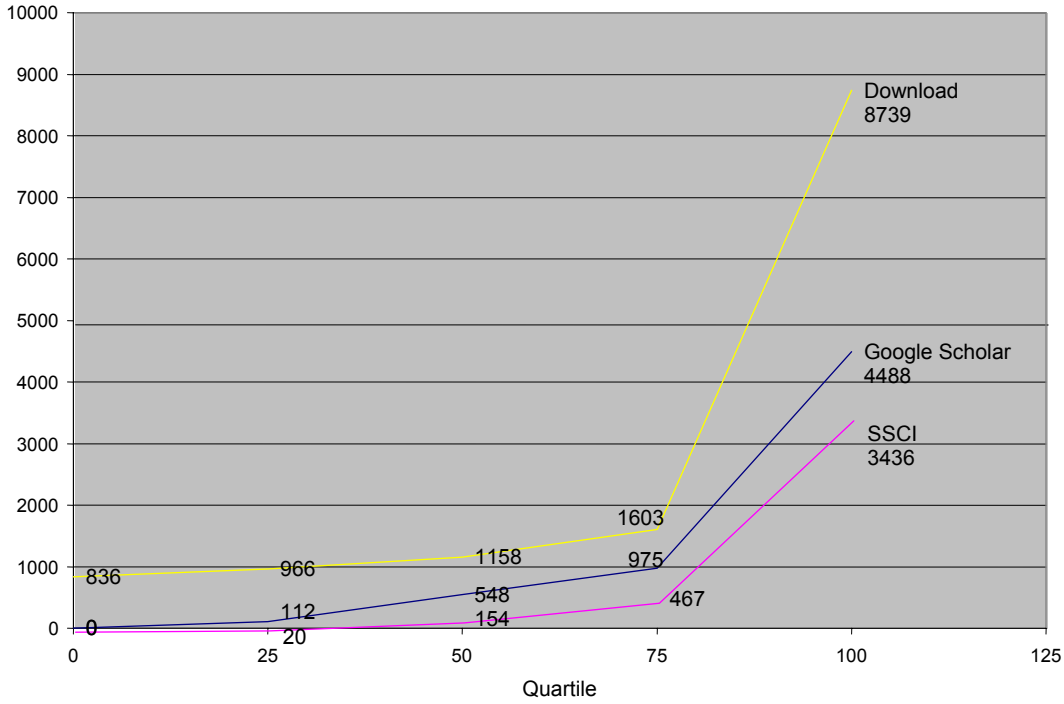

Figure 1: Download and Citation Frequencies in Quartile

First, download frequency for the 200 selected documents is on average greater than citation count, over three times more than the citation count from SSCI and twice as many as that from Google Scholar. The difference between the download and citation data can easily be explained by the fact that a document must be downloaded (or read) before cited while a document could be downloaded but never read or cited. One related note is that citations to the documents selected for this study also include those made to them from sources beyond RePEc (e.g., printed journals, e-journal sites). Otherwise, the discrepancy between downloads and citations would be even larger.

Second, a comparison of citation frequencies between SSCI and Google Scholar clearly indicates that the top 200 downloaded documents are cited more often at Google Scholar than in SSCI. Each of the 200 documents on average receives over twice as many citations from Google Scholar as that from SSCI.[4]

[3] All numbers are rounded to whole digits.
[4] The actual number is 2.07 (=713/344).

Since Google Scholar was established much later than SSCI, how could the former generates more citations than the latter? Possible answers to this question would be the scope and nature of both citation services. It is unarguable that Google Scholar is one of the newest "kids" on the "block of citation services". However, Google Scholar covers a much large pool of citing sources, which differs entirely from the highly selective set of source publications (mostly journals) SSCI bases on for extracting citation data. Citations from as well as by preprints, unpublished manuscripts and other scholarly writings could all be included in Google Scholar as long as they are accessible over the Internet. In contrast, citations would be recorded in SSCI only if they are from source publications selected by ISI.

Third, Google Scholar apparently exhibits more variation in its citation frequency distribution than that of SSCI according to their respective standard deviations. That is, some of the 200 documents were cited many times while others got fewer citations at Google Scholar. Other summary values (e.g., mean and maximum) in Table 1 for Google Scholar and SSCI also attest that different variations exist in each set of the citation data. In addition, as shown in Figure 1, most of the differences between the pair of the citation data occurred in the top 25 percent range. Likewise, the top 25 percent of the documents account for most of the variations in the download data.

