The Hybrid Open Access Citation Advantage: How Many More Cites is a $3,000 Fee Buying You?

Frank Mueller-Langer, Max Planck Institute for Intellectual Property and Competition Law,
Munich Center for Innovation and Entrepreneurship Research*
Richard Watt, University of Canterbury, Department of Economics and Finance

May 10, 2013

Preliminary version - Please do not circulate without the authors' permission

Abstract

In this paper, we empirically analyze the effect of author-choice hybrid open access (HOA) at the paper level by comparing citation rates for HOA papers to closed access (CA) papers that appear in the same journal. In contrast to previous literature, our approach allows us to control for journal quality, article quality as well as institution quality. We use a unique, hand-collected dataset containing information on 7,932 articles published from January 2000 to December 2011 in 15 HOA economics journals. We find three main positive and significant explanatory variables for net citations: months since publication, number of references and HOA. Our results suggest that HOA generates, on average, between 0.3 and 0.4 additional net cites per article over a period of up to 76 months since publication. We also compare HOA and CA publishing at the journal level, i.e. in terms of the number of articles published, by setting up a control group of 15 CA journals (9,891 articles), that exhibit the highest degree of relatedness to the HOA journals under study. We find that HOA and CA journals exhibit similar trends in terms of number of papers published and impact factors.

*This research has benefitted from the Sloan Economics of Knowledge Contribution and Distribution Grant 2013. The authors are most grateful for this financial aid. We thank Patrick Andreoli-Versbach, Henri de Belsunce, Bob Reed, Joel Waldfogel, Michael Waldman and participants of the 2012 Congress of the Society for Economic Research on Copyright Issues for their valuable comments and suggestions. We thank Jonas Rathfelder, Janis Sussick and Paul Walker for excellent research assistance.

1 Introduction

Well-established economics journals such as Public Choice or Economic Theory have recently introduced hybrid open access (HOA) as their publication format. HOA journals give authors the option of paying an open access publication fee to make their papers freely available online in addition to the print-version. After submission, papers go through the standard peer-review procedure. Once accepted, the author can choose whether or not to use the open access option. If the author chooses HOA, she has to pay the HOA fee, which is on average about $3,000. In exchange, the copyright remains with the author. She has the right to publish the final version of the article in institutional repositories without any embargo period. Also, she has the right to publish the article freely online, and therefore allow it to be freely downloadable by anyone. The exact conditions, i.e. license type, fees etc., are set by the publishers and tend to be standardized for all participating journals of a particular publisher. HOA has recently been pushed by major scientific publishers as a possible gradual transition path between the traditional closed access (CA) publishing model and open access to scientific publications. In addition, the number of HOA journals across disciplines has increased from 2,017 in October 2009 to 4,381 in February 2012 (Björk, 2012).

We use a unique, hand-collected dataset containing information on 7,932 articles published from January 2000 to December 2011 in 15 HOA economics journals to analyze the impact of HOA at the paper and the journal level. In particular, we address the question of whether the number of cites generated by HOA papers is significantly different from that generated by CA papers. Besides, we explore how other factors such as the months since publication, number of references and authors and the position in an issue drive citation rates. Our work is inspired by McCabe and Snyder (2011) who find that online access does not have a significant positive effect on citations to economics articles. They criticize previous literature that found significant positive citations effects such as Curti at al. (2001), Lawrence (2001) and Walker (2004) for

their failure to control for article quality. In particular, McCabe and Snyder (2011, p. 5) state that "much of the estimated effect of online or open access from the previous literature can be attributed to bias due to omitted quality." Our approach allows us to control for journal quality, article quality as well as institution quality. Article quality is proxied by author quality which we measure with the H-index, total cites and average cites of authors. Intuitively, high-quality authors have sufficient experience in doing research, established networks of co-authors and the necessary resources, e.g. to hire student assistants or Ph.D. students, in order to write high-quality papers. Additional motivating factors are high reference points for quality research, peer pressure as well as reputation concerns. Institution quality is measured by the position of institutions authors are at in the Academic Ranking of World Universities (Shanghai Ranking) 2011. Hence, we account for McCabe and Snyder (2011)'s critique and provide an analysis that is distinct from existing works on open-access related citation effects such as Curti et al. (2001), Davis (2009) and Lawrence (2001).

McCabe and Snyder (2005) suggest that author-pays publication formats such as HOA may have a negative effect on journal quality. Intuitively, HOA publishers may accept more papers of lower quality as compared to the traditional reader-pays model in order to increase revenue. To address this issue at the journal level, we analyze the effect of HOA on the number of papers published and the impact factors of HOA journals as compared to a control group of CA journals.

At the paper level, we analyze the effect of HOA by comparing citation rates for HOA papers to CA papers published in the very same journal. A basic hypothesis is that authors may choose to pay for HOA because they expect that by doing so their paper will be more visible, and will end up generating more cites. Given that we have a price for HOA, our data analysis will shed light on the value of the authors' willingness to pay for cites.

To the best of our knowledge, this paper is the first attempt to empirically analyze the relationship between HOA and citation rates in the economics literature. The paper

relates to Eysenbach (2006) and Davis (2009). Eysenbach (2006) finds a significant positive OA citation advantage for articles published in the Proceedings of the National Academy of Sciences between June 2004 and December 2004. Davis (2009) conducts a broader analysis of citation rates under HOA for the medical and biological literature and finds a significantly positive citation effect which is decreasing over the analyzed time period (2004-2007). We transfer this approach to economics and focus on a longer and more recent time period (2006-2011). Furthermore, we extend the research question to incorporate broader quality effects on journals as well as changes in citation rates and impact factors of journals pre and post HOA.

