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

University of Zurich[UZH]

Institute of Sociology

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

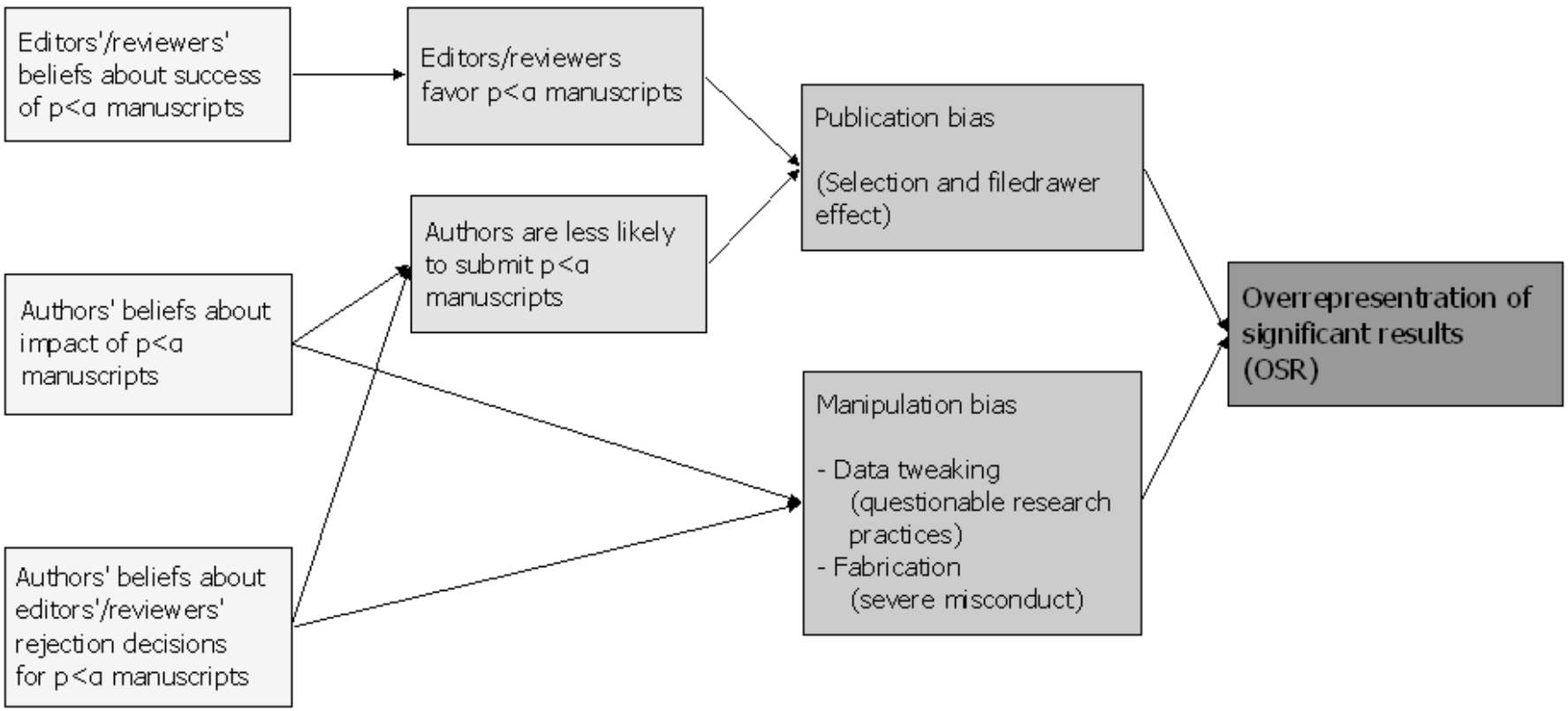
Two mechanisms

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



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)

University of Zurich[UZH]

Institute of Sociology

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 α levels (e.g. 5% or 1%) are in fact arbitrary chosen values
- results just below and just above the α 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)

$\rightarrow$ focus on z-/t-values

$\rightarrow$ compares the occurrence of test values just above and just below the critical threshold (in the case of the normal distribution: 1.96 for α=0.05 and 2.56 for α=0.01)

