
Open Access Determinants and the Effect on Article Performance

Sumiko Asai

School of Political Science and Economics, Meiji University, Tokyo, Japan

Email address:

asai@meiji.ac.jp

To cite this article:

Sumiko Asai. Open Access Determinants and the Effect on Article Performance. *International Journal of Business and Economics Research*. Vol. 6, No. 6, 2017, pp. 145-152. doi: 10.11648/j.ijber.20170606.11

Received: November 23, 2017; **Accepted:** December 11, 2017; **Published:** January 10, 2018

Abstract: Although open access has steadily developed with the continuous increase in subscription journal price, the effect of open access articles on citations remains a controversial issue. The present study empirically examines the factors determining authors' choice to provide open access and the effects of open access on downloads and citations in hybrid journals. This study estimates author's choice of open access using a probit model, and the results show that the cost of open access is an important factor in the decision. After a test for endogeneity of open access choice, the equation for downloads is estimated with the variables representing characteristics of articles and authors. The results of estimating downloads by ordinary least squares show that open access increases the number of downloads in hybrid journals. On the other hand, from citation estimations using a negative binomial model, this study found that the effect of open access on the number of citations differs among hybrid journals. It is a good practice for authors to consider a balance between article processing charges and the benefits that will be gained from open access when deciding whether to provide open access.

Keywords: Hybrid Journal, Open Access, Download, Citation

1. Introduction

Whereas a continuous increase in subscription journal price has been a serious problem in academic circles, everyone can download articles from open access journals free of charge. Open access journals that are distributed electronically have become an important tool for obtaining information and knowledge through the reach of the internet. In addition to open access journals, academic journal publishers such as Elsevier and Springer replaced traditional subscription journals with hybrid open access journals in the early 2000s. Hybrid open access journals give authors of the articles the option to retain their copyright by paying article processing charges for open access or to transfer the copyright to the publishers without paying. If authors pay the article processing charges to the publishers, they can post the articles in their institutional repositories and on their own websites without any condition, and every reader can access and download the articles free of charge. In contrast, if they do not pay the article processing charges, the articles are only accessed by subscribers of the journal or readers who pay a fee to access each individual article. Thus, a hybrid journal

includes open access articles and non-open access articles and, as its name shows, has characteristics of both open access and traditional subscription journals. In the case of hybrid journals provided by Springer, the article processing charge for open access is USD 3,000, or EUR 2,200, for an article, and other traditional publishers generally have similar article processing charges for open access.

Open access articles collected in hybrid journals are easy to access and might be cited more frequently, although the authors must pay article processing charges. If researchers who select open access do enhance their fame through the many resulting downloads or citations of their articles, it would be worthwhile for the authors to pay article processing charges for open access. However, whereas several studies asserted that open access has a positive impact on citations, [1] and [2] found that the positive impact of open access is small and insignificant. A consensus regarding the effect of open access has not been reached.

The present study empirically investigates the effect of open access in hybrid journals on the numbers of downloads and citations. Although several previous studies compared the number of citations of subscription journals with that of open

access journals, the quality of the journals differ. Therefore, the present study compares the numbers of downloads and citations between open access and non-open access articles in a hybrid journal in order to avoid the problem of quality difference among journals.

The present study is unique as it considers whether the choice for open access is endogenous in the equations for downloads and citations. The quality of articles may differ not only among multiple authors but also among the articles written by an individual author. An author may deliberate upon the open access option for each individual article from the standpoint of balancing the costs of and benefits gained from open access, because the article processing charge traditional publishers set for open access is high. When an author writes an article that is expected to acquire many downloads and/or citations, this author may select open access and pay the article processing charge. Conversely, when an author is not confident about the article quality, the author may opt not to pay the article processing charge in hybrid journals. If the author takes into account the quality of the article at the time the decision to provide open access is made, the decision may influence the numbers of downloads and citations. In this case, the decision is regarded as an endogenous variable in the equations for downloads and citations. The present study first investigates the determining factors in choosing open access in hybrid journals, and then tests the endogeneity of the open access option in download and citation equations before these two equations are estimated.