To distinguish between open access journals and HOA journals is crucial for the following reasons. The vast majority of open access journals have offered open access since the first issue. This makes a comparison of citation rates before and after the adoption of the open access business model simply impossible. Additionally, the recently established open access economics journals typically have relatively low academic prestige. In contrast, HOA journals are often well-established economics journals that publish only some articles as open access articles upon payment of a publication fee. This difference allows us to compare citation rates of open access papers and CA papers published in the same journal and thus to control for journal quality.

Our paper also addresses the question of how HOA publishing may affect the quality standards of journals. Our underlying assumption is that the flow of submissions is similar over time and has a similar quality distribution. Under this assumption, a journal would have to accept more low-quality papers for publication after the introduction of HOA in order to increase expected HOA revenue. An overall decrease in the quality standards may harm authors because it would reduce the journals' ability to function as a means for peer recognition. Readers may then incur a higher total reading cost because they would have to read more low-quality papers to gather the same quality-adjusted amount of scientific knowledge.

The remainder of the paper is organized as follows. Section 2 describes the data.

In Section 3, we analyze the impact of HOA publishing at the journal level. Section 4 provides an analysis of the effect of HOA publishing at the article level. Section 5 concludes.

2 The Data

We hand-collected a unique dataset containing information on 7,932 articles published from January 2000 to December 2011 in 15 HOA economics journals including 208 HOA articles. For most journals the starting year for HOA is 2007, although some journals have HOA articles since December 2006. We consider journals that are marked as "ECON" by the "Keele List of 442 Economics Journals" which is published by the Keele University, UK. Applying this general filter, we searched for publishers which have a HOA publishing option for their journals and which mark those articles that are published under these licenses. This search strategy enabled us to identify two major HOA publishers: Springer Science+Business Media and Oxford University Press (OUP). For these publishers, we systematically went through all economics publications to identify the time of the first occurrence of HOA articles. Starting with the first issue that contains at least one HOA article we collected detailed information for 1,329 articles (of which 1,121 are CA articles) published since then in issues with at least one HOA publication. Only original articles are included in our analysis. Editorials, errata, letters etc. are excluded. The dataset contains the name(s) of the author(s), number of authors, the position of institutions in the Shanghai Ranking 2011, title of the paper, journal, date, pages, volume, issue, months since print publication, numbers of HOA and CA articles for each issue, numbers of citations and self-cites. For issues that do not contain HOA articles, we only collected general data, such as the volume and issue number and total number of articles (in total, 3,200 articles since the introduction of HOA) in order to be able to calculate the share of HOA articles on total articles in a given set of journals for a given period of time. The required information is mainly

available on the websites of the publishers. The number of citations and further article metadata were retrieved from the Journal Citation Reports provided by the Thomson Reuters (formerly ISI) Web of Knowledge, while the number of self-cites had to be counted manually by comparing authors of citing papers and cited papers. A citation is regarded as a self-citation if at least one author of the cited article is (co-) authoring the citing paper. We calculate net citations as the difference between total citations and self-cites to correct for self-cites.

In order to identify which articles are author-choice HOA articles, it was necessary to manually review the journals' websites because Thomson Reuters Web of Knowledge does not indicate the access status of an article in its database. For the HOA and the control group journals, we also gathered information on the impact factors and total number of articles and issues published from 2000 to 2011, as given by the Thomson Reuters Web of Knowledge Journal Citation Reports. In addition, we gathered information on the H-index, total citations and average cites per year of authors registered with Research Papers in Economics (RePec) to include further quality measures. Publication fee structures and data about HOA licenses are retrieved from the publishers' webpages.

Table 1 summarizes the 15 HOA journals subject to our analysis, i.e. HOA starting dates and the share of HOA articles on the total number of articles since the first HOA issue. The share of HOA articles ranges from a minimum of 3.02 per cent for Empirical Economics to a maximum of 18.06 per cent for the Journal of Evolutionary Economics. Overall, 6.5 per cent of the articles published in the 15 journals under study since the introduction of HOA are HOA articles.

Table 1: Summary of HOA journals and share of HOA articles

HOA Journals	Dates covered	Articles published	HOA articles	First HOA issue	Total articles since first HOA	HOA share since first HOA	HOA fee	Publisher
Economic Theory	Jan 00 - Oct 11	1,130	23	Mar 07	410	5.61%	$3000 or €2000	Springer
Empirical Economics	Feb 00 - Dec 11	647	8	Jun 08	265	3.02%	$3000 or €2000	Springer
Finance and Stochastics	Jan 00 - Dec 11	312	4	Apr 09	64	6.25%	$3000 or €2000	Springer
International Journal of Game Theory	Feb 00 - Nov 11	450	24	Apr 07	203	11.82%	$3000 or €2000	Springer
Journal of Economic Inequality	Apr 03 - Dec 11	188	6	Dec 06	133	4.51%	$3000 or €2000	Springer
Journal of Economics	Feb 00 - Dec 11	405	6	Jun 07	170	3.53%	$3000 or €2000	Springer
Journal of Evolutionary Economics	Feb 00 - Dec 11	344	28	Feb 07	155	18.06%	$3000 or €2000	Springer
Journal of Financial Econometrics	Apr 03 - Oct 11	198	4	Oct 07	91	4.4%	£1700 or $3000	OUP
Journal of Population Economics	Mar 00 - Dec 11	533	8	Oct 08	190	4.21%	$3000 or €2000	Springer
Metrika	Apr 00 - Nov 11	542	6	Nov 08	149	4.03%	$3000 or €2000	Springer
Networks and Spatial Economics	Mar 01 - Dec 11	246	6	Dec 07	122	4.92%	$3000 or €2000	Springer
Population Research and Policy Review	Feb 00 - Dec 11	383	15	Dec 06	201	7.46%	$3000 or €2000	Springer
Public Choice	Jan 00 - Dec 11	1,261	34	Jun 07	502	6.77%	$3000 or €2000	Springer
Social Choice and Welfare	Jan 00 - Oct 11	723	15	Jun 07	298	5.03%	$3000 or €2000	Springer
The Annals of Regional Science	Mar 00 - Dec 11	570	21	Dec 07	247	8.5%	$3000 or €2000	Springer
TOTAL		7,932	208		3,200	6.5%		