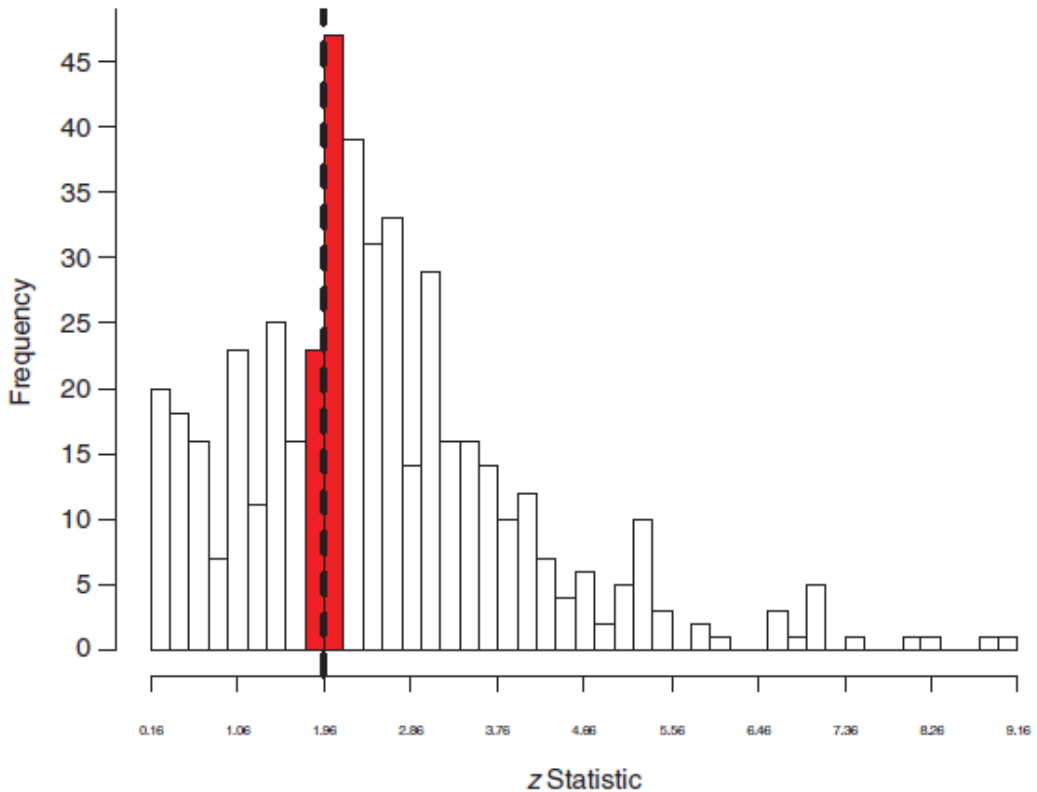
Methodology

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 coeffcients were tested, the lower the risk of ORS (*coefficients*)

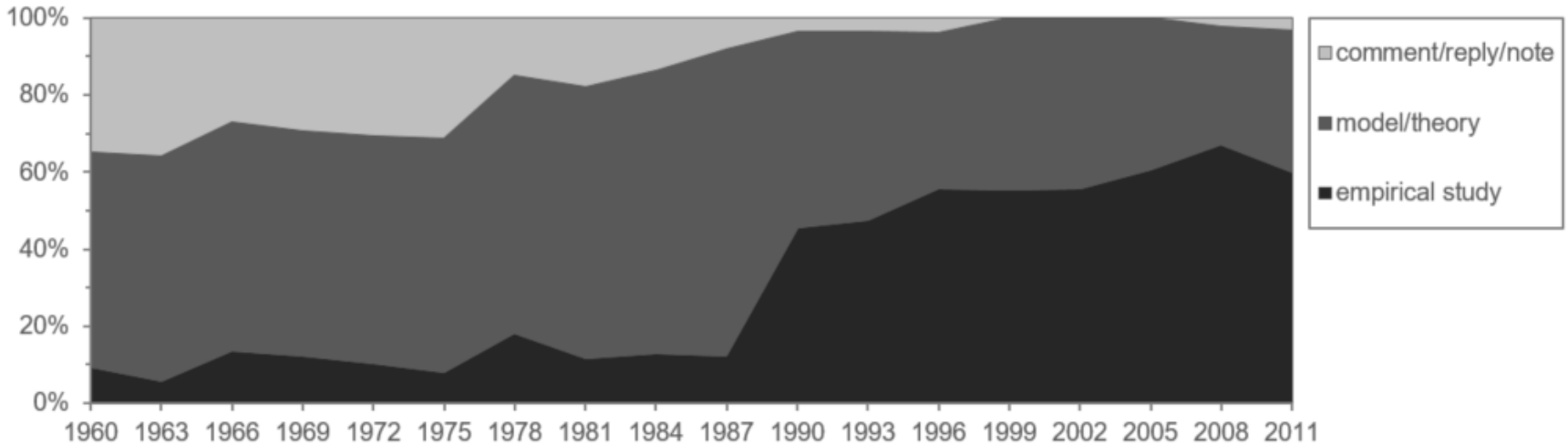
H5: Experiments facilitate moderate data tweaking, potentially resulting in higher ORS (*experiment*)

University of Zurich[UZH]

