A survey-based structural equation model analysis on influencing factors of non-citation

Zewen Hu[1], Yishan Wu[2,*] and Jianjun Sun[3]

[1]School of Management Science and Engineering, Nanjing University of Information Science and Technology, Nanjing, Jiangsu 210044, China
[2]Chinese Academy of Science and Technology for Development, Beijing 100038, China
[3]School of Information Management, Nanjing University, Nanjing, Jiangsu 210093, China

Although bibliometric approach has been frequently utilized to analyse reasons behind non-citation and show relations between uncitedness and impact factors, the survey-based structural equation model approach is not usually used. Therefore, a Likert scale questionnaire was designed to collect data on non-citation and its various types of determinants. The survey-based structural equation model was used to analyse mutual relations and correlation degrees between non-citation rate and its various determinants. As a result, the categories 'academic status of journal' and 'personal features of papers' were found to be two extremely significant determinants of non-citation rate. Their path coefficients reached 0.83 and 0.43 respectively. Accordingly, the category 'contents and topics of papers' was shown to have extremely significant indirect influence on non-citation rate through 'academic status of journal'. The three observed variables of 'academic status of journal' including 'public praise of journal', 'impact factor of journal', and 'member of SCI, EI and Scopus Journals', showed the highest values of indirect effect on the non-citation rate. Furthermore, there were weaker correlations among 'academic status of journal', 'personal features of papers', 'contents and topics of papers' and 'publicity and recommendation' except between 'contents and topics of papers' and 'academic status of journal'. Meanwhile, the six observed variables of 'publicity and recommendation' and 'contents and topics of papers' show the smaller values at ≤0.12 of indirect effect on non-citation rate. Our empirical results suggest some significant determinants of non-citation rate that might enlighten researchers on how to improve the chance of having their works cited, and assist them in expanding their research impact. These findings can also help journal editors to identify contributions with high-citation potential.

Keywords: Determinants, impact factor, influencing factors, non-citation, questionnaire, structural equation model, uncitedness.

THE 'non-citation rate (NCR)' refers to the proportion of articles that do not receive a single citation within a given timeframe following their publication. This is a common phenomenon in science publication domain. The future citation analysis and research indicators should take uncitable articles into account[1]. Empirical analysis results on factors influencing uncitedness may help such entities as researchers, universities, countries, editorial staff of journals, and administrators of institutes to identify contributions that have high-citation potential. These results may also help academic organizations increase the chance of having papers cited and lower the percentage of uncited papers, thus raising their performance in impact assessments and overall research quality.

Studies on uncitedness from the Web of Science, Google Scholar and Scopus database can be mainly classified into two categories – bibliometric analysis and survey-based analysis. Bibliometric analysis has been frequently utilized to reveal reasons or determinants of uncitedness. Although survey-based analysis has not been widely used, it can be an important complementary method to reveal various determinants causing uncitedness, especially those that are hard to be quantified and revealed through bibliometric analysis. That means, survey-based analysis results from subjective perceptions of small group can also provide complementary and mutually verified evidence for objective results from bibliometric analysis.

In the present article, a survey-based approach was applied to analyse the mutual relationships between non-citation rate (NCR) and its various determinants. We first take NCR as a cognitive concept, not the concrete value, and design a Likert scale questionnaire considering NCR and its various types of determinants. A survey-based structural equation model (SEM) was then employed to verify and reveal the mutual relations and correlation degrees between the uncitedness factor and its various types of determinants including the related factors presented by various authors[2–11], such as journal impact factor, accessibility and internationality of journal, quality and type of publication, topic and length of publication, number and reputation of authors, and other determinants that are hard to be quantified.

Literature review

Quite some time ago, the reasons behind the phenomenon of non-citation were explained. Garfield[2] considered that non-citation of papers may occur due to their mediocre,

*For correspondence. (e-mail: wuyishan@istic.ac.cn)

low-quality, low-impact source journal, and further argued that such papers may be unintelligible, irrelevant, or valuable yet undiscovered or forgotten Years later, Garfield[3] further listed a series of possible reasons leading to non-citation of articles, such as language, type of publication, being 'premature,' delayed recognition, bibliographic plagiarism, or other variations of misconduct. However, these descriptions on reasons for non-citation were seldom supported by empirical studies.

Fortunately, after many years, a series of bibliometric studies that focused on influencing factors of non-citation were reported[12–16] which revealed a decreasing or negative *S*-shaped relationship between impact factor and uncitedness factor. Similar studies[17–19] also pointed journal impact factor as the most important determinant of future citation impact for articles. Accordingly, a positive correlation between *h*-index and the number of uncited papers was verified on a sample of 75 top researchers from different fields[20].

Further, the most important determinant of citation score of paper – top-author – was identified[4] from its influencing variations including top-author's number of references, language, journal category and journal influence through multiple regression analysis. The quality and impact of publications were reported[21] to have an influence on their citation through analysing different citation patterns of publications. A comparison for several variables – average number of pages, references and authors – between cited and uncited papers was made[22], and their influence on the citation of papers was studied. A bibliometric analysis found[23] the length of a paper to greatly influence its citeability. Another meaningful and comprehensive study utilized a zero-inflated negative binomial regression model and negative binomial-logit hurdle model on the data set from Web of Science to verify a series of determinants of citation counts of articles, covering the journal impact factor, the impact of references, the internationality of authors, journals and references, the number of authors, institutions and references. This study identified the impact on journals and references as the most effective determinants of citation counts of articles[7,8].

