OPEN ACCESS 2020 DE

Nationaler Open-Access-Kontaktpunkt

APCs – Mirroring the impact factor or legacy of the subscription-based model?

3rd ESAC Workshop „On the Effectiveness of APCs“
Munich, 28.06.2018

Dr. Nina Schönfelder

OPEN ACCESS 2020 DE

Nationaler Open-Access-Kontaktpunkt

Agenda

- Background
- Data
- Method
- Results
- Limitations and potential weaknesses
- Conclusion

OPEN ACCESS 2020 DE

Nationaler Open-Access-Kontaktpunkt

NOAK Working Package 4

- Analysing financial flows, shaping financial models, and consultation with funders

OPEN ACCESS 2020 DE

Nationaler Open-Access-Kontaktpunkt

Aim

- Analysing the determinants for APC-levels
- Projecting APCs for currently closed-access journals
- Comparing projected total APC-spending with libraries budgets‘ for each German university/research institute after a hypothetical full journal flipping
- Similar approach as in the “Pay It Forward”-study conducted at the University of California Libraries

Data

- OpenAPC data set (part of the INTACT project at the Bielefeld University Library, Germany)
  - APCs actually paid (in contract to catalogue prices)
  - country, period, journal type (hybrid/oa), journal title, pubisher
- CWTS Journal Indicators (calculated by Leiden University's Centre for Science and Technology Studies based on the Scopus bibliographic database produced by Elsevier)
  - "source normalized impact per paper" (SNIP)
  - subject area of the journal

OPEN ACCESS 2020 DE

Nationaler Open-Access-Kontaktpunkt

Summary statistics

country

```
GBR    :24572
DEU    :14054
AUT    : 4244
SWE    : 1532
NOR    : 1171
CAN    :  929
(Other): 1240
```

institution

```
UCL                           : 4526
FWF - Austrian Science Fund: 4205
Wellcome Trust                : 3782
MPG                           : 3465
University of Cambridge       : 2044
University of Oxford          : 1506
(Other)                       :28214
```

period

```
2016   :16210
2015   :12892
2014   :11178
2013   : 3253
2012   : 1472
2017   :  905
(Other): 1832
```

publisher

```
Elsevier BV                          : 6838
Springer Nature                      : 6484
Public Library of Science (PLoS)     : 5690
Wiley-Blackwell                      : 4265
Springer Science + Business Media: 3627
Frontiers Media SA                   : 2718
(Other)                              :18120
```

journal_full_title

```
PLOS ONE                  : 4789
Scientific Reports        : 1388
New Journal of Physics    :  983
Frontiers in Psychology:  680
Nature Communications     :  630
BMJ Open                  :  437
(Other)                   :38835
```

is_hybrid

```
Mode :logical
FALSE:26755
TRUE :20987
```

SNIP

```
Min.    : 0.000
1st Qu.: 1.050
Median : 1.230
Mean   : 1.435
3rd Qu.: 1.620
Max.   :15.870
NA's   :5013
```

