Time trends and risk-factors in publication bias

Julia Jerke & Heiko Rauhut
University of Zurich, Institute of Sociology

Seminar Analytical Sociology
Venice International University, San Servolo
November 20 - 23, 2017
Background
Terminology

What is publication bias?

“[…] publication bias occurs when the probability that a result is published depends on the estimates produced by the study, holding the methodological quality of the study fixed.”

(Gerber & Malhotra 2008)

– publication of a manuscript depends on the therein reported results
– tendency to publish papers reporting significant and positive results only
Terminology
Two mechanisms

1. Publication bias
   - Publishers rejecting manuscripts with insignificant respectively negative results (*selection effect*)
   - Researchers not submitting manuscripts with insignificant respectively negative results (*filedrawer effect*)

2. Manipulation bias
   - trimming of data and results to achieve statistical significance
   - p-hacking: unjustified restriction to certain model or subgroup specifications
   - optional stopping: sampling until significance is achieved
   - fabrication or manipulation of data

Consequences
   - Biased research literature
   - **Overrepresentation of significant results** (OSR)
Theoretical embedding

- Editors'/reviewers' beliefs about success of p<α manuscripts
- Authors' beliefs about impact of p<α manuscripts
- Authors' beliefs about editors'/reviewers' rejection decisions for p<α manuscripts

Editors/reviewers favor p<α manuscripts

Authors are less likely to submit p<α manuscripts

Publication bias
- (Selection and filedrawer effect)

Overrepresentation of significant results (OSR)

Manipulation bias
- Data tweaking (questionable research practices)
- Fabrication (severe misconduct)
Some findings

– reviewers seem to favor manuscripts with positive results (Mahoney 1977, Dickersin et al. 1992)
– studies with negative results are less likely to be published (Turner et al. 2008, Franco et al. 2014)
– published manuscripts suffer from inconsistencies regarding the reported test results (Nuijten et al. 2016, Leggett at al. 2013)
– several prominent cases of data fabrication/manipulation demonstrate that severe scientific misconduct is a real problem
– overall, negative results seem to disappear from the scientific literature (Fanelli 2012)
The research project
Main research questions

1. Is there a publication bias in a leading journal of economics?
2. If so, did the publication bias increase over time?
3. What influences the magnitude of publication bias?
Methodology

The caliper test (Gerber & Malhotra 2008)

Using the principle of regression discontinuity:

- the $\alpha$ levels (e.g. 5% or 1%) are in fact arbitrary chosen values
- results just below and just above the $\alpha$ levels should be equally likely

“[…]comparing the number of observations in equal-sized intervals just below and just above the threshold value for statistical significance. If there are an unusually large number of observations just over the critical value, this is taken as evidence of publication bias.” (Gerber & Malhotra 2008)

→ focus on z-/t-values
→ compares the occurrence of test values just above and just below the critical threshold (in the case of the normal distribution: 1.96 for $\alpha=0.05$ and 2.56 for $\alpha=0.01$)
Methodology
The caliper test (Gerber & Malhotra 2008)

– in absence of publication bias:
  the frequency of reported estimates just below the critical significance level should equal the frequency of reported results just above the critical significance level

– overrepresentation of estimates in the interval just above the critical threshold is then assumed to be an evidence for publication bias (most likely due to manipulation)
Methodology
Example for the caliper test

(Gerber, A. S., Malhotra, G., 2008. Publication Bias in Empirical Sociological Research: Do Arbitrary Significance Levels Distort Published Results?)
Data collection

Sample

- The Quarterly Journal of Economics (impact factor 6.654)
- 1960 to 2013 (articles $N \approx 2700$)
- criteria of inclusion
  - quantitative article reporting an empirical study
  - must rely on either explicit or implicit hypothesis
  - inferential statistics
Data collection

Process of data collection

– Screening of all eligible articles
– extracting z-or t-values, respectively
– Classification into over- and under-caliper
  – over-caliper: values in a narrow interval just above the critical threshold
  – under-caliper: values in a narrow interval just below the critical threshold
  – narrow means: x% below or above the threshold (e.g. x=5, 10, or 15)
– further information collected: number of authors, experiment vs. study, explicit vs. implicit hypothesis, sample size, number of coefficients per paper, funding, number of citations
Working hypotheses

Increasing publication pressure may drive researchers to engage in QRP

H1: Overrepresentation of significant results increased over time (time)

H2: Funding of study may motivate authors to present positive results, thus increasing OSR (funding)

Author group size may have two opposing effects: 1) more social control in larger teams; 2) diffusion of responsibility in larger teams

H3b: The higher the number of authors, the lower the risk of questionable research practices, resulting in lower OSR (social control)

H3a: The higher the number of authors, the higher the risk of questionable research practices, resulting in lower OSR (diffusion)
Working hypotheses

Several study characteristics may either facilitate or complicate mild forms of data tweaking

H4: The more coefficients were tested, the lower the risk of ORS (coefficients)

H5: Experiments facilitate moderate data tweaking, potentially resulting in higher ORS (experiment)
Composition of the QJoE over time

- no editorial shifts around 1990
- seems to document the shift towards experimental and empirical studies in economics
Are negative results really disappearing?