However, fewer survey-based analyses on reasons of citation found. Yue[24,25] explored the effects of various external factors on journal citation impact by combining SEM and empirical data from 41 research journals in clinical neurology. The results revealed that accessibility and internationality of the journal, and its perceived quality have large, medium, and small effects respectively on journal citation impact. This study provided a perspective at periodical level, bearing upon our empirical analysis about various types of determinants of non-citation. Following this, eight major causes were revealed[10] facilitating easy citation of one's papers through a questionnaire containing a series of subjective judgements: research hotspots and novel topics of content, longer intervals after publication, research topics similar to citers work, high quality of content, reasonable self-citation, highlighted title, prestigious authors, and academic tastes and interests similar to citers.

Structural equation modelling

SEM can be used to explore and verify multiple and interrelated causal relationships among different variables including observed (measurement) and unobserved (latent) variables through factor analysis, path analysis, multiple correlation and regression analysis. Furthermore, SEM can also measure errors in estimation and define a concept model explaining an entire set of relationships among variables[26,27]. In this paper, we present an SEM constructed to explore and verify the mutual relations and the relationship between uncitedness factor and various related factors.

SEM involves latent and observed variables, and reveals the relationships among these variables. Latent variables are abstract and cannot be observed directly or quantized, but they can be measured indirectly by observed variables. In contrast, observed variables can be measured or quantized by using the item rating scale in a questionnaire[28]. Measurement and structural components are two core components in SEM. The former comprises measurement errors of observed variables, as well as mutual relationships between observed variables and the represented latent variable. The latter describes the relationships among latent variables.

Based on hypotheses on causal relationships among latent variables, a sketch of SEM is shown in Figure 1. It contains one structural component and four measurement components. The structural component contains hypothetical relationships among latent variables.

In Figure 1, squares that contain χ and y represent observed variables, while ellipses that contain ξ and η represent latent variables. The latent variable ξ and its corresponding observed variables χ are called exogenous variables that only play an explanatory role or have a one-way influence on other variables in the model. Meanwhile, the latent variable η and its corresponding observed variables y are called endogenous variables that

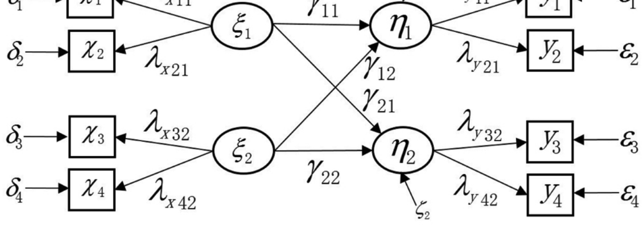

Figure 1. Sketch map of the structural equation model.

Table 1. Latent variables and their observed variables

Latent variables	Abbr.	Observed variables	Abbr.
Academic status of journal	AS	Public praise of journal	Jvalue1
		Impact factor of journal	Jvalue2
		Member of SCI, EI, and Scopus Journals	Jvalue3
		Age of journal	Jvalue4
Contents and topics of papers	CT	Novelty of paper topic	Topic1
		Quality of paper content	Topic2
		Interdisciplinarity of paper content	Topic3
Personal features of papers	PF	Number of authors	Self1
		Length of paper	Self2
		Native English-speaking authors	Self3
		Number of references	Self4
Publicity and recommendation	PR	Recommendation by self-citation	Pub1
		Active recommendation of papers to peers	Pub2
		Recommendation by social media such as blogs and forums	Pub3
Non-citation rate	NCR	Academic reputation of journal's sponsor	Ncr1
		Popularity of first author or corresponding author	Ncr2
		Funded by national foundation	Ncr3
		Open and free access	Ncr4

not only have an influence on the other variables, but can also be affected. Furthermore, δ represents the error of variables χ; ε represents the error of variables y; and ζ represents residuals or disturbances that cannot be explained by exogenous variables. In addition, λ_x represents the mutual relations between the latent variable ξ and its corresponding observed variables χ; λ_y represents the mutual relations between latent variable η and its corresponding observed variables y; and γ represents the mutual relations between exogenous variables ξ and endogenous variables η.

The SEM as shown in Figure 1 contains the following three matrix equations.

$$y = \Lambda_y \eta + \varepsilon, \quad (1)$$

$$\chi = \Lambda_\chi \xi + \delta, \quad (2)$$

$$\eta = B\eta + \Gamma\xi + \zeta. \quad (3)$$

Here, Λ_y is a factor loading matrix consisting of a series of λ_{yij} that reflects the mutual relations between η_j and its corresponding observed variables y_i. Further, Λ_x is a factor loading matrix assembling λ_{xij} that reflects the mutual relations between ξ_j and its corresponding observed variables χ_i. Finally, B is the path coefficient that reflects the correlation degree among η, and Γ is the path coefficient that reflects the influencing degree of ξ on η.

Methodology

Classification and definition of influencing factors

To examine the influence of all possible factors on non-citation of papers, we first collected all possible influencing factors through surveying literature and consulting experts, researchers and authors. We then classified these factors into four main categories: 'academic status of journal' (AS), 'contents and topics of papers' (CT), 'personal features of papers' (PF) and 'publicity and recommendation' (PR). These categories were combined with 'non-citation rate' (NCR) to make up a total of five latent variables. Where NCR is a core endogenously latent variable, that is reflected by the four other one-order exogenously latent variables. Each latent variable contained at least three sub-items (these are also referred to as the latent variable's observed variables). The five latent variables and their observed variables are shown in Table 1.

Hypotheses of influencing relations among variables

To verify the influencing relations and degrees of the four exogenously latent variables on the endogenously latent variable NCR, four basic hypotheses were defined as follows.

H1. The 'academic status of journal' (AS) and its four sub-indicators have significant influence on 'non-citation rate' (NCR). The improvement of AS can lower NCR, i.e. percentage of uncited papers in this journal.

H2. The 'contents and topics of papers' (CT) and its three sub-indicators have significant influence on NCR. The novelty, quality, and interdisciplinarity of papers' contents and topics are reflective of CT. Papers involving more novel topics and high-quality, interdisciplinary work in journals may receive more chances to get cited over an extended period of time.