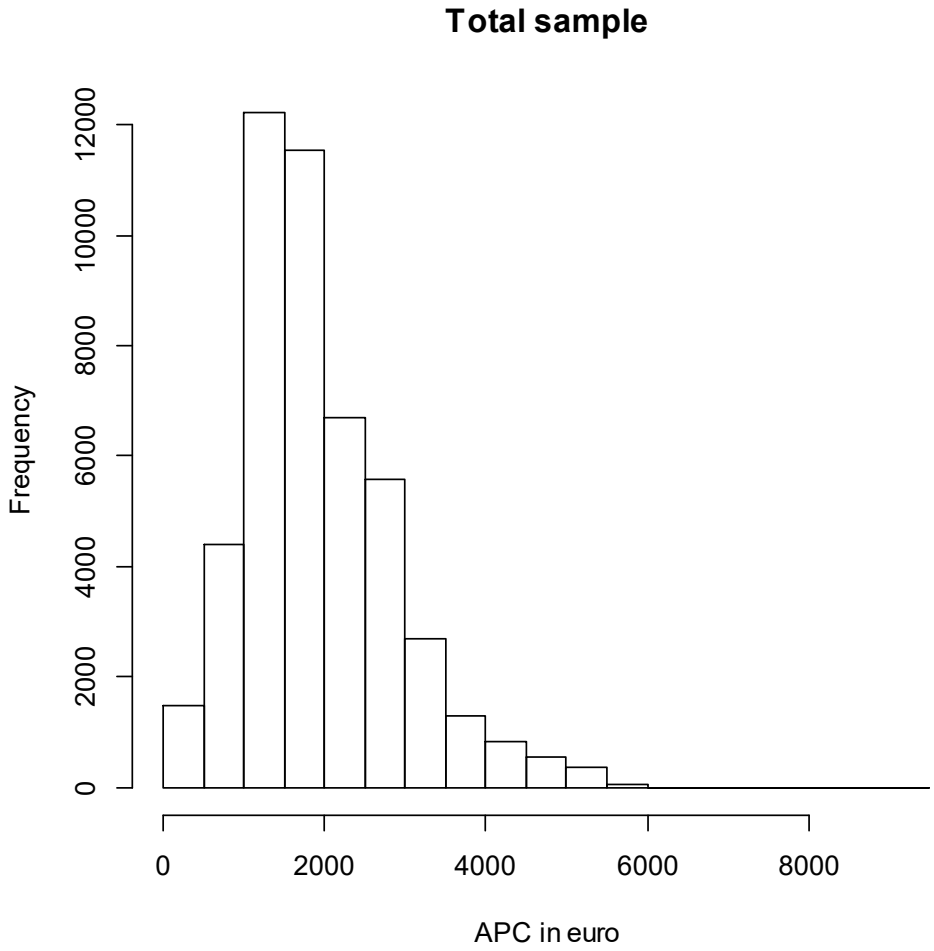
Subject.area

```
Health Sciences              :10616
Life Sciences                :20312
Physical Sciences            : 9462
Social Sciences & Humanities: 2339
NA's                         : 5013
```

euro

```
Min.   :  40
1st Qu.:1255
Median :1738
Mean   :1924
3rd Qu.:2450
Max.   :9079
```