This study examines three hybrid economic journals published by Springer: *Theory and Decision*, *Small Business Economics*, and *European Journal of Health Economics*. *Theory and Decision* collects theoretical studies, as indicated by the title. Most articles collected in *Small Business Economics* are empirical studies about small and medium-sized enterprises and their activities. *European Journal of Health Economics* is an interdisciplinary journal as compared with the other two journals, and the authors of the journal include medical doctors and pharmacologists in addition to health economics researchers in universities. Thus, these three journals have different characteristics, which is the reason for selecting them as samples. The second reason is that these three hybrid journals have a relatively high number of open access articles and high impact factors. In addition, the reason for selecting journals published by Springer is that the publisher notes the numbers of downloads and citations for each individual article on the journal's web site.

2. Related Literature

[3] compared the citations between offline and online articles in computer science. He reported that the average number of citations to offline articles is 2.74, whereas the number of citations to online articles that are freely available is 7.03. Since [3] found that freely available online articles are more highly cited, the issue of knowledge distribution has been addressed by researchers in a wide range of fields. [4] compared the number of citations to open access and

non-open access articles in four disciplines: philosophy, political science, electrical and electronic engineering, and mathematics. [4] reported that open access articles had more citations in these four disciplines. However, most studies conducted in the early 2000s compared the average citations of open access and non-open access articles without considering the differences in article quality and/or author performance, posing a problem in the analytical method.

Empirical studies conducted since the late 2000s considered characteristics of the journal and the author to examine the impact of open access on citation. [5] empirically investigated the impact of open access on citation in ophthalmology through a general linear model using variables representing the number of authors, funding, and region in which the article was published. [5] found that open access was not a statistically significant factor in citation, although the average number of citations for open access articles was larger than that for articles collected in subscription journals. [6] examined the impact of open access on citation in a hybrid journal titled *Proceedings of the National Academy of Sciences* using a stepwise backward logistic regression model. [6] found that open access articles are more frequently cited in their early stages as compared with non-open access articles and pointed out the difference in the timing of the citation. [7] investigated the impact of open access on citations in three fields (biology, mathematics, and pharmacy and pharmacology) using regression analysis and found that the impacts are different among varying fields of science.

[8] investigated the impact of open access on citation for 11 hybrid journals in biological and medical fields using regression analysis. [8] showed that two of the 11 journals see positive effects on the number of citations due to open access, although the citation advantage appears to weaken over time. [9] compared the number of citations between open access and non-open access articles in hybrid journals published by Elsevier and Springer. They found that open access articles in health science and natural science are cited more than non-open access articles, whereas open access articles in social science and humanities are cited slightly more than non-open access articles. However, [10] stated that the open access advantage in social science humanities disappears when the number of citations of the single most-cited article is eliminated from the data set in [9], suggesting that careful examination of data is needed for analyzing the advantage of open access.

[11] investigated the impact of open access and non-open access articles in *Nature Communications* on page views, citations, and social media, and found that open access has positive effects on the three aspects. [1] analyzed the impact of open access on download and citation numbers in scientific hybrid journals, and found that the number of downloads for open access articles is greater than that for non-open access articles, whereas the effect on citation is insignificant. [1] concluded that open access publishing may reach more readers, but additional readers may not generate more citations. [12] investigated the relationship between the numbers of downloads and citations in an organic chemistry journal titled *Tetrahedron Letters* and reported that the

Spearman rank correlation, at 0.22, is small. In contrast, [13] calculated the correlation between numbers of downloads and citations in an electronic open access journal called *Journal of Vision*, which is published by the Association of Research in Vision and Ophthalmology, and reported that the correlation is 0.74. Thus, findings regarding the relationship between downloads and citations also differ among empirical studies.

[2] investigated the effect of open access on citation in hybrid journals in the field of science. [2] found that open access increases the number of citations by eight percent on average and that the effect is not large. Further, [14] investigated the impact of open access on citation in hybrid economic journals using the Poisson quasi-maximum likelihood regression and reported that the effect is not significant. Thus, we do not reach a consensus on the effect of open access performance.

3. Data

The subjects of the present study are articles in three hybrid economic journals published by Springer: *Theory and Decision*, *Small Business Economics*, *European Journal of Health Economics*. Original articles published during the period from 2010 to 2015 are our sample, and editorials and

reviews are excluded. Further, original articles on special issues are also not included, because the topic of the special issue, which was set by the editors, may influence the numbers of downloads and citations.