H3. The 'personal features of papers' (PF) and its four sub-indicators have significant influence on NCR. Stern[29]

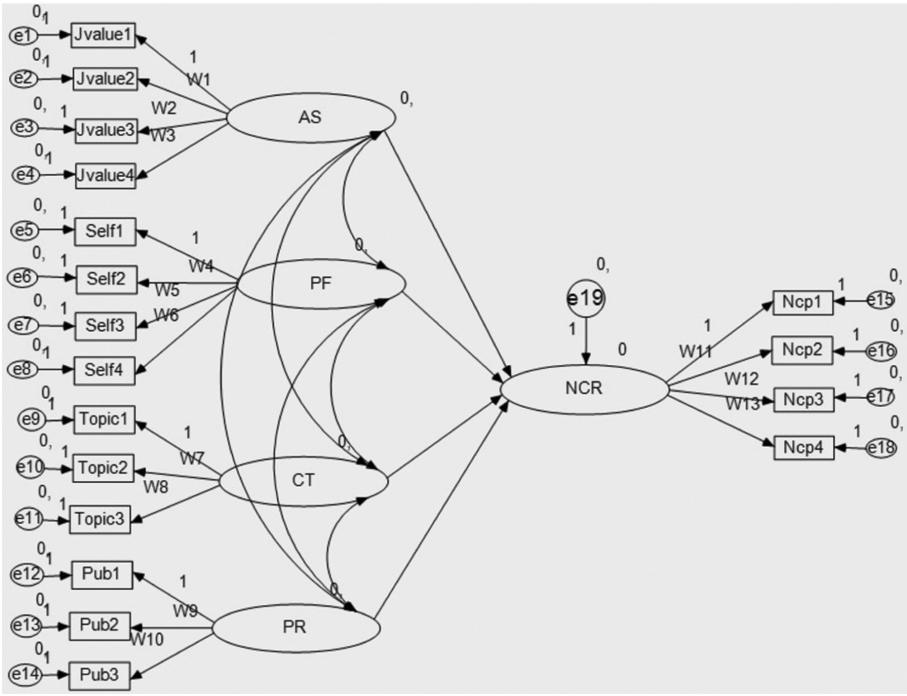

Figure 2. The initial sketch map of structural equation model.

suggested that such personal features as number of authors, words of a title, and keywords and references had some influence on whether a paper is cited or not, and found that the influence of the number of references was far higher than other personal features. However, there still needs a new method to explore the influence of other sub-indicators on NCR, as well as to verify existing findings.

H4. The 'publicity and recommendation' (PR) methods and its three sub-indicators have significant influence on NCR. Evans and Reimer[30] have even found that free internet access can widen the circle of those who read and make use of scientists' studies, and can increase the citations of an article by 8%. Another empirical study found that free access can only increase the download counts of articles, but does not appear to increase instances of citation[31]. However, regarding the influence of other PR methods on NCR, we have not yet discovered related studies.

Construction of SEM

Based on the aforementioned hypotheses on the influencing relations among four exogenously latent variables (AS, CT, PF and PR) and the core endogenously latent variable (NCR), we constructed an initial sketch map of SEM using the AMOS (Analysis of Moment Structures) software, as shown in Figure 2.

NCR represents the dependent endogenous variable, while the abbreviated AS, CT, PF and PR are the four respective types of influencing variables or independent exogenous variables of NCR. The abbreviated Jvalue1–Jvalue4, Self1–Self4, Topic1–Topic3, Pub1–Pub3 and Ncr1–Ncr4 respectively represent observed variables of AS, CT, PF, PR and NCR. The symbols e1–e19 respectively represent the error term of each variable. The symbols W1–W13 respectively represent the regression weight of paths among variables, and the number 1 is the initial default value of the regression weight of some paths.

Collection of sample data

To verify the initially developed SEM, we designed a Likert scale questionnaire that reflects the influencing degree or correlation degree of five kinds of observed variables on latent variables as shown in Table 1 and Figure 2. We rated the correlation degree (or called influencing degree) of eighteen observed variables along a five-point scale (1 = negative correlation, 2 = non-correlation, 3 = weak correlation, 4 = general correlation, and 5 = strong correlation). The actual survey questions in Likert scale questionnaire form the concrete correlation degree at five-point scale for each observed variable on its latent variable. We placed the Likert scale questionnaire on the biggest Chinese academic website–*sciencenet* (http://www.sciencenet.cn/english/) on 1 November 2014 for 6 months, then randomly attracted and invited Chinese researchers to answer problems in questionnaire that will not be listed here one by one. As a result, 240 valid

Table 2. Distribution of degrees and professional titles from 198 participants

Degrees	Percentage	Titles	Percentage
Bachelor	6.06	Assistant lecturer or lecturer	11.62
Master	42.93	Senior lecturer	35.86
Ph D	42.93	Associate professor or professor	29.29
Postdoctoral	7.07	No title	23.23

Table 3. Distribution of research output and major domains from 198 participants

Research output	Number of participants	Percentage	Major domain	Percentage
≤5 papers	66	33.50	Management	57.58
6–10 papers	38	19.29	Engineering and technology	11.11
11–15 papers	41	20.81	Biology	6.06
≥16 papers	52	26.40	Medicine	5.56
			Chemistry	5.56

Table 4. Distribution of research history from 198 participants (defined as the years since they published their first paper)

History	Percentage	History	Percentage
≤5 years	38	11–15 years	12
6–10 years	39	≥16 years	11

Table 5. Reliability test results

Categories	Number of observed variables	Cronbach's alpha
AS	4	0.818
CT	3	0.842
PF	4	0.758
PR	3	0.859
NCR	4	0.645
Total	18	0.88

questionnaires were obtained after deleting the duplicated questionnaires by checking respondents' IP addresses, names, and contact information, and the rate of valid questionnaires was 87.59%. There were 198 questionnaires answered by respondents who were aware of the citation situation of their papers. After eliminating 16 questionnaires that failed to capture significant data, the remaining 182 questionnaires were coded as sample data. 182 questionnaires is sufficient for our survey-based analysis that meets the requirement of at least 100 questionnaires and at least 5 times the number of variables recommended in earlier studies[32–34].

