DZHW

German Centre for Higher Education Research and Science Studies

Bibliometric Measurement Variation and University Rankings

HRK-Serviceprojekt „Internationale Hochschulrankings“: NWV 2021
Stephan Stahlschmidt (with *Marion Schmidt*)

University of Göttingen and *SHELX*

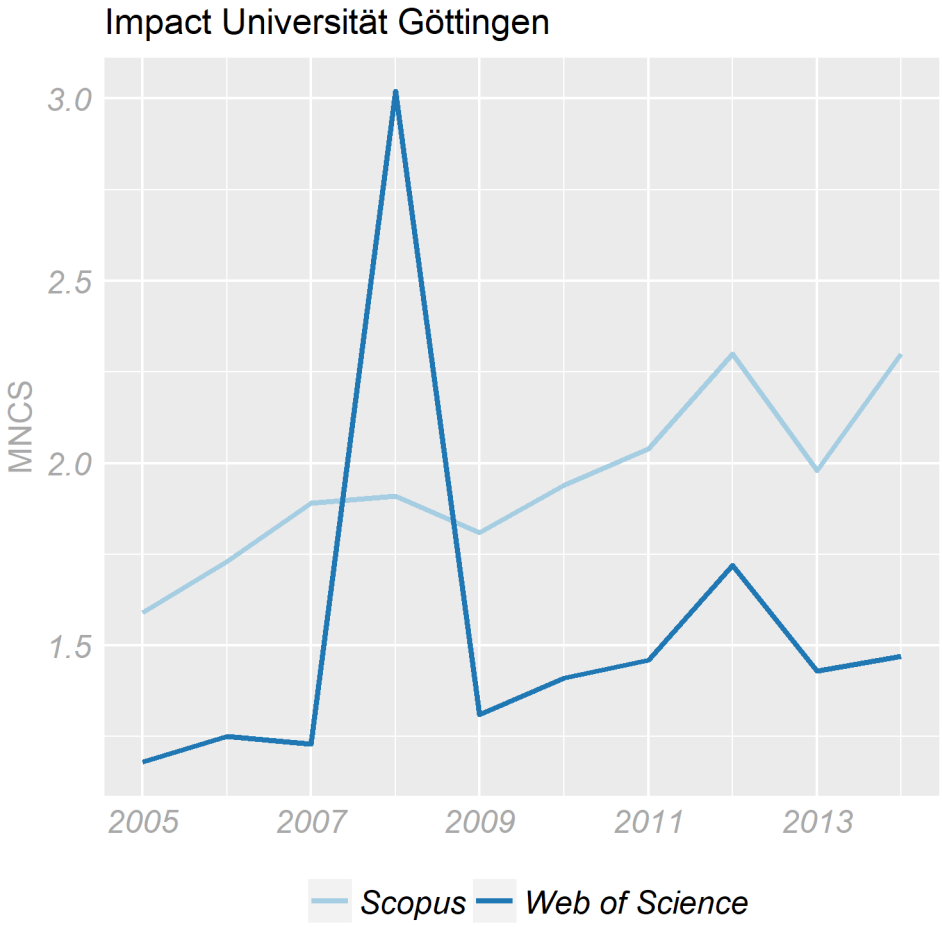
Sheldrick (2008). A short history of SHELX. *Acta Crystallographica Section A*, 64, 112 – 122.

Web of Science:

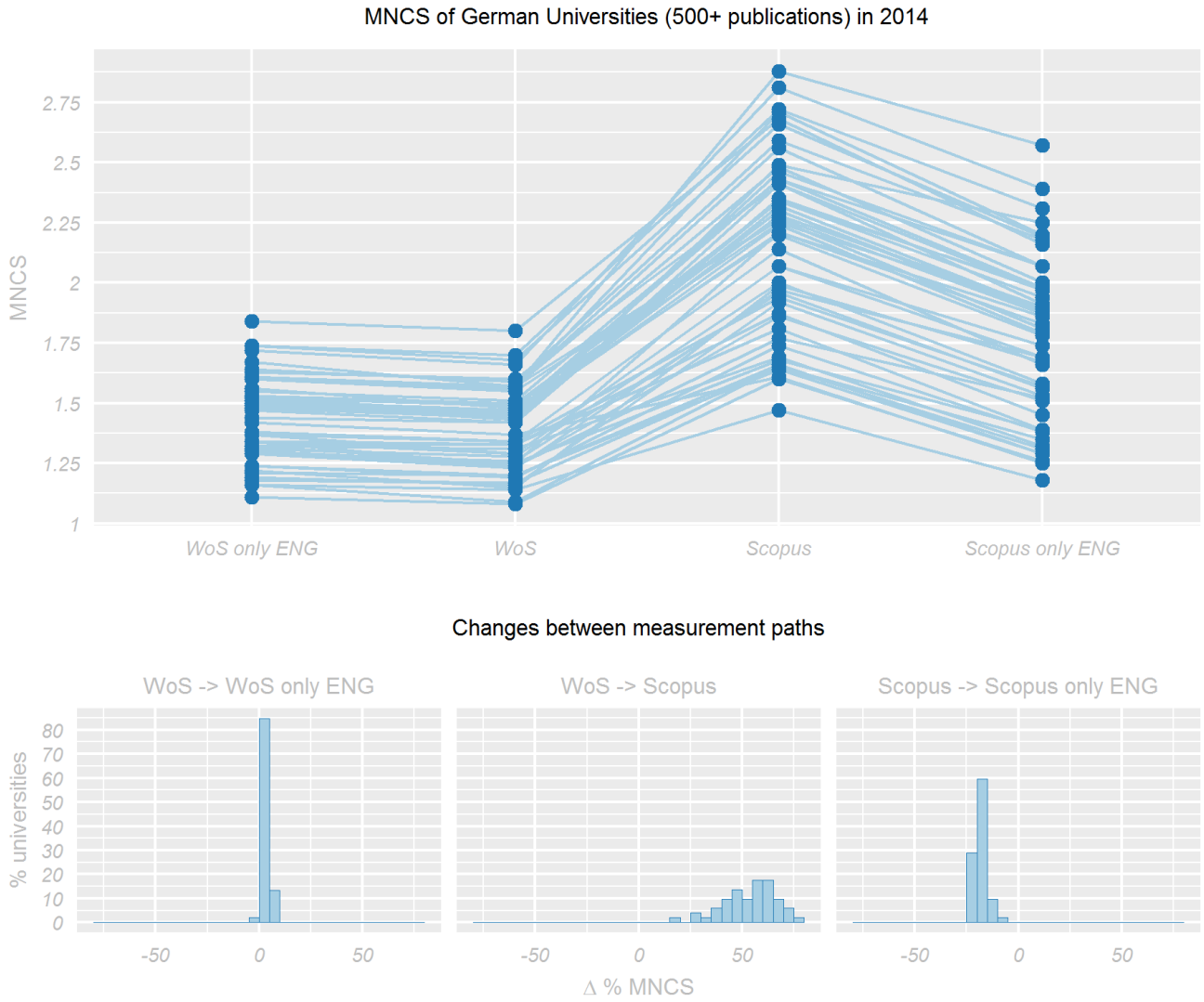
- Increased JIF 20-fold
- Made *Universität Göttingen* the most cited / impactful research institution in 2008