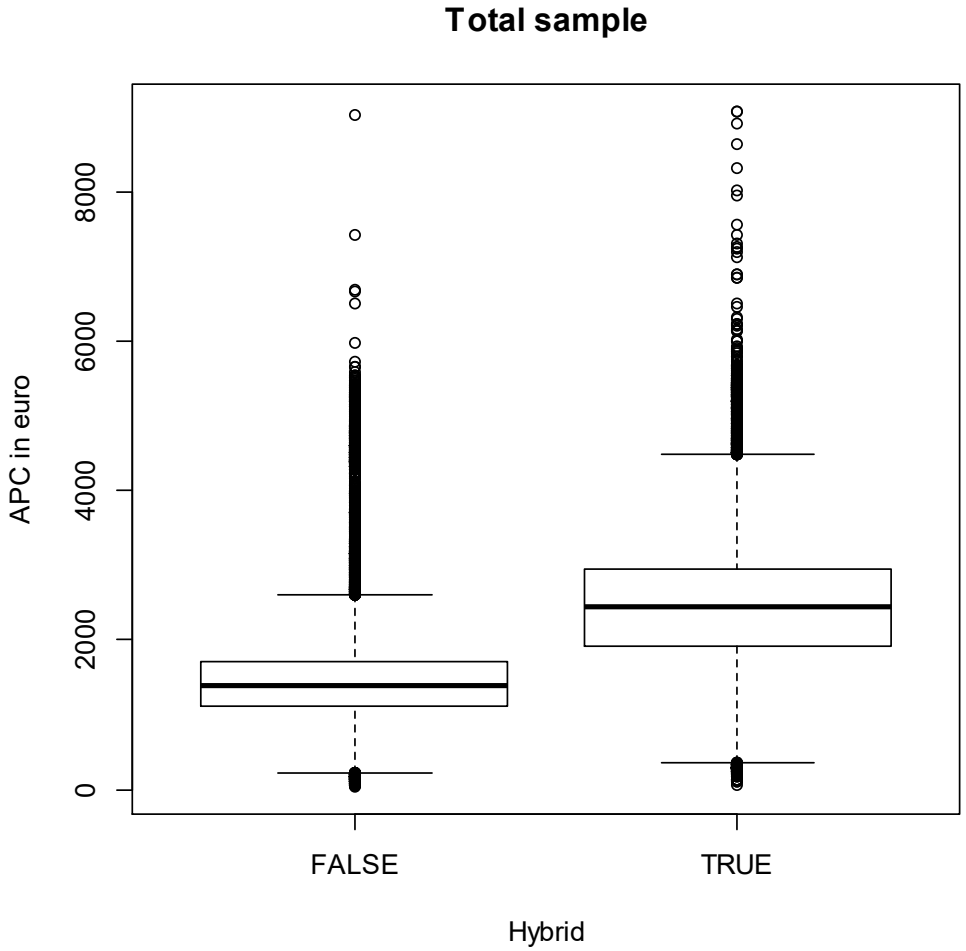
Histogram of APC in euro



Article Processing Charges

- range mostly between 1,000 – 3,000 EUR.
- but amount sometimes to 4,000 – 6,000 EUR.

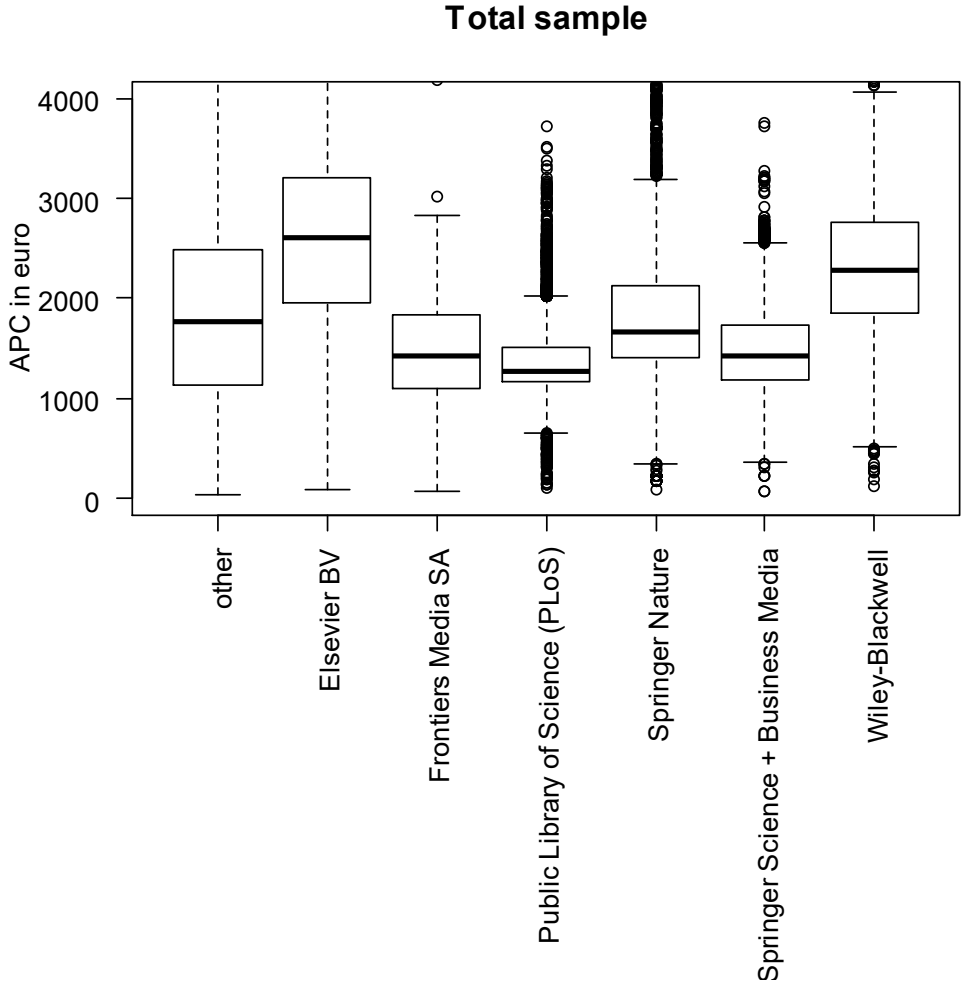
Box plots of APC in euro for OA and hybrid journals