Levels of significance

Results of separate logistic regressions for $\alpha = 0.01/0.05/0.1$
(DV: significant; 1-yes 0-no)

- no time trend
- significant negative effect of number of coefficients
- tendency towards a slightly positive effect of funding (but not significant on 5%)
Distribution of empirical z scores
Distribution of empirical z scores

1960 - 1989

1990 - 2013

Density

z_rekursiv

0.16 1.06 1.96 2.86 3.76 4.66

0.16 1.06 1.96 2.86 3.76 4.66
### Distribution of empirical z scores

<table>
<thead>
<tr>
<th>5%-significance level</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OC</td>
<td>UC</td>
<td>p value</td>
</tr>
<tr>
<td>1960 - 2013</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5%-caliper</td>
<td>18</td>
<td>24</td>
<td>0.86</td>
</tr>
<tr>
<td>10%-caliper</td>
<td>45</td>
<td>43</td>
<td>0.458</td>
</tr>
<tr>
<td>15%-caliper</td>
<td>63</td>
<td>60</td>
<td>0.428</td>
</tr>
<tr>
<td>1990 - 2013</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5%-caliper</td>
<td>109</td>
<td>90</td>
<td>0.101</td>
</tr>
<tr>
<td>10%-caliper</td>
<td>195</td>
<td>167</td>
<td>0.078</td>
</tr>
<tr>
<td>15%-caliper</td>
<td>280</td>
<td>239</td>
<td>0.04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1%-significance level</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OC</td>
<td>UC</td>
<td>p value</td>
</tr>
<tr>
<td>1960 - 2013</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5%-caliper</td>
<td>20</td>
<td>17</td>
<td>0.371</td>
</tr>
<tr>
<td>10%-caliper</td>
<td>41</td>
<td>41</td>
<td>0.544</td>
</tr>
<tr>
<td>15%-caliper</td>
<td>57</td>
<td>63</td>
<td>0.739</td>
</tr>
<tr>
<td>1990 - 2013</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5%-caliper</td>
<td>95</td>
<td>90</td>
<td>0.384</td>
</tr>
<tr>
<td>10%-caliper</td>
<td>185</td>
<td>203</td>
<td>0.833</td>
</tr>
<tr>
<td>15%-caliper</td>
<td>260</td>
<td>314</td>
<td>0.989</td>
</tr>
</tbody>
</table>
Distribution of empirical z scores over time

Proportion of test statistics in the OC for the 5% level of significance
Caliper width 15%
Distribution of empirical z scores

Proportion of test statistics in the OC for the 1% level of significance
Caliper width 15%
Influencing factors

- restriction to 1990 - 2013
- logistic regression with DV: indicator for over-caliper
  1 - over-caliper
  0 - under-caliper
- independent variables:
  - funding (H2)
  - author group size (H3a/b) (one/two authors vs. three or more authors)
  - number of coefficients (H4)
  - type of hypothesis (explicit vs. implicit)
- different regressions for $\alpha = 0.05/0.01$ and the caliper sizes
- results, in a nutshell
  - no effects
  - except for author group size!
Influencing factors

Author group size effect

logistic regression coefficients
team size on publication bias, 5%

logistic regression coefficients
team size on publication bias, 1%

-1  -0.5  0   0.5   1

team effect

-0.6  -0.4  -0.2  0

team effect
Influencing factors

Author group size effect
Conclusion and discussion
Summary

- no (continuous) time trend observable, neither for levels of significance in general nor publication bias
- nevertheless, comparing the distribution of z-score before and after 1990 suggests mild publication bias after 1990
- the magnitude of publication bias does not seem to be influenced by funding, type of hypothesis, number of coefficients
- however, author group size seems to matter, suggesting that larger research teams are more prone to publication bias

Prospect

- further analyses
- experiment vs. empirical study
- citation rates
- effect sizes
- etc.
Literature


Fanelli, Daniele. 2012. Negative results are disappearing from most disciplines and countries. Scientometrics 90: 891-904.