Scopus:

- Paper's effect hardly visible on institutional level



Language barrier?



Motivation

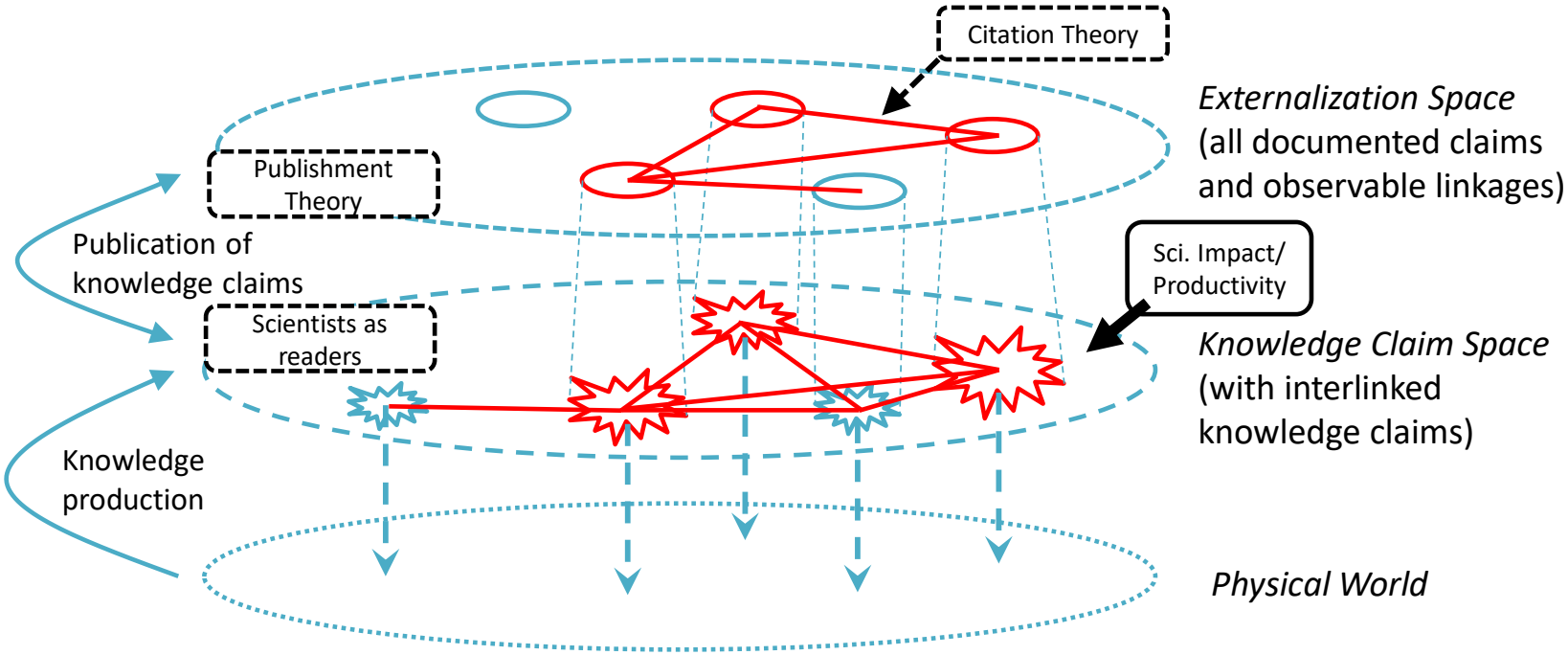
Computed bibliometric impact depends on:

1. Database
2. Processing of data, e.g. any subsets or partitioning

Measurement process generates, mostly unobserved, measurement variation:

- How strong is this variation?
- Can we observe any structure in the variation?
- What explains the measurement variation?

