Title characteristics and citations in economics

Wohlrabe, Klaus and Gnewuch, Matthias

30 November 2016

Title characteristics and citations in economics

Matthias Gnewuch*   Klaus Wohlrabe†

Abstract

We investigate the relationship between article title characteristics and citations in economics using a large data set from Web of Science. Our results suggest that articles with a short title that also contains a non-alphanumeric character achieve a higher citation count.

JEL Code: A12, A14
Keywords: articles, title characteristics, citations, non-alphanumeric characters, title length

---

*University of Bonn

†Ifo Institute for Economic Research at the University of Munich, Poschingerstr. 5, 81679 Munich, Germany, Phone: +49(0)89/9224-1229, wohlrabe@ifo.de

1 Introduction

Citations have always been handy to get a quick and quantitative measure of a researcher's performance. They further play a central role when it comes to decisions about tenure and the allocation of research funds. Several online resources, such as Scopus, Google Scholar, RePEc and Web of Science, readily supply citation counts for journals, papers and authors. Numerous journal and author rankings are based on citations.

Since citations are crucial not only for the reputation, but for the career of any researcher in the first place, it is comprehensible, that everybody targets a maximal count. Naturally, the safest way to be cited frequently is to write a great and influential paper. Nevertheless, attributes surrounding the mere text of a publication also influence the citation count, positively just as negatively. Some of these attributes are the journal prestige, the number of authors, the time of publication and of course the title. In this study, we focus on the latter attribute and seek to find out, what constitutes good title characteristics for a research paper in economics.

In particular, we evaluate how title length and non-alphanumeric characters affect the success of a publication, i.e. the citation count, using a sample of 312,879 articles from 430 economic journals retrieved from Web of Science. There are several previous studies discussing the titles of scientific papers, which we draw on.

Subotic and Mukherjee (2014) and Nair and Gibbert (2016) provide good overviews of previous research on title characteristics and citations. Regarding the relationship between title length and citations, there is no consensus on whether there is a relationship at all and if there was, whether it would be positive or negative. Results vary across scientific fields and even among studies in the same field. For instance, Rostami et al (2014) find that the length of the title has no impact on the citation rate. Similar conclusions were documented in Alimoradi et al (2016). In contrast, Paiva et al (2012)

state that shorter titles gather more citations. We contribute to this strand of literature by examining the relationship in the field of economics, which has not been subject to analysis before. Merely the work by Guo et al (2015) has studied titles in economics, but not linked them to citations. In addition, we employ a larger data set than most other studies.

Regarding non-alphanumeric characters, Buter and van Raan (2011) find that they are used in 68% of titles across different fields of study and that this share is relatively stable from 1999 to 2008. Further, they find the usage of non-alphanumeric characters in the title to be correlated positively with the impact of the respective paper, although this result does not hold for some individual disciplines. Hartley (2007) does not find any relationship between colons in the title and citations by means of several small-scale studies. We contribute by examining and quantifying the impact of question marks, exclamation marks and colons, as well as non-alphanumeric characters in general.

We proceed as follows. Section 2 describes the data and contains some basic descriptive statistics. Section 3 presents the empirical approach and our main findings. These findings are related to previous research in Section 4.

2 Data

We utilize data from Web of Science (WoS). Our data set contains 312,879 articles that were published between 1980 and 2015 in 430 journals listed in the *Economics* category of WoS. The article title, journal, publication year, length (number of pages), number of authors and citation count as of 2015 were extracted.

In contrast to Nair and Gibbert (2016) we do not evaluate each title manually, but rely on automated evaluation using statistical software. Naturally, the size of the sample requires this approach. For each article, we counted the number of words and

the number of characters in the title. In addition, we checked, whether the title contains a non-alphanumeric character and in particular, whether it contains a question mark, an exclamation mark or a colon.

In our sample, the title length varies from 1 to 53 words or 3 to 366 characters. The left panel of Figure 1 shows that over the years, the average number of words in a title has increased from 9 to 10.5, in line with an increase in the average number of characters from 65 to 77. This finding is consistent with Guo et al (2015). The rise in the average number of words in a title coincides with a rise in the average number of authors per paper, as the right panel of Figure 1 shows. Yitzhaki (1994) analyzes this relationship more thoroughly and offers explanations for why more authors might lead to longer titles. Figure 2 depicts the frequency distribution of the number of words in a title, which is slightly right-skewed (Skewness: 0.71) with a mode of 9.