The distributions of degrees and professional titles, research output and major domains, as well as research history from 198 participants are shown in Tables 2–4 respectively.

From Table 2, we see that among the 198 participants who answered the questionnaire, most have masters or Ph D degree, as well as the titles of senior lecturer or professor. A total of 170 participants hold masters or Ph D degrees (85.86%), 70 (35.86%) as senior lecturer and 58 (29.29%) as associate professor or professor.

As shown in Table 3, among the 198 participants who answered the questionnaire, 40.1% have published between six and fifteen papers, 26.40% have ≥ sixteen papers, while the other 33.50% of the participants have ≤ five papers. Table 3 also shows the distribution of main domains from 198 participants, with 114 (57.58%) participants from management and 22 (11.11%) participants from engineering and technology expressing their views on various reasons of non-citation. In comparison, there are only 12 (6.06%), 11(5.56%) and 11(5.56%) participants who are respectively from biology, medicine and chemistry domain.

Table 4 shows that 38% of participants have engaged in research for less or equal to five years since publication of their first academic paper, and 39% have research history for 6 to 10 years, 12% have devoted themselves to research for 11 to 15 years, and 11% have longer research history of more than 16 years.

Reliability test

Before verifying the initial developed SEM, the Cronbach's α test method[35,36] was used to examine the reliability and internal consistency of five kinds of observed variables by the statistical tool referred to as the Statistical Package for Social Sciences (SPSS) in combination with data collected through the survey. Moreover, we also utilized Kruskal–Wallis (KW) method[37] to test the significant difference of coincident attitudes of respondents from different groups to determinants of non-citation. There is no significant difference if the significance value (Sig) is ≥0.05, while there is significant difference if the significance value is <0.05.

Table 6. Significant difference of coincident attitudes from different groups of respondents

Group variables	Determinants	Number of no significance variables with sig > 0.05	Number of significance variables with sig < 0.05
Different disciplines	18	17	1
Different number of papers	18	17	1
Different education levels	18	15	3
Different research history	18	18	0
Different titles	18	18	0

The test results are shown in Tables 5 and 6. The testing results in Table 5 show that the Cronbach's α values for each type of observed variable vary from 0.645 to 0.859, with 0.88 as the total alpha value. Studies by Nunnally[38] and Cortina[39] show that the Cronbach's α value exceeding 0.7 represents a high reliability and internal consistency. Therefore, it can be concluded that our sample is reliable and five kinds of observed variables are basically reliable for verifying the SEM.

Table 6 shows the significant difference of coincident attitudes to 18 determinants of non-citation from different groups of respondents. The test results show that 198 respondents from different disciplines, education levels, research history, academic titles, and with different number of papers all express identical or quite similar views with no significant difference to most determinants of non-citation.

Results

Descriptive statistical analysis

Based on the sample data collected, we conducted a descriptive statistical analysis of standard deviations and means of sample data on influencing degrees of each type of factor (or observed variables) on NCR, as shown in Table 7. We also conducted a descriptive statistical analysis of the percentages of assentors (who support or agree with statements in the questionnaire) on the different influencing degree of each factor, as shown in Table 8.

Table 7 shows that the standard deviations of sample data on influencing degrees for each type of factors are between 0.954 and 1.193, with smaller values accounting for a smaller deviation degree. While the standard deviation for data on Ncr2 is the biggest, the smallest deviation value is for Pub1. This indicates that attitudes of respondents regarding influencing degrees of Ncr2 on NCR show a bigger fluctuation. On the contrary, there is a smaller fluctuation for the attitudes of Pub1. Table 7 also shows that the means of sample data on influencing degrees for each type of factor are between 2.45 and 4.54. Finally, the smallest overall mean is 2.45 for Self1, and the biggest overall mean is 4.54 for Jvalue1. Thus, the influence degree of Self1 on NCR is far lower than that of Jvalue1.

Table 8 shows that most of the respondents consider Pub1, Self2 and Self1 as three factors with lower influence degree values of 1 and 2 on NCR, where accumulated percentages of assentors reach 36.20, 42.30 and 54.90 respectively. Secondly, there are 70.9%, 71.5%, 80.7%, 81.3%, 82.4%, 86.8%, 89% and 90.6% of respondents who respectively, consider Ncr1, Ncr2, Topic1, Topic3, Jvalue3, Jvalue2, Jvalue1 and Topic2 as eight factors with higher influence degree values of 4 and 5 on NCR. Thirdly, among four influencing factors belonging to AS category, the largest percentage of respondents (79.1%) considered Jvalue1 as the factor with the highest influence degree values of 5 on NCR. Among three influencing factors belonging to CT category, the largest percentage of respondents (73.6%) considered Topic2 as the factor with highest influence degree values of 5 on NCR. Fourthly, among four influencing factors belonging to NCR category, the largest percentage of respondents (39.60%) considered Ncr1 as the factor with the highest influence degree values of 5 on NCR. Furthermore, majority of respondents (90%) did not consider the four factors belonging to PF category and the three factors belonging to PR category to be the biggest influencing factors on NCR.

SEM analysis of influencing factors on NCR

Based on the sample data obtained with high reliability, we first executed an initial fitting analysis to SEM (Figure 2) by utilizing the AMOS software platform. The maximum likelihood estimation method and the standardized path coefficient (also called loading coefficient) were used. The results indicated that skewness and kurtosis reflecting the normality of eighteen variables in the model were all under the guidelines of less than 3 and 10 respectively, as recommended by Keline[27]. The skewness for 67% variables and kurtosis for 78% variables were <1 approaching the optimal value of zero, and the skewness between 1.278 and 2.399, while the kurtosis is between 1.699 and 5.03 for the remainder. Furthermore, some paths with higher modification indices were modified through constructing the co-variation relations among them. The modified model and fit results are shown in Figure 3.