A brief summary of samples from the three journals is provided in Table 1. The average ratio of open access articles is low at 8.9 percent, although Springer seems to promote hybrid journals more compared with other major publishers, and the three journals include many open access articles compared with hybrid economic journals published by Springer. As [6] reported that the ratio of open access articles in *Proceedings of the National Academy of Sciences* was 14 percent as of December 2004, it seems that the open access choice had not dispersed to economic journals.

Data on downloads and citations were collected on the journals' web sites from late February to early March 2017. Table 1 shows that the values of the skewness of downloads and citations in the three journals are positive and large, implying that the distribution of downloads and citations is skewed to the right. The minimum value of downloads for each article is 33 in *Theory and Decision*, and the values of downloads are all positive integers. In contrast, 157 of a total of 1,079 articles in the three journals have not been cited as of March 2017.

Table 1. Summary of Economic Journals.

	Articles		Download		Citation		Impact Factor (5 years)
	Total	Open Access	Mean	Skewness	Mean	Skewness	
<i>Theory</i>	335	24 (7.2%)	254	2.87	3.27	4.05	0.936
<i>Small</i>	356	25 (7.0%)	962	3.84	10.77	2.53	2.318
<i>Health</i>	388	47 (12.1%)	494	3.53	6.49	1.78	2.266

Theory: Theory and Decision; Small: Small Business Economics; Health: European Journal of Health Economics.

Table 2 shows the means of the numbers of downloads and citations, based on a classification of open access and non-open access. Welch *F* test is used to test the equality of the means of downloads and citations between open access and non-open access, because the hypothesis that variances are equal is rejected. The test results show that open access articles have more downloads from the three journals at the 1 percent significance level. In contrast, the hypothesis that the means of the number of citations are equal is not rejected for *Theory and Decision* and *Small Business Economics* at the 10 percent significance level, and the number of citations for open access articles is larger than that for non-open access articles collected in *European Journal of Health Economics* at the 1 percent significance level. It seems that the effect of open access differs in downloads and citations among the journals.

Table 2. Means of Numbers of Downloads and Citations.

	Download		Citation	
	Open	Non-Open	Open	Non-Open
<i>Theory</i>	598	228***	3.9	3.2
<i>Small</i>	2194	869***	15.0	10.4
<i>Health</i>	1223	394***	10.4	5.9***

***1% significance level; *Theory*: Theory and Decision; *Small*: Small Business Economics; *Health*: European Journal of Health Economics.

The correlation values between the numbers of downloads and citations reported by [12] and [13] are 0.22 and 0.74, respectively. Whereas the correlations in *Small Business Economics* and *European Journal of Health Economics* are almost similar at 0.561 and 0.521, respectively, the correlation in *Theory and Decision* is small at 0.373.

4. Estimation

4.1. Open Access Option

After a submitted article is accepted by editors, the author is offered a choice between open access and non-open access in a hybrid journal. First, the present study investigates the determinants driving authors' choice between open access and non-open access using a probit model. The dependent variable *Open* is set to 1 if the article is an open access article, and is 0 otherwise. The independent variables denote the characteristics of the article and author. These are similar to the variables previous studies used to represent the characteristics of articles and authors. The variable *Page* is given by number of pages, and the variable *Reference* denotes the number of references for an article. The two variables represent the characteristics of the article. The variable *Person* denotes the number of authors for an article. It is assumed that

more authors enable them to share the article processing charge for open access. The variable *Fund* is set to 1 if the article is financially supported by a fund, and is 0 otherwise. Information relating to financial support is available from the acknowledgements in the article. The variable *Performance* is stated by the number of articles written by the first author and published in any academic journal until the submitted article was accepted by the journal editors. The data are collected from Web of Science provided by Clarivate Analytics (formerly Thomson Reuters). The variable *Performance* represents the author’s achievement at the time of making the decision regarding the option for open access.

The present study classifies the locations of the research institutions to which the first authors belong as the US or Canada, Europe, and other regions. Open access policy and the capability to pay article processing charges may differ among countries and regions. Whereas several open access journals set high article processing charges for authors in developed countries, lower charges are required for authors in developing countries, in consideration of their ability to pay.

In contrast, article processing charges for open access in hybrid journals set by Springer are static worldwide and high at USD 3,000 or EUR 2,200 for an article. Therefore, the present study uses the variables *US Canada* and *Others* representing the capability to make payment. The variable *US Canada* is set to 1 if the location of the research institute to which the first author belongs is in the US or Canada, and is 0 otherwise. The variable *Others* is set to 1 if the location is other than the US, Canada, or Europe, and is 0 otherwise.