Figure 1: Words, characters and authors
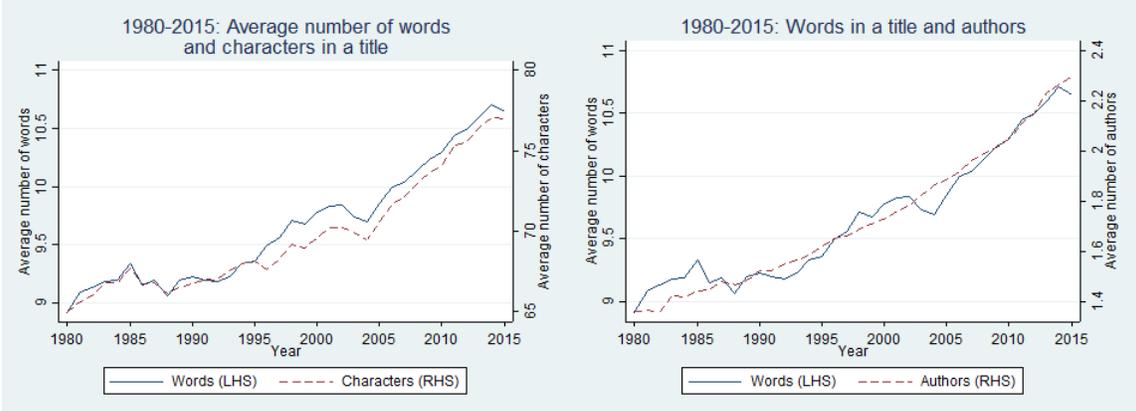
Regarding non-alphanumeric characters, there is an important limitation to our analysis. As the left panel of Figure 3 shows, titles in the data contain hardly any question marks, exclamation marks or colons before 1996. A casual inspection of the data reveals however that there were titles with those characters, but that these characters were not transferred to WoS. Furthermore, the right panel of Figure 3 shows, that

Figure 2: Relative frequency of number of words in a title
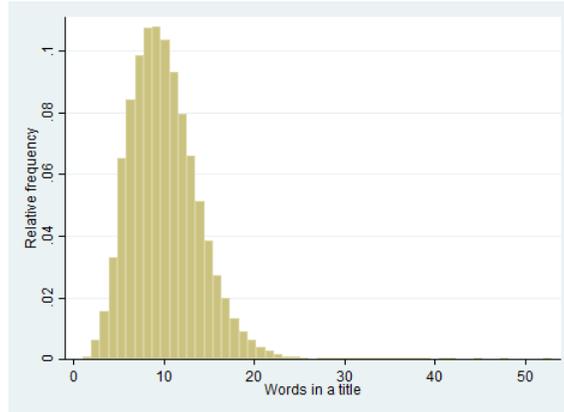
despite the rise in question marks, exclamation marks and colons in 1996, the relative frequency of non-alphanumeric characters fell substantially in that year. A close look at the data unveils, that this development is driven by a decline in utilized hyphens by 58% in 1996. Therefore, we assume that articles before 1996 were processed differently in WoS. In consequence, we only regard articles from 1996 on in our subsequent analysis. Between 1996 and 2015, 11.0% of titles contained a question mark, 0.1% an exclamation mark and 34.2% a colon. 64.5% of titles contained some non-alphanumeric character in general. This percentage for economic papers is slightly below the percentage across fields of 68% found by Buter and van Raan (2011).

Figure 3: Non-alphanumeric characters
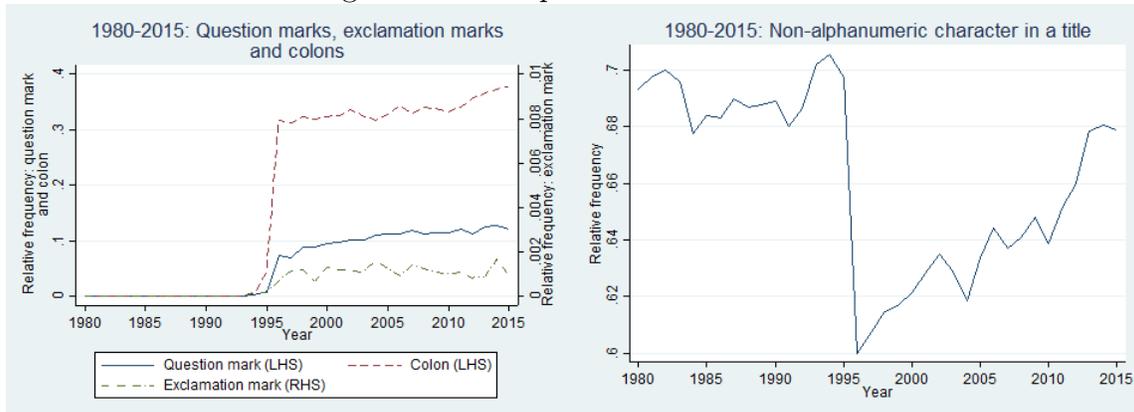
3   Empirical approach and results

We turn to the analysis of the impact of various title characteristics on citations in economic papers. In particular, we examine the effect of the number of words[1] and the presence of non-alphanumeric characters (*nonalpha*), while controlling for variables that previous studies found to have an effect on the citation count. These include the length of the article in pages, the number of authors, and the article age in years (*age*). Our basic regression equation is

$$citations_i = \beta_0 + \beta_1 words_i + \beta_2 pages_i + \beta_3 authors_i + \beta_4 age_i$$
$$+ \beta_5 nonalpha_i + \gamma_j + \delta_t + \epsilon_i$$

where γ_j and δ_t capture journal and time fixed effects. The former account for the general quality or reputation of a journal whereas the latter take changing citations patterns over time into account.