Source: Thomson Reuters Journal Citation Reports and Web of Knowledge; publishers' websites.
Note: The Journal of Economic Inequality and the Journal of Financial Econometrics were first issued in 2003. Networks and Spatial Economics was first issued in 2001.

We also set up a control group of 15 CA journals as reported in Table 2. We chose the CA journals for the control group on the basis of a journal's degree of "relatedness" to the respective HOA journal. We used the maximum relatedness ($Rmax$) of journals measure provided by the Thomson Reuters Journal Citation Report.[1] Econometrica exhibits the highest degree of relatedness with the following four HOA journals: Economic Theory, International Journal of Game Theory, Journal of Financial Econometrics and Social Choice and Welfare. We assigned Econometrica to Economic Theory. For the latter three HOA journals, we assigned the CA journal with the second highest value

[1] This measure takes into account the relationship between cited and citing journals, i.e. in terms of the number of citations from the citing journal to the cited journal or the total number of citations from the citing journal.

of *Rmax* in the control group. That is, we assigned the Review of Economic Studies to the International Journal of Game Theory, the Journal of Finance to the Journal of Financial Econometrics and the American Journal of Political Science to Social Choice and Welfare, respectively.

Table 2: Summary of CA journals control group

Control Group CA Journal	*Rmax*	Dates Covered	Articles Published	2011 Impact Factor
Econometrica (Economic Theory)	2284.80	Jan 2000 - Dec 2011	713	2.976
Journal of Business and Economic Statistics (Empirical Economics)	372.65	Jan 2000 - Dec 2011	464	1.779
Mathematical Finance (Finance and Stochastics)	2043.70	Jan 2000 - Dec 2011	328	1.246
Review of Economic Studies* (International Journal of Game Theory)	300.38	Jan 2000 - Dec 2011	529	2.810
Applied Economics Letters (Journal of Economic Inequality)	15.37	Jan 2000 - Dec 2011	2709	0.226
Journal of Industrial Economics (Journal of Economics)	547.35	Jan 2000 - Dec 2011	325	1.04
Journal of Business Venturing (Journal of Evolutionary Economics)	306.04	Jan 2000 - Dec 2011	423	3.062
Journal of Finance* (Journal of Financial Econometrics)	1091.88	Jan 2000 - Dec 2011	981	4.218
Journal of Labour Economics (Journal of Population Economics)	965.48	Jan 2000 - Dec 2011	335	1.644
Biometrika (Metrika)	412.32	Jan 2000 - Dec 2011	936	1.912
Transportation Science (Networks and Spatial Economics)	1109.82	Jan 2000 - Dec 2011	370	1.507
Population and Development Review (Population Research and Policy Review)	636.63	Jan 2000 - Dec 2011	402	2.224
Journal of Political Economy (Public Choice)	801.22	Jan 2000 - Dec 2011	472	2.902
American Journal of Political Science* (Social Choice and Welfare)	399.44	Jan 2000 - Dec 2011	701	2.756
Economic Geography (The Annals of Regional Science)	345.69	Jan 2000 - Dec 2011	203	3.975

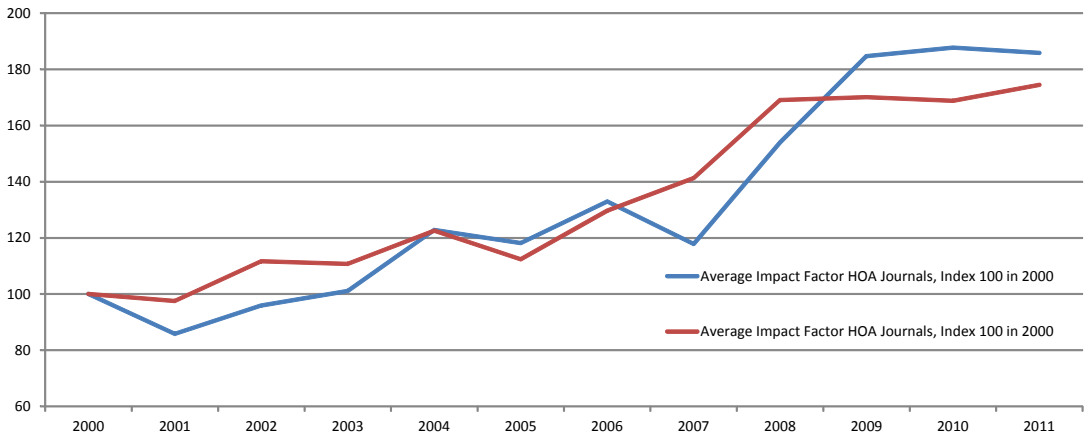
Source: Thomson Reuters Journal Citation Reports and Web of Knowledge.
Note: *Rmax* is a measure for the relatedness of journals. Most closely related HOA journals are given in brackets. Econometrica exhibits the highest degree of relatedness with four HOA journals. In three cases indicated by (*), CA journals with the second highest *Rmax* are used in the control group.