Conceptual Model



Evaluative Bibliometrics: Citation Theory

Citation behaviour differs between disciplines and over time:

- Normalization facilitates valid comparison

Noise and Signal:

- Signal:
  - Mertonian citations: give credit where credit is due (Merton, 1957)
- Noise:
  - Not all influences are cited (e.g. MacRoberts and MacRoberts, 1989)
  - Rhetorical citations (e.g. Gilbert, 1977)
  - Social citations (e.g. Latour and Woolgar, 1986)
  - Citations are treated as binary signals (e.g. Small, 1982)

→ Analyse of higher aggregates is more informative:

- More publications allow to separate signal from noise
- Statistics: any difference in means (signal) is not obscured by potentially large variance (noise)

Evaluative Bibliometrics: Measurement Variation

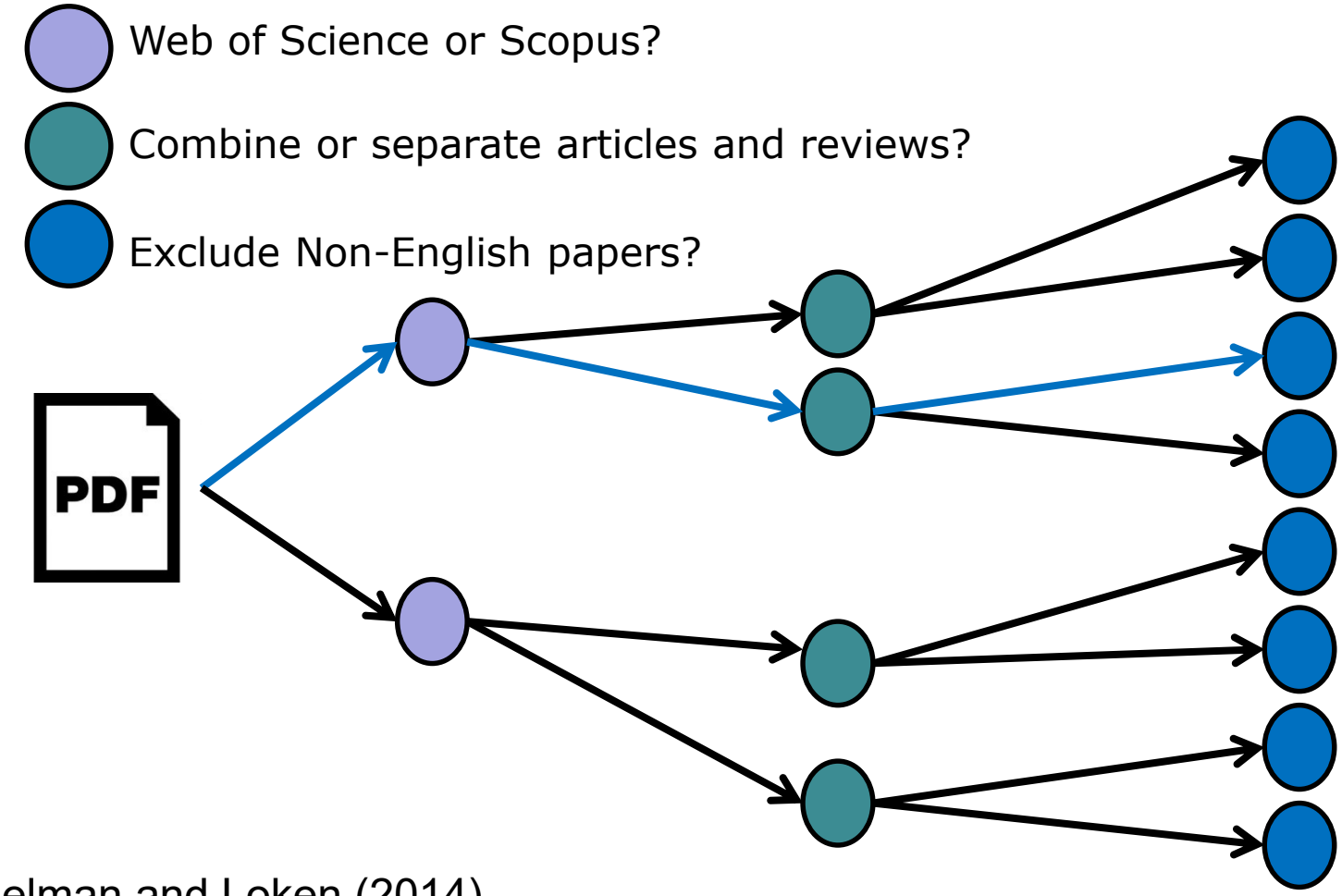
Bibliometric measurement process:

- Implementation: Translating abstract mathematical indicator formula to actual data
- Measurement paths: Act of implementation allows for different measurement approaches:
  - Productivity: Transfers publications into bibliometric indicators
  - Impact: Transfers references into bibliometric indicators

Impact:

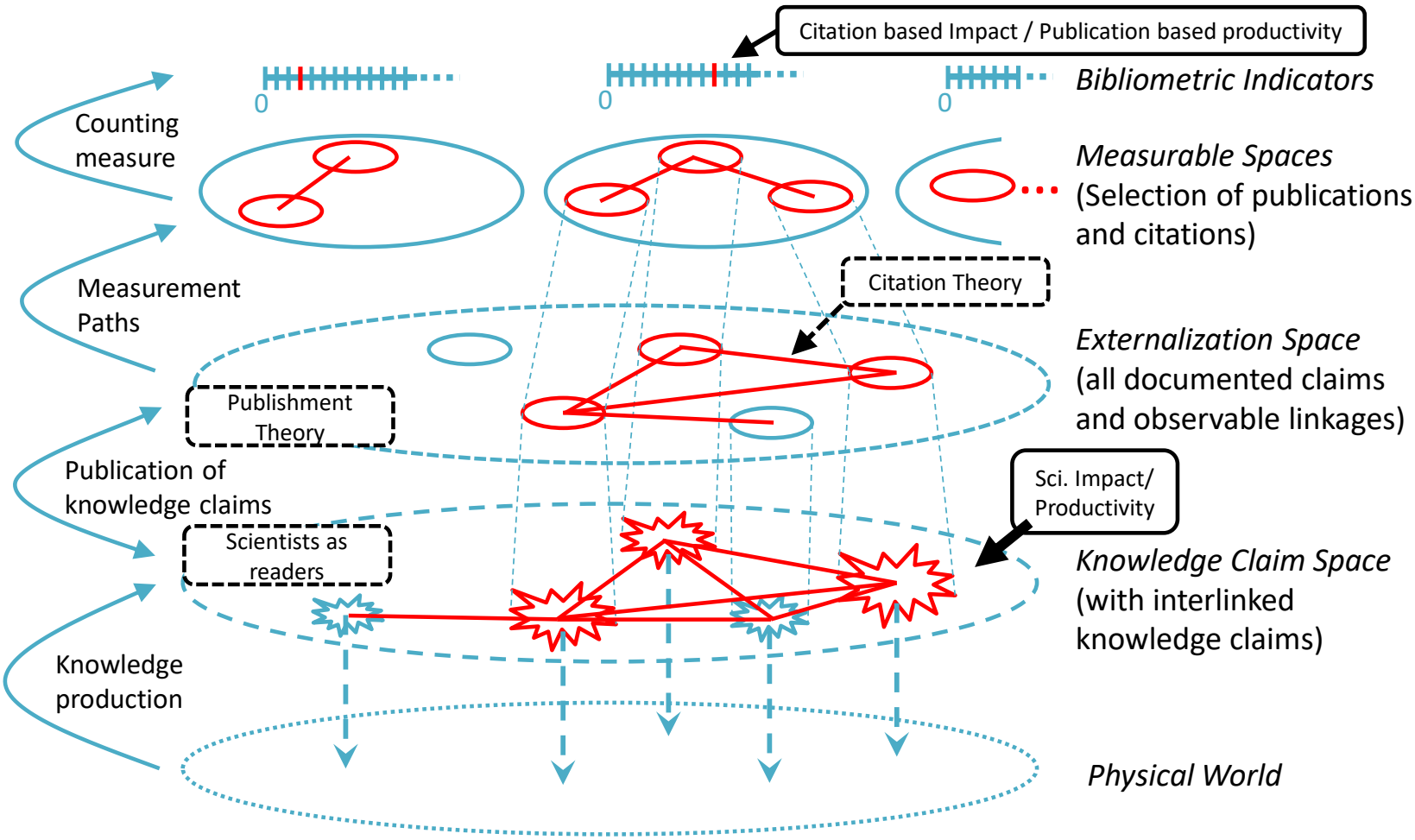
- Scientific impact as a latent construct
- Citations and scientific impact overlap, but are not congruent
- No “optimal” measurement path identifiable

Garden of forking *Measurement* paths



Gelman and Loken (2014)

Conceptual Model



Context

Researcher's degree of freedom:

- *p*-hacking, replication crisis, mixed results,…

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

Alan Turing Institute (2019) The Turing Way

Answers:

1. Sensitivity Analysis:
   - Recalculating outcomes under alternative assumptions
2. Meta-Analysis:
   - Observe weighted average effect among several studies
     → Publication Bias
   - Contrary we control data generating process and cause variation at will
3. Standards:
   - Deliberately exclude variation by defining single prevailing measurement path
     → highly political
   - Contrary we embrace variation and exploit it to detail the effect of individual measurement decisions

Analysing Measurement Variation

Application:

- Bibliometric raw data: Follow several measurement paths
- Compare results
- Infer influence of measurement process on results

Measurement Decisions:

- Choice of database: Web of Science or Scopus
- Include or exclude Non-English publications
- Combine or separate reviews and articles
- Include or exclude self-citations
- Include or exclude Social Sciences and Humanities
- Multi-author papers: Apply fractional or whole counting
- Counting citations: Three-year or five-year citation window
- Normalize by discipline classification: classification by database provider or OECD Disciplines of Sciences

Analysing Measurement Variation

Draw random sample:

- 25% of all potentially 256 parallel bibliometric worlds

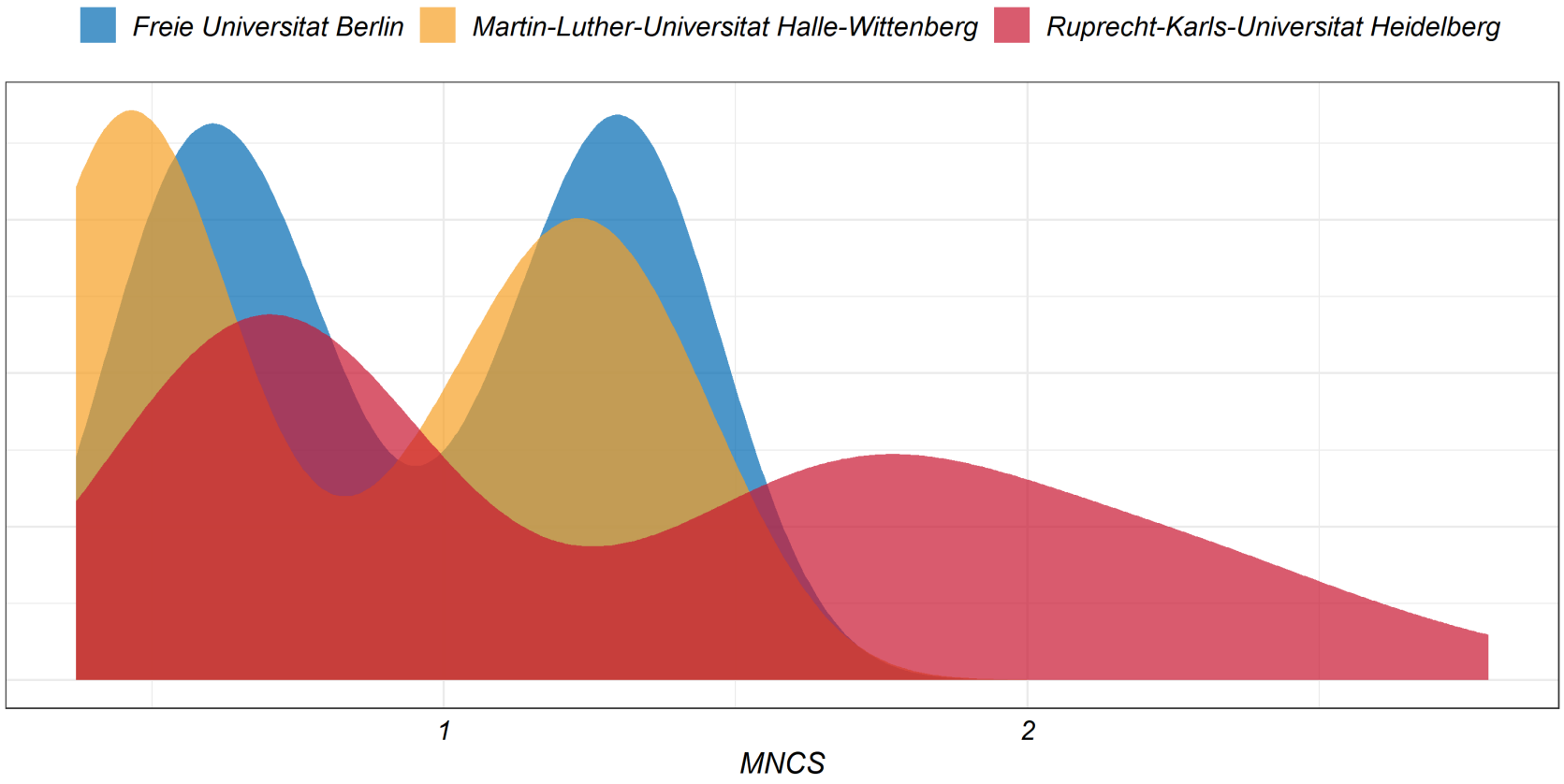
Compute impact:

- for 37 German universities (with > 1000 publications in 2012)
- *Mean Normalised Citation Score (MNCS)* for German university j with publications $i \in [1, \ldots, I]$:

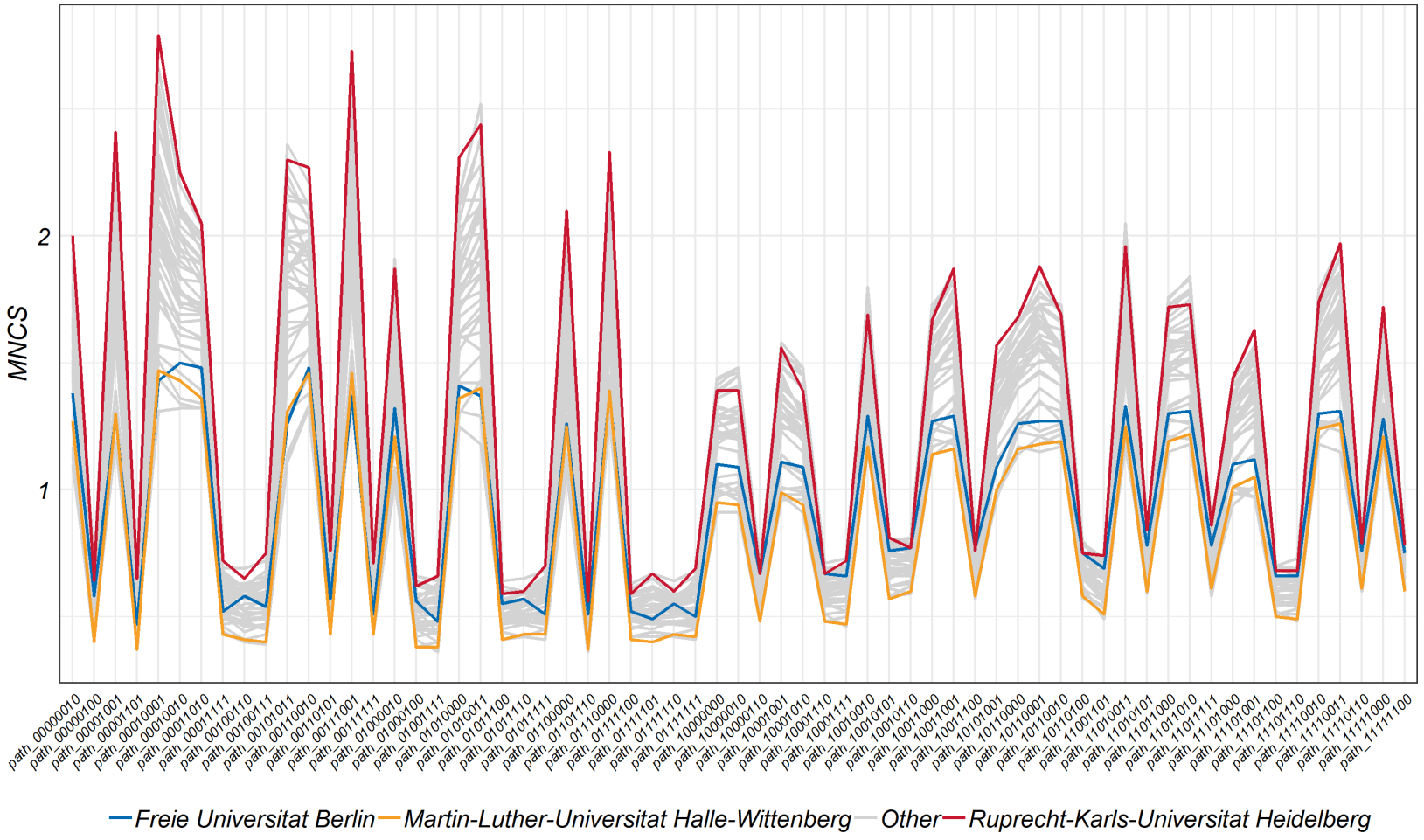
$$MNCS_j = \frac{1}{I} \sum_{i=1}^{I} \frac{obtained\ citations_i}{expected\ citations_i}$$

where $expected\ citations_i$ depends on the publication year and respective discipline.