Article Processing Charges

- range mostly between 1,000 – 3,000 EUR.
- but amount sometimes to 5,000 – 6,000 EUR.
- are (on average) more expensive in hybrid journals.

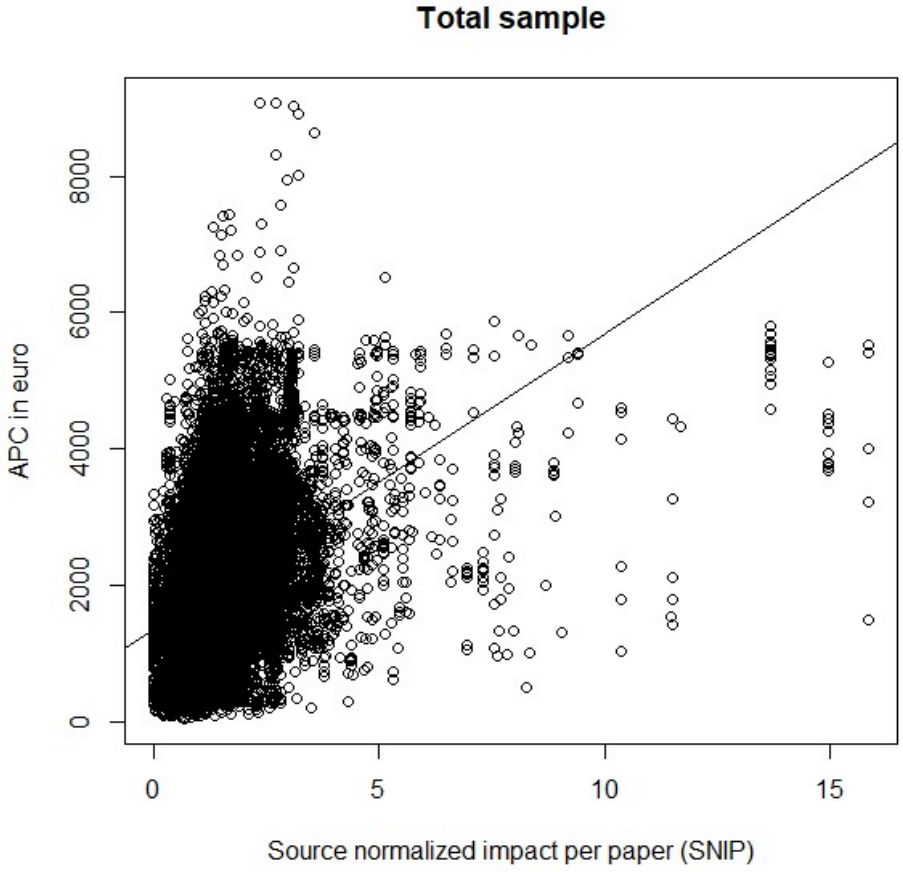
Box plots of APC depending on publisher



Article Processing Charges

- range mostly between 1,000 – 3,000 EUR.
- but amount sometimes to 5,000 – 6,000 EUR.
- are (on average) more expensive in hybrid journals.
- are quite different depening on publisher.