Institute of Sociology

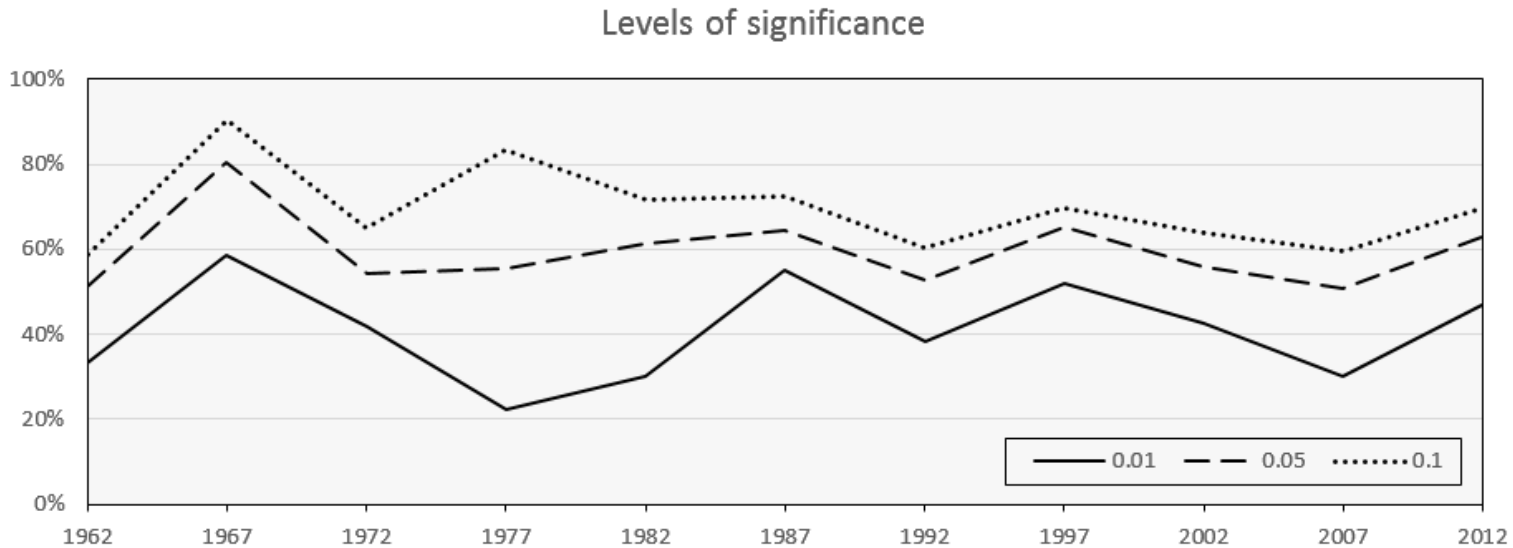
First results

Composition of the QJoE over time



- no editorial shifts around 1990
- seems to document the shift towards experimental and empirical studies in economics