Path coefficients among variables. Critical ratio (CR) is the ratio of estimated value and its standard deviation of

Table 7. Standard deviation and mean of sample data of influencing factors (or observed variables)

Categories of influencing factors	Influencing factors (or observed variables)	Std. deviation	Mean
AS	Public praise of journal (Jvalue1)	1.040	4.54
	Impact factor of journal (Jvalue2)	1.062	4.34
	Member of SCI, EI, and Scopus Journals (Jvalue3)	1.116	4.25
	age of journal (Jvalue4)	1.183	3.51
CT	Novelty of paper topic (Topic1)	1.133	4.12
	Quality of paper content (Topic2)	0.965	4.52
	Interdisciplinarity of paper content (Topic3)	1.072	4.13
PF	Number of authors (Self1)	1.101	2.45
	Length of paper (Self2)	0.964	2.71
	Native English-speaking authors (Self3)	1.015	2.98
	Number of references (Self4)	1.001	2.83
PR	Recommendation by self-citation (Pub1)	0.954	2.73
	Active recommendation of papers to peers (Pub2)	1.025	3.07
	Recommendation by social media (Pub3)	1.051	3.17
NCR	Academic reputation of journal's sponsor (Ncr1)	1.115	3.93
	Popularity of first author or corresponding author (Ncr2)	1.193	3.87
	Funded by national foundation （NCR3）	1.092	3.08
	Open and free access （Ncr4）	1.186	3.45

Table 8. Percentages of assentors on varying degrees of influencing factors

Categories	Influencing factors	Influencing degrees and their respective percentage				
		1	2	3	4	5
AS	Jvalue1	4.40	2.20	4.40	9.90	79.10
	Jvalue2	4.40	1.60	7.10	24.20	62.60
	Jvalue3	5.50	2.70	9.30	23.60	58.80
	Jvalue4	4.90	13.70	28.60	26.40	26.40
CT	Topic1	4.40	6	8.80	29.10	51.60
	Topic2	3.30	2.20	3.80	17	73.60
	Topic3	3.30	4.40	11	32.40	48.90
PF	Self1	21.40	33.50	29.70	8.20	7.10
	Self2	9.90	32.40	37.90	16.50	3.30
	Self3	8.20	17.60	45.10	22	7.10
	Self4	8.20	17.60	45.10	22	7.10
PR	Pub1	11.50	24.70	44	16.50	3.30
	Pub2	6.60	21.40	33.50	31.90	6.60
	Pub3	7.10	17.60	32.40	34.10	8.80
NCR	Ncr1	4.40	4.90	19.80	31.30	39.60
	Ncr2	6	8.20	14.30	34.10	37.40
	Ncr3	8	23	34	26	10
	Ncr4	7	13	31	26	23

parameters W_i. CR and its corresponding probability p are major criteria testing the significance of path coefficients. If $P < 0.01$, the path coefficient among each pair of variables shows an extremely significant difference.

As shown in Figure 3, our model achieved a good fit as the aforementioned hypothesized causal relationships were replaced by correlations and path coefficients reflecting the degrees of correlation. The estimated parameters reflecting the model performance show that the other path coefficients are all extremely significant at $P < 0.01$ except for the two path coefficients between CT and NCR, and between PR and NCR. The χ^2, degrees of freedom and probability levels are 261.072, 119 and 0.000 respectively, indicating that an extremely significant model at $P < 0.01$ is obtained. Further, the goodness-of-fit indices also obtain better values. As an illustration, the comparative fit index (CFI) and normed fit index (NFI) respectively reach 0.923 and 0.869, approaching the level of the optimal value of 1. The root mean square error of approximation (RMSEA) of model shows a small value

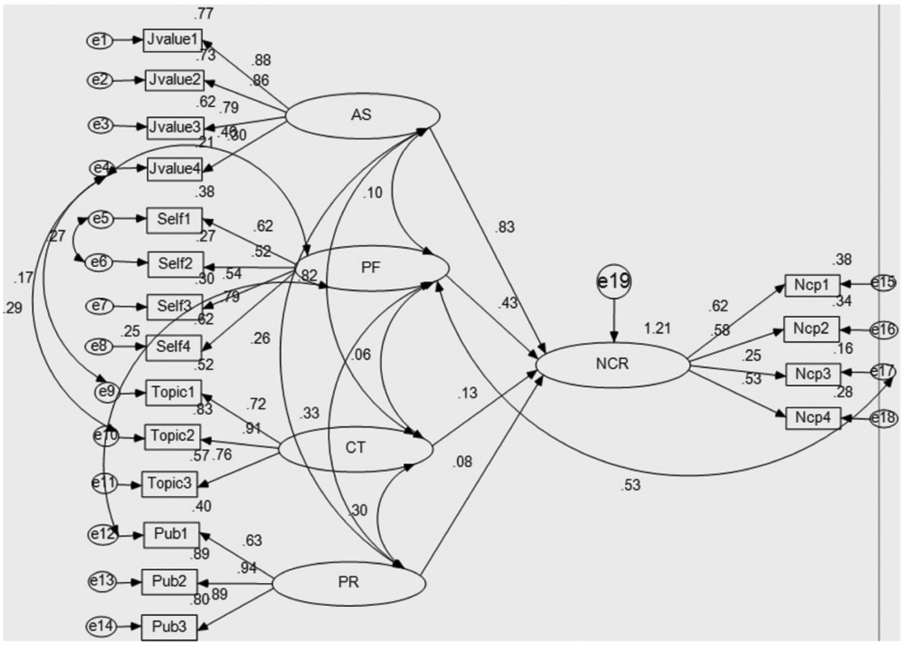

Figure 3. Fit results of the modified structural equation model.

of 0.078, which is close to an optimal standard value of zero[40].