In the case of hybrid journals published by Springer, authors who select open access pay USD 3,000, or EUR 2,200, excluding tax. However, if the institution to which the corresponding author belongs has an agreement with Springer regarding open access choice, the author can choose open access at no charge. The variable *Springer* is set to 1 if the corresponding author belongs to an institution that has an agreement with Springer when the article is accepted, and is 0 otherwise.

The determinants of open access choice can be examined using equation (1). Equation (1) is estimated using a probit model.

$$Open = \alpha_0 + \alpha_1 Page + \alpha_2 Reference + \alpha_3 Person + \alpha_4 Fund + \alpha_5 Performance + \alpha_6 US\ Canada + \alpha_7 Others + \alpha_8 Springer \tag{1}$$

Table 3. Estimation Results of Open Access Choice.

	Theory	Small	Health
α_0 Constant	-2.358 (0.589) ***	-1.689 (0.563) ***	-1.798 (0.365) ***
α_1 Page	-0.024 (0.022)	-0.030 (0.029)	0.007 (0.028)
α_2 Reference	-0.007 (0.010)	-0.005 (0.004)	0.003 (0.005)
α_3 Person	0.287 (0.134) **	0.257 (0.099) ***	0.042 (0.041)
α_4 Fund	-0.062 (0.274)	-0.119 (0.254)	0.114 (0.182)
α_5 Performance	0.004 (0.023)	-0.006 (0.019)	-0.003 (0.008)
α_6 US Canada	0.429 (0.507)	0.195 (0.345)	-0.004 (0.345)
α_7 Others	0.848 (0.395) **	-0.132 (0.471)	-0.203 (0.360)
α_8 Springer	2.099 (0.359) ***	1.431 (0.281) ***	1.138 (0.213) ***
McFadden R ²	0.343	0.259	0.120
AIC	0.392	0.427	0.696

***1%, **5%, and *10% significance levels; Theory: Theory and Decision; Small: Small Business Economics; Health: European Journal of Health Economics; AIC: Akaike Information Criterion. The standard errors are in parentheses.

Table 3 reports the estimation results for open access choice. All estimated coefficients of the variable *Springer* for the three economic journals are significantly positive at the 1 percent significance level. Twenty-four articles in *Theory and Decision* are open access, and the corresponding authors for 17 of the 24 open access articles (71 percent) did not have to pay open access processing charge, because their institutions signed an agreement for open access with Springer. Moreover, 16 of the 25 open access articles (64 percent) in *Small Business Economics* and 20 of the 47 open access articles (43 percent) in *European Journal of Health Economics* were not subject to article processing charges for the same reason. The estimated coefficients of the variable *Person* are all positive, and the values in the two journals are statistically significant, suggesting that sharing article processing charges among authors leads to more frequent adoption of open access. The positive coefficients for the variables *Springer* and *Person* imply that cost is an important

factor affecting open access choice. In other words, authors in research institutions that do not sign an agreement with Springer generally hesitate to choose the open access option.

The estimated coefficients of the variable *Fund* for the three journals (α_4) are not significant at the 10 percent significance level. The fund is given when the research begins, and it usually takes several years from the start of research to the acceptance of the article. The fact that financial support does not generally cover activities performed after the completion of the article may make the estimated values insignificant. The estimated coefficients of the variable *Performance* for the two journals are negative, although the hypothesis that the value is equal to zero is not rejected at the 10 percent significance level. The negative value may indicate that authors who have already achieved high performance and built fame do not have incentive to disperse their articles by selecting open access. In other words, the estimated value suggests that researchers who intend to get positions or obtain promotions at research institutions tend to

select open access with the expectation that open access tends to disseminate the authors' names and their articles.

4.2. Download

The present study estimates the number of downloads using variables representing characteristics of authors and articles to examine the effect of open access on downloads. Equation (2) is a logarithmic function, since all articles have a positive number of downloads.