The above discussion demonstrates that the top downloaded documents from RePEc are cited, many in fact highly cited, by the scholarly community. But are there any relationships between downloads and citations? The next section tries to address questions of this nature.

Downloads vs. Citations: Correlations

With the descriptive overview provided in the previous section, Pearson's r is calculated to examine correlations between downloads and each type of citations (i.e., SSCI and Google Scholar). Figure 2 presents the results obtained. All correlations are significant at the 0.01 level.

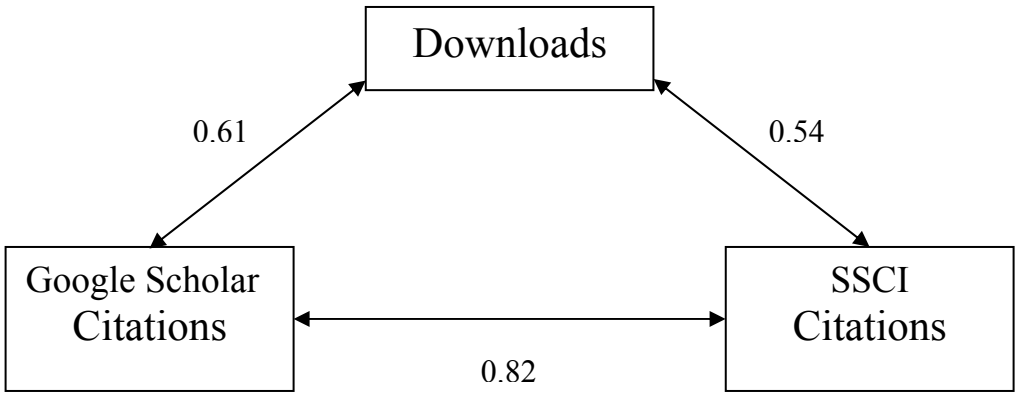

Figure 2: Correlation among Downloads and Citations

As exhibited in Figure 2, there exists little difference between the two correlation coefficients for downloads vs. Google Scholar citations (0.61) and downloads vs. SSCI counterparts (0.54). Both correlations fall within the moderate range. In other words, downloads only have a moderate correlation with citations, regardless of the source of citations being SSCI or Google Scholar, even though the coefficient value is slightly higher in the case of Google Scholar. A stronger correlation (0.82) was, however, observed between the two types of citations, perhaps reflecting the homogeneity of SSCI and Google Scholar as citation services.

A further look at the download and citation averages in Table 1 reveals that there is one citation at SSCI for around every four (4.26 to be exact) downloads from RePEc while the download number is reduced to 2.06 for each citation recorded at Google Scholar. Needless to say, both the download vs. citation ratios are much higher than what Moed (2005) reported in his study: about one citation for every 100 downloads. This huge difference in ratio could be due to the fact that this part of Moed's study covered only 25 months of time whereas the oldest downloaded document examined in the present research was published in 1897, and the average age for the top 200 downloaded documents is 15.4 years. In any case, the afore-described ratios between downloads and citations (i.e., 4.26:1 with SSCI and 2.06:1 with Google Scholar) suggest a relationship beyond the moderate range reported.

Both measures, the correlation coefficients and ratios, are computed using aggregated download and citation data. Taking a different approach, Figure 3 illustrates the connection of each downloaded document with its corresponding citations at SSCI and Google Scholar (GS). The three distributions depicted in Figure 3 appear by and large parallel. On the other hand, most of the documents positioned above the download curve are among the top 10 highly cited, which will be discussed in the following section.