3 Journal Level Analysis

In the following, we analyze the trends in impact factor and total number of articles published from 2000 to 2011 for the HOA journals and the control group of CA journals. Figure 1 illustrates the average impact factor trend for the set of HOA journals and the

control group using the 2000 impact factor as the respective benchmark. It suggests that the average impact factors of the two groups of journals follow similar upward trends over time with an increase in slopes from 2007 on.

Figure 1: Average impact factor over time



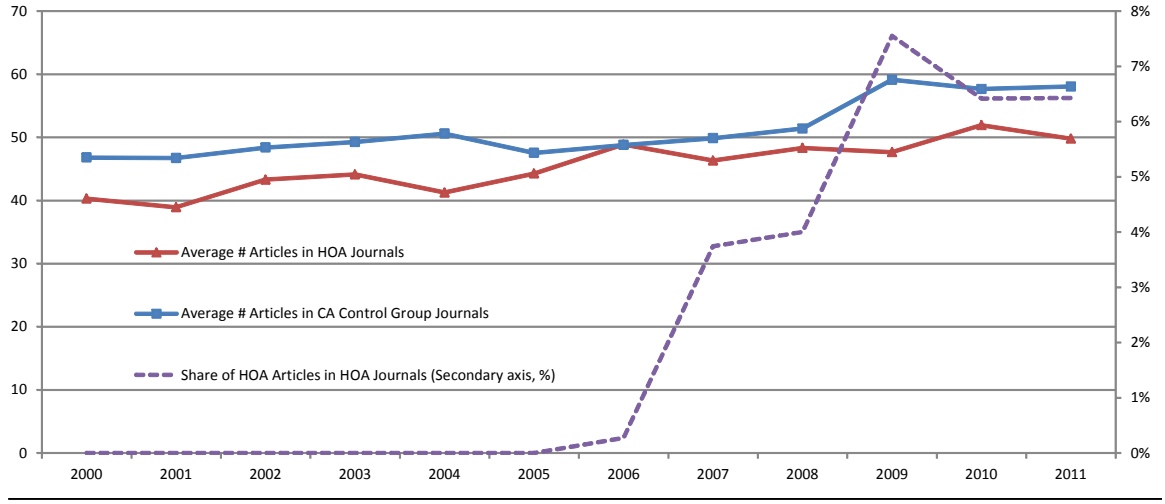
Source: Thomson Reuters Journal Citation Reports, Web of Knowledge.
Note: Impact factor data is not fully available for all years for all HOA journals as some journal were first issued after 2000 (Journal of Economic Inequality, Journal of Financial Econometrics and Networks and Spatial Economics) or were not listed in all Journal Citation Reports from 2000 to 2011 (Empirical Economics, Finance and Stochastics and Journal of Economics). We calculate the yearly averages for the HOA journals accordingly.

Figure 2 illustrates the average number of articles published in HOA journals and in the CA journals of the control group from 2000 to 2011 as well as the share of HOA articles in HOA journals from 2006 to 2011 as given by the dotted line. After a sharp increase in 2006 and 2008 the frequency of HOA articles seems to stabilize between 6% and 7%. Figure 2 suggests that the average number of articles published in HOA journals and CA journals seems to follow a similar trend over time. In addition, there is no evidence for the existence of a structural break for HOA journals in terms of the average number of papers published after the introduction of HOA.[2] It rather seems that

[2]For robustness' sake, we use the average number of articles per quarter and journal and apply the Chow test for HOA-driven structural changes over time. In particular, we test whether the slope coefficients for the time variable have changed after the introduction of HOA. For 12 out of 15 HOA journals, we do not reject the null hypothesis at significance levels between 11,43% and 97,62%. We reject the null at the 5% significance level for Finance and Stochastics and at the 1% level for the Journal of Population Economics and Annals of Regional Science. However, the slope differences for the three latter journals are not due to HOA-driven structural breaks but rather due to special conference issues post HOA that contain unproportionally large numbers of articles.

the average number of articles published in CA journals has increased relatively to that in HOA journals since the first emergence of HOA in 2006. These findings are important for the following reason. In a theoretical model on open access publishing, McCabe and Snyder (2005) suggest that author-pays business models might incentivize journal publishers to publish more papers at the expense of journal quality. However, our empirical evidence of the 15 HOA journals under study does not support the hypothesis of a HOA-driven increase in the number of papers published.

Figure 2: Average number of articles in HOA and CA journals



Source: Thomson Reuters Journal Citation Reports, Web of Knowledge, Jstor and journals' websites.
Note: Only original articles are included in our analysis. Editorials, errata, letters etc. are excluded.

Figures 1 and 2 suggest that the HOA journals under study are of similar characteristics and nature as the most related CA journals. In addition, the introduction of HOA in 2006/2007 did not have a significant impact on the publishing strategies and characteristics of the HOA journals. Due to these fundamental similarities between HOA and CA journals we can therefore focus the following analysis of citation patterns on articles published in HOA journals only.

4 Article Level Analysis

Henceforth, we consider only the HOA dataset as reported in Table 1, i.e. we exclusively take into account articles published in HOA journals. We focus our analysis on the 1,329 articles published in HOA journals since the first occurrence of HOA in December 2006. 1,121 CA papers published in HOA journals serve as control group for 208 HOA papers.