Analysis results of direct effects among variables: The path coefficients among different variables are equal to values of direct effects among them. As shown in Figure 3, AS has an extremely significant influence on NCR of papers. The path coefficient between AS and NCR reaches 0.83, which means that NCR will drop 0.83 units, when AS increases by one unit. Among four measured variables of AS, Jvalue1 is the most significant observed variable as the path coefficient reaches 0.88, which means that AS will improve by 0.88 units when the Jvalue1 increases by one unit. In contrast, Jvalue4 is the weakest observed variable of AS with a path coefficient of 0.46.

Results also revealed that PF has significant influence on NCR. The path coefficient among these two variables is 0.43, which means that NCR will drop by 0.43 units when PF improves by one unit. Among the four measured variables of PF, Self4 reflects PF most accurately as the path coefficient reaches 0.79.

However, CT and PR do not have a significant influence on NCR of papers. The path coefficients among them are only 0.13 and 0.08 respectively, which means that NCR will drop by only 0.13 and 0.08 units when CT and PR improve by one unit respectively. Further, among the three observed variables of CT, Topic2 is the most significantly influential variable with a 0.91 path coefficient, while among the three observed variables of PR, Pub2 has the biggest influence on improvement of PR as the path coefficient reaches 0.94.

Also, among the four observed variables of NCR, Ncr1 is the most significantly influential variable with a path coefficient of 0.62, while Ncr3 is the weakest influencing variable of NCR with a path coefficient of 0.25.

Mutual relations among four exogenously latent variables: As shown in Figure 3, the relation between CT and AS is most significant as their path coefficient reaches 0.82. Following that, the weaker relations are shown between PF and PR, and between CT and PR, as well as between AS and PR, with their correlation coefficients being 0.33, 0.3 and 0.26 respectively. Finally, the correlations between PF and CT, and between PF and AS are least significant with the smallest path coefficients at 0.06 and 0.1 respectively.

Analysis results of indirect effects among variables: The path coefficient can only reflect the direct effect of mutual influence among different variables, but cannot reflect the indirect effect among variables. Indirect effect refers to the causal variable as having an indirect influence on the independent variable by one or more intervening variables. When there is only one intervening variable, the value of indirect effect among two variables is equal to the result of multiplying the path coefficient between one variable and an intervening variable with the path coefficient between another variable and the intervening variable. Based on the values of direct effects among variables (Figure 3), we determined the values of indirect effect between the endogenous latent variable (NCR) and fourteen observed variables of AS, PF, CT and PR, as shown in Table 9.

Table 9. Values of indirect effects between NCR and fourteen observed variables

Observed variables	Indirect effects	Observed variables	Indirect effects
Jvalue1	0.73	Self1	0.27
Jvalue2	0.71	Self2	0.22
Jvalue3	0.65	Self3	0.23
Jvalue4	0.38	Self4	0.34
Pub1	0.05	Topic1	0.10
Pub2	0.08	Topic2	0.12
Pub3	0.07	Topic3	0.10

Table 9 shows conclusions similar to those shown in Figure 3. The three observed variables respectively named Jvalue1–Jvalue3 have larger indirect effects on NCR. Their values of indirect effect are 0.73, 0.71 and 0.65 respectively, indicating that NCR will drop by 0.73, 0.71 and 0.65 units respectively, when the three observed variables improve by one unit. This survey-based finding on the decreasing relationship of 0.71 between impact factor (Jvalue2) and NCR is similar to conclusions obtained in various earlier studies[12–16,41].

On the other hand, the six observed variables of PR (Pub1–Pub3) and CT (Topic1–Topic3) show smaller values ≤0.12 of indirect effect on the NCR. Of these, the quality of paper content (Topic2) has a relatively larger value of 0.12 for indirect effect on the NCR. Further, among the four observed variables of PF (Self1–Self4), Self4 has a larger value of 0.34 for indirect effect on NCR, and this finding is similar to conclusions obtained by Stern[29] and Webster *et al.*[42] who also found the influence of number of references on NCR to be far higher than other personal features.

Finally, we also figured out the values of indirect effects between NCR and three exogenously latent variables PF, CT and PR through the intervening variable AS, and their values are 0.083, 0.681 and 0.216 respectively, highlighting the fact that CT and PR have larger indirect influence on NCR through AS, although they are shown to have the smallest direct influence on NCR.

Conclusion

Although some scholars have analysed the determinants of non-citation through bibliometric analysis, very few empirical studies can comprehensively show and discuss all types of influencing factors and their respective influencing degree on NCR using survey-based SEM methods. In this paper, we designed a Likert scale questionnaire and performed SEM analysis on mutual relations and correlation degrees between NCR and its various influencing factors in combination with data collected through the questionnaire. Through this process, we suggest the following conclusions.

(i) Our Likert scale questionnaire has shown a positive effective rate and high reliability. The rate of valid questionnaires reaches 87.59%. The Cronbach's α values reflecting the reliability of data on each influencing factor vary from 0.645 to 0.859, and the total α value is 0.88, representing high reliability and internal consistency.

(ii) Our SEM shows positive performance of fit. In the model as shown in Figure 3, all performance indicators of model approach the optimal values. The χ^2, degrees of freedom and probability level reveal that the model reaches an extremely significant level at $P < 0.01$, and most of path coefficients are also extremely significant at $P < 0.01$, except for the two insignificant path coefficients ($P > 0.05$). Furthermore, the goodness-of-fit indices including CFI and NFI also obtain the approximately ideal values approaching the level of 1. And RMSEA is also close to an optimal standard value of zero.

