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

Müller-Langer, Frank and Watt, Richard

Max Planck Institute for Innovation and Competition

31 January 2014

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

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

January 31, 2014

Abstract

We study the hybrid open access (HOA) citation effect. Under HOA pilot agreements, HOA is assigned for all articles of eligible authors. We use unique data on 208 (1,121) HOA (closed access) economics articles. We control for the quality of journals, articles and institutions and citations to RePec pre-prints. Performing Poisson quasi-maximum likelihood regressions, HOA turns out to be a significant predictor of citations with marginal effects ranging between 22% and 26%. However, once we additionally control for institution quality and citations to RePEc pre-prints, the marginal HOA citation advantage turns out to insignificant and drops to 0.4%.

Keywords: Hybrid open access, diffusion processes, citation effects
JEL codes: L17, O33, A11

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 (OA) 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 OA option. If the author chooses HOA, she has to pay the HOA fee, which is typically 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.

In this study, we benefit from the recent introduction of HOA pilot agreements between commercial publishers, i.e., Springer and Oxford University Press (OUP), and UC California, the Universities of Hong-Kong and Goettingen and all Dutch universities as well as all Max Planck Institutes. Under these agreements, the HOA status of articles is typically exogenously assigned to all articles of authors affiliated with HOA pilot institutions. This unique set-up helps us to avoid the author-driven selection bias brought forward in recent literature on the (H)OA citation effect in science (Davis, 2009; Gaulé and Maystre, 2011). In addition, we control for institution quality to address the concern that HOA pilot articles might be picking up the quality of HOA pilot partner institutions.

HOA has recently been pushed by major scientific publishers as a possible gradual transition path between the traditional closed access (CA) publishing model and OA to scientific publications. The number of HOA journals across disciplines has increased from 2,017 in October 2009 to 4,381 in February 2012 (Björk, 2012). However, McCabe et al. (2013, p.14) find that a commercial monopoly journal would *not* find HOA more attractive than traditional pricing and explain the emergence of HOA as a response to "the threat or actual imposition of an OA mandate by governments, foundations and other institutions." For instance, researchers funded by the EU Research Framework Programmes are required to publish their results in OA journals, following similar OA mandates by the United States, the United Kingdom and several funding agencies such as the Wellcome Trust or the U.S. National Institutes of Health (The Economist, 2012; European Commission, 2012; McCabe and Snyder, forthcoming (a)).

We use a unique, hand-collected dataset containing information on 1,329 articles

published from December 2006 to December 2011 in 15 HOA economics journals to analyze the citation effect of HOA. 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 online publication, number of references and authors, the availability of OA pre-prints and the position of an article in an issue drive citation rates. Our work is inspired by McCabe and Snyder (forthcoming (b)) 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 citation effects as large as several hundred percent such as Harnad and Brody (2004), Lawrence (2001) and Walker (2004) for their failure to control for article quality. In particular, McCabe and Snyder (forthcoming (b), p. 3) 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." Controlling for unobservable article quality by including fixed-effects for journal volumes, McCabe and Snyder (forthcoming (b)) provide evidence that the citation effect of online-access drops to zero. In addition, 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". The unique set-up of HOA pilot agreements helps us to address these concerns. Our approach also 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 (forthcoming (b))'s critique and provide an analysis that is distinct from existing works on OA related citation effects such as Curti et al. (2001), Davis (2009), Harnad and Brody (2004), and Lawrence (2001).

To the best of our knowledge, this paper is the first attempt to empirically analyze the relationship between HOA and citations in economics. The (H)OA citation literature has so far focused on science journals as, in contrast to economics, OA pre-prints are uncommon in science. We analyze the impact of HOA on citations to published articles in economics while controlling for the availability of OA pre-prints on Social Science Research Network (SSRN) and RePec as well as controlling for the citations to these pre-prints. 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). Our paper differs from this strand of literature in several aspects. First, we consider HOA pilot agreements. Second, we study economics journals. Third, we focus on a longer time period (2006-2011).

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

The remainder of the paper is organized as follows. Section 2 describes the data. In Section 3, we provide an analysis of the effect of the HOA-Pilot status on net citations.

Section 4 concludes.

2 Data

We hand-collected a unique dataset containing information on 1,329 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 online 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-citations to correct for self-citations.

In order to identify which articles are 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. 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. Information about HOA fees and licenses are retrieved from the publishers' webpages. We gathered information on the availability of OA pre-prints from the SSRN and RePec websites of the authors under study. We used the online publication date as indicated in the published articles as the benchmark to calculate the months of availability of OA pre-prints.

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.

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

Source: Thomson Reuters Journal Citation Reports and Web of Knowledge; publishers' websites.
Note: The HOA publication fee for all journals is either $3,000 or €2,000. 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.

Table 1: Summary of HOA journals and articles

3 Analysis of the Effect of HOA on Net Citations

We focus our analysis on the 1,329 articles published in the HOA journals in our data set since the first occurrence of a HOA article in December of 2006. The 1,121 CA papers published in the same HOA journals over the same time period serve as a control group for the 208 HOA papers. We identify HOA pilot agreements between Springer (14 out of 15 HOA journals in our sample) and the Dutch Consortium of University Libraries (UKB), Max Planck Society (MPG), UC California, University of Goettingen and Hong-Kong University. Under these agreements, all articles that are accepted for publication by Springer during the pilot period written by authors who are affiliated with one of

the HOA pilot partner institutions are made openly available immediately.[1] Appendix A provides an overview of the HOA pilot agreements and periods under study. During the pilot-period the fees were included in the subscription fees that the institutions paid to the publishers. As the HOA status is typically assigned automatically under HOA pilot agreements and authors do not have to pay the HOA option from their own research budgets, we argue that the existence of an author-driven selection bias can be ruled out when we run the regressions only with the HOA pilot articles. However, we include the quality of institutions to address the concern that the coefficient for HOA pilot articles might be picking up the quality of HOA pilot institutions.

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 2 provides an overview of the variables under study.

[1] To better understand the functioning of HOA pilot agreements, we gathered the email addresses of all 208 HOA contact authors and conducted a short email survey on 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? 16 authors 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 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)."