If authors choose open access for an article that is expected to gain many downloads, the choice of open access may influence the number of downloads. In this case, the decision regarding the open access option is an endogenous variable in the download equation; therefore, equations (1) and (2) should be estimated simultaneously. In contrast, if the decision about the open access option is an exogenous variable in the download equation, only the equation for downloads should

$$\begin{aligned} \ln \text{Download} = & \alpha_0 + \alpha_1 \ln \text{Page} + \alpha_2 \ln \text{Reference} + \alpha_3 \ln \text{Person} + \alpha_4 \text{Fund} \\ & + \alpha_5 \ln H \text{ index} + \alpha_6 \text{US Canada} + \alpha_7 \text{Others} \\ & + \alpha_8 \ln \text{Past Month} + \alpha_9 \ln \text{Past Month}^2 + \alpha_{10} \text{Open} \end{aligned} \quad (2)$$

In this case, \ln denotes the natural logarithm. The dependent variable *Download* is given by the number of downloads for an article. The variables *Page*, *Reference*, *Person*, *Fund*, *US Canada*, and *Others* are the same as those in equation (1). However, the implication of the variables *US Canada* and *Others* used in equation (2) is different from that in equation (1). The two variables used in equation (1) represent the capability to pay article processing charges for open access. On the other hand, these variables in equations (2) and (3) are used to control the influence of the topic the article addresses on the numbers of downloads and citations. Theoretical studies are universal, and the topics of articles generally do not reflect regional characteristics. In contrast, the two empirical economic studies address social systems such as small and medium-sized enterprise business and the health and medical system. Whereas social systems have global commonalities, they also have differences across countries, reflecting the political and economic circumstances of each country. In general, a reader may be interested in the empirical article that deals with a social system that has similarities with the system in the reader's country, and vice versa. In this case, the numbers of downloads and citations to an empirical article may be influenced by the country or region targeted by the article.

The variable *H index* is calculated using the numbers of both articles written by the first author and citations for the author, and the data are available at Web of Science. The variable *H index* for the first author of an article is used in equation (2), instead of the variable *Performance* adopted in equation (1), for the following reason. Though authors are aware of the number of their own articles published in academic journals, most authors do not frequently check the number of citations of their articles and/or H index. Therefore, the number of articles published in academic journals is used

be estimated using ordinary least squares (OLS). In order to test the endogeneity, the present study added the ordinary residuals of equation (1) estimated using the probit model to equation (2) as an independent variable and estimated equation (2) with the ordinary residuals using OLS. When the hypothesis that the estimated coefficient of the ordinary residuals in equation (2) is equal to zero is rejected, the variable *Open* is justified as an endogenous variable, and thereby equations (1) and (2) are estimated simultaneously. In contrast, when the hypothesis is not rejected, the variable *Open* is exogenous and only equation (2) is estimated using OLS. The estimated results show that the coefficients of the ordinary residuals were not significant at the 10 percent significance level for all three economic journals. Therefore, the decision regarding open access is justified as exogenous in equation (2), and the equation for download is estimated for the three journals using OLS.

as the independent variable in equation (1) to identify the determinants of open access. In contrast, people who download an article are readers of the article. When the factors affecting the number of downloads are investigated, the variable denoting the author's achievement needs to incorporate evaluation from the viewpoint of readers. Since the H index provided at Web of Science is calculated using the numbers of both articles written by the author and citations by readers of the article, the index is used as the variable denoting the performance of the author in equation (2). However, the correlation coefficients between the variables *Performance* in equation (1) and *H index* in equation (2) are approximately 0.9 for the three journals. When the two variables are used in equation (2) as independent variables, a multi-collinearity problem arises. Furthermore, equation (1) with the variable *Performance* fits more than this equation with the variable *H index*, whereas equation (2) with the variable *H index* fits more than equation (2) with the variable *Performance*.

The variable *Past Month* denotes the number of months after the article was first published on the journal's web site. The present study added the squared value as the independent variable in consideration of the possibility of non-linear relationships between the numbers of downloads and months from the first publication.

Table 4 reports the estimation results of the number of downloads specified by equation (2). The estimated coefficients for the variable *Reference* for the three journals are all positive at the 1 percent significance level. An article with more references may imply that the article covers a wider range of topics, which leads to an increase in related researchers. The estimated coefficients for the variable *Past Month* are positive, and the squared values are negative at the 1 percent significance level. The results are common to the three journals. It implies that the larger the number of elapsed

months, the higher the number of downloads, but this increase rate gradually decreases. The three estimated coefficients for the independent variable *Open* are all positive at the 1 percent significance level, implying that open access articles have a larger number of downloads. The results imply that it is reasonable for authors who wish to acquire many downloads to choose open access. The estimated coefficients for the variable *Fund* are all negative, although these are not significant. It seems that financial support for research does not necessarily lead to an author writing articles that are

heavily downloaded. Furthermore, the estimated coefficients for the variable *H index* are all positive, and articles written by authors who already show high performance tend to be more frequently downloaded. However, the estimated coefficient in *Theory and Decision* is not significant. The scope of this journal covers theoretical studies, and the authors tend to be younger, whereas the main scope of the other two journals is empirical studies. It seems that the difference in the characteristics of journals is reflected in the estimated coefficients of the variable *H index*.

