



Universiteit
Leiden
The Netherlands

The deliverance of open access books : examining usage and dissemination

Snijder, R.

Citation

Snijder, R. (2019, January 29). *The deliverance of open access books : examining usage and dissemination*. Retrieved from <https://hdl.handle.net/1887/68465>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/68465>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/68465> holds various files of this Leiden University dissertation.

Author: Snijder, R.

Title: The deliverance of open access books : examining usage and dissemination

Issue Date: 2019-01-29

3 The influence of open access on monograph sales : The experience at Amsterdam University Press

Snijder, R. (2014). The Influence of Open Access on Monograph Sales : The experience at Amsterdam University Press. *LOGOS: The Journal of the World Book Community*, 25(3), 13–23. <https://doi.org/10.1163/1878-4712-11112047>

3.1 Introduction

For years, decreasing sales have threatened the sustainability of monograph publishing and this has led to a search for alternative models. Most of these proposed models are hybrid: they contain an open access component combined with selling other versions of the book. In this paper, the experiences of Amsterdam University Press (AUP) in using a hybrid model are analysed by looking at the effect of open access publishing on monographs sales. However, several other influences may also affect sales, and these will be taken into account as well. The goal is to find what effect making books freely available online has on sales. To achieve this, I shall apply statistical methods to the sales data of 513 books from a period of three years.

The economic problems concerning monographs have been discussed by Wasserman and Thompson. Wasserman (1998) discusses the costs of publishing monographs and the dramatic effects of declining library sales. Thompson (2005) extensively reviews the challenges – including financial challenges – facing monograph publishing.

Others look at the possibilities of digital publishing in an open access model. Greco & Wharton (2008) conclude that university presses cannot survive on a ‘print-only’ business model and should consider open access publishing. Steele (2008) draws more or less the same conclusion and describes the open access model as ‘a viable alternative when placed within institutional settings’. Houghton *et al.* (2009) discuss the costs of scholarly publishing – including the costs of monographs – and conclude that open access publishing is beneficial for society. Withey *et al.* (2011) acknowledge a trend towards more open access publishing, but stress the need for sustainable business models. Cross urges academic libraries to support small

academic publishers by purchasing their monographs (Cross, 2011). Pinter (2012) also discusses the financial perils of publishing monographs and proposes a solution in which a consortium of libraries fund an open access version of a title, enabling the publisher to sell enhanced digital or paper versions of the book.

Recently, Ferwerda (2014) listed the current business models for open access and monographs, ranging from a hybrid publication model to crowd-funding. Jackson (2014) – a publisher at Oxford University Press – describes the current lack of demand for a publishing model in which all costs are met before publication. At this moment, there is no consensus regarding a ‘proven’ business model.

Some authors also try to find evidence of whether free digital versions of a book have an effect on sales. Hilton & Wiley (2011) conclude that a correlation exists between a free e-book and increased print sales. Their research used an experimental group of eight books and a control group of six books, both fiction and non-fiction. Snijder (2010) set up an experiment on monographs, using three experimental groups of 100 titles each and a control group of 100 titles. One result was that making a book freely available did not affect the number of copies sold. Based on the same principles, the Dutch-based OAPEN Foundation set up a two-year experiment: OAPEN-NL. During that period, 50 books were published on open access and also as a paper monograph. Several aspects – sales among them – were monitored and compared with a control group of comparable titles published in the traditional way. The results were similar to Snijder’s results: the number of copies sold was not affected by publishing on open access (Ferwerda *et al.*, 2013). In the UK, JISC set up an experiment called OAPEN-UK. Here, 29 ‘matched pairs’ of monographs are compared: one title in each pair is made available on open access while no changes are made to the other’s publication model (Collins & Milloy, 2012).

This paper does not follow the same controlled arrangement used by Snijder (2010) or the OAPEN-NL experiment. Instead of investigating carefully balanced data sets, I use all titles published under one imprint published by AUP. Whereas the experiment of 2010 used data selected over nine months of the year 2009, here the time frame is much larger: 36 months, the years 2010 to 2012.