	Obs.	Mean	Std. Dev.	Min	Max
Net Citations	1,329	0.8585403	2.377005	0	31
Months Since Online Publication	1,329	37.16704	14.23689	1	74
Months Since Online Publication, squared	1,329	1583.926	1119.211	1	5,476
Number of References	1,329	32.9842	20.24538	1	200
HOA Article	1,329	0.1565087		0	1
HOA Article Subject to HOA Pilot Agreement	1,329	0.1309255		0	1
Number of Authors	1,329	1.930775	0.8580024	1	4
Lead Article	1,329	0.1068473		0	1
Number of Pages	1,329	19.26787	7.330445	2	52
Availability of Open Access Pre-Prints	1,329	0.3619263		0	1
SSRN Open Access Pre-Prints	1,329	0.1866065		0	1
RePec Open Access Pre Prints	1,329	0.3054929		0	1
H-Index of Best Author	806	7.849876	6.37998	1	39
Total Number of Cites of Best Author	806	382.4491	667.808	1	7,645
Average Number of Cites of Best Author	717	22.01674	28.49933	1	232
Rank of Best Institution	924	185.8506	141.5706	1	450
Months of Availability of RePec Open Access Pre-Prints	406	32.20936	19.76374	1	117
Months of Availability of RePec Open Access Pre-Prints, squared	406	1,427.086	1764.66	1	13,689
Total Citations to RePec Open Access Pre-Prints	406	3.768473	7.726535	0	109
UKB HOA Pilot	1,329	0.0948081		0	1
UC California HOA Pilot	1,329	0.020316		0	1
MPG HOA Pilot	1,329	0.0158014		0	1
Goettingen HOA Pilot	1,329	0.0037622		0	1
Hong-Kong HOA Pilot	1,329	0.0007524		0	1

Source: Thomson Reuters Journal Citation Reports; Web of Knowledge; RePec, websites of 15 HOA journals. Information about HOA pilot agreements between HOA publishers and academic institutions, e.g., agreement periods, are collected from webpages of participating pilot partners, i.e. the Dutch Association of University Librabries and the National Library (UKB), University of California, Max Planck Society (MPG), University of Goettingen and University of Hong-Kong.
Note: One paper is one observation.

Table 2: Summary statistics

Our dependent variable, $NetCit_{ijt}$, is given by the total number of net cites (as of March 2012) which is the total number of cites minus the number of self cites of article i published in journal j in year t. $MonthPub_{ijt}$ ($MonthPubSq_{ijt}$) indicates the (squared) number of months since the paper under study was published online. $NumRef_{ijt}(NumAut_{ijt})$ is given by the number of references (authors) of an article. HOA_{ijt} is a binary variable that indicates whether the article under study was published under the HOA format (1) and thus available online without log-in as subscriber or not (0).[2] $HOAPilot_{ijt}$ is a binary variable that indicates whether the article under study

[2]We identified HOA articles by manually going through the journal websites of HOA publishers without log-in as subscriber. In this case, HOA publishers clearly indicate the HOA status of articles on the journal websites. In addition, HOA and CA articles are characterized by different copyright notices.

was subject to a HOA pilot agreement between the publisher and the institution the authors under study are affiliated with.[3] We include $NumAut_{ijt}$ as co-authors might have an incentive to split the HOA fee. $LeadArt_{ijt}$ is a binary variable indicating whether an article was the lead article in the issue under study. 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. $OAPrePrint_{ijt}$ is a binary variable that indicates whether an OA pre-print of the paper under study is available on either SSRN ($SSRNOAPrePrint_{ijt}$) and/or RePec ($RePEcOAPrePrint_{ijt}$).[4] We include journal dummy variables. $BestH_{ijt}$ is given by the RePec H-index[5] of a single author or the highest individual 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 in the case of single-authored articles or the number of citations (average citations per year) of the best co-author in the case of multiple authors, respectively. RePec data for at least one author is available for 806 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 924 out of 1,329 articles as only the top 500 universities are listed in the Shanghai Ranking. If anything we therefore overstate the quality of institutions in our data set. $MonthsRePecPrePrint_{ijt}$ ($MonthsRePecPrePrintSq_{ijt}$) indicates the (squared) number of months since an OA pre-print was available prior to the publication of the final version of the article under study. The total number of cites to RePec OA pre-prints is given by $RePEcCitPrePrints_{ijt}$. In the case of multiple OA pre-prints (74 out of a total of 406 cases), we consider the sum of citations to all

[3] Four of the 34 HOA articles that were not subject to a HOA pilot agreement were published in OUP's Journal of Financial Econometrics. In contrast to Springer, OUP does not have HOA pilot agreements.

[4] We manually gathered information on OA pre-prints from the author websites on SSRN and IDEAS/RePEc. We define an article as OA pre-print if a working paper or discussion paper was made available on SSRN or IDEAS/RePEc prior to the publication of the final version of the article in the journal under study. Note that this definition excludes published journal articles that are often also made available on SSRN or RePec.

[5] H is the number of articles with at least H citations.

available OA pre-prints. HOA pilot articles written by UKB-researchers are the most common case with 126 observations. Appendix B reports the histogram of $NetCit_{ijt}$. It indicates 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 Poisson quasi-maximum likelihood (PQML) as well as the 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 online publication, the number of references, the number of authors, article quality, availability of OA pre-prints etc. More specifically, we estimate the following specification:

$$\begin{aligned} NetCit_{ijt} &= Const + \beta_1 MonthPub_{ijt} + \beta_2 MonthPubSq_{ijt} + \beta_3 NumRef_{ijt} \\ &\quad + \beta_4 HOAPilot_{ijt} + \beta_5 NumAut_{ijt} + \beta_6 LeadArt_{ijt} + \beta_7 NumPage_{ijt} \\ &\quad + \beta_8 OAPrePrint_{ijt} + \beta_9 SSRNOAPrePrint_{ijt} + \beta_{10} RePecOAPrePrint_{ijt} \\ &\quad + \beta_{11} BestH_{ijt} + \beta_{12} BestCit_{ijt} + \beta_{13} BestAvCit_{ijt} + \beta_{14} InstRank_{ijt} \\ &\quad + \beta_{15} MonthsRePecPrePrint_{ijt} + \beta_{16} MonthsRePecPrePrintSq_{ijt} \\ &\quad + \beta_{17} RePEcCitPrePrints_{ijt} + u. \end{aligned}$$