(iii) The results reveal that the AS and the CT are two extremely significant determinants of NCR. Among five latent variables, the AS has the most significant influence on NCR of papers. The path coefficient among them is 0.83. However, CT and PR do not have significant direct influence on NCR of papers with very small path coefficients, viz. 0.13 and 0.08, contrary to our hypotheses. Interestingly, the CT is of more significant indirect influence on NCR through the intervening variable – AS. So, it should not be strange that there is the most significant relation between AS and CT with a path coefficient of 0.82. Further, our survey-based analysis result about the decreasing perceived relationship between impact factor (Jvalue2) and NCR (0.71) was also found through bibliometric analysis[7,8,12–17,19,41].

(iv) Among 18 measured variables, 'public praise of journal' (Jvalue1), 'number of references' (Self4), 'quality of paper content' (Topic2), 'active recommendation of papers to peers' (Pub2) and 'academic reputation of journal's sponsor' (Ncr1) are the most significant observed variables respectively for the latent variables AS, PF, CT and PR and NCR, while 'age of journal' (Jvalue4), 'length of paper' (Self2), 'novelty of paper topic' (Topic1), 'recommendation by self-citation' (Pub1) and 'funded by national foundation' (NCR3) are the weakest observed variables respectively for the latent variables AS, PF, CT, PR an NCR.

(v) Three observed variables of AS, viz. 'public praise of journal' (Jvalue1), 'impact factor of journal' '(Jvalue2)', 'member of SCI, EI, and Scopus Journals' '(Jvalue3)' have the highest values of indirect effect at more than 0.64 on the NCR. Further, all measured variables of PR and CT show the lowest values of indirect effect, with less than or equal to 0.12 for path coefficients.

Some insights can be drawn from our empirical study. First, to reduce NCR and improve the academic influence of scholars, journals, and research organizations, papers are better published in journals that receive good public praise, have high impact factors, and are indexed by prominent academic databases. This also happens to be a common behaviour of most scholars. However, it does not mean that papers issued by journals with a low academic status (AS) are all of a low quality. Another result about strongest correlation between AS and CT also indicates that for improving the AS of journals, written and published papers should involve original topics and high-quality, interdisciplinary content of CT. The highest indirect influence of CT on NCR also indicates that we cannot ignore the importance and role of CT on lowering NCR of papers. In addition, the impact and promotion of journals should also be enhanced by encouraging journals' sponsors to try their best to promote the public's awareness and recognition of their journals. The editorial offices of journals and researchers may publicize and recommend their studies to peers or the wider academic audience through such outlets as email, academic forums and conferences, and open source academic websites and blogs. Accordingly, the important influence of the four PF factors on NCR warn us that bibliometric features of papers should be kept in a standard and normalized format.

The SEM method based on a Likert scale questionnaire is an intuitive and effective approach to analyse the mutual relations between non-citation and the various factors contributing to causes of non-citation. However, it must be recognized that the sample data from this study's questionnaire may be influenced by some subjective biases from respondents and the difference in the group's subject fields, education levels, research history and level, academic titles and rate of non-citation. Although, we have proved through the Kruskal–Wallis (KW) method that different groups of respondents from our current sample data have no significant difference in the coincident attitudes to 18 determinants of non-citation, we cannot guarantee that there will be no significant difference when larger and more individualized samples from different groups of respondents, subject fields and nations are involved.

Therefore, in future, we shall attempt to design a more comprehensive questionnaire focusing on microcosmic and individualized objects and questions to explore the reasons leading to non-citation and low-citation through investigating more varied groups on a global level. Accordingly, we shall also try to design a panel data model and combine the objective statistics data to analyse all types of influencing factors of NCR, and conduct a comparative analysis on the results by using two types of models and sample data.