Amsterdam University Press is an academic publisher – owned by the University of Amsterdam – that publishes monographs and journals, mostly in the field of humanities and social sciences (AUP, 2012). The Press has gained extensive experience with open access publishing. The open access monographs are always made available via a hybrid model in which the

print version of the book is sold and a digital version is made available free. Since 2010, the open access titles published under the imprint 'Amsterdam University Press' have been released not only through AUP's repository but also via the OAPEN Library. The OAPEN Library is an important dissemination channel for AUP's open access books: in April 2014, the Library contained 447 titles published by AUP.

The OAPEN Library (<http://www.oapen.org>) was officially launched in September 2010 (OAPEN Consortium, 2011). It is a web-based collection of open access monographs published by dozens of publishers. In April 2014, the collection contained over 2100 titles by 68 publishers. The OAPEN Library offers several ways to access its contents: it enables searching and browsing, readers can share book descriptions via social media, and it contains several data feeds (Snijder, 2013a). Amsterdam University Press is part of the board of the OAPEN Foundation, which maintains the OAPEN Library.

3.2 The data set

In this paper, the following research question will be discussed: what is the influence of open access on monograph sales and how large is the influence of open access publishing compared with other influences on monograph sales? The data set consists of 513 books published under the imprint 'Amsterdam University Press'. All books published under this imprint are subject to peer review. The group of books consists of 69 published in 2010, 68 published in 2011, 62 published in 2012, and 319 published between 1995 and 2009. Over 70 per cent of those books – 378 titles – were published on open access and are available in the OAPEN Library (Table 1).

Table 1. Titles available and not available on open access

	<i>Number of titles</i>	<i>Percentage</i>	<i>Number of copies sold</i>	<i>Sales percentage</i>	<i>Average sales per title</i>
On open access	378	73.7	67 210	65.6	66.3
Not on open access	135	26.3	35 170	34.4	105.6
Total	513	100	102 380	100	76.0

Of the total number of copies sold during the years 2010–2012, over 65 per cent were open access titles. However, the average number of copies sold per title was lower, than for titles not published in open access. The

turnover associated with these sales will not be discussed in this paper. Below, other influences on monograph sales are highlighted.

3.3 Influences on monograph sales

On the basis of the expertise of employees of AUP, several other possible influences on sales were defined: commercial potential, frontlist and backlist, and language. Each influence will be discussed below.

3.3.1 Commercial potential

An important part of the publishing process is determining how well a title will sell. The publisher will take into account several properties of the book and predict the number of copies that will sell. This then informs the print run, the number of copies made available for sale. The print run of the titles under consideration ranges from zero – no copies are printed beforehand – to 5000. The average print run of the books available on open access is lower than the average print run of books that are not available in this way. The average print run for books on open access is 459, whereas the average print run for the other titles is 652 (Table 2). In other words, the expected sales of books that are not available on open access are over 140 per cent of the expected sales of books on open access.

Table 2. Mean print run

	Number of titles	Mean print run
On open access	378	458.8
Not on open access	135	652.2
Total	513	509.7

Using a mean print run for a set of 513 titles is rather a crude instrument, which hides the complex decisions made for each title. Furthermore, if print runs are declining, this may also influence the data. First of all, the decline in print runs is not very clear in my data. The average print runs per publication year range from 286 to 782, without any clear trend. Secondly, the statistical analysis will take into account the sales data of each individual book per year and not use the averages described here. However, these averages give us a first clue about the commercial expectations of

the titles that have been made available on open access, compared with the other titles.

3.3.2 Frontlist and backlist

Publishers refer to the titles published in the current year as the ‘frontlist’; all other titles are referred to as the ‘backlist’. Experience shows that sales in the first year are generally higher than sales figures in subsequent years. This is the case with the books in our data set. The average number of copies sold of books published in 2011 was twice the average number the next year. The same holds true for books published in 2010: the average sales in 2010 are almost 2.5 times higher than the average number of copies sold in 2011 (Table 3).