3.1 Poisson Quasi-Maximum Likelihood Regression Model

Following Santos Silva and Tenreyro (2006, 2010), we perform a PQML regression to predict net citation outcomes. It checks for the existence of estimates and drops regressors that may cause the non-existence of estimates.[6] We cluster robust standard errors around journals. Henceforth, we drop 34 observations of HOA articles that were not subject to a HOA pilot agreement. Otherwise, our analysis would suffer from

[6] We use Santos Silva and Tenreyro (2011)'s implementation of the QML estimator in Stata. See also Wooldridge (1999) on PQML estimation for panel data.

the self-selection bias brought forward by Davis (2009) and Gaulé and Maystre (2011) that authors are more likely to select more citable papers as HOA.[7] In this respect, HOA pilot articles and HOA articles that are not subject to a HOA pilot agreement are to be treated as fundamentally different.[8] The regression results are reported in Table 3. 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 together with $BestH_{ijt}$. In specification (7) we additionally control for the months of availability of OA pre-prints and citations to OA pre-prints.

Our main findings are as follows. First, $MonthPub_{ijt}$ is a strong predictor for $NetCit_{ijt}$. The coefficient is positive and significant at the 1% level in specifications (1) to (6) and at the 5% level in specification (7) with a p-value of 0.023. Intuitively, the longer a paper is available online the more cites it generates. 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. Table 4 reports the marginal effects at the means of the PQML regression. The statistically significant positive marginal effect of $MonthPub_{ijt}$ on $NetCit_{ijt}$ ranges between 5% and 7% across specifications. Second, articles that cite more other papers generate higher net cites. The coefficient is positive and significant at the 1% level throughout all specifications. This interesting result could be due to a number of reasons. Perhaps by having more references more Google (Scholar) 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. The statistically significant positive marginal effect of

[7] Clearly, it would also not be admissible to include HOA articles that are not subject to HOA pilot agreements as CA articles in the regressions as this would bias the CA citation effect.

[8] We ran the regressions with and without the 34 currently omitted papers. The results are practically identical which suggests that the papers are not really all that different in terms of selection bias. This contrasts with Davis (2009) and Gaulé and Maystre (2011).

$NumRef_{ijt}$ on $NetCit_{ijt}$ ranges between 0.4% and 0.7% across specifications.

Dependent Variable:	Net Citations						
Regression Model:	PQML Regression						
Specification:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Months Since Online Publication	0.218***	0.207***	0.206***	0.217***	0.216***	0.196***	0.144**
	(0.0460)	(0.0605)	(0.0610)	(0.0614)	(0.0622)	(0.0698)	(0.0636)
Months Since Online Publication, sqrd	-0.00148***	-0.00138**	-0.00136**	-0.00148**	-0.00147**	-0.00118	-0.000912
	(0.000483)	(0.000649)	(0.000659)	(0.000652)	(0.000659)	(0.000745)	(0.000697)
Number of References	0.0125***	0.0191***	0.0193***	0.0198***	0.0201***	0.0234***	0.0118***
	(0.00376)	(0.00401)	(0.00433)	(0.00428)	(0.00431)	(0.00480)	(0.00369)
HOA Article Subject to HOA Pilot Agreement	0.562**	0.618**	0.619**	0.659*	0.643**	0.472	0.0107
	(0.254)	(0.307)	(0.307)	(0.338)	(0.299)	(0.350)	(0.147)
Number of Authors	0.268**	0.279**	0.289**	0.260**	0.267**	0.355***	-0.0578
	(0.135)	(0.124)	(0.141)	(0.122)	(0.122)	(0.135)	(0.134)
Lead Article	0.340	0.249	0.251	0.239	0.233	0.134	0.506**
	(0.253)	(0.216)	(0.217)	(0.230)	(0.230)	(0.231)	(0.224)
Number of Pages	0.00189	0.00764	0.00711	0.00947	0.00870	0.00479	0.0224
	(0.00877)	(0.0164)	(0.0168)	(0.0172)	(0.0166)	(0.0187)	(0.0152)
OA Pre-Print	-0.333	-0.325	-0.335	-0.551	-0.638	-0.291	[dropped]
	(0.347)	(0.354)	(0.362)	(0.346)	(0.414)	(0.389)	
SSRN OA Pre-Print	-0.253	-0.260	-0.237	-0.242	-0.239	-0.186	-0.142
	(0.273)	(0.213)	(0.215)	(0.227)	(0.232)	(0.259)	(0.209)
RePec OA Pre-Print	0.333	0.191	0.196	0.343	0.405*	-0.000294	[dropped]
	(0.208)	(0.269)	(0.267)	(0.225)	(0.241)	(0.312)	
H-Index of Best Author		-0.00184			-0.0106	-0.0155	0.0238
		(0.00962)			(0.0476)	(0.0141)	(0.0169)
Overall Citations of Best Author			-0.0000674		-0.000437		
			(0.000175)		(0.000983)		
Average Citations of Best Author				-0.000293	0.0106		
				(0.00240)	(0.0107)		
Rank of Best Institution						-0.000110	-0.00007
						(0.000365)	(0.00098)
Months of Availability of RePec OA Pre-Prints							-0.0229
							(0.0143)
Months of Availability of RePec OA Pre-Prints, sqrd							0.000007
							(0.000173)
Citations to RePec OA Pre-Prints							0.0141***
							(0.00513)
Constant	-7.762***	-7.864***	-7.853***	-8.057***	-8.050***	-7.847***	-4.940***
	(1.234)	(1.458)	(1.445)	(1.496)	(1.420)	(1.546)	(1.147)
Observations	1,295	783	783	695	695	565	246
R-squared	0.304	0.347	0.353	0.375	0.404	0.405	0.625