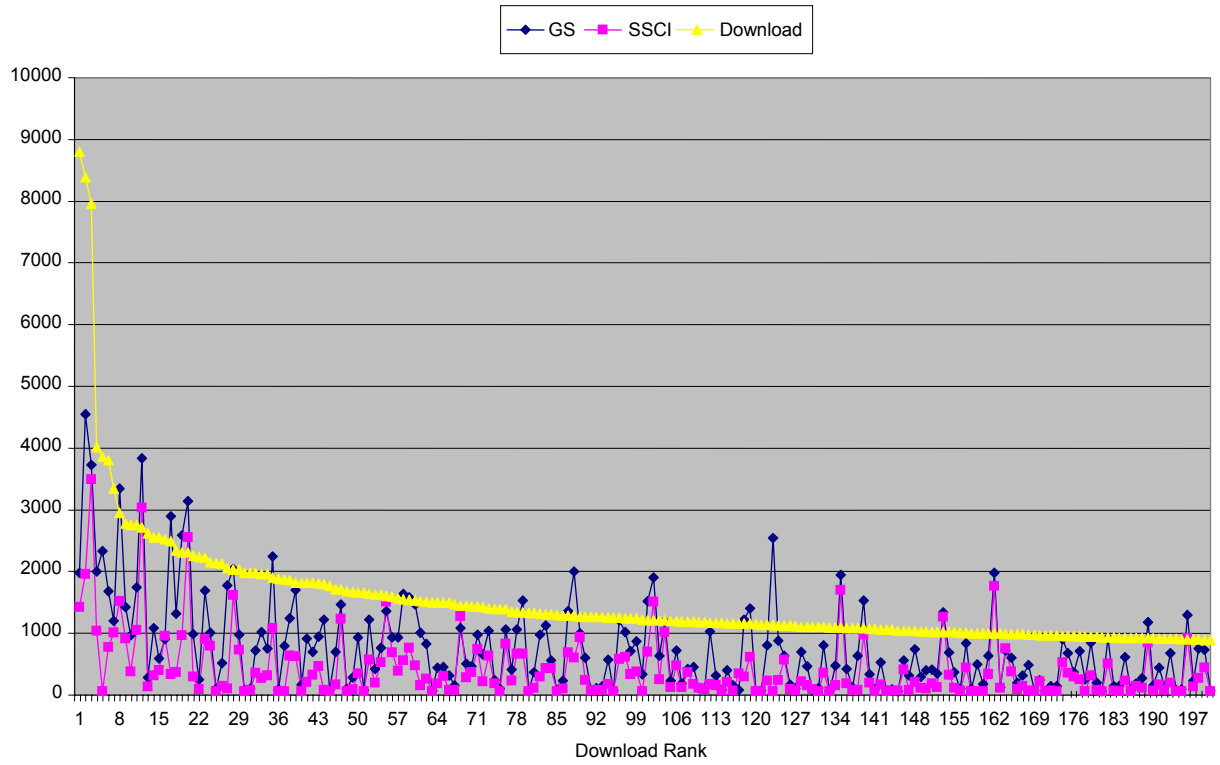

Figure 3: Download vs. Citation: A One-to-One View

Although the correlations between downloads and each type of citations seem moderate, the ratio for the two parameters at the aggregated level as well as the visualization of the three sets of data at the individual level (see Figure 3) demonstrate that downloads and citations are related in a degree stronger than moderate. A further scrutiny of that relationship is intended in a separate project with the consideration of such factors as document age to explore the question whether highly downloaded documents eventually receive more citations than those downloaded fewer times.

Downloads vs. Citations: The Top 10s

In order to gain a better understanding of why certain papers are highly downloaded and also heavily cited, the top 10 papers from each of the three sources (i.e., RePEc downloads, SSCI and Google Scholar citations) are to be examined and contrasted below from the perspective of their subject contents. The reason for choosing only the top 10 from each category is to keep the analysis manageable. As shown in Table 2, there are a total of 21 unique documents in the three sets of top 10s.