Table 3. Front and backlist sales

<i>Publishing date</i>	<i>Number of titles</i>	<i>Mean sales in 2010</i>	<i>Mean sales in 2011</i>	<i>Mean sales in 2012</i>
2009 and before	314	49.8	31.1	21.4
2010	69	355.9	145.4	95.1
2011	68	–	195.2	96.0
2012	62	–	–	150.5
Total	513			

We could also look at the average sales per year. This seems to reveal a downward trend: in 2010, the average number of copies sold was 104; in 2011, 73; and in 2012, 56. Nevertheless, three years’ data cannot be used to make conclusions about long-term developments. Moreover, it is possible that the decline in sales is spread evenly over all titles, regardless of whether they are published on open access. The study is set up to answer the question of whether publishing on open access makes a difference, taking into account the differences in frontlist and backlist sales, and the average sales per year.

3.3.3 Language

The analysed books were either published in Dutch or in English. Dutch-language books are more likely to be sold only in Dutch-speaking countries, whereas English-language books may be sold globally. The differences in

potential markets may influence the number of copies sold. In the examined group of titles, 63 per cent are published in English (Table 4).

Table 4. Titles per language

	<i>English</i>	<i>Dutch</i>	<i>Total</i>
On open access	268	110	378
Not on open access	54	81	135
Total	322	191	513

3.4 Data and Results

Here I measure the effect of the four influences – open access publishing, commercial potential, front and backlist, language – on sales. The ANOVA statistical method (analysis of variance) is used to check whether each influence has a significant effect. As a second step, the influences are combined to see in what way the interaction of these influences affects the number of books sold. The data are summarized in Table 5.

Table 5. Data: mean sales per influence

Influence		Mean sales	N	Percentage total sales
Open access publishing	With	66.28	1014	65.6
	Without	105.62	333	34.4
Commercial potential	Print run: 0	27.76	430	11.7
	Print run: 1–1000	69.62	782	53.2
	Print run: 1001–2000	21.71	105	22.3
	Print run: 2001–3000	4.54	24	10.6
	Print run: 3001–4000	3.24	3	0.9
	Print run: 4001–5000	4.46	3	1.3
Front and backlist	Frontlist	236.94	199	46.1
	Backlist	48.11	1148	53.9
Language	English	57.36	829	46.4
	Dutch	105.85	518	53.6

3.4.1 Separate influences

The effect of each influence is measured using the ANOVA procedure. This tests whether the differences among the mean sales of the books can be explained by chance. The results of each individual test are summarized in Appendix 1: ANOVA results per influence.

It is clear from the results that each influence by itself correlates with monograph sales in our data. So, while it is true that open access publishing is connected to sales, this is also true for commercial potential, front and backlist, and language. We can use two parameters to estimate the size of the effect: ω^2 and F -ratio. The first indicates that both commercial potential and front and backlist sales have a more profound effect on the number of copies sold than do open access or language. If we take into account the F -ratio, there is one outlier: front and backlist sales, whose F -ratio is 51.016, almost three times larger than the second-highest F -ratio. The difference between the mean sales of the frontlist and the mean sales of the backlist also reflect this large effect.

3.4.2 Combining influences

We did see that each influence is statistically significant, and this makes it harder to single out the effects of open access. It also became clear that there is a large difference between sales of the frontlist and sales from the backlist. The mean of all frontlist sales is almost 237, whereas the mean of all backlist sales is just over 48. In order to compensate for this large difference, the data are split into frontlist sales and backlist sales. Statistical methods are applied to these two data sets to measure the effect of open access publishing, combined with the influence of commercial potential and language.