Source: Thomson Reuters Journal Citation Reports; Web of Knowledge; RePec, websites of 15 HOA journals.
Note: Following Santos Silva and Tenreyro (2006, 2010), we use Santos Silva and Tenreyro (2011)'s implementation of the PQML estimator. Robust standard errors in parentheses. Robust standard errors clustered around journals. One article is one observation. We drop 34 observations of HOA articles that are not subject to a HOA pilot agreement. We also include journal dummys. 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 and $BestH_{ijt}$. An increase in $InstRank_{ijt}$ represents a decrease in the ranking of the institution. Thus, a negative coefficient stands for higher net citations for papers of researchers affiliated with higher-ranked institutions. In specification (7) we additionally control for the months of availability of OA pre-prints on RePec and cites to these pre-prints. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 3: PQML regression results

Third, the coefficient for $HOAPilot_{ijt}$ is positive throughout all specifications.[9] It

[9]The effect of the HOA pilot status of articles on net cites is more pronounced for older articles as

is significant at the 5%-level in specifications (1), (2), (3) and (5) and at the 5.1%-level in specification (4). However, the HOA pilot status turns out to be insignificant in specification (6) with a p-value of 0.178 when we control for institution quality and the H-index of the best author. In addition, it is insignificant with a high p-value of 0.942 in specification (7) where we additionally control for the months of availability of OA pre-prints and the citations to these pre-prints. As for the marginal citation effect of HOA, the HOA pilot status of articles increases net cites by about 22% to 26% across specifications (1) to (5) as compared to CA articles published in the same group of journals. This result suggests that, in the absence of HOA pilot agreements, a $\$3,000$ HOA fee would be buying HOA authors between 22% and 26% more net cites. However, once we control for institution quality together with the H-index of the best author the effect turns out to be insignificant and drops down to about 17%. When we additionally control for the months of availability of OA pre-prints on RePec and citations to these pre-prints, the statistically insignificant marginal effect of HOA on net citations drops to 0.4% with a high p-value of 0.942. Fourth, the number of authors appears to be a significant predictor for $NetCit_{ijt}$ as the coefficient is positive and statistically significant at least at the 5% level across specifications (1) to (6). However, once we control for the months of availability of OA pre-prints and citations to these pre-prints in specification (7) the coefficient turns out to be insignificant.

Appendix C illustrates. The HOA citation advantage appears to be relatively low for articles between 20 months since publication (when the first cites occur) and about three years since publication and then increases over the next years.

	Dependent Variable: Net Citations						
	Marginal Effects at the Means after PQML Regression						
Specification:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Months Since Online Publication	0.0684***	0.0642***	0.0637***	0.0668***	0.0657***	0.0614***	0.0502***
	(0.00984)	(0.0162)	(0.0169)	(0.0163)	(0.0178)	(0.0194)	(0.0184)
Months Since Online Publication, sqrd	-0.000464***	-0.000426**	-0.000420**	-0.000455**	-0.000447**	-0.000369*	-0.000317
	(0.000119)	(0.000186)	(0.000193)	(0.000185)	(0.000194)	(0.000221)	(0.000223)
Number of References	0.00391***	0.00591***	0.00597***	0.00610***	0.00613***	0.00734***	0.00412***
	(0.00115)	(0.00114)	(0.00117)	(0.00124)	(0.00117)	(0.00177)	(0.00154)
HOA Article Subject to HOA Pilot Agreement	0.219**	0.240**	0.241**	0.259**	0.248**	0.172	0.00373
	(0.106)	(0.119)	(0.117)	(0.130)	(0.106)	(0.124)	(0.0513)
Number of Authors	0.0840**	0.0862***	0.0894***	0.0801***	0.0814***	0.111***	-0.0201
	(0.0349)	(0.0277)	(0.0329)	(0.0272)	(0.0256)	(0.0333)	(0.0467)
Lead Article	0.123	0.0850	0.0856	0.0809	0.0776	0.0442	0.216
	(0.0988)	(0.0903)	(0.0910)	(0.0951)	(0.0934)	(0.0864)	(0.136)
Number of Pages	0.000591	0.00236	0.00220	0.00292	0.00265	0.00150	0.00781
	(0.00275)	(0.00522)	(0.00534)	(0.00553)	(0.00528)	(0.00596)	(0.00542)
OA Pre-Print	-0.100	-0.101	-0.104	-0.174	-0.200	-0.0917	[dropped]
	(0.108)	(0.112)	(0.114)	(0.114)	(0.133)	(0.124)	
SSRN OA Pre-Print	-0.0733	-0.0755	-0.0691	-0.0706	-0.0688	-0.0559	-0.0493
	(0.0663)	(0.0528)	(0.0555)	(0.0595)	(0.0614)	(0.0727)	(0.0684)
RePec OA Pre-Print	0.112	0.0600	0.0613	0.108	0.126	-0.00009	[dropped]
	(0.0833)	(0.0890)	(0.0886)	(0.0791)	(0.0841)	(0.0976)	
H-Index of Best Author		-0.000570			-0.00324	-0.00485	0.00828
		(0.00292)			(0.0148)	(0.00390)	(0.00640)
Overall Citations of Best Author			-0.00002		-0.000133		
			(0.00005)		(0.000292)		
Average Citations of Best Author				-0.00009	0.00324		
				(0.000726)	(0.00327)		
Rank of Best Institution						-0.00003	-0.000026
						(0.000112)	(0.000340)
Months of Availability of RePec OA Pre-Prints							-0.00799
							(0.00546)
Months of Availability of RePec OA Pre-Prints, sqrd							0.000002
							(0.00006)
Citations to RePec OA Pre-Prints							0.00492***
							(0.00165)
Observations	1,295	783	783	695	695	565	246