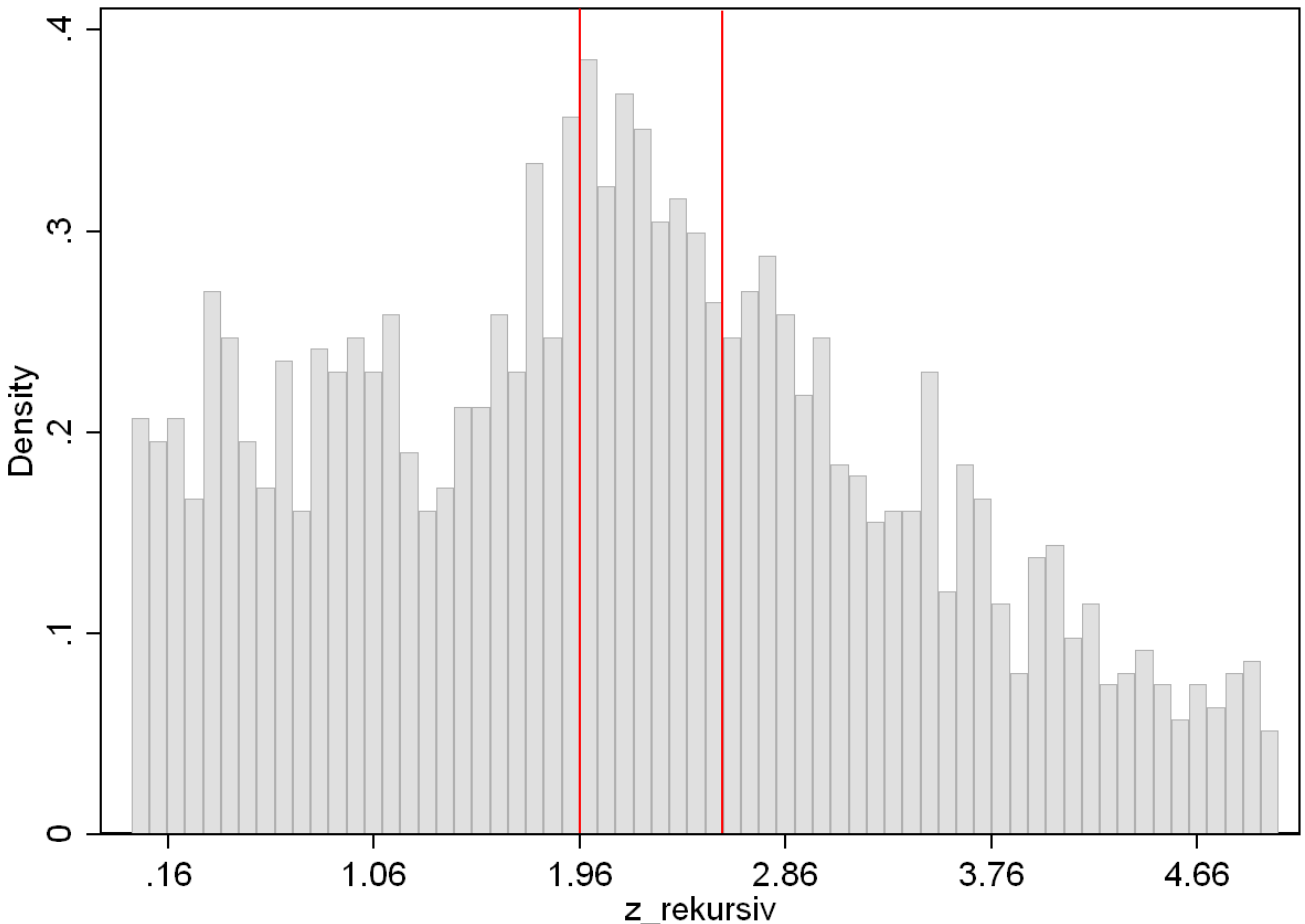
Are negative results really disappearing?



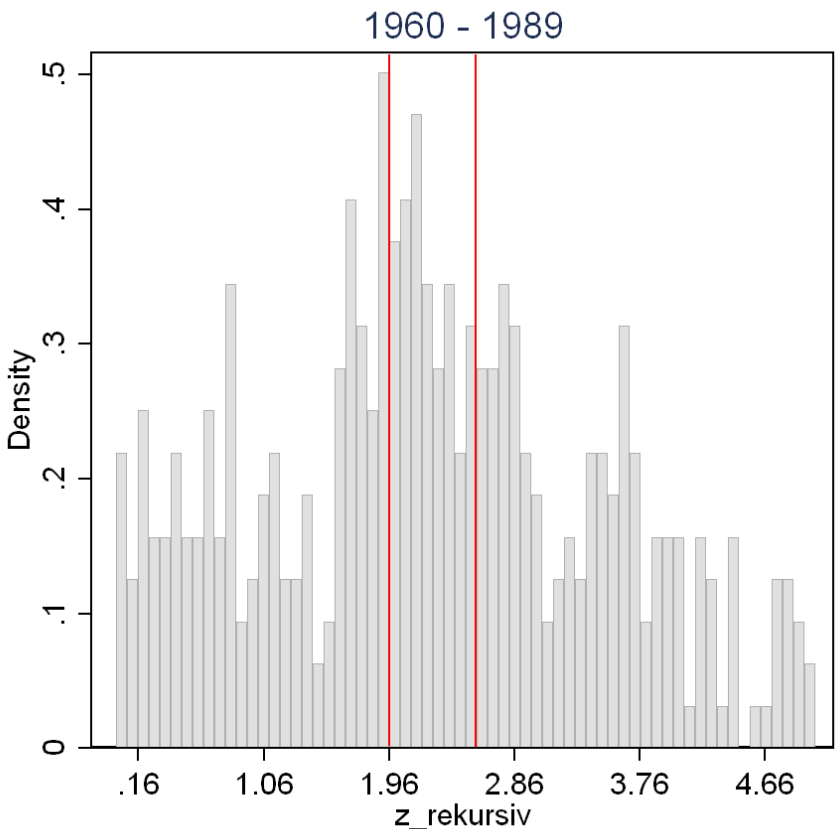
Results of separate logistic regressions for $\alpha = 0.01/0.05/0.1$ (DV: significant; 1-yes 0-no)