Table 2. 21 Unique Documents among the Top 10s

Ranking			**Document**
Download	SSCI	GS	
1	11	14	Akerlof, George A. (1970). The Market for 'Lemons': Quality Uncertainty and the Market Mechanism. Quarterly Journal of Economics, 84(3), 488-500.
2	4	1	Black, Fischer, and Scholes, Myron S. (1973). The Pricing of Options and Corporate Liabilities. Journal of Political Economy, 81(3), 637-654.
3	1	3	Kahneman, Daniel, and Tversky, Amos. (1979). Prospect Theory: An Analysis of Decision under Risk. Econometrica, 47(2), 263-291.
4	17	12	Stiglitz, Joseph E., and Weiss, Andrew. (1981). Credit Rationing in Markets with Imperfect Information. American Economic Review, 71(3), 393-410.
5	175	9	Mankiw, N Gregory, Romer, David, and Weil, David. (1992). A Contribution to the Empirics of Economic Growth. Quarterly Journal of Economics, 107(2), 407-437.
6	30	22	Cox, John C., Ingersoll, Jonathan E., and Ross, Stephen A. (1985). A Theory of the Term Structure of Interest Rates. Econometrica, 53(2), 385-407.
7	19	42	Fama, Eugene F. (1980). Agency Problems and the Theory of the Firm. Journal of Political Economy, 88(2), 288-307.
8	8	4	Romer, Paul M. (1986). Increasing Returns and Long-run Growth. Journal of Political Economy, 94(5), 1002-1037.
9	25	30	Kydland, Finn E., and Prescott , Edward C. (1977). Rules Rather Than Discretion: The Inconsistency of Optimal Plans. Journal of Political Economy, 85(3), 473-491.
10	65	61	Krugman, Paul. (1979). A Model of Balance-of-Payments Crises. Journal of Money, Credit and Banking, 11(3), 311-325.
12	**2**	2	Engle, Robert F., and Granger, Clive W.J. (1987). Co-integration and Error Correction: Representation, Estimation, and Testing. Econometrica, 55(2), 251-276.
20	**3**	5	Heckman, James J. (1979). Sample Selection Bias as a Specification Error. Econometrica, 47(1), 153-161.
162	**5**	15	Hausman, Jerry A. (1978). Specification Tests in Econometrics. Econometrica, 46(6), 1251-1271.
135	**6**	16	Engle, Robert F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica, 50(4), 987-1007.
28	**7**	11	Alchian, Armen A., and Demsetz, Harold. (1972). Production, Information Costs, and Economic Organization. American Economic

			Review, 62(5), 777-795.
102	**9**	17	Hansen, Lars Peter. (1982). Large Sample Properties of Generalized Method of Moments Estimators. Econometrica, 50(4), 1029-1054.
55	**10**	32	Stigler, George J. (1971). The Theory of Economic Regulation. Bell Journal of Economics, 2(1), 3-21.
17	74	**6**	Romer, Paul M. (1990). Endogenous Technological Change. Journal of Political Economy, 98(5), s71-s102.
19	21	**7**	Barro, Robert J. (1991). Economic Growth in a Cross Section of Countries. Quarterly Journal of Economics, 106(2), 407-443.
123	0	**8**	Friedman, Milton. (1997). John Maynard Keynes. Economic Quarterly, Spring, 1-24.
35	15	**10**	Jensen, Michael C. (1986). Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers. American Economic Review, 76(2), 323-329.

Simply looking at the three columns of ranking data in Table 2, we can see that Black & Scholes 1973, Kahneman & Tversky 1979, and Romer 1986 are positioned among the top 10s in all measures while Engle & Granger 1987 and Heckman 1979 only appear in the top 10 lists of citation counts. Besides, Mankiw, Romer & Weil 1992 is ranked in both the top 10 download and Google Scholar series. The remaining documents are all placed in one of the three top 10s. In addition, there are two authors (i.e., Paul M. Romer and Robert F. Engle) who each contributed two papers to the list of 21 unique documents under discussion in this section. The download series overall shows little agreement in ranking with either the SSCI (Spearman's rho = -0.27) or Google Scholar (Spearman's rho = -0.21) citation counterparts. In other words, the 21 documents were frequently downloaded or cited for different reasons. This also indicates that the relationship between downloads and citations is hardly causal.

Among the 21 documents, nine authors are Nobel Laureates in economics although some (i.e., Robert F. Engle, Clive W.J. Granger, James J. Heckman, Finn E. Kydland, and Myron S. Scholes) won the prize mainly for their papers selected in this study while others (i.e., George A. Akerlof, Milton Friedman, Daniel Kahneman, and George J. Stigler) received the award for their works in addition to what is listed in Table 2. More specifically, five out of the nine Nobel Laureates authored papers among the top 10 downloads (i.e., #1, #2, #3, #4 & #9). In contrast, six papers from the top 10 SSCI citations (i.e., #1, #2, #3, #4, #6 & #10) were published by Nobel Prize winners whereas five papers from the Google Scholar set (i.e., #1, #2, #3, #5 & #8) belong to the same category. It becomes evident from the above amplification that the top 3 documents in each of the three measures are, as shown below, all authored by at least one Nobel Laureate.