1. Thelwall, M., Are there too many uncited articles? Zero inflated variants of the discretised lognormal and hooked power law distributions. *J. Infor.*, 2016, **10**(2), 622–633.
2. Garfield, E., Uncitedness III – the importance of not being cited. *Curr. Contents*, 1973, **8**, 5–6.
3. Garfield, E., To be an uncited scientist is no cause for shame. *The Scientist*, 1991, **5**(6), 12.
4. Peters, H. P. F. and Raan, A. F. J. V., On determinants of citation scores: a case study in chemical engineering. *J. Am. Soc. Inform. Sci.*, 1994, **45**(1), 39–49.
5. Yue, W., Predicting the citation impact of clinical neurology journals using structural equation modeling with partial least squares. Dissertation, University of New South Wales, Sydney, 2004.
6. Yue, W. and Wilson, C. S., An integrated approach for the analysis of factors affecting journal citation impact in clinical neurology. *Proc. Am. Soc. Inform. Sci. Technol.*, 2004, **41**(1), 527–536.
7. Didegah, F. and Thelwall, M., Determinants of research citation impact in nanoscience and nanotechnology. *J. Am. Soc. Inform. Sci. Technol.*, 2013, **64**(55), 055–1064.
8. Didegah, F. and Thelwall, M., Which factors help authors produce the highest impact research? Collaboration, journal and document properties. *J. Inform.*, 2013, **7**(4), 861–873.
9. Zhao, S. X., Uncitedness of reviews. *Curr. Sci.*, 2015, **109**(8), 1377–1378.
10. Hu, Z. W. and Wu, Y. S., A probe into causes of non-citation based on survey data. *Soc. Sci. Inform.*; http://arxiv.org/pdf/1507.06879
11. Zhou, P. and Leydesdorff, L., A comparative study of the citation impact of Chinese journals with government priority support. *Frontiers in Research Metrics and Analytics*, 2016; http://journal.frontiersin.org/article/10.3389/frma.2016.00003/full.
12. Van Leeuwen, T. N. and Moed, H. F., Characteristics of journal impact factors: the effects of uncitedness and citation distribution on the understanding of journal impact factors. *Scientometrics*, 2005, **63**(2), 357–371.
13. Egghe, L., The mathematical relation between the impact factor and the uncitedness factor. *Scientometrics*, 2008, **76**(1), 118–123.
14. Egghe, L., The distribution of the uncitedness factor and its functional relation with the impact factor. *Scientometrics*, 2010, **83**(3), 689–695.
15. Hsu, J. W. and Huang, D. W., A scaling between impact factor and uncitedness. *Phys. A – Stat. Mech. Appl.*, 2012, **391**(5), 2129–2134.
16. Burrell, Q. L., A stochastic approach to the relation between the impact factor and the uncitedness factor. *J. Inform.*, 2013, **7**(3), 676–682.
17. Bornmann, L. and Daniel, H. D., Multiple publication on a single research study: does it pay? The influence of number of research articles on total citation counts in biomedicine. *J. Am. Soc. Inform. Sci.*, 2007, **58**(8), 1100–1117.
18. Boyack, K. W. and Klavans, R., Predicting the importance of current papers. In Proceedings of ISSI 2005 (eds Ingwersen, P. and Larsen, B.), Karolinska University Press, Stockholm, pp. 335–342.
19. Kulkarni, A. V., Busse, J. W. and Shams, I., Characteristics associated with citation rate of the medical literature. *PLOS ONE*, 2007, **2**(5), e403.
20. Egghe, L., Guns, R. and Rousseau, R., Thoughts on uncitedness: nobel laureates and Fields medalists as case studies. *J. Am. Soc. Inform. Sci. Technol.*, 2011, **62**(8), 1637–1644.
21. Li, J. and Ye, F. Y., A probe into the citation patterns of high-quality and high-impact publications. *Malaysian J. Lib. Inform. Sci.*, 2014, **19**(2), 17–33.
22. Liang, L. M., Zhong, Z. and Rousseau, R., Uncited papers, uncited authors and uncited topics: a case study in library and information science. *J. Inform.*, 2015, **9**, 50–58.
23. Hu, Z. W. and Wu, Y. S., Regularity in the time-dependent distribution of the percentage of never-cited papers: An empirical pilot study based on the six journals. *J. Inform.*, 2014, **8**(1), 136–146.
24. Yue, W., Predicting the citation impact of clinical neurology journals using structural equation modeling with partial least squares. Dissertation, University of New South Wales, Sydney, 2004.
25. Yue, W. and Wilson, C. S., An integrated approach for the analysis of factors affecting journal citation impact in clinical neurology. *Proc. Am. Soc. Inform. Sci. Technol.*, 2004, **41**(1), 527–536.

26. Cho, K., Hong, T. and Hyun, C., Effect of project characteristics on project performance in construction projects based on structural equation model. *Exp. Syst. Appl.*, 2009, **36**(7), 10461–10470.
27. Keline, R. B. (ed.), *Principles and Practice of Structural Equation Modeling*, The Guilford Press Inc, New York, 2005.
28. Byrne, B. M. (ed.), *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming*, Routledge, New York, 2010.
29. Stern, R. E., Uncitedness in the biomedical literature. *J. Am. Soc. Inform. Sci.*, 1990, **41**, 193–196.
30. Evans, J. A. and Reimer, J., Open access and global participation in science. *Science*, 2009, **323**(5917), 1025.
31. Davis, P. M. and Walters, W. H., The impact of free access to the scientific literature: a review of recent research. *J. Med. Libr. Assoc. Jmla*, 2011, **99**(3), 208.
32. MacCallum, R. C., Widaman, K. F., Zhang, S. and Hong, S., Sample size in factor analysis. *Psychol. Meth.*, 1999, **4**, 84–99.
33. Bryant, F. B. and Yarnold, P. R., Principal components analysis and exploratory and confirmatory factor analysis. In *Reading and Understanding Multivariale Statistics* (eds Grimm, L. G. and Yarnold, R. R.), American Psychological Association, Washington, DC, 1995, pp. 99–136.
34. Kline, P., *Psychometrics and Psychology*, Academic Press, London, 1979.
35. Cronbach, L. J., Coefficient alpha and the internal structure of tests. *Psychometrika*, 1951, **16**(3), 297–334.
36. Hair, J. F., Tatham, R. L., Anderson, R. E. and Black, W., *Multivariate Data Analysis*, Upper Saddle River: Prentice Hall, 1998.
37. Breslow, N., A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship. *Biometrika*, 1970, **57**(3), 579–594.
38. Nunnally, J. C., Assessment of reliability. In *Psychometric Theory*, McGraw-Hill, New York, 1978, 2nd edn.
39. Cortina, J. M., What is coefficient alpha? An examination of theory and applications. *J. Appl. Psychol.*, 1993, **78**, 98–104.
40. Bollen, K. A. and Long, J. S., Testing structural equation models. Newbury Park, Sage Publications, 1993.
41. Egghe, L., The functional relation between the impact factor and the uncitedness factor revisited. *J. Inform.*, 2013, **7**(1), 183–189.
42. Webster, G. D., Jonason, P. K. and Schember, T. O., Hot topics and popular papers in evolutionary psychology: analyses of title words and citation counts in evolution and human behavior, 1979–2008. *Evolutionary Psychol.*, 2009, **7**(3), 348–362.

ACKNOWLEDGEMENTS. This study is supported by the National Natural Science Foundation of China (Grant No. 71603128 and 71373252), the Natural Science Foundation of Jiangsu Province of China (Grant No. BK20160974 and BK20150928), the Humanity and Social Science Youth Foundation of Ministry of Education of China (Grant No. 15YJC870011), the Project of Philosophy and Social Science Research in Colleges and Universities in Jiangsu Province (Grant No. 2015SJB068).

Received 4 December 2017; revised accepted 9 February 2018

doi: 10.18520/cs/v114/i11/2302-2312