Source: Thomson Reuters Journal Citation Reports; Web of Knowledge; RePec, websites of 15 HOA journals.
Note: Following Santos Silva and Tenreyro (2006, 2010), we use Santos Silva and Tenreyro (2011)'s implementation of the PQML estimator. Robust standard errors in parentheses. Robust standard errors clustered around journals. One article is one observation. We drop 34 observations of HOA articles that are not subject to HOA pilot agreements. We also include journal dummys. 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 and $BestH_{ijt}$. An increase in $InstRank_{ijt}$ represents a decrease in the ranking of the institution. Thus, a negative coefficient stands for higher net citations for papers of researchers affiliated with higher-ranked institutions. In specification (7) we additionally control for the months of availability of OA pre-prints on RePec and cites to these pre-prints. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 4: PQML regression marginal effects

Fifth, the coefficient for the RePec OA pre-print dummy is positive and statistically significant at the 10%-level in specification (5). It is statistically insignificant in specifications (1), (2), (3), (4) and (6). The RePec OA pre-print dummy is dropped

in specification (7) where we control for citations to RePec OA pre-prints. The coefficient for citations to OA pre-prints on RePec is positive and significant at the 1% level in specification (7). The statistically significant positive marginal effect of $RePEcCitPrePrints_{ijt}$ on $NetCit_{ijt}$ is 0.5%. Intuitively, more successful working papers in terms of citations may ceteris paribus result in more cited published articles as they have already attracted interested (and citing) readers. In addition, it is not the mere existence of RePec OA pre-prints that has a statistically significant positive effect on net citations but citations to these pre-prints. Sixth, $LeadArt_{ijt}$ is positive and significant at the 5% level in specification (7). However, $NumPage_{ijt}$, $OAPrePrint_{ijt}$ and $SSRNOAPrePrint_{ijt}$ are insignificant across all specifications.[10] $BestH_{ijt}$, $BestCit_{ijt}$, $BestAvCit_{ijt}$ and $InstRank_{ijt}$ are (where included) statistically insignificant.[11] Finally, $MonthsRePecPrePrint_{ijt}$ and $MonthsRePecPrePrintSq_{ijt}$ are statistically insignificant in specification (7).

3.2 Negative Binomial Regression Model

As an alternative model, we use the negative binomial regression model with robust standard errors clustered around journals to estimate the relation between net citations and the set of variables under study. The regression results are reported in Table 5.

[10] As the coefficient for $SSRNOAPrePrint_{ijt}$ is negative and insignificant across all specifications, we focus our analysis on the availability of OA pre-prints on RePec and citations to these pre-prints.

[11] 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 0.4827. We cannot reject the null hypothesis and conclude that the three variables are jointly insignificant.

Dependent Variable:	Net Citations						
Regression Model:	Negative Binomial Regression						
Specification:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Months Since Online Publication	0.211***	0.192**	0.192**	0.200***	0.201***	0.173**	0.168**
	(0.0599)	(0.0754)	(0.0764)	(0.0776)	(0.0754)	(0.0852)	(0.0782)
Months Since Online Publication, sqrd	-0.00143**	-0.00125	-0.00124	-0.00135	-0.00136*	-0.000983	-0.00122
	(0.000640)	(0.000827)	(0.000841)	(0.000855)	(0.000823)	(0.000941)	(0.000887)
Number of References	0.0146***	0.0195***	0.0197***	0.0195***	0.0192***	0.0236***	0.0146***
	(0.00414)	(0.00357)	(0.00368)	(0.00422)	(0.00399)	(0.00343)	(0.00449)
HOA Article Subject to HOA Pilot Agreement	0.432*	0.454**	0.458**	0.444**	0.436**	0.278	0.0885
	(0.221)	(0.218)	(0.224)	(0.218)	(0.206)	(0.270)	(0.193)
Number of Authors	0.220**	0.221***	0.229***	0.206**	0.204***	0.301***	0.0491
	(0.0954)	(0.0840)	(0.0880)	(0.0902)	(0.0780)	(0.0875)	(0.134)
Lead Article	0.226	0.232	0.235	0.240	0.242	0.0644	0.450**
	(0.206)	(0.192)	(0.192)	(0.231)	(0.231)	(0.168)	(0.216)
Number of Pages	-0.00141	0.00525	0.00509	0.00615	0.00610	0.00486	0.0105
	(0.00789)	(0.0134)	(0.0136)	(0.0129)	(0.0128)	(0.0118)	(0.0134)
OA Pre-Print	-0.326	-0.273	-0.280	-0.363	-0.378	-0.188	[dropped]
	(0.294)	(0.294)	(0.296)	(0.346)	(0.317)	(0.327)	
SSRN OA Pre-Print	-0.161	-0.163	-0.152	-0.165	-0.156	-0.153	-0.127
	(0.197)	(0.157)	(0.151)	(0.166)	(0.161)	(0.173)	(0.190)
RePec OA Pre-Print	0.518**	0.379	0.389	0.415	0.427*	0.113	[dropped]
	(0.244)	(0.278)	(0.268)	(0.287)	(0.256)	(0.326)	
H-Index of Best Author		0.0118*			0.000194	0.00711	0.0277*
		(0.00604)			(0.0311)	(0.0107)	(0.0149)
Overall Citations of Best Author			0.00008		-0.000126		
			(0.000067)		(0.000593)		
Average Citations of Best Author				0.00242	0.00519		
				(0.00154)	(0.0102)		
Rank of Best Institution						0.00008	-0.00002
						(0.000270)	(0.000990)
Months of Availability of RePec OA Pre-Prints							-0.0278*
							(0.0144)
Months of Availability of RePec OA Pre-Prints, sqrd							0.00007
							(0.000167)
Citations to RePec OA Pre-Prints							0.0152***
							(0.00391)
Constant	-7.543***	-7.502***	-7.453***	-7.545***	-7.554***	-7.423***	-5.468***
	(1.445)	(1.727)	(1.749)	(1.773)	(1.733)	(1.832)	(1.455)
ln(a)	0.0587	0.0464	0.0473	0.126	0.123	0.0705	-0.639*
	(0.193)	(0.225)	(0.226)	(0.224)	(0.215)	(0.219)	(0.361)
Observations	1,295	787	787	698	698	566	249