- Top 3 downloads: #1-Akerlof 1970, #2-Black & Scholes 1973, #3-Kahneman & Tversky 1979
- Top 3 SSCI citations: #1-Kahneman & Tversky 1979, #2-Engle & Granger 1987, #3-Heckman 1979
- Top 3 Google Scholar citations: #1-Black & Scholes 1973, #2-Engle & Granger 1987, #3-Kahneman & Tversky 1979

From the viewpoint of citation frequency, what is listed above does not appear surprising at all as Garfield (1986), in one of his several studies on the same topic, pointed out that Nobel Prize winners often authored citation classics based on his analysis of ISI citations to the works of 125 Nobelists in chemistry, physics and physiology or medicine from 1965 to 1984. Citation classics at that time were defined as publications that received 300 or more citations (Garfield, 1985). Although the threshold for determining citation classics undoubtedly need to be modified as time goes by, the top 3 documents were cited at least 2,496 times and automatically fit into the category of citation classics even if the criterion is raised several times higher. Likewise, the top 3 downloads can be regarded as download masterpieces. In the age of e-

publishing and open access, Nobel winners not only write citation classics but also author download masterpieces.

On the other hand, three documents appeared more than once among the top 3 lists that consist of only five unique papers. This unusual concentration of top ranking downloads and citations possibly indicates that such papers are consistently used and recognized in the scholarly community. For instance, Black & Scholes 1973, ranked #2 in downloading and #1 at Google Scholar, presented a solution to the problem of option pricing, which no one knew how before its publication. Engle & Granger 1987, positioned #2 at both SSCI and Google Scholar, reported an econometric technique that was dominating macroeconometrics during the 1990s. Kahneman & Tversky 1979, the only document appeared in all top 3s (i.e., #1 at SSCI, and #3 in downloading as well as at Google Scholar), introduced the prospect theory as an alternative theory to expected value of utility maximization for economic agents. While Daniel Kahneman was awarded the Nobel Prize for this and many of his other works, the prominent ranking of Kahneman & Tversky 1979 seems to some extent unanticipated because the prospect theory is often considered as an alternative rather than a standard theory. In this case, being the first one in proposing the theory probably outweighs the alternative factor.

Akerlof 1970, although not ranked among the top 10s at either citation measure, is the most downloaded paper out of all the ones included in this study. It is a seminal research paper that uses simple mathematics, and therefore becomes an ideal teaching document. This elucidates why it received top downloads but was positioned lower than the 10^{th} place in terms of citation ranking. The only paper remains to be discussed among the top 3s is Heckman 1979. The author described a two-stage technique that deals with the sampling bias occurring in certain studies. It was ranked #3 at SSCI perhaps because documents of this kind on econometric methodology appear more likely to get cited at SSCI than other papers included in this research.

In addition to the five individual papers that comprise the top 3s, other papers listed in Table 2 of course have particular merits that earned them a place among the top 10s. Those documents normally are the first or the most crucial in developing a theory, model or methodology. For example, Mankiw, Romer & Weil 1992, positioned #5 in downloading and #9 at Google Scholar, established a theory of economic growth which has been a popular basis for other refinements. Paul M. Romer created the theory of endogenous growth in his 1986 paper which was ranked among the top 10 in all three categories (i.e., #4 at Google Scholar, and #8 in downloading as well as at SSCI). The author later extended the theory in his 1990 document, positioned #6 at Google Scholar. The test described in Hausman 1978, ranked #5 at SSCI, is now named as the Hausman specification test. Similar accounts can be provided for the remaining individual documents listed in Table 2. However, we decide not to do so in consideration of presentation clarity.

The above presentation reveals the following points in light of the subject contents of the top 10s examined in this section. First, papers of pedagogical nature (e.g., Akerlof 1970) would increase downloads but not necessarily citations. Second, papers about econometric methodology (e.g., Engle & Granger 1987) are usually cited more often than others at SSCI. Third, Google Scholar appears in favor of papers that are less technical and/or published more recently (e.g., Friedman 1997). Out of all the top 10 papers published in 1990s, only Mankiw, Romer & Weil 1992 is also ranked #5 in downloading. The remaining three documents (i.e., Romer 1990, Barro 1991, and Friedman 1997) all appear just in the top 10 of Google Scholar citations (i.e., #6, #7 and #8 respectively).