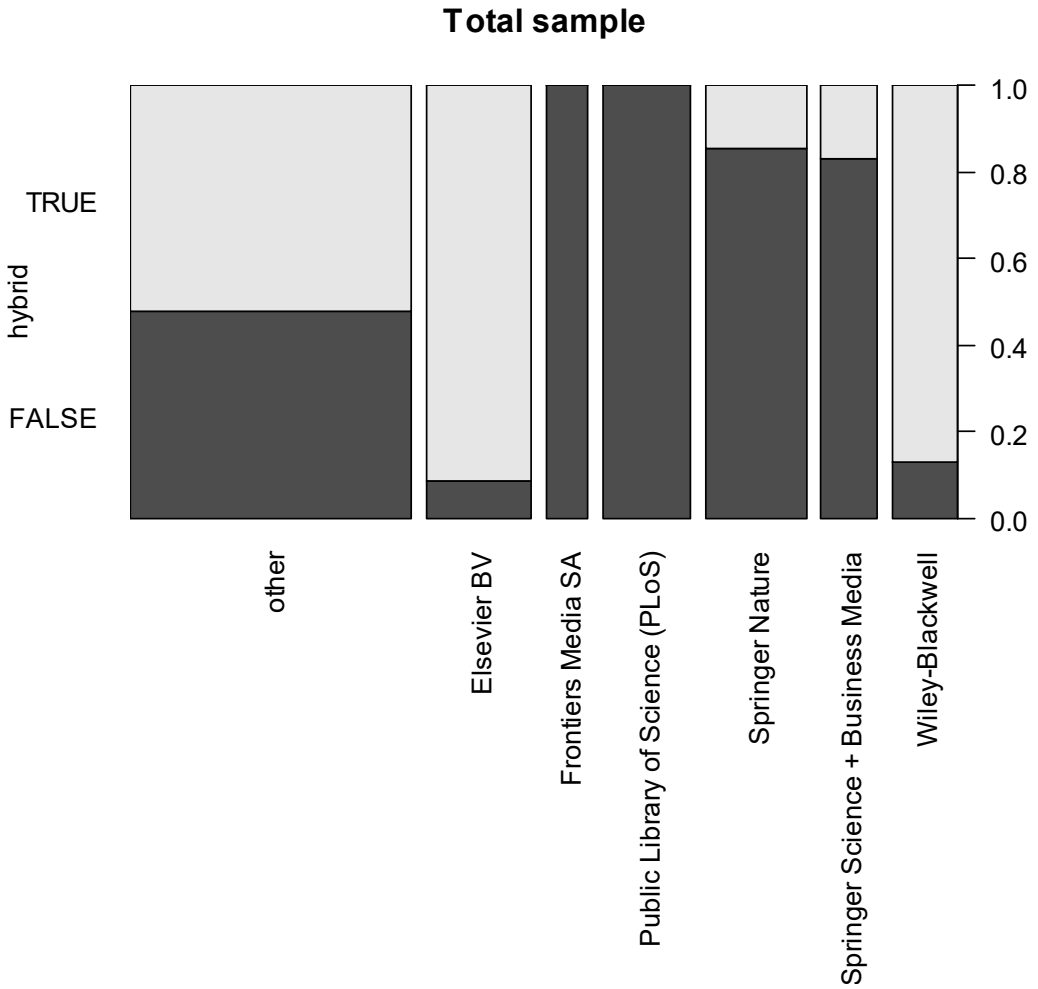
Scatter plot of APC vs. SNIP



Article Processing Charges

- range mostly between 1,000 – 3,000 EUR.
- but amount sometimes to 5,000 – 6,000 EUR.
- are (on average) more expensive in hybrid journals.
- are quite different depending on publisher.
- are related to the impact factor (SNIP).