4.1 The HOA Citation Advantage

We analyze citation rates of HOA and CA articles published in the same set of journals. This approach allows us to control for journal quality. Table 3 provides an overview of the variables under study.

Table 3: Summary statistics

	Obs.	Mean	Std. Dev.	Min	Max
Net Citations	1,329	.8585403	2.377005	0	31
Months Since Publication	1,329	41.14898	15.85525	15	76
Months Since Publication, squared	1,329	1944.439	1389.265	225	5776
Number of References	1,329	32.9842	20.24538	1	200
Number of Authors	1,329	1.933032	.8603725	1	4
HOA Article	1,329	.1565087	.3634737	0	1
HOA Article Subject to HOA Pilot Agreement	1,329	.1324304	.3390857	0	1
HOA Article Not Subject to HOA Pilot Agreement	1,329	.0240783	.1533499	0	1
Lead Article	1329	.1068473	.3090352	0	1
Number of Pages	1,329	19.26787	7.330445	2	52
H-Index of Best Author	806	7.849876	6.37998	1	39
Total Number of Citations of Best Author	806	382.4491	667.808	1	7,645
Average Number of Citations of Best Author	717	21.98755	28.49397	1.09	231.67
Rank of Best Institution	957	142.8892	117.8079	1	401

Source: Thomson Reuters Journal Citation Reports; Web of Knowledge; RePec, websites of 15 HOA journals.
Note: One paper is one observation.

Our dependent variable, $NetCit_{ijt}$, is given by the number of net cites which is the number of total cites minus the number of self cites of article i in journal j in year t. As of April 2013, $MonthPub_{ijt}$ ($MonthPubSq_{ijt}$) indicates the (squared) number of months since the paper under study was published in print. $NumRef_{ijt}$($NumAut_{ijt}$) is given by the number of references (authors) of an article. We include $NumAut_{ijt}$ as co-authors might have an incentive to split the HOA fee. HOA_{ijt} is a binary variable that indicates whether the article under study was published under the HOA format (1) or not (0). $LeadArt_{ijt}$ is a binary variable indicating whether an article was the lead article in the issue under study (1) or not (0). Intuitively, articles positioned at the beginning of an issue may be more visible and thus read and cited more often than articles at the end of an issue. $NumPage_{ijt}$ indicates the number of pages of an article. We include journal fixed effects. $BestH_{ijt}$ is given by the RePec H-index of a single author or the highest H-index of all authors in the case of multiple authors, respectively. $BestCit_{ijt}$ ($BestAvCit_{ijt}$) is given by the total number of citations (average cites per year) retrieved from RePec of a single author or the number of citations (average cites per year) of all authors in the case of multiple authors, respectively. RePec data for at least one author is available for about 60 per cent of the articles under study. $InstRank_{ijt}$ indicates the position in the Shanghai Ranking 2011 of the best institution authors are affiliated with. Rank data is available for 957 out of 1,329 articles as only the top 500 universities are listed in the Shanghai Ranking.[3] If anything, we therefore overstate the quality of institutions in our data set. $HOAPilot_{ijt}$ is a binary variable indicating whether a HOA article was subject to a HOA pilot agreement (1) or not (0).[4] The HOA pilot agreements enable authors from participating institutions to choose the

[3]The last 100 universities are ranked as "401-500". For these universities, we take the best possible rank (401) into consideration.

[4]Data about HOA pilot agreements between HOA publishers and academic institutions are collected from webpages of participating pilot partners, i.e. the Dutch Association of University Librabries and the National Library, Max Planck Society, University of Goettingen, University of California and University of Hong-Kong, as well as from the publishers' webpages.

HOA option without having to pay the fee by themselves.[5] During the pilot-period the fees were included in the subscription fees that the institutions paid to the publishers. $NoPilotHOA_{ijt} \equiv HOA_{ijt} - HOAPilot_{ijt}$ is a binary variable indicating those HOA articles that were not subject to a HOA pilot agreement. In our dataset, 176 HOA articles were submitted under HOA pilot-agreements – 32 were submitted without such programs. We include the (possibly overstated) quality of institutions to address the concern that $HOAPilot_{ijt}$ might be picking up the quality of institutions.

Figure 3 reports the histogram of $NetCit_{ijt}$. One purpose of this figure is to indicate how many articles in our data set generated zero net cites over the period of time under study. It is, however, important to note that many of the zeros might turn into positive numbers given sufficient time since publication. As $NetCit_{ijt}$ is neither normally distributed nor normally distributed with a truncation we use the unconditional fixed-effects negative binomial regression model for the over-dispersed count data. To analyze the question whether choosing HOA does have an effect on citations we have to control for all other effects that might have an impact on citation rates such as the months since publication, the number of references, the number of authors, article

[5] To better understand the factors influencing authors' decisions to choose the HOA option when HOA pilot agreements were in place, we gathered the email addresses of 208 HOA contact authors and conducted a short email survey on the incentives to choose HOA and HOA-related budgets. In particular, we asked the following three HOA-related questions: *1) The reason for your use of this option? In particular, does your employer require or encourage you to choose the HOA option if available?, 2) What was the source of funding for payment of the fee for the HOA publication, and if it was your employer, would you have been willing to pay from your own research funds had your employer not paid?, 3) In the case of your most recent co-authored HOA publication, did you split the HOA fee between all authors?* From the 16 authors that replied, ten authors answered that their employer had paid the HOA fee. The HOA fee was typically not split between authors. One author from the University of Amsterdam said that he "was neutral" with respect to the HOA option. One author from the University of Brussels replied that before he signed the copyright form he "didn't even know that the option existed". One author from the University of California stated: "I didn't see any costs to the option". A federal employee of the USA stated that "all publications by a federal employee needs to be open access". He also stated: "I wouldn't have paid (of course depending on the amount, I'd have paid up to $10, maybe)." One author from Florida State University stated that "the open-access publication was not required, but it was a free option for us as authors since one of us had European support".