Source: Thomson Reuters Journal Citation Reports; Web of Knowledge; RePec, websites of 15 HOA journals.
Note: Robust standard errors in parentheses. Robust standard errors clustered around journals. One article is one observation. We drop 34 observations of HOA articles that are not subject to HOA pilot agreements. We also include journal dummys. 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 and $BestH_{ijt}$. An increase in $InstRank_{ijt}$ represents a decrease in the ranking of the institution. Thus, a negative coefficient stands for higher net citations for papers of researchers affiliated with higher-ranked institutions. In specification (7) we additionally control for the months of availability of OA pre-prints on RePec and cites to these pre-prints. The log-transformed over-dispersion parameter ln(a) is estimated and displayed. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 5: Negative binomial regression results

The following findings support our main results of the PQML regressions. First, the months since online publication are a strong predictor for net citations. The coefficient is positive and significant at least at the 5% level throughout all specifications. Table 6

reports the marginal effects at the means of the negative binomial regression.[12] As in the PQML regression, the statistically significant positive marginal effect of $MonthPub_{ijt}$ on $NetCit_{ijt}$ ranges between 5% and 7% across specifications. Second, the number of references is a strong predictor for net citations. The coefficient is positive and significant at the 1%-level throughout all specifications. In addition, the statistically significant positive marginal effect of $NumRef_{ijt}$ on $NetCit_{ijt}$ ranges between 0.4% and 0.8% across specifications. Third, the coefficient for $HOAPilot_{ijt}$ is positive throughout all specifications. It is significant at the 5%-level in specifications (2), (3), (4) and (5) and at the 5.1%-level in specification (1). However, as in the PQML regressions, the HOA pilot status of articles turns out to be insignificant in specification (6) with a p-value of 0.303 when we control for institutional quality and the H-index of the best author. In addition, it is insignificant with a high p-value of 0.647 in specification (7) where we additionally control for the months of availability of OA pre-prints and the citations to these pre-prints. As for the marginal citation effect of HOA, the HOA pilot status of articles increases net cites by about 17% across specifications (1) to (5) as compared to CA articles published in the same group of journals. However, once we control for institution quality together with the H-index of the best author the effect turns out to be insignificant and drops down to about 10%. When we also control for the months of availability of OA pre-prints on RePec and citations to these pre-prints, the (statistically insignificant) marginal effect of HOA on net citations drops to 2.4%.

[12] We also run the PQML and negative binomial regressions including the five dummy variables for HOA pilot institutions (results not reported here). Including these variables does not qualitively change our results. Notably, however, none of these dummy variables has a statistically significant positive marginal effect on net citations across all specifications. This result suggests that $HOAPilot_{ijt}$ is not positively picking up the quality of HOA pilot institutions which are slightly better ranked (on average, 177) than Non-HOA pilot institutions (on average, 189).

Dependent Variable: Net Citations

Marginal Effects at the Means after Negative Binomial Regression

Specification:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Months Since Online Publication	0.0698***	0.0596***	0.0594***	0.0645***	0.0645***	0.0591**	0.0452***
	(0.0129)	(0.0176)	(0.0180)	(0.0184)	(0.0178)	(0.0230)	(0.0169)
Months Since Online Publication, sqrd	-0.000473***	-0.000389*	-0.000386*	-0.000436*	-0.000438*	-0.000336	-0.000330
	(0.000166)	(0.000221)	(0.000226)	(0.000233)	(0.000224)	(0.000290)	(0.000212)
Number of References	0.00484***	0.00605***	0.00610***	0.00628***	0.00617***	0.00807***	0.00393***
	(0.00138)	(0.00123)	(0.00127)	(0.00147)	(0.00130)	(0.00156)	(0.00142)
Article Subject to HOA Pilot Agreement (‡)	0.169*	0.166*	0.168*	0.168*	0.164*	0.104	0.0244
	(0.0992)	(0.0904)	(0.0935)	(0.0939)	(0.0856)	(0.107)	(0.0542)
Number of Authors	0.0728**	0.0685***	0.0712***	0.0665**	0.0656***	0.103***	0.0132
	(0.0302)	(0.0233)	(0.0250)	(0.0271)	(0.0224)	(0.0299)	(0.0357)
Lead Article(‡)	0.0820	0.0787	0.0801	0.0848	0.0853	0.0226	0.146
	(0.0785)	(0.0721)	(0.0722)	(0.0898)	(0.0900)	(0.0611)	(0.0931)
Number of Pages	-0.000468	0.00163	0.00158	0.00198	0.00196	0.00166	0.00282
	(0.00263)	(0.00409)	(0.00416)	(0.00406)	(0.00403)	(0.00397)	(0.00371)
OA Pre-Print(‡)	-0.104	-0.0848	-0.0870	-0.118	-0.122	-0.0646	[dropped]
	(0.0945)	(0.0921)	(0.0927)	(0.110)	(0.101)	(0.114)	
SSRN OA Pre-Print(‡)	-0.0508	-0.0484	-0.0454	-0.0512	-0.0483	-0.0506	-0.0340
	(0.0553)	(0.0430)	(0.0416)	(0.0493)	(0.0470)	(0.0524)	(0.0489)
RePec OA Pre-Print(‡)	0.192*	0.121	0.125	0.137	0.141*	0.0388	[dropped]
	(0.101)	(0.0917)	(0.0883)	(0.0917)	(0.0818)	(0.113)	
H-Index of Best Author		0.00366*			0.0000622	0.00243	0.00746*
		(0.00206)			(0.00998)	(0.00383)	(0.00436)
Overall Citations of Best Author			0.000025		-0.00004		
			(0.00002)		(0.000193)		
Average Citations of Best Author				0.000781	0.00166		
				(0.000573)	(0.00342)		
Rank of Best Institution						0.000028	-0.000006
						(0.000096)	(0.000267)
Months of Availability of RePec OA Pre-Prints							-0.00749*
							(0.00387)
Months of Availability of RePec OA Pre-Prints, sqrd							0.000019
							(0.00005)
Citations to RePec OA Pre-Prints							0.00409***
							(0.00112)
Observations	1,295	787	787	698	698	566	249