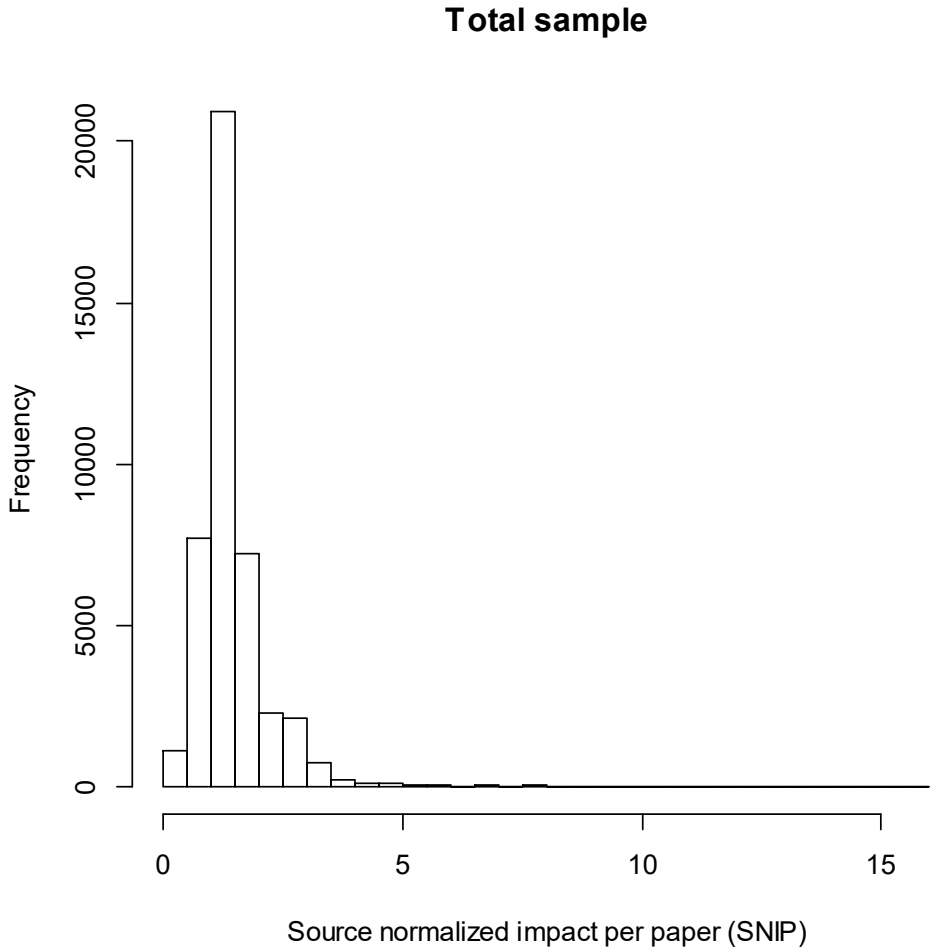
Share of articles published in OA-/hybrid journals



Articles are published

- often in hybrid journals at Elsevier and Wiley-Blackwell.
- often in open-access journals at Springer and Nature.