Table 4. Estimation Results for Download.

	<i>Theory</i>	<i>Small</i>	<i>Health</i>
α_0 Constant	-11.109 (3.161) ***	-6.931 (2.181) ***	-18.392 (3.399) ***
α_1 Page	0.054 (0.088)	0.026 (0.121)	0.086 (0.091)
α_2 Reference	0.329 (0.062) ***	0.476 (0.071) ***	0.250 (0.062) ***
α_3 Person	0.148 (0.063) **	-0.014 (0.070)	0.063 (0.044)
α_4 Fund	-0.034 (0.064)	-0.007 (0.064)	-0.013 (0.051)
α_5 H index	0.051 (0.051)	0.138 (0.042) ***	0.104 (0.040) **
α_6 US Canada	-0.006 (0.087)	-0.008 (0.078)	0.230 (0.090) **
α_7 Others	-0.064 (0.080)	-0.037 (0.096)	0.002 (0.082)
α_8 Past Month	8.093 (1.591) ***	5.847 (1.116) ***	11.803 (1.724) ***
α_9 Past Month ²	-1.075 (0.198) ***	-0.743 (0.144) ***	-1.509 (0.218) ***
α_{10} Open	1.127 (0.116) ***	0.903 (0.123) ***	1.147 (0.076) ***
Adjusted R ²	0.407	0.313	0.469
AIC	1.630	1.752	1.405

***1%, **5%, and *10% significance levels; *Theory*: *Theory and Decision*; *Small*: *Small Business Economics*; *Health*: *European Journal of Health Economics*; AIC: Akaike Information Criterion. The standard errors are in parentheses.

4.3. Citation

This subsection discusses the impact of open access on the number of citations. Ninety-seven articles (29.0 percent) in *Theory and Decision*, 15 articles (4.2 percent) in *Small Business Economics*, and 45 articles (11.6 percent) in the *European Journal of Health Economics* have not been cited as of late February or early March 2017. The present study estimates the number of citations using the count regression model, because the numbers of citations are small values, including zero. Although the traditional count regression model is the Poisson model, the property of Poisson, that is the mean and variance are equal, is rejected by an over dispersion test for the three journals. Therefore, the present study uses a negative binominal regression model that relaxes the constraint. The negative binominal model includes two types, according to [15]. When α is a parameter to be estimated and ω and μ denote the variance and mean of the samples, respectively, the variance function of the first type is defined by $\omega = (1 + \alpha)\mu$, and that of the second type is defined by $\omega = (1 + \alpha\mu)\mu$. The citation equations for the two types of

models are estimated using a quasi-maximum likelihood method, and the type is selected from the viewpoint of fitness. *Theory and Decision* and *Small Business Economics* adopt the second type of negative binominal model, whereas *European Journal of Health Economics* adopts the first type. However, the differences in the estimation results are small between the two model types.

The Hausman-type test conducted for the download equation was also implemented in the citation estimation to test whether the open access option is an endogenous variable in the citation equation. The results indicate that the open access option is an exogenous variable for the three journals, which is similar to the download equation. Therefore, the present study estimates a single equation for citation using the negative binominal regression model.

The dependent variable *Citation* is given by number of citations for an article. Although the independent variables are common to those for download estimation, the logarithms were not taken for the variables. The citation function is specified by equation (3).

$$Citation = \alpha_0 + \alpha_1 Page + \alpha_2 Reference + \alpha_3 Person + \alpha_4 Fund + \alpha_5 H Index + \alpha_6 US Canada + \alpha_7 Others + \alpha_8 Past Month + \alpha_9 Past Month^2 + \alpha_{10} Open \quad (3)$$

The estimation results are reported in Table 5. The articles with many more references tend to be more frequently cited, and the articles written by high performance authors also generate more citations. The estimated coefficients for the variables *US Canada* and *Others* in *Small Business Economics* are negative and significant at the 1 percent significance level.