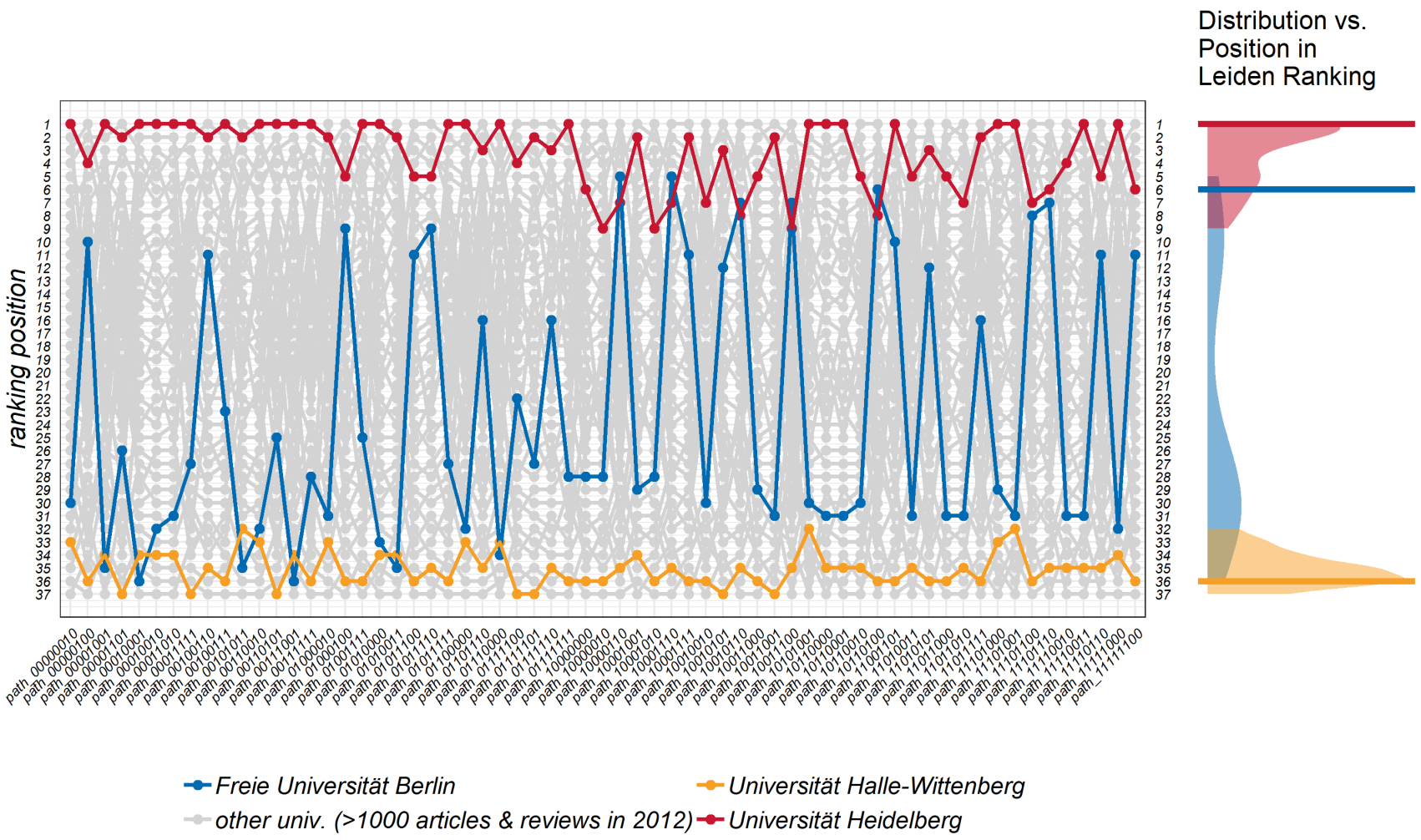
Measurement Variation: Institution



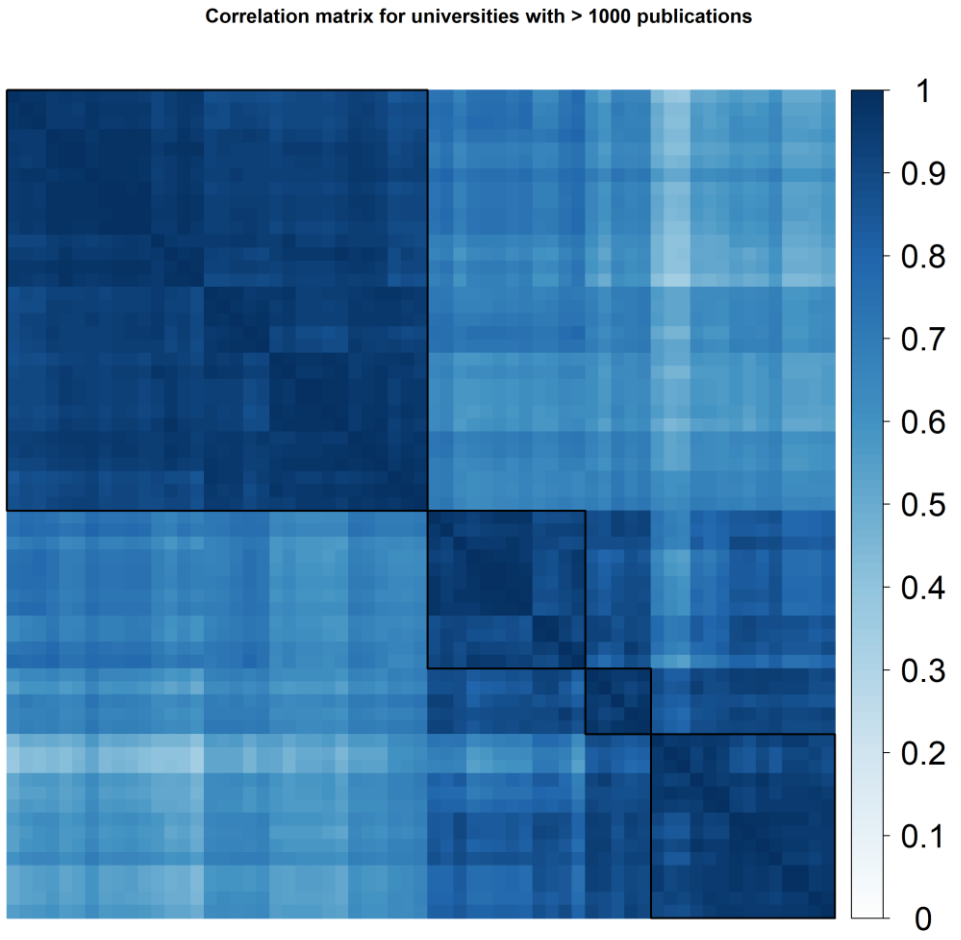
Measurement Variation: System

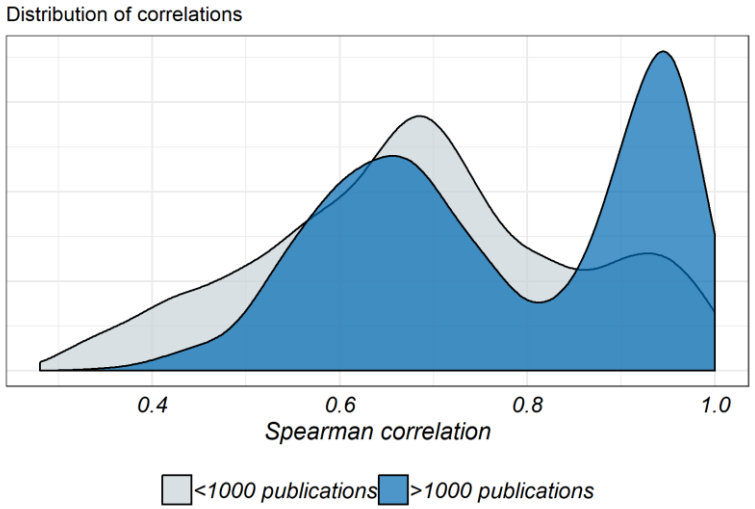


Measurement Variation: Ranking



Measurement Variation: Correlation





Measurement Variation: Modelling

How does each measurement decision affect the measured impact?