3.4.3 Frontlist: data and results

The frontlist data consist of the sales data of 199 titles: the titles that were published in 2010, 2011, and 2012. Only the sales of the first year of publication are taken into account. At first glance, the difference is not very large between the mean sales of open access books (240.15) and those of books not available on open access (227.88). This contrasts strongly with language, where the mean sales of English-language books is approximately 40 per cent of the mean sales of books published in Dutch. The effects of commercial potential are visible: books with a higher print run did sell better

on average. A multifactor ANOVA procedure is used to test the effect of the combined influences.

When we look at the statistical analysis in Appendix 2: Frontlist results, the results of the frontlist can be explained by a combination of commercial potential and language. Open Access publishing does not have an effect in this situation. When we look at the effect size – measured by partial η^2 – of both print run and language, it becomes clear that commercial potential (measured by print run) plays the largest role. Of course, this is hardly surprising.

3.4.4 Backlist: data and results

The amount of available data from the backlist is much larger. Firstly, it contains the data of the 314 books published in 2009 and earlier which were sold during 2010–2012. On top of that, it contains the sales data for the years 2011 to 2012 of the 69 books published in 2010. And, lastly, the 2012 sales of the 68 books published in 2011 are also part of this set. The mean sales of the backlist are much lower: the backlist sales are on average 21 per cent of the frontlist sales mean. Compared with frontlist sales, the difference between mean sales of books on open access and mean sales of books not on open access is much larger: 82 for titles not on open access versus just under 37 for open access books. Still, the total number sold of backlist books not on open access is roughly 70 per cent of the number of backlist open access books sold.

Using the same procedure as before, the results for the backlist can be explained by a combination of commercial potential and open access. Language does not play a significant role. The results are listed in Appendix 3: Backlist results. Still, to get meaningful results from a multifactor ANOVA procedure, several preconditions must be met. The most important precondition is homogeneity of variance. In other words, the means used in the procedure should be evenly spread. The backlist data did not meet this condition, and so we must interpret the results with caution.

As a possible solution to overcome the statistical problems, the data can be split into smaller samples based on ‘print run groups’. This creates four subsets of books with the same commercial potential, where each subset contains books that are published on open access and books that are not. Creating these subsets enables us to measure the effect of open access while controlling for the effects of commercial expectations. As a consequence, the subsets contain fewer data; this is most noticeable with the set ‘print run 2001–3000’, where the number of data items is as low as 20 ($N = 20$). For

each set, a one-way ANOVA procedure is performed to test the influence on open access on sales.

Table 6 presents the mean sales of books in the backlist, sorted by print run. The most striking difference between open access books and books not on OA can be found for print runs between 1001 and 2000. In this relatively small group ($N = 92$), the mean sale of titles not on open access is 201.94, compared with 68.21 for titles published on open access. Consistent with the discussion in the introduction, most of the titles published have small print runs: 1000 or less. In the group of titles with a print run of zero and the group of titles with a print run between 1 and 1000, the difference in mean sales between open access books and books not on open access is smaller.

Table 6. Backlist data: commercial potential and open access

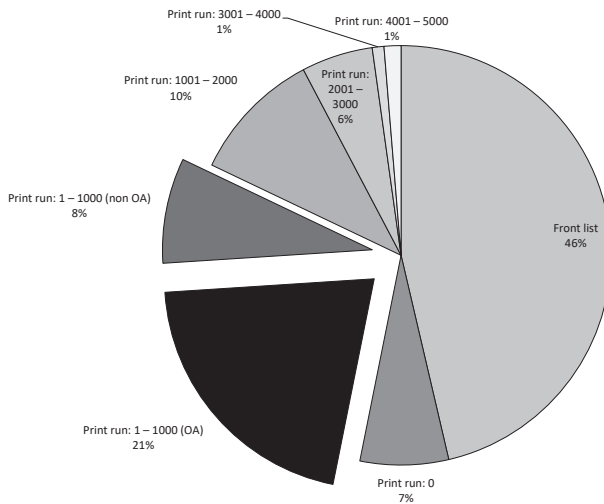
Commercial potential		Mean sales	N	Percentage total sales
Print run: 0	On open access	17.08	290	4.8
	Not on open access	21.65	95	2.0
Print run: 1–1000	On open access	41.84	509	20.8
	Not on open access	62.38	136	8.3
Print run: 1001–2000	On open access	68.21	61	4.1
	Not on open access	202.94	31	6.1
Print run: 2001–3000	On open access	213.57	7	1.5
	Not on open access	321.38	13	4.1
Print run: 3001–4000	Not on open access	324	3	0.9
Print run: 4001–5000	Not on open access	446	3	1.3