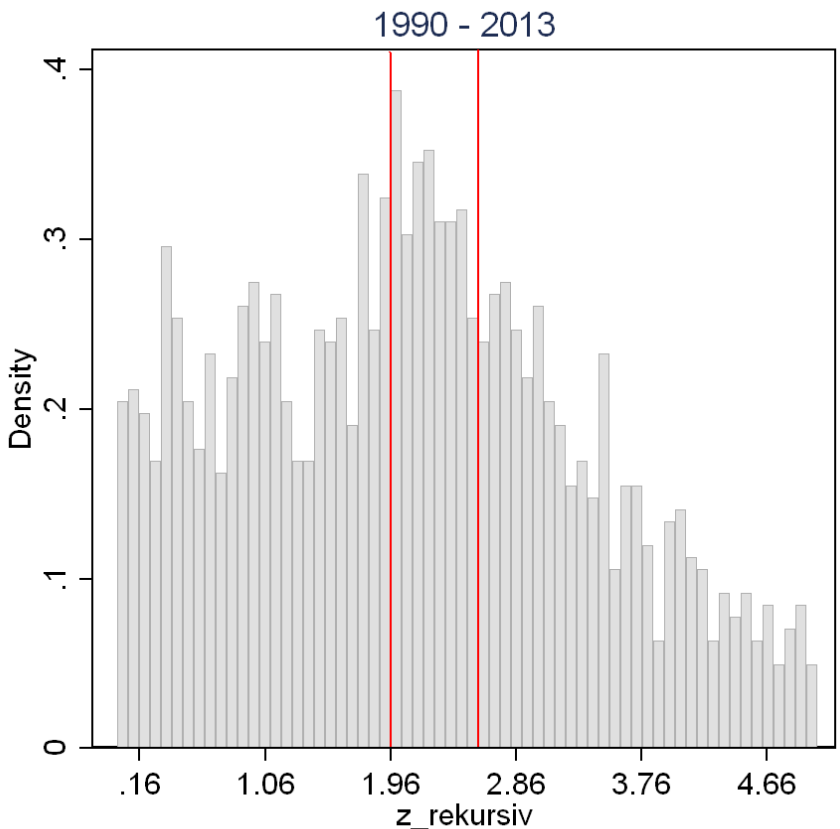
- no time trend
- significant negative effect of number of coefficents
- tendency towards a slightly positive effect of funding (but not significant on 5%)

Distribution of empirical z scores



Distribution of empirical z scores





Distribution of empirical z scores

5%-significance level

		OC		UC		p value
1960 - 2013	5%-caliper	18	42.9%	24	57.1%	0.86
	10%-caliper	45	51.1%	43	48.9%	0.458
	15%-caliper	63	51.2%	60	48.8%	0.428
1990 - 2013	5%-caliper	109	54.8%	90	45.2%	0.101
	10%-caliper	195	53.9%	167	46.1%	0.078
	15%-caliper	280	53.9%	239	46.1%	0.04