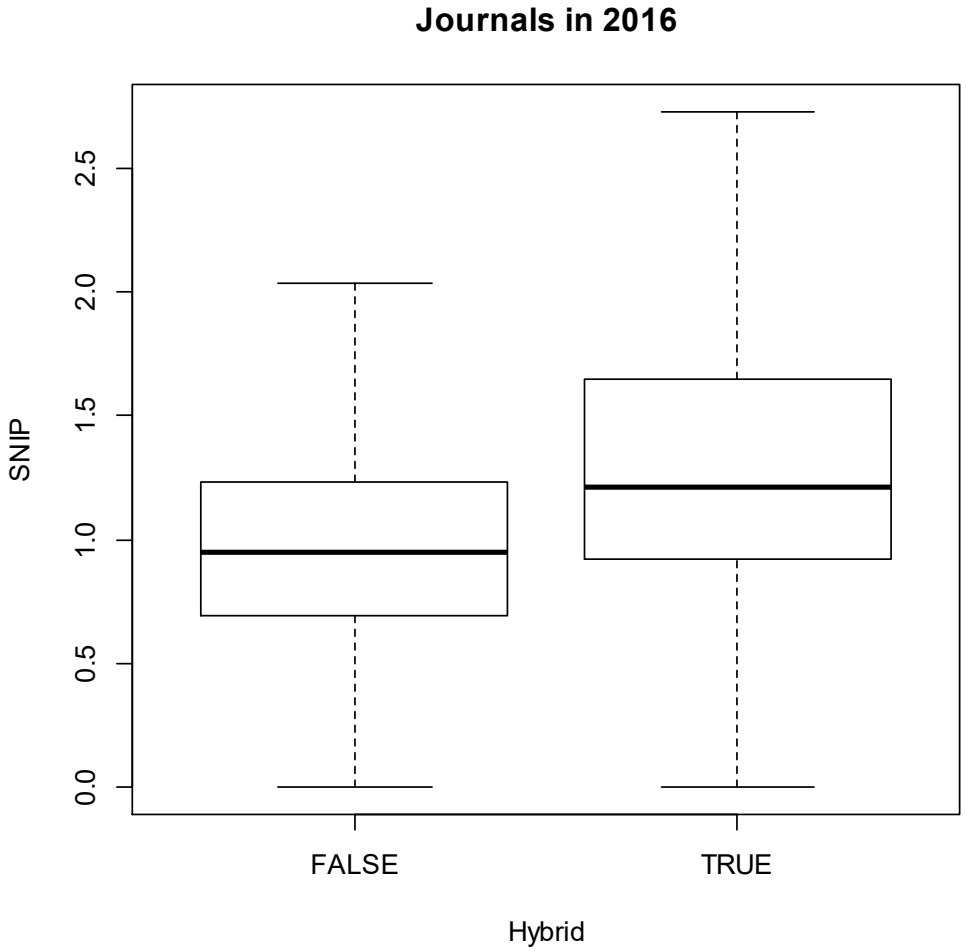
Histogram of SNIP for openAPC-records



Articles are published

- often in hybrid journals at Elsevier and Wiley-Blackwell.
- often in open-access journals at Springer and Nature.
- rarely in high-impact journals.

Box plots of SNIP for open-access or hybrid journals



Journals that are

- open access tend to have lower impact.
- hybrid tend to have higher impact.

Method and statistical model

- Static linear regression with random and time effects based on *T* successive cross sections

$$APC_{it} = \alpha_i + \beta_1 SNIP_{it} + \beta_2 Hybrid_{it} + \beta_3 SNIP_{it} \times Hybrid_{it} + \mathbf{Big\_publisher}'_{it}\boldsymbol{\beta}_4 + \mathbf{Subject\_area}'_{it}\boldsymbol{\beta}_5 + \gamma_t + \epsilon_{it}$$

- Ordinary least squares (OLS)
- Heteroscedasticity-robust standard errors
- Software: R
- Sub-sample: UK, 2014–2016,
  without outliers (1%-quantile< X <99%-quantile)

OPEN ACCESS 2020 DE

Nationaler Open-Access-Kontaktpunkt

Results

	Model 1	Model 2	Model 3	Model 4
(Intercept)	1797.19 (19.95)***	1800.70 (10.39)***	727.92 (40.98)***	519.38 (40.96)***
SNIP	320.42 (12.98)***		788.60 (31.82)***	728.07 (29.74)***
is_hybrid		702.61 (12.42)***	1475.81 (43.96)***	1395.93 (43.07)***
SNIP:is_hybrid			−603.29 (33.19)***	−539.69 (31.32)***
Elsevier BV				225.06 (15.76)***
Frontiers Media SA				−114.05 (31.03)***
Public Library of Science (PLoS)				−328.48 (20.28)***
Springer Nature				235.59 (22.34)***
Springer Science + Business Media				145.00 (20.60)***
Wiley-Blackwell				−29.11 (15.19)*
Life Sciences				179.48 (13.62)***
Physical Sciences				−146.77 (15.10)***
Social Sciences and Humanities				−374.95 (26.47)***
period 2015				312.13 (14.28)***
period 2016				283.40 (13.45)***
R^2	0.10	0.12	0.24	0.31
Adj. R^2	0.10	0.12	0.24	0.31
Num. obs.	22310	23818	22310	22310
RMSE	888.05	878.87	818.79	777.41