When the sales data are analysed using statistical methods, we see that open access publishing is a significant – negative – influence on the average number of copies sold in certain cases only: the subsets of books whose print run is between 1 and 1000 or between 1001 and 2000. No significant effect on books with a print run of zero or between 2001 and 3000 could be measured. Furthermore, the measured effect of open access on sales is much higher for the books with a print run between 1001 and 2000 than for the books with a print run between 1 and 1001. The results are fully described in Appendix 3: Backlist results.

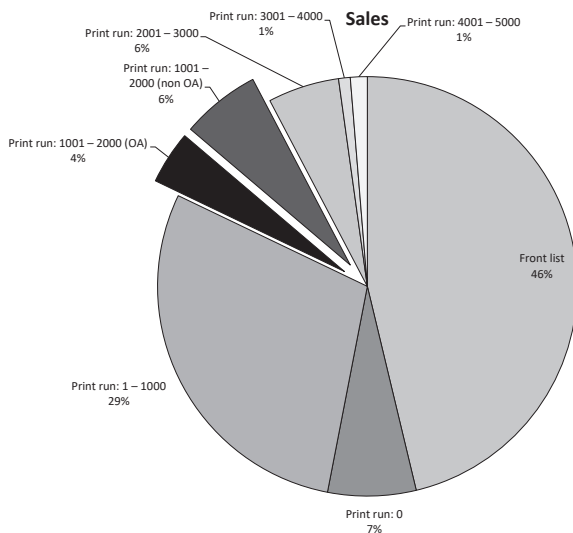
At first glance, the outcomes of this paper run counter to the results of Snijder (2010) and OAPEN-NL (Ferwerda *et al.*, 2013). Here we see that making books available on open access has affected sales in certain circumstances. However, when we look at the total number of copies sold, the effects are not as strong as might be expected. This is best explained using the following illustrations.

In Figure 1, the backlist sales of titles with a print run between 1 and 1000 are highlighted. The mean sales of books published in closed access are higher than those of books published in open access. However, the total number of copies sold is much lower: it amounts to 8 per cent of all sales. In contrast, the number of copies sold of titles under open access is 21 per cent of all sales.

Figure 1. Backlist sales, print run 1–1000.



The backlist sales of titles with a print run between 1001 and 2000 are highlighted in Figure 2. Here, the difference between mean sales of open access titles and those of titles not on open access is quite large: 68 versus 202. This does not lead to an equally large difference in total sales. We can see that 10 per cent of all copies sold are have a print run between 1001 and 2000. Here, four per cent are published on open access and six per cent that are not made available in this way.

Figure 2. Backlist sales, print run 1001–2000.

3.5 Discussion

Sales of monographs are influenced by several factors, and this paper tries to measure the effects of open access publishing, commercial potential, front and backlist sales, and language. The results show that all factors are influential, but the strengths of the measured effect are not equal. The difference between front and backlist sales is by far the greatest, and it was necessary to split the data to find meaningful results for the other influences.

The data used here did not come from a carefully set up experiment. Instead there was a bias in commercial expectations: the mean print run of the open access titles was 70 per cent of that of the books not published on open access. The difference is reflected in the mean sales of the two groups: the mean sales of open access titles are over 63 per cent of the mean sales of books not on open access. We might conclude that the lower commercial expectations of OA books is reflected in the mean sales. However, the number of open access titles available is larger, and the *total* number of copies sold is also larger: over 65 per cent of all sales.