Source: Thomson Reuters Journal Citation Reports; Web of Knowledge; RePec, websites of 15 HOA journals.
Note: Robust standard errors in parentheses. Robust standard errors are clustered around journals. One article is one observation. We drop 34 observations of HOA articles that were not subject to a HOA pilot agreement. We also include journal dummys. 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 and $BestH_{ijt}$. An increase in $InstRank_{ijt}$ represents a decrease in the ranking of the institution. In specification (7) we additionally control for the months of availability of OA pre-prints on RePec and cites to these pre-prints. (‡) Marginal effect for discrete change of dummy variable from 0 to 1. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Negative binomial regression marginal effects

Fourth, the coefficient for citations to RePec OA pre-prints is positive and significant at the 1%-level in specification (7). The statistically significant positive marginal effect

of $RePEcCitPrePrints_{ijt}$ on $NetCit_{ijt}$ is 0.4%. Fifth, the number of authors appears to be a significant predictor for $NetCit_{ijt}$ as the coefficient is positive and significant at least at the 5% level across specifications (1) to (6). However, as in the PQML regressions, once we control for the months of availability of OA pre-prints and citations to these pre-prints in specification (7) the coefficient turns out to be insignificant. Sixth, the coefficient for the availability of OA pre-prints on RePec is statistically significant at the 5%-level in specification (1) and at the 10%-level in specification (5). However, once we control for the H-index of the best author together with institutional quality the coefficient is insignificant with a p-value of 0.73 in specification (6). Seventh, $LeadArt_{ijt}$ is positive and significant at the 5% level in specification (7). Eighth, $NumPage_{ijt}$, $OAPrePrint_{ijt}$ and $SSRNOAPrePrint_{ijt}$ are insignificant across all specifications. Ninth, $BestCit_{ijt}$ and $BestAvCit_{ijt}$ are insignificant in specifications (3), (4) and (5), respectively. Finally, there are two main differences between the results of the PQML regressions and the negative binomial regressions. The coefficient for $BestH_{ijt}$ is positive and statistically significant at the 10% level in specification (2) with a p-value of 0.051 and in specification (7) with a p-value of 0.063 in the negative binomial regression whereas it is insignificant in the PQML regression. In addition, the coefficient for $MonthsRePecPrePrint_{ijt}$ is negative and statistically significant at the 10% level in specification (7) in the negative binomial regression whereas it is negative and statistically insignificant in the PQML regression. A possible interpretation for the negative sign of the coefficient in both models is that rejections increase the time between the upload of a working paper and the publication date of the final article. In addition, repeatedly rejected papers are ceteris paribus more likely to be published in (less read and cited) journals at the lower end of the quality ladder.

4 Conclusion

We use a unique, hand-collected dataset containing information on 1,329 CA and HOA articles published in the same journals to analyze the effect of HOA on net citations. Our analysis benefits from the recent introduction of HOA pilot agreements under which the HOA status is typically assigned to all articles of authors affiliated with HOA pilot institutions. We find that the months since online publication, the number of references and citations to RePec OA pre-prints are strong predictors for net citations. Performing PQML regressions, the HOA status turns out to be a significant predictor of net citations in five out of seven specifications with marginal effects ranging between 22% and 26%. However, once we additionally control for institution quality and citations to RePEc OA pre-prints, the HOA citation advantage turns out to insignificant and the marginal citation effect of HOA drops to 0.4%. Our additional results from negative binomial regressions provide strong support for these findings. Our empirical evidence therefore suggests that HOA does not constitute a significant advantage in the competition for citations. Finally, the analysis of the impact of the introduction of the HOA publication format on journal impact factors appears to be a promising idea for further research. In addition, HOA pilot agreements can be exploited to analyze whether author-driven self-selection is in fact present in HOA publishing as suggested in the science literature on OA citation effects.