***p < 0.01, **p < 0.05, *p < 0.1

Results

	Model 4
(Intercept)	519.38 (40.96)***
SNIP	728.07 (29.74)***
is_hybrid	1395.93 (43.07)***
SNIP:is_hybrid	−539.69 (31.32)***
Elsevier BV	225.06 (15.76)***
Frontiers Media SA	−114.05 (31.03)***
Public Library of Science (PLoS)	−328.48 (20.28)***
Springer Nature	235.59 (22.34)***
Springer Science + Business Media	145.00 (20.60)***
Wiley-Blackwell	−29.11 (15.19)*
Life Sciences	179.48 (13.62)***
Physical Sciences	−146.77 (15.10)***
Social Sciences and Humanities	−374.95 (26.47)***
period 2015	312.13 (14.28)***
period 2016	283.40 (13.45)***
R^2	0.31
Adj. R^2	0.31
Num. obs.	22310
RMSE	777.41

***p < 0.01, **p < 0.05, *p < 0.1

Base groups

- Year: 2014
- Publisher: other / smaller
- Subject area: health sciences
- Journal type: open access

Equation

- for PLoS-articles in life sciences in 2016

$$APC = (519 - 328 + 179 + 283) + 728 \times SNIP$$

$$APC = 653 + 728 \times SNIP$$

- for Elsevier hybrid-journal, else as above

$$APC = (519 + 225 + 179 + 283 + 1{,}396) + (728 - 540) \times SNIP$$

$$APC = 2{,}602 + 188 \times SNIP$$

Example for in-sample prediction (SNIP=1, life sciences, 2016)

- PLOS ONE article in life sciences in 2016

$$APC = 653 + 728 = 1{,}381\ €$$

- Article in “Journal of Neuroscience Methods” (Elsevier hybrid-journal)

$$APC = 2{,}602 + 188 = 2{,}790\ €$$

Limitations and potential weaknesses

- Sample selection
  - openAPC possibly not a representative sample
  - due to selective recoding and reporting behavior
  - e.g. due to price cap in Germany (2,000 €)
  - ➢ Use UK sub-sample!
- Missing data
  - missing SNIP-values for 6-11% of the observations (and missing subject area)
  - related to coverage of Scopus
  - related to maturity of a journal
  - ➢ “Worst-case” analysis shows that the bias is not severe.

Limitations and potential weaknesses

- Sample selection
  - openAPC possibly not a representative sample
  - due to selective recoding and reporting behavior
  - e.g. due to price cap in Germany (2,000 €)
  - ➢ Use UK sub-sample!
- Missing data
  - missing SNIP-values for 6-11% of the observations (and missing subject area)
  - related to coverage of Scopus
  - related to maturity of a journal
  - ➢ “Worst-case” analysis shows that the bias is not severe.

Conclusion

- APCs – Mirroring the impact factor?
  - In open-access journals!
  - At genuine open-access publishers!
- APCs – Legacy of the subscription-based model?
  - In hybrid journals!
  - Often at Elsevier, Springer and co.!

OPEN ACCESS 2020 DE

Nationaler Open-Access-Kontaktpunkt

Questions?

Dr. Nina Schönfelder
National Open-Access Contact Point OA2020-DE
Bielefeld University Library
Universitätsstr. 25 | D-33615 Bielefeld
Tel.: +49 (0) 521/106-2546 | email: nina.schoenfelder@uni-bielefeld.de

www.oa2020-de.org
@oa2020de