The characteristics of small business may differ among countries or regions. If an empirical article in *Small Business Economics* focuses on small business that is specific to a country or region, researchers in other countries or regions may not be highly interested in this study. In the case of *Small Business Economics*, articles written by first authors who

belong to research institutions in Europe account for 67 percent, and the estimation results show that these articles are more frequently downloaded and cited. In contrast, articles written by first authors who belong to institutions in Asia, Africa, South America, and Australia, that is, areas that correspond to the variable *Others*, account for only 12 percent. Those articles tend to have smaller numbers of downloads and citations. On the other hand, the estimated coefficients for the variables *US Canada* and *Others* are not significant in *Theory and Decision*, although they are negative. One reason for the insignificant values is that *Theory and Decision* deals with theoretical studies, and the articles do not generally have region-specific characteristics.

The estimated coefficients for the variable *Past Month* are positive, and the squared values are negative for all three journals, which is similar to the download equation. The results imply that the numbers of elapsed months and citations have a positive relationship, but the increase rate gradually decreases. Impact factor is calculated by the number of citations for articles published in the past three or five years. The estimation results suggest that setting a three- or five-year period for calculation may provide a reasonable balance between the level of effort and accuracy of the calculations.

The impact of the open access option on citation is different among the three journals. Open access articles in *Theory and Decision* and *Small Business Economics* do not influence the number of citations, whereas open access in *European Journal*

of Health Economics increases the number of citations significantly. If universities and other research institutions make Big Deal contracts with Springer, in which the publisher provides all electronic journals under the condition that the research institutions continue to purchase the subscription journals, researchers belonging to the organizations that make Big Deal contracts can access and download all articles collected in electronic journals published by Springer. Since articles in hybrid journals are also electronically distributed, the researchers can access all articles in hybrid journals without additional charges, irrespective of whether the article is open access or non-open access. In other words, readers who benefit more from open access articles in hybrid journals are researchers belonging to institutions that do not have Big Deal contracts. Almost all authors in *Theory and Decision* and *Small Business Economics* are researchers who belong to universities. Perhaps, many readers of the two journals also belong to universities that make Big Deal contracts with major publishers. In contrast, with regard to the *European Journal of Health Economics*, several authors work for hospitals or medical centers that may not enter Big Deal contracts. The finding that open access significantly increases the numbers of downloads and citations in the *European Journal of Health Economics* may suggest that the impact of open access depends on the circumstances surrounding readers, who may become future authors in this journal.

Table 5. Estimation Results for Citations.

	<i>Theory</i>	<i>Small</i>	<i>Health</i>
α_0 Constant	-2.987 (0.750) ***	-1.125 (0.346) ***	-1.012 (0.596) *
α_1 Page	0.016 (0.009)	0.007 (0.009)	0.011 (0.015)
α_2 Reference	0.020 (0.004) ***	0.008 (0.001) ***	0.009 (0.003) ***
α_3 Person	0.029 (0.071)	0.055 (0.041)	0.046 (0.023) **
α_4 Fund	0.013 (0.137)	0.058 (0.085)	0.076 (0.103)
α_5 H index	0.069 (0.023) ***	0.054 (0.009) ***	0.034 (0.006) ***
α_6 US Canada	-0.151 (0.187)	-0.293 (0.107) ***	0.134 (0.180)
α_7 Others	-0.043 (0.191)	-0.391 (0.131) ***	-0.329 (0.174) *
α_8 Past Month	0.066 (0.021) ***	0.061 (0.009) ***	0.042 (0.019) **
α_9 Past Month ²	-0.0003 (0.0001) **	-0.0003 (0.00007) ***	-0.0001 (0.0001)
α_{10} Open	0.171 (0.229)	-0.026 (0.160)	0.479 (0.145) ***
Log likelihood	-887.21	-1122.25	-1082.37
AIC	5.362	6.367	5.636

***1%, **5%, and *10% significance levels; *Theory*: *Theory and Decision*; *Small*: *Small Business Economics*; *Health*: *European Journal of Health Economics*; AIC: Akaike Information Criterion. The standard errors are in parentheses.

