EVALUATING POPULARITY DATA FOR RELEVANCE RANKING IN LIBRARY INFORMATION SYSTEMS

Kim Plassmeier, Timo Borst
German National Library of Economics
k.plassmeier;t.borst@zbw.eu

Christiane Behnert, Dirk Lewandowski
Department of Information, Hamburg University of Applied Sciences
christiane.behnert;dirk.lewandowski@haw-hamburg.de

INTRODUCTION

In this poster, we present our work in progress to develop a relevance model for ranking in Library Information Systems (LIS), which takes non-textual factors into account. Here we focus on three types of popularity data: citation counts, author metrics and usage data. The data were gathered from our test environment – EconBiz, a search portal for economics hosted by the German National Library of Economics (ZBW) – and other external sources.
The combination of multiple criteria in a linear model requires that the data are "comparable". However, this is not fulfilled by the raw data, in general. Further, the raw data might contain biases. We address these problems by transforming the raw data via the Characteristic Scores and Scales (CSS) method.