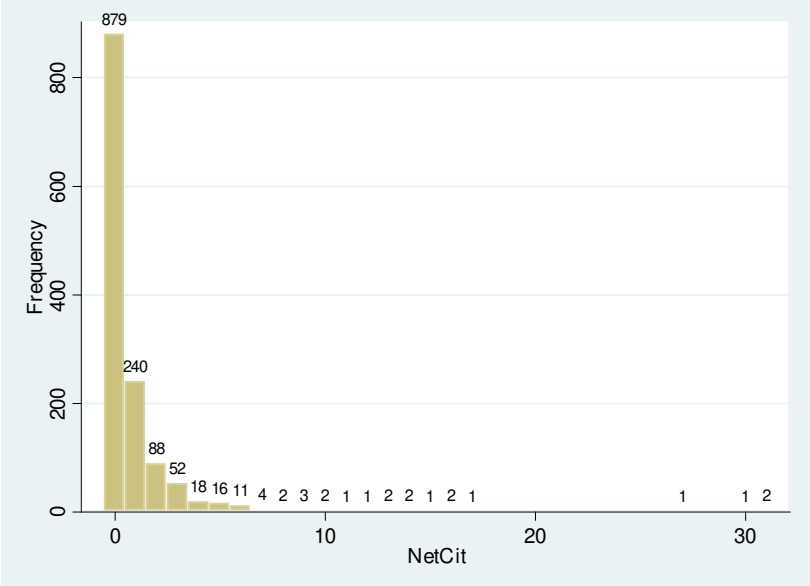
5 Appendix

Appendix A: HOA pilot agreements and periods

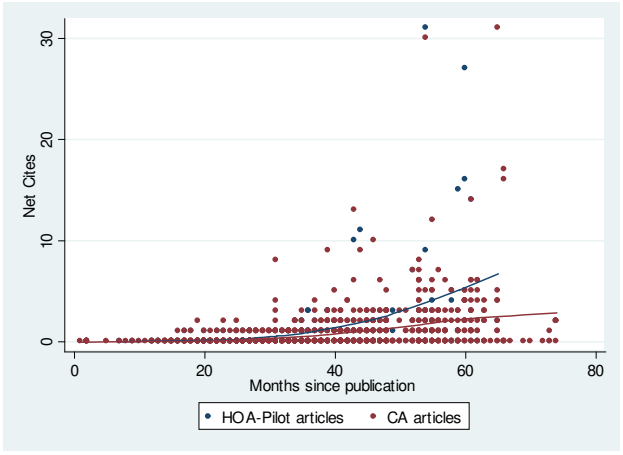
Institutions	Pilot Periods	Members
Dutch Consortium of University Libraries (UKB)	01/01/2007 – 06/31/2012	Universities of Amsterdam, Delft, Eindhoven, Groningen, Leiden, Maastricht, Nijmegen, Rotterdam, Tilburg, Twente, Utrecht and Wageningen; Vrije University
University of California	11/01/2008 – 12/31/2010	UC-campuses: Berkeley, Davis, Irvine, Los Angeles, Merced Riverside, San Diego San Francisco, Santa Barbara, Santa Cruz
Max Planck Society (MPG)	01/01/2008 – 12/31/2009	All 78 Max Planck Institutes (MPIs), e.g., MPI for Economics, MPI for Research on Collective Goods and MPI for Intellectual Property and Competition Law
University of Goettingen	01/01/2007 – 12/31/2012	Single member
University of Hong-Kong	03/01/2010 – 03/01/2012	Single member

Source: UC Berkeley (2009), UKB (2007), Wageningen University and Research Centre (2008), Schmidt and Shearer (2012) and personal communications with the head of the Max Planck Virtual Library.

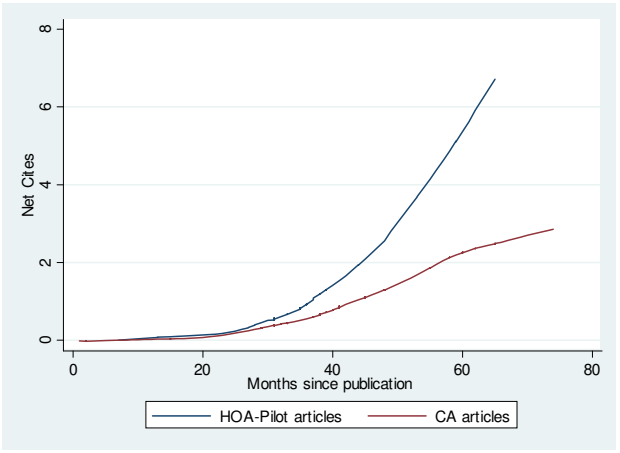
Appendix B: Histogram of net citations



Appendix C: Net cites to HOA pilot articles and CA articles



a) Scatterplot



b) Locally weighted scatterplot smoothing

Acknowledgement: *We gratefully acknowledge financial support from the Sloan Foundation under the Economics of Knowledge Contribution and Distribution Grant 2013. We thank Patrick Andreoli-Versbach, Henri de Belsunce, Georg von Graevenitz, Dietmar Harhoff, Ariel Katz, Mark McCabe, Ivan Png, Bob Reed, Ralf Schimmer, Florian Stellner, Jerry Thursby, Stefan Wagner, Joel Waldfogel and Michael Waldman for their helpful comments and suggestions. We also thank participants of the 2013 Conference of the Scientific Commission "Technology, Innovation and Entrepreneurship", the Congress of the Society for Economic Research on Copyright Issues 2012 and 2013, the Fourth Workshop for Junior Researchers on the Law & Economics of Intellectual Property and Competition Law as well as seminar participants at LMU Munich, Hamburg University and the University of Waikato for their valuable comments. We thank Michael Gerstenberger, Julian Hackinger, Benjamin Heisig, Jonas Rathfelder, Janis Sussick and Paul Walker for excellent research assistance.*

References

[1] 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, 63(8). 1496-1504.

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

[3] 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.

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

[5] The Economist (2012), Brought to Book: Academic Journals Face a Radical Shake-Up, July 21. Available at: http://www.economist.com/node/21559317 (last accessed on December 12, 2013).

[6] European Commission (2012). Towards Better Access to Scientific Information: Boosting the Benefits of Public Investments in Research. European Commission: Brussels.

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

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

[9] Harnad, S., Brody, T. (2004), Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals, D-Lib Magazine, 10(6).

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

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

[12] McCabe, M. J., Snyder, C. M., forthcoming (a), Identifying the Effect of Open Access on Citations Using a Panel of Science Journals, Economic Inquiry.

[13] McCabe, M. J., Snyder, C. M., forthcoming (b), Does Online Availability Increase Citations? Theory and Evidence from a Panel of Economics and Business Journals, Review of Economics and Statistics.

[14] McCabe, M. J., Snyder, C. M., Fagin, A. (2013), Open Access versus Traditional Journal Pricing: Using a Simple "Platform Market" Model to Understand Which Will Win (and Which Should), Journal of Academic Librarianship, 39, 11-19.

[15] Schmidt, B., Shearer, K. (2012), Licensing Revisited: Open Access Clauses in Practice, LIBER Quarterly, 22(3), 176-189.

[16] Santos Silva, J. M. C., Tenreyro, S. (2006), The Log of Gravity, Review of Economics and Statistics, 88(4), 641-658.

[17] Santos Silva, J. M. C., Tenreyro, S. (2010), On the Existence of the Maximum Likelihood Estimates in Poisson Regression, Economics Letters, 107(2), 310-312.

[18] Santos Silva, J. M. C., Tenreyro, S. (2011), Poisson: Some Convergence Issues, Stata Journal, 11(2), 207-212.

[19] UC Berkeley (2009), UC libraries and Springer Sign Pilot Agreement for Open Access Journal Publishing. Available at: http://www.universityofcalifornia.edu/news/article/19335 (last accessed on December 12, 2013).

[20] UKB (2007), Springer and Dutch Library Consortium to Cooperate in Open Access Initiative. Available at: http://www.dans.knaw.nl/en/content/categorieen/nieuws/springer-and-dutch-library-consortium-cooperate-open-access-initiative (last accessed on December 12, 2013).

[21] Wageningen University and Research Centre (2008), Springer Open Choice also Free in 2008 and 2009. Available at: http://library.wur.nl/WebQuery/nieuws/desktop/712 (last accessed on December 12, 2013).

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

[23] Wooldridge, J. M. (1999), Distribution-Free Estimation of Some Nonlinear Panel Data Models, Journal of Econometrics, 90(1), 77-97.