5. Conclusion

The average ratio of open access articles to total articles for the three hybrid journals is low, at 8.9 percent from 2010 to 2015, and more than half of the open access articles were not subject to the article processing charges. The present study first examined the determinants involved in choosing the open access option using the probit model and found that the cost is an important determining factor in choosing open access. This may make the open access decision an exogenous variable in the equations for downloads and citations.

The skewness of downloads and citations for these journals

is positive and large, implying that a few articles have large numbers of downloads and citations. Although many universities have made Big Deal contracts with major publishers that enable them to access all electronic journals provided by the publisher, they have recently seen a troubling increase in the annual payments for those contracts. The skewness of downloads and citations suggests that there is room to reconsider the renewal of Big Deal contracts from the standpoint of a balance between the costs of these contracts and the benefits generated by access to a large number of electronic journals.

Furthermore, although the present study found that publishing an article as open access increases the number of

downloads in the three examined journals, it does not lead to an increase in citations for two of these journals, and the impact of open access on the number of citations differs among journals. For a few decades, new academic journals, including open access journals, have been released, and the number of research articles collected in journals has, in turn, increased drastically. Under the circumstances, if the purpose of choosing open access is the prevalence of the article, article processing charges for open access may be worth paying. However, if the author's purpose is an increase in the number of citations, which is a representative index for evaluating researchers' performance, the author needs to consider the readership and characteristics of the academic journal in advance of this decision.

Finally, we should note that the impact of open access on article citations and downloads from hybrid journals may change due to the influence of Big Deal contracts. Recently, several universities terminated their Big Deal contracts due to consistent fee increases. As more universities withdraw from their Big Deal contracts, the benefits that readers gain from open access articles in hybrid journals may grow, and the effect on citations may be large. Reader environment needs to be considered when evaluating the impact of open access in hybrid journals.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 15K03470.

References

- [1] Davis, P. M. (2011). Open access, readership, citations: a randomized controlled trial of scientific journal publishing. *The FASEB Journal*, 25 (7), 2129–2134.
- [2] McCabe, M. J. & Snyder, C. M. (2014). Identifying the effect of open access on citations using a panel of science journals. *Economic Inquiry*, 52 (4), 1284–1300.
- [3] Lawrence, S. (2001). Online or invisible? *Nature*, 411 (6837), 521.
- [4] Antelman, K. (2004). Do open-access articles have a greater research impact? *College & Research Libraries News*, 65 (5), 372–382.
- [5] Lansingh, V. C. & Carter, M. J. (2009). Does open access in ophthalmology affect how articles are subsequently cited research? *Ophthalmology*, 116 (8), 1425–1431.
- [6] Eysenbach G. (2006). Citation advantage of open access articles. *PLoS Biology*, 4 (5) <http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0040157> (last access on 1 March 2017).
- [7] Frandsen T. F. (2009). The integration of open access journals in the scholarly communication system: Three science fields. *Information Processing and Management*, 45 (1), 131–141.
- [8] Davis, P. M. (2009). Author-choice open-access publishing in the biological and medical literature: A citation analysis," *Journal of the American Society for Information Science and Technology*, 60 (1), 3–8.
- [9] Sotudeh, H., Ghasempour, Z., & Yaghtin, M. (2015). The citation advantage of author-pays model: the case of Springer and Elsevier OA journals. *Scientometrics*, 104 (2), 581–608.
- [10] Wray, K. B. (2016). No new evidence for a citation benefit for author-pay open access publications in the social science and humanities. *Scientometrics*, 106 (3), 1031–1035.
- [11] Wang, X., Liu, C., Mao, W., & Fang, Z. (2015). The open access advantage considering citation, article usage and social media attention. *Scientometrics*, 103 (2), 555–564.
- [12] 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 and Technology*, 56 (10), 1088–1097.
- [13] Watson, A. B. (2009). Comparing citations and downloads for individual articles. *Journal of Vision*, 9 (4), 1–4.
- [14] Müller-Langer, F. & Watt, R. (2014). The hybrid open access citation advantage: How many more cites is a \$3000 fee buying you? *Munich Personal RePEc Archive, Paper No. 61801*, <https://mpra.ub.uni-muenchen.de/61801/> (last accessed 20 March 2017).
- [15] Cameron, A. C., & Trivedi, P. K. (1986). Econometric models based on count data: Comparisons and applications of some estimators and tests. *Journal of Applied Econometrics*, 1 (1), 29–53.