quality etc. More specifically, we analyze:

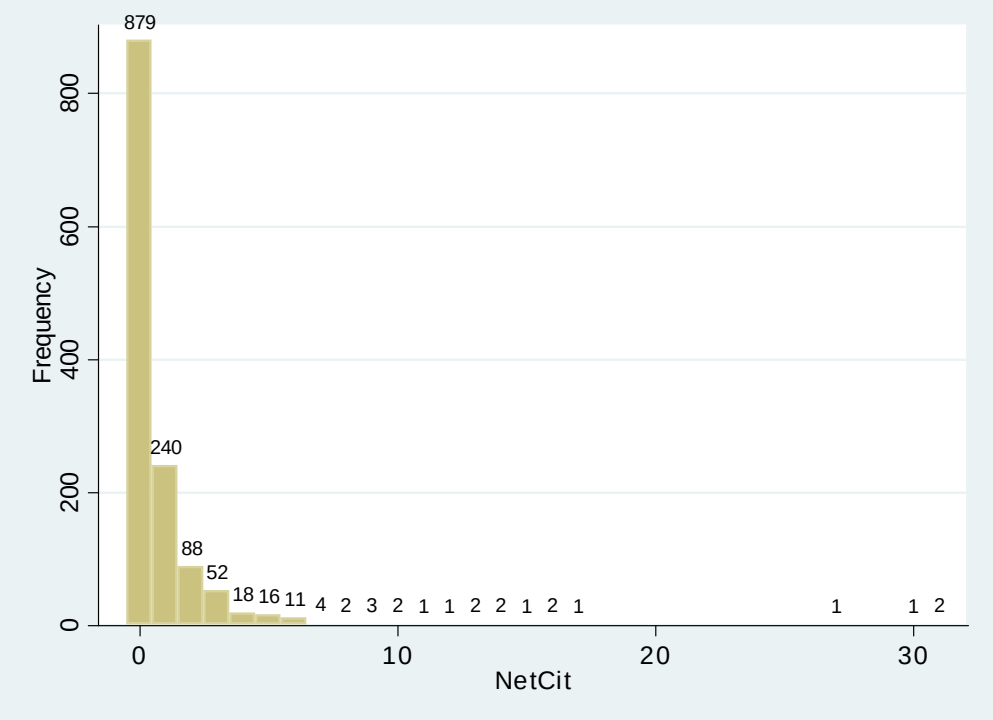
$$NetCit_{ijt} = Const + \beta_1 MonthPub_{ijt} + \beta_2 MonthPubSq_{ijt} + \beta_3 NumRef_{ijt} \quad (1)$$

$$+\beta_4 HOA_{ijt} + \beta_5 NumAut_{ijt} + \beta_6 LeadArt_{ijt} + \beta_7 NumPage_{ijt}$$

$$+\beta_8 BestH_{ijt} + \beta_9 BestCit_{ijt} + \beta_{10} BestAvCit_{ijt} + u. \quad (2)$$

Following Allison and Waterman (2002), we use the unconditional fixed-effects negative binomial regression model to estimate the relation between net cites and the set of variables presented above.

Figure 3: Histogram of net citations



The regression results are reported in Table 6. In specification (1), we refrain from including RePec quality measures and institution quality. In specification (2), we include $BestH_{ijt}$, in (3) $BestCit_{ijt}$ and in (4) $BestAvCit_{ijt}$ to control for article quality. In specification (5), we include all three quality measures. In specification (6), we include the institution quality measure. This approach is crucial for the following reasons.

Table 6: Regression results

Dependent variable:	Net Citations											
Regression model:	Unconditional fixed-effects negative binomial regression											
Specification:	(1)		(2)		(3)		(4)		(5)		(6)	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Months Since Publication	0.204***	0.026	0.190***	0.033	0.187***	0.033	0.198***	0.036	0.202***	0.036	0.208***	0.030
Month Since Publication, squared	-0.001***	0.000	-0.001***	0.000	-0.001***	0.000	-0.001***	0.000	-0.001***	0.000	-0.001***	0.000
Number of References	0.015***	0.003	0.020***	0.004	0.020***	0.004	0.020***	0.004	0.020***	0.004	0.016***	0.003
HOA Article	0.325**	0.127	0.387**	0.153	0.399***	0.153	0.378**	0.164	0.371**	0.164	0.286**	0.145
Number of Authors	0.192***	0.057	0.168**	0.080	0.181**	0.079	0.159*	0.085	0.154*	0.085	0.229***	0.067
Lead Article	0.253*	0.148	0.218	0.191	0.226	0.191	0.219	0.205	0.234	0.205	0.282	0.179
Number of Pages	-0.001	0.008	0.008	0.010	0.008	0.010	0.007	0.011	0.007	0.011	-0.012	0.009
H-Index of Best Author			0.010	0.010					0.014	0.028		
Total Number of Citations of Best Author					0.000	0.000			0.000	0.000		
Average Number of Citations of Best Author							0.001	0.002	0.006	0.008		
Rank of Best Institution											0.000	0.000
Constant	-8.089***	0.673	-8.186***	0.894	-8.074***	0.882	-8.265***	0.957	-8.414***	0.971	-8.082***	0.801
ln(a)	-0.025	0.107	-0.114	0.148	-0.117	0.148	-0.041	0.152	-0.051	0.152	-0.067	0.128
a	0.976	0.104	0.892	0.132	0.890	0.132	0.959	0.146	0.950	0.145	0.935	0.119
LR chi2(21)	582.140											
LR chi2(22)			349.320		348.540		306.860				425.620	
LR chi2(24)									308.110			
Prob > chi2	0.000		0.000		0.000		0.000		0.000		0.000	
Pseudo R2	0.184		0.186		0.185		0.180		0.180		0.186	
Observations	1,329		806		806		717		717		957	