The main question to answer is whether open access publishing is affecting the sales of monographs and how it compares with other influences

on monograph sales. The results for frontlist sales are clear: no significant effect of open access on sales could be found, after controlling for the effect of print run and language.

The answer for backlist sales is a bit more nuanced. Owing to limitations of the data, it is not possible to run the same procedure as used for the frontlist. Instead the backlist titles were split into four groups based on the commercial expectations. Taking into account this division, we can conclude that open access publishing has no effect on books with a print run of zero or between 2001 and 3000. Moreover, where an effect could be found, the effect size is quite different for categories of print run: a small effect for books with a print run between one and 1000 and a large effect for books with a print run between 2001 and 3000.

In the subsets of books with a print run higher than 0 but below 2000, open access has a negative effect. Nevertheless, the group of books with a print run between one and 1000 is very heavily skewed towards open access books: the data analysed contain over 3.6 times more data items ($N = 509$) for titles available under OA than data items for closed access titles ($N = 136$). The difference between the mean sales is much less: the mean sales of books not on open access is almost 1.5 times the mean sales of open access books in this group. This amounts to a much higher sale of copies of books on open access: almost 21 per cent of the total sales, compared with just over eight per cent for books not on open access. In an economic sense, the negative effect is not very important. The differences in the group of books with a print run between 1001 and 2000 are much more dramatic. But in this case the groups are small (with open access: $N = 61$; without open access: $N = 31$) and the number of books sold is 10 per cent of all sales. It is likely that the influence on revenue will not be very large.

From these data, it is not clear why sales of paper monographs are so lightly affected by free online versions. For a possible answer, we could look again at Snijder (2010) and OAPEN-NL (Ferwerda *et al.*, 2013; Snijder, 2010). There, academic libraries are described as a major purchaser of monographs. As long as availability on open access is not taken into account when paper monographs are acquired, the same pattern keeps emerging. However, we also saw that sales of monographs – whether available on open access or not – are far from soaring. Lack of budget at academic libraries is probably a major factor, as illustrated in (“Association of Research Libraries (ARL) :: ARL Statistics 2009-10,” 2012).

The introduction to this paper discussed the sustainability of the current monograph publication system. From the results in the paper, we can conclude that using a hybrid model or closed access only does not make a large

difference to sales. The model under research does not lead to more sales of open access monographs, and the loss of sales is negligible. The data suggest that a hybrid model in which open access versions are made available in combination with paid-for print versions does not change the status quo. If the status quo is considered to be a broken publication system, a hybrid model is not an option to change things for the better. However, publishers who do well from selling paper monographs could consider making their titles available on open access as a way to enlarge the number of readers. Publishers who are making losses on monographs may want to change their business model in a more radical way than adopting a hybrid model.

3.6 Limitations

The data set used in this paper is large: it contains the sales data of 513 titles sold over a period of three years. Results from a large data set are less prone to be influenced by outliers, which helps to validate the outcomes. Nevertheless, the results are derived from one publisher and this makes it hard to establish whether other aspects – such as reputation or marketing budget – have influenced the results. Owing to the properties of the used sales data, the analysis was carried out on smaller subsets. Further research could establish whether the role of open access publishing in a hybrid model really is so small.

3.7 Acknowledgements

The author would like to thank Managing Director Jan-Peter Wissink of AUP for facilitating the research, and Professor Paul Wouters and Ludo Waltman of the Centre for Science and Technology Studies (CWTS) for commenting on a draft version of the paper.