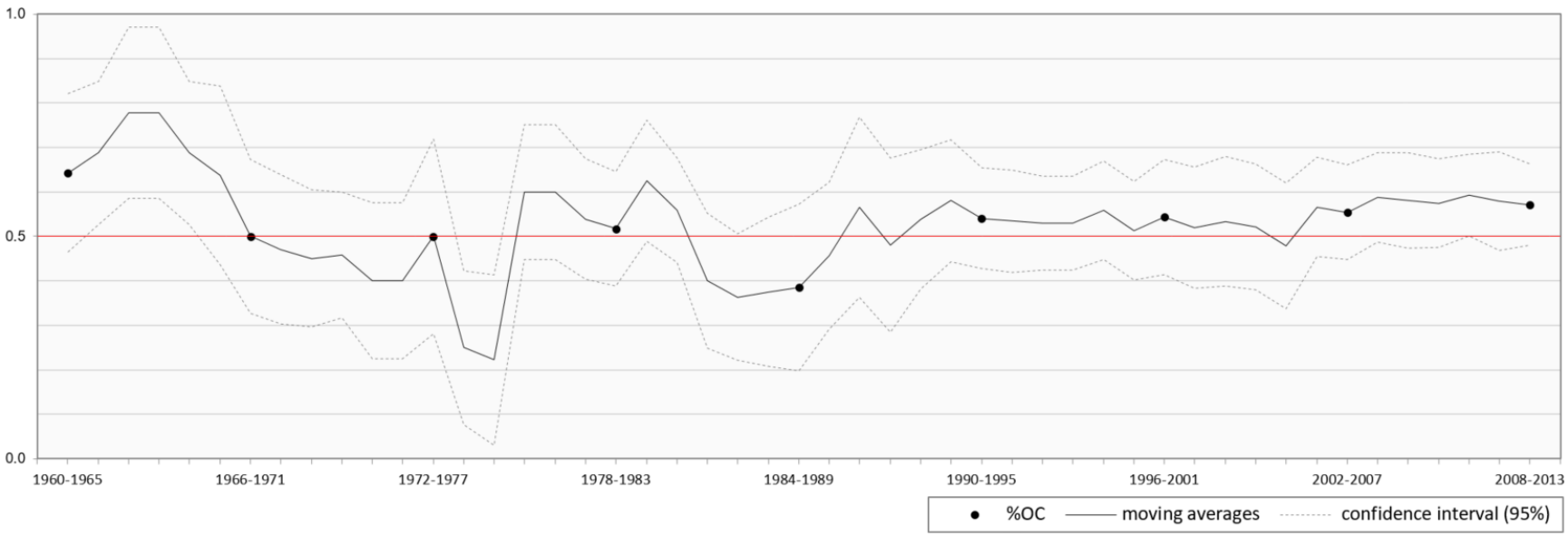
1%-significance level

		OC		UC		p value
1960 - 2013	5%-caliper	20	54.1%	17	45.9%	0.371
	10%-caliper	41	50.0%	41	50.0%	0.544
	15%-caliper	57	47.5%	63	52.5%	0.739
1990 - 2013	5%-caliper	95	51.4%	90	48.6%	0.384
	10%-caliper	185	47.7%	203	52.3%	0.833
	15%-caliper	260	45.3%	314	54.7%	0.989