Source: Thomson Reuters Journal Citation Reports; Web of Knowledge; RePec, websites of 15 HOA journals.
Note: One article is one observation. We also include journal journal fixed-effects. We drop the journal dummy for Public Choice which corresponds to the most common category with 294 articles in the data set. We include $BestH_{ijt}$ in specification (2), $BestCit_{ijt}$ in (3) and $BestAvCit_{ijt}$ in (4) and all quality measures together in (5). In (6), we control for institution quality. The log-transformed [untransformed] over-dispersion parameter ln(a) [a] is estimated and displayed. The Wald chi-square statistic with 21, 22 and 24 degrees of freedom is given by LR chi2(21) for specification (1), by LR chi2(22) for specification (2), (3), (4) and (6) and by LR chi2(24) for specification (5), respectively, followed by the p-value for the chi-square. The significance levels are reported as follows: *** p<0.01, ** p<0.05, * p<0.1.

Davis (2009) and Gaulé and Maystre (2011) suggest that it is not the type of access that explains the HOA citation advantage, but rather it is self-selection. In particular, Davis (2009, p. 6) suggests that "more citable articles have a higher probability of being made freely accessible". In addition, McCabe and Snyder (2011) suggest that the citation advantage associated with online publishing found in previous studies such as Curti et al. (2001) and Lawrence (2001) "were an artifact of their failure to control for article quality".

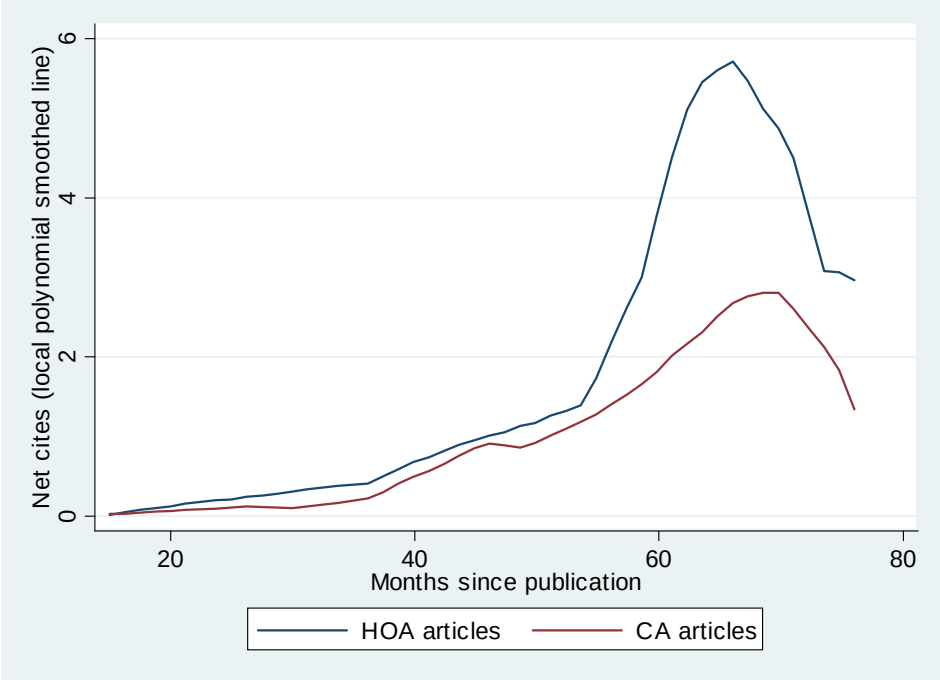
We find four main significant explanatory variables for $NetCit_{ijt}$ in the four specifications: $MonthPub_{ijt}$ ($MonthPubSq_{ijt}$), $NumRef_{ijt}$ and HOA_{ijt}. Intuitively, the longer a paper is available in print the more cites it generates. The coefficient is positive and significant at the 1% level throughout all specifications. We test whether the coefficients for $MonthPub_{ijt}$ and $MonthPubSq_{ijt}$ are jointly different from 0. The p-value is 0.0000 in all specifications. We reject the null hypothesis and conclude that both variables have a significant effect on net citations. Second, articles that cite more other papers generate higher net cites. The coefficient is positive and significant at the 1% level throughout all specifications.[6]

Third, HOA articles generate more cites than CA articles in the same set of journals. The coefficient is positive throughout all specifications. It is significant at the 5%-level in specification (1) with a p-value of 0.01 and at the 5% level in specification (2) with a p-value of 0.011. In addition, it is significant at the 1% level in specification (3) with a p-value of 0.009, at the 5% level in specification (4) with a p-value of 0.021 and at the 5% level in specification (5) with a p-value of 0.024. It is also significant at the 5% level in specification (6) with a p-value of 0.049. Doing a two-sample t-test with equal variances with 208 (1,121) observations for $NetCit_{ijt}$ if $HOA_{ijt} = 1$ ($HOA_{ijt} = 0$) we conclude that HOA_{ijt} is statistically significant at the 1 % level with a p-value of .0009. Controlling for article quality, our results suggest that HOA generates additional net