3.8 Appendix 1: ANOVA results per influence

Table A1.1. ANOVA results per influence

<i>Influence</i>	<i>Results</i>
Open access	There was a significant effect of open access on monograph sales, $F(1, 529.828) = 10.974$, $p = 0.001$, $\omega^2 = 0.01$
Commercial potential	There was a significant effect of print run on monograph sales, $F(5, 11.083) = 16.727$, $p < 0.001$, $\omega^2 = 0.16$
Front and backlist	There was a significant effect of front and backlist on monograph sales, $F(1, 202.781) = 51.016$, $p < 0.001$, $\omega^2 = 0.14$
Language	There was a significant effect of language on monograph sales, $F(1, 660.003) = 17.216$, $p < 0.001$, $\omega^2 = 0.02$

Note: The assumption of homogeneity of variance was violated; therefore, the Welch *F*-ratio is reported for 'open access'; 'commercial potential'; 'front and backlist'; 'language'.

3.9 Appendix 2: Frontlist results

Table A2.1. Frontlist data: mean sales[Q31]

Influence		Mean sales	N	Percentage of total sales
Open access publishing	With	240.15	147	34.5
	Without	227.88	52	11.6
Commercial potential	Print run: 0	109.49	45	4.8
	Print run: 1–1000	179.99	137	24.1
	Print run: 1001–2000	949.46	13	12.1
	Print run: 2001–3000	1305.75	4	5.1
Language	English	158.14	134	20.7
	Dutch	399.4	65	25.4

Multifactor ANOVA

The results convey that the covariate print run was significantly related to sales, $F(1, 195) = 81.651$, $p < 0.001$, partial $\eta^2 = 0.295$. Also, the covariate language was significantly related to sales, $F(1, 195) = 22.577$, $p < .001$, partial $\eta^2 = 0.104$. However, no significant effect of open access on sales could be found after controlling for the effect of print run and language, $F(1, 195) = 2.83$, $p = 0.094$, partial $\eta^2 = 0.014$.

3.10 Appendix 3: Backlist results

Table A3.1. Backlist data: mean sales

Influence		Mean sales	N	Percentage total sales
Open access publishing	With	36.8	867	31.2
	Without	82	281	22.5
Commercial potential	Print run: 0	18.21	385	6.8
	Print run: 1–1000	46.17	645	29.1
	Print run: 1001–2000	113.61	92	10.2
	Print run: 2001–3000	283.65	20	5.5
	Print run: 3001–4000	324	3	0.9
	Print run: 4001–5000	446	3	1.3
Language	English	37.93	695	25.7
	Dutch	63.73	453	28.2

Multifactor ANOVA

Using the same procedure as deployed on the frontlist leads to the following result: the covariate print run was significantly related to sales, $F(1, 1144) = 234.618$, $p < 0.001$, partial $\eta^2 = 0.17$. However, the covariate language was not significantly related to sales, $F(1, 1144) = 2.17$, $p = 0.141$, partial $\eta^2 = 0.002$. Open Access has a significant effect on sales after controlling for the effect of print run and language, $F(1,1144) = 27.948$, $p < 0.001$, partial $\eta^2 = 0.024$.

Table A3.2. Backlist data: commercial potential

Commercial potential	Results
Print run: 0	No significant effect of open access on monograph sales could be found, $F(1, 126.225) = 1.25$, $p = 0.291$, $\omega^2 = 0.00$
Print run: 1–1000	Open access had a significant negative effect on monograph sales, $F(1, 179.348) = 7.364$, $p = 0.007$, $\omega^2 = 0.01$
Print run: 1001–2000	Open access had a significant negative effect on monograph sales, $F(1, 36.510) = 9.795$, $p = 0.003$, $\omega^2 = 0.13$
Print run: 2001–3000	No significant effect of open access on monograph sales could be found, $F(1, 18) = 0.449$, $p = 0.511$, $\omega^2 = 0.00$

Note: The assumption of homogeneity of variance was violated for 'Print run: 0', 'Print run: 1–1000', and 'Print run: 1001–2000'; therefore, the Welch *F*-ratio is reported.