Conclusions

This research is conducted to address several questions regarding downloads and citations. Using citations obtained from SSCI and Google Scholar that correspond to the top 200 downloaded papers at RePEc, we

find that those documents are cited or used when citing is considered as an indicator of usage. On average, a single downloaded paper receives twice as many citations from Google Scholar as that from SSCI although SSCI has been established much earlier in time.

While the relationship between downloading and citation appears to be moderate according to correlation coefficients, other measures such as the download vs. citation ratio indicate a stronger connection between the two. Therefore, we intend to explore this association further in a separate study by taking into consideration of other factors like document age. In addition, different parameters (e.g., targeted readers and subject content) seem accounting for the documents that are repeatedly downloaded or cited while an author's reputation simultaneously increases both measures.

What would the findings of this study imply for RePEc as a digital library in economics? In a nutshell, an infrastructure that encourages downloading at RePEc would eventually lead to higher usage of its resources. Such an infrastructure includes greater coverage of research materials that are available via open access. As the OA movement advances, more documents from digital libraries like RePEc can be downloaded to facilitate better communication in the scholarly community.

References

Barrueco Cruz, J.M., & Krichel, T. (2000). Cataloging economics preprints: An introduction to the RePEc project. *Journal of Internet Cataloging, 3(3)*, 227 -241. Retrieved June 5, 2006 from: http://openlib.org/home/krichel/papers/shankari.html.

Bauer, K., & Bakkalbasi, N. (2005). An examination of citation counts in a new scholarly communication environment. D-Lib Magazine, 11(9). Retrieved June 5, 2006 from: http://www.dlib.org/dlib/september05/bauer/09bauer.html.

Bollen, J., Van de Sompel, H., Smith, J.A., & Luce, R. (2005). Toward alternative metrics of journal impact: A comparison of download and citation data. *Information Processing & Management, 41,* 1419-1440.

Brown, C. (2001) The E-volution of preprints in the scholarly communication of physicists and astronomers. *Journal of the American Society for Information Science & Technology, 52(3),* 187-200.

Charbonneau, L. (March 2006). Google Scholar service matches Thomson ISI citation index. *University Affairs*. Retrieved June 5, 2006 from: http://www.universityaffairs.ca/issues/2006/march/google_scholar_01.html.

Coats, A.J.S. (2005). Top of the charts: Download versus citations in the International Journal of Cardiology. *International Journal of Cardiology, 105(2),* 123-125.

Darmoni, S.J., Roussel, F., Benichou, J., Faure, G.C., Thirion, B., & Pinhas, N. (2000). Reading factor as a credible alternative to impact factor: A preliminary study. *Technology and Health Care, 8(3-4),* 174-175.

Davis, P.M., & Fromerth, M.J. (2006). Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles? Retrieved June 5, 2006 from: http://arxiv.org/abs/cs.DL/0603056.

Fosmire, M. (2004). Scan it and they will come … But will they cite it? *Science & Technology Libraries, 25(1/2),* 55-72.

Garfield, E. (November 4, 1985). Contemporary classics in the life sciences: An autographical feast. *Current Contents, (44),* 3-8.

Garfield, E. (June 9, 1986). Do Nobel Prize winners write citation classics? *Current Contents, (23),* 3-8.

Jasco, P. (2005). As we may search – Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced database. *Current Contents, 89(9)*, 1537-1547.

Kaplan, N.R., & Nelson, M.L. (2002). Determining the publication impact of a digital library. *Journal of the American Society for Information Science & Technology, 51(4),* 324-339.

Moed, H.F. (2005). Statistical relationships between downloads and citations at the level of individual documents within a single journal. *Journal of the American Society for Information Science & Technology, 56(10),* 1088-1097.

Noruzi, A. (2005). Google Scholar: The next generation of citation indexes. *Libri, 55(4),* 170-180.

Roth, D.L. (2005). The emergence of competitors to the Science Citation Index and the Web of Science. *Current Science, 89(9),* 1531-1536.