We expect the coefficient for *words* to be negative, because shorter titles are generally considered more concise and convey a clear focus (Nair and Gibbert, 2016). These title properties encourage to read the paper and also make it more memorable, contributing positively to its success. Further, we expect the coefficients for *pages*, *authors* and *age* to be positive in line with previous research, see Nair and Gibbert (2016) among others. Regarding the analysis of non-alphanumeric characters, we expect the coefficients for question marks, exclamation marks and colons to be positive. The idea is that questions (in the title) provoke answers, while exclamations make the title memorable. Further, colons structure the title, making it more concise. Therefore all of these characters should contribute positively to the citation count.

---

[1]We conducted the same analysis with the number of characters instead of the number of words and obtained similar results. We focus on words to make the results more intuitive.

Table 1 shows the results of our analysis. In specification (1), without any non-alphanumeric characters, all variables are statistically significant and have the expected sign. An additional word in the title decreases the citation count ceteris paribus by 0.1.[2] In specification (2), all coefficients remain substantially the same and significant, while the presence of a non-alphanumeric character increases the citation count c.p. by 0.47. In specification (3), a question mark increases the citation count c.p. by 1.64 and a colon by 0.90, while an exclamation mark has no significant effect. The marginal effect of another word in the title amplifies to -0.15. Across all specifications, another page increases the citation count c.p. by 0.36, another author by 1.31 and one more year since publication by 1.21.

Table 1: Regression results

	Dependent variable: article citations					
	(1)		(2)		(3)	
	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.
constant	-8.98	(-13.49)	-9.09	(-13.71)	-8.99	(-13.53)
words	-0.10	(-6.09)	-0.11	(-7.37)	-0.15	(-9.13)
pages	0.36	(9.93)	0.36	(9.93)	0.36	(9.94)
authors	1.31	(19.29)	1.31	(19.29)	1.31	(19.34)
age	1.21	(41.67)	1.21	(41.72)	1.21	(41.84)
nonalpha			0.47	(3.54)		
question					1.64	(9.01)
exclamation					*0.21*	*(0.23)*
colon					0.90	(6.92)
R^2	0.19		0.19		0.19	
N	229,827		229,827		229,827	

Notes: Data from 1996-2015. Applying robust standard errors, all coefficients are statistically significant at the 5%-level, except for *exclamation* (Specification 3, italic)

---

[2]We also performed regressions including the quadratic term of words and obtained substantially the same results.

4 Concluding remarks

While previous studies found mixed evidence on the relationship between title length and the success of a publication, we find that a short title is clearly preferable in economics. This conflicts with the trend of longer titles displayed in Figure 1. Further, we confirm the finding of Buter and van Raan (2011) that the use of a non-alphanumeric character in the title increases the publication's citation count. In contrast to Hartley (2007) we do also find a positive effect of having a colon in the title on citations. Exclamation marks do not significantly affect citations.

Although a good title will not turn a bad into a great paper, we conclude from our analysis, that a good title for a scientific article in economics should be rather short and may very well contain a question mark or other non-alphanumeric character. With regard to our former recommendation, the forthcoming AER article "Beeps" by Ely (2016) should qualify to be successful in terms of the citation count.

References

Alimoradi F, Javadi M, Mohammadpoorasl A, Moulodi F, Hajizadeh M, et al (2016) The effect of key characteristics of the title and morphological features of published articles on their citation rates. Annals of Library and Information Studies (ALIS) 63(1):74–77

Buter R, van Raan AF (2011) Non-alphanumeric characters in titles of scientific publications: An analysis of their occurrence and correlation with citation impact. Journal of Informetrics 5(4):608–617

Ely JC (2016) Beeps. American Economic Review forthcoming

Guo S, Zhang G, Ju Q, Chen Y, Chen Q, Li L (2015) The evolution of conceptual diversity in economics titles from 1890 to 2012. Scientometrics 102(3):2073–2088

Hartley J (2007) Planning that title: Practices and preferences for titles with colons in academic articles. Library and Information Science Research 29(4):553 – 568, DOI http://dx.doi.org/10.1016/j.lisr.2007.05.002

Nair LB, Gibbert M (2016) What makes a good title and (how) does it matter for citations? a review and general model of article title attributes in management science. Scientometrics 107(3):1331–1359

Paiva CE, Lima JPdSN, Paiva BSR (2012) Articles with short titles describing the results are cited more often. Clinics 67(5):509–513

Rostami F, Mohammadpoorasl A, Hajizadeh M (2014) The effect of characteristics of title on citation rates of articles. Scientometrics 98(3):2007–2010

Subotic S, Mukherjee B (2014) Short and amusing: The relationship between title characteristics, downloads, and citations in psychology articles. Journal of Information Science 40(1):115–124

Yitzhaki M (1994) Relation of title length of journal articles to number of authors. Scientometrics 30(1):321–332