[6]This interesting result could be due to a number of reasons. Perhaps by having more references more Google hits are achieved in a search, leading more future authors to notice the paper. Also possible is that authors are reciprocally citing each other in an attempt to artificially increase cites.

cites ranging from a minimum of about 0.3 per article in specification (6) to a maximum of about 0.4 per article in specification (3) as compared to CA articles published in the same group of journals. Hence, a $\$3,000$ HOA fee is buying HOA authors, on average, less than half a net citation over a period of up to 76 months since publication. However, the effect of HOA on net cites changes over the lifetime of an article as Figure 4 illustrates. The HOA citation advantage appears to be relatively low for articles between 20 months since publication (when the first cites occur) and about 54 months since publication. It increases over the next year and then decreases again.

Figure 4: Change of net cites over time



The coefficient for the position of an article in an issue is positive and significant at the 10 % level in specification (1). Once we control for article quality, however, this positive effect turns out to be insignificant. Interestingly, the coefficient for $NumAut_{ijt}$ is positive and significant at the 1% level in specification (1) and (6). However, the coefficient for $NumAut_{ijt}$ and the significance levels decrease when we control for article quality. We may interpret this result in the sense that, in contrast to the mere number

of authors, the quality-adjusted number of authors has a less significant effect on cites. In addition, the probability that at least one author of a paper is registered with RepEc ceteris paribus increases in the number of authors. Hence, in specifications (2), (3), (4) and (5) articles written by a relatively large number of authors might be overrepresented as compared to specification (1). For the 523 papers for which no author is registered with RePec the average number of authors is 1.78. For the entire sample (1,329 papers) the average number of authors is 1.93. Surprisingly, the quality of a single author (or the best author in case of coauthored papers) in terms of $BestH_{ijt}$, $BestCit_{ijt}$ or $BestAvCit_{ijt}$ does not have a significant positive effect on net cites in specifications (2), (3) and (4), respectively. In specification (5), we test whether the coefficients for $BestH_{ijt}$, $BestCit_{ijt}$ and $BestAvCit_{ijt}$ are jointly different from 0. The p-value is 0.6576. We cannot reject the null hypothesis and conclude that the three variables are jointly insignificant. $NumPage_{ijt}$ is insignificant at the 10 % level throughout all specifications. Finally, $InstRank_{ijt}$ is insignificant in specification (6) with a high p-value of 0.866. Note that it is important to control for institution quality as one may argue that $HOAPilot_{ijt}$ might be picking up the quality of institutions author(s) are at. Taking this aspect into consideration and controlling for institution quality, however, the HOA citation advantage remains significant.

5 Conclusions

Our analysis of a set of 15 HOA journals and a control group of 15 closely related CA journals suggests that HOA and CA journals are similar in terms of impact factor trends and the number of published articles from 2000 to 2011. We also find that there is no obvious change in publishing strategies of formerly CA journals after the introduction of HOA in 2006/2007. In contrast to prior theoretic literature that points to a quality-eroding effect of author-pays publication formats due to an increase in published papers, we do not find empirical support for the hypothesis of a HOA-driven

increase in the number of papers published. We use a unique, hand-collected dataset containing information on 1,329 articles published in HOA journals to analyze the effect of HOA on cites at the article level. Controlling for journal quality, article quality as well as institution quality, our analysis suggests that the HOA publication format has a significant positive effect on net cites. However, the HOA citation advantage is rather small: on average, between 0.33 and 0.4 additional net cites per article over a period of up to 76 months since publication. Taking into account the typical HOA fee of $3,000, our analysis suggests that an additional average net cite over a period of up to 76 months "costs" between $7,500 and $9,000. It is, however, important to note that the HOA citation advantage changes over the lifetime of an article and is more pronounced for articles older than about five years.

References

[1] Allison, P.D., Waterman, R.P. (2002), Fixed-effects Negative Binomial Regression Models, in: R.M. Stolzenberg (ed.), Sociological Methodology. Oxford: Basil Blackwell.

[2] Björk, B.-C. (2012), The Hybrid Model for Open Access Publication of Scholarly Articles: A Failed Experiment?, Journal of the American Society for Information Science and Technology, in press.

[3] Curti, M., Pistotti, V., Gabutti, G., Klersy, C. (2001), Impact Factor and Electronic Versions of Biomedical Scientific Journals, Haematologica 86, 1015-20.

[4] 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.

[5] Diamond, A. (1986), What is a Citation Worth?, Journal of Human Resources 21, 200-215.

[6] Eysenbach, G. (2006), Citation Advantage of Open Access Articles, PLoS Biology 4, e157.

[7] Gaulé, P., Maystre, N. (2011), Getting Cited: Does Open Access Help?, Research Policy 40, 1332-1338.

[8] Lawrence, S. (2001), Free Online Availability substantially Increases a Paper's Impact, Nature 411, 521.

[9] McCabe, M. J. and C. M. Snyder (2005), Open Access and Academic Journal Quality, American Economic Review Papers and Proceedings, 95(2); 453-58.

[10] McCabe, M.J., Snyder, C.M. (2011), Did Online Access to Journals Change the Economics Literature?, Available at SSRN: http://ssrn.com/abstract=1746243.

[11] Walker, T. (2004), Open Access by the Article: An Idea Whose Time Has Come?, Nature Web Focus Article 13, April 15.