Linear Mixed Model:

$$Y_{ij} = \alpha_i + x_{ij}^t \beta + u_{ij}^i \gamma_i + \epsilon_{ij}$$

where

- Y_{ij} indicates the *MNCS* of university i corresponding to measurement path j
- $i \in [1, \ldots, m]$ denotes the $m = 37$ clusters of German universities
- $j \in [1, \ldots, n_i]$ states the balanced size of $n_i = 64$ observation per university
- α_i denotes the university specific (random) intercept
- β describes the fixed effects the 8 binary measurement decisions and
- γ_i details the random effects.

Modelling: Limitations and coefficients

Current limitations:

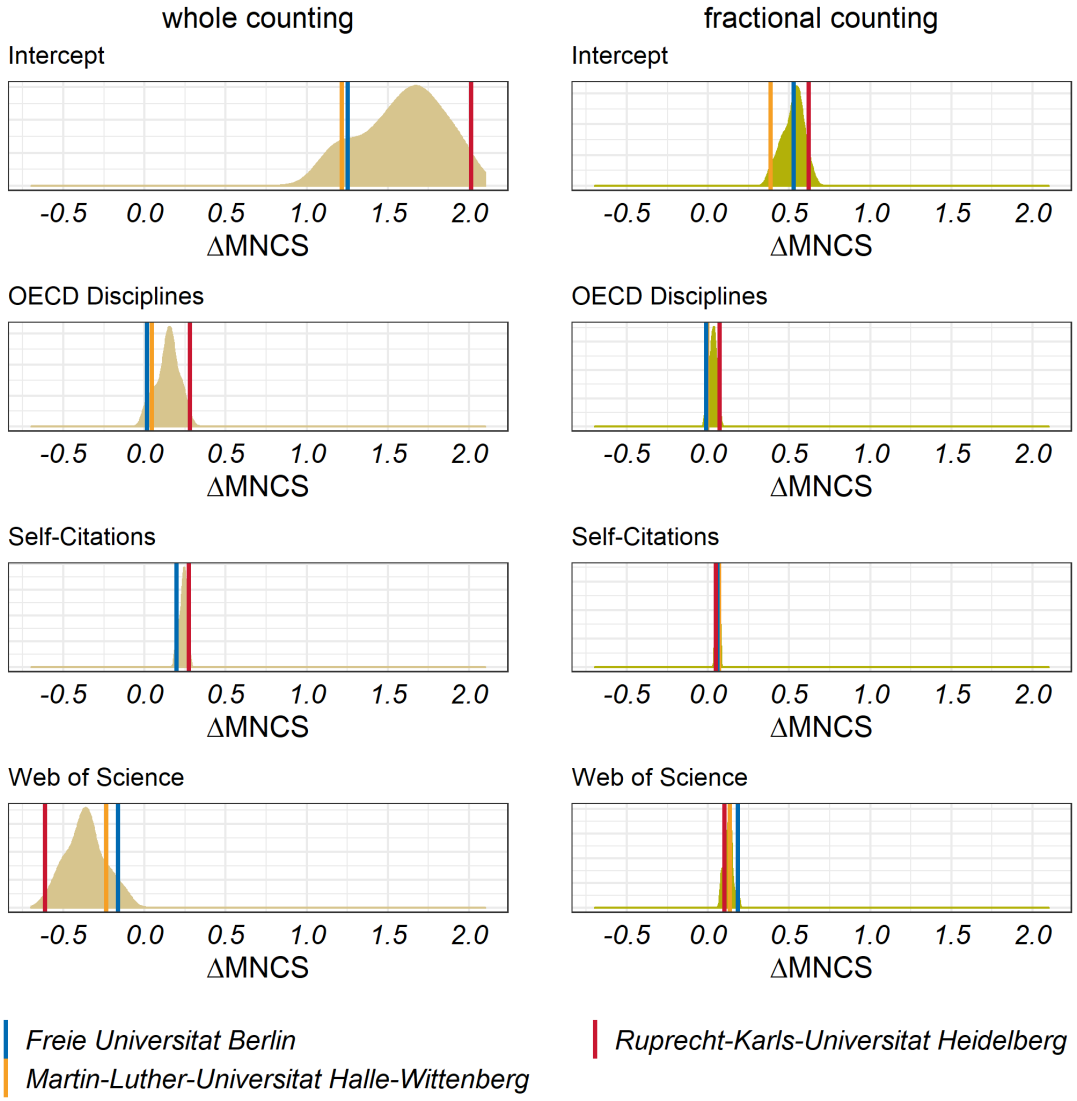
- Separate analysis for each counting method due to computation feasibility

Negligible measurement decisions (centred on zero):

- Excluding Non-English paper
- Separating reviews
- Citation window
- Excluding Social Sciences & Humanities
  (for whole counting; slight negative effect for fractional counting)

→ No effect in regression framework

Modelling: Coefficients



Conclusions

Measurement variation as the last chain link:

$$\textit{Measured Impact} = f(\textit{impact theory}) + f(\textit{choice of indicator} \mid \textit{impact theory}) + f(\textit{implementation on data} \mid \textit{indicator})$$

Spurious precision:

- Measurement process influences universities' impact values
- Citation-based scientific impact is less precise than it seems
  - Consequences for researchers: Funding, promotion, salary,…
- Constant structure, albeit with less precision, is observable

Thank you!

Acknowledgements:

- Daniel Sirtes for inspiration
- Jesper Schneider for productive discussions
- Brooke Strucke / Science-Metrix for hospitality