Distribution of empirical z scores over time

Proportion of test statistics in the OC for the 5% level of significance

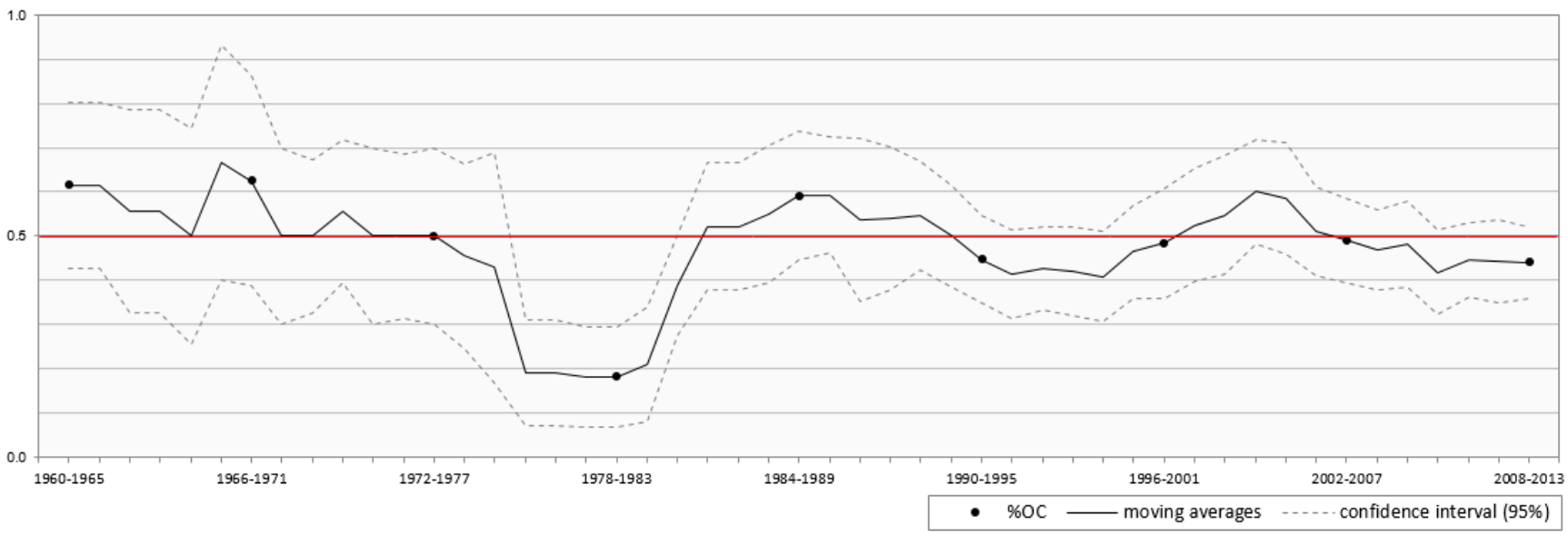
Caliper width 15%



Distribution of empirical z scores

Proportion of test statistics in the OC for the 1% level of significance

Caliper width 15%

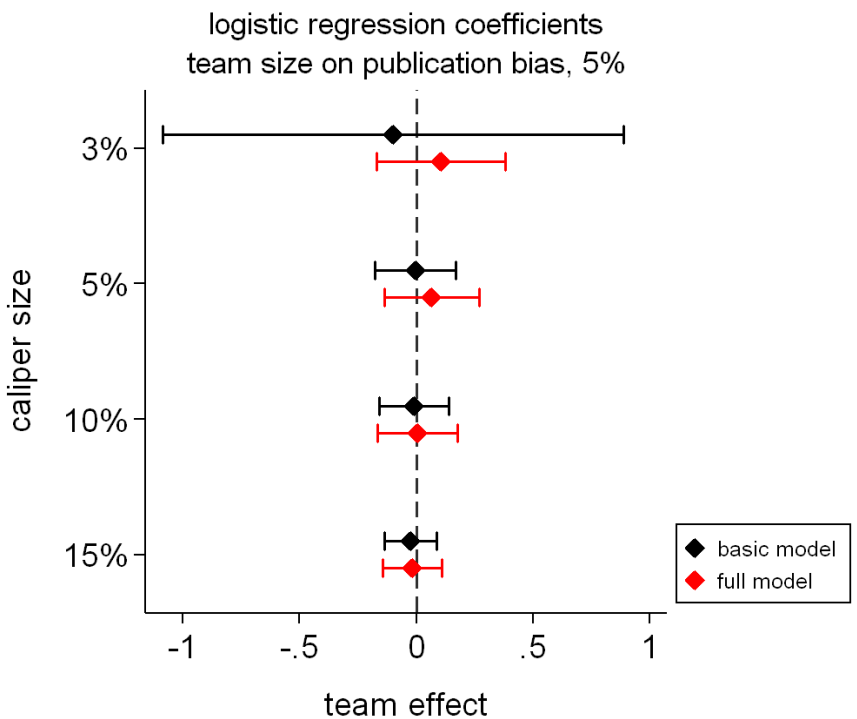


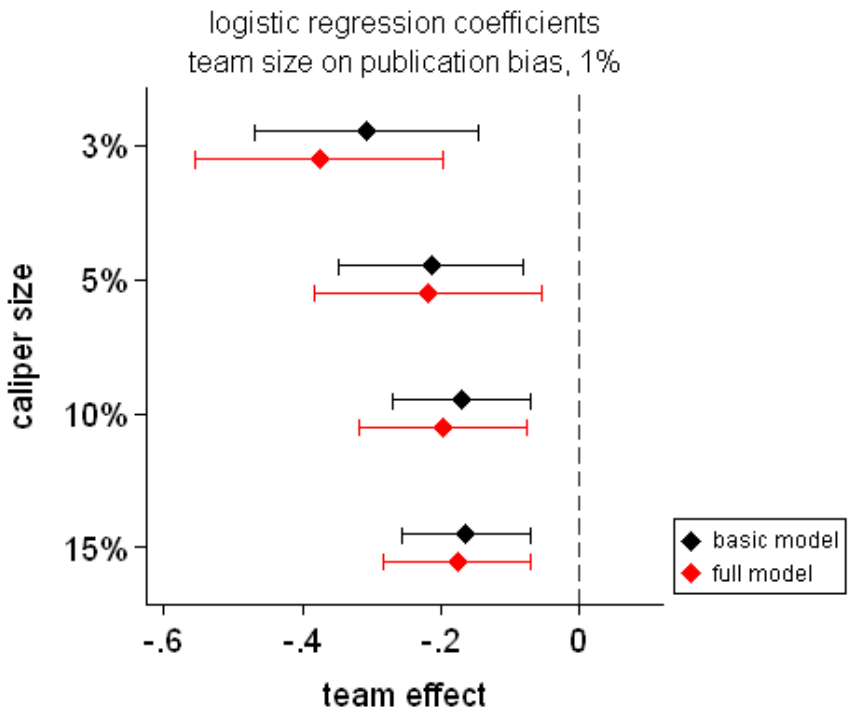
Influencing factors

- 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

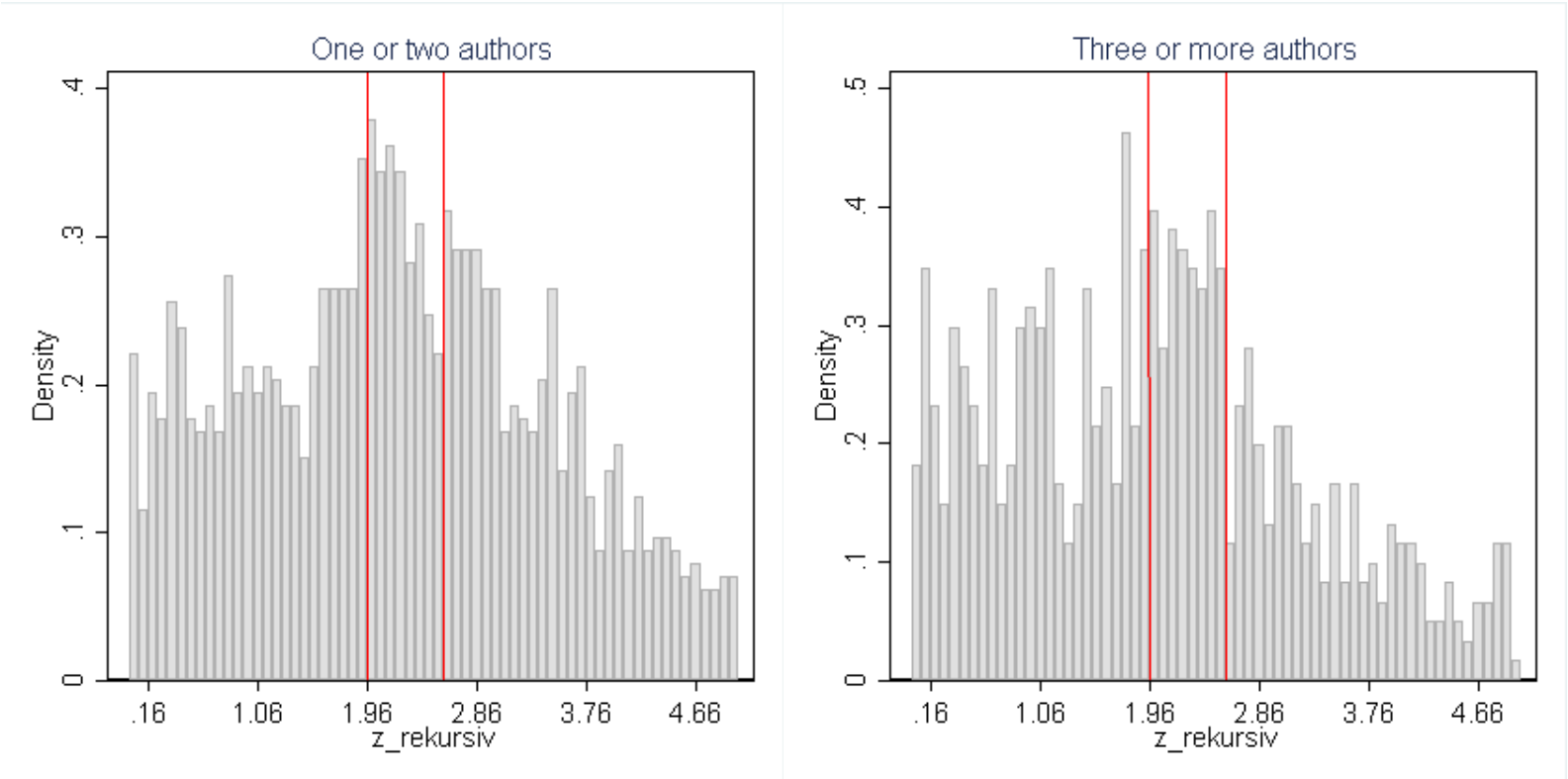
Author group size effect





Influencing factors

Author group size effect



University of Zurich[UZH]

Institute of Sociology

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

Dickersin, Kay, Yuan-I Min and Curtis L. Meinert. 1992. Factors Influencing Publication of Research Results. Follow-up of Applications Submitted to Two Institutional Review Boards. Journal of the American Medical Association 267 (3): 374-378.

Gerber, Alan and Neil Malhotra, G., 2008. Publication Bias in Empirical Sociological Research : Do Arbitrary Significance Levels Distort Published Results? Sociological Methods & Research 37: 3-30.

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

Franco, Annie, Neil Mahotra and Gabor Simonovits. 2014. Publication bias in the social sciences: Unlocking the file drawer. Science 345 (6203): 1502-1505.

Leggett, Nathan C., Nicole A. Thomas, Tobias Loetscher and Michael E.A. Nicholls. 2013. The life of p: 'Just significant' results are on the rise. The Quarterly Journal of Experimental Psychology 66 (12): 2303-2309.

Mahoney, Michael J. 1977. Publication Prejudices: An Experimental Study of Confirmatory Bias in the Peer Review System. Cognitive Therapy and Research 1 (2): 161-175.

Nuijten, Michèle B., Chris H. J. Hartgerink, Marcel A. L. M. van Assen, Sacha Epskamp and Jelte M. Wicherts. 2016. The prevalence of statistical reporting errors in psychology (1985–2013). Behavior Research Methods 48 (4): 1205-1226.

Turner, Erick H., Annette M. Matthews, Eftihia Linardatos, Robert A. Tell and Robert Rosenthal. 2008. Selective Publication of Antidepressant Trials and Its Influence on Apparent Efficacy. The New England Journal of Medicine 358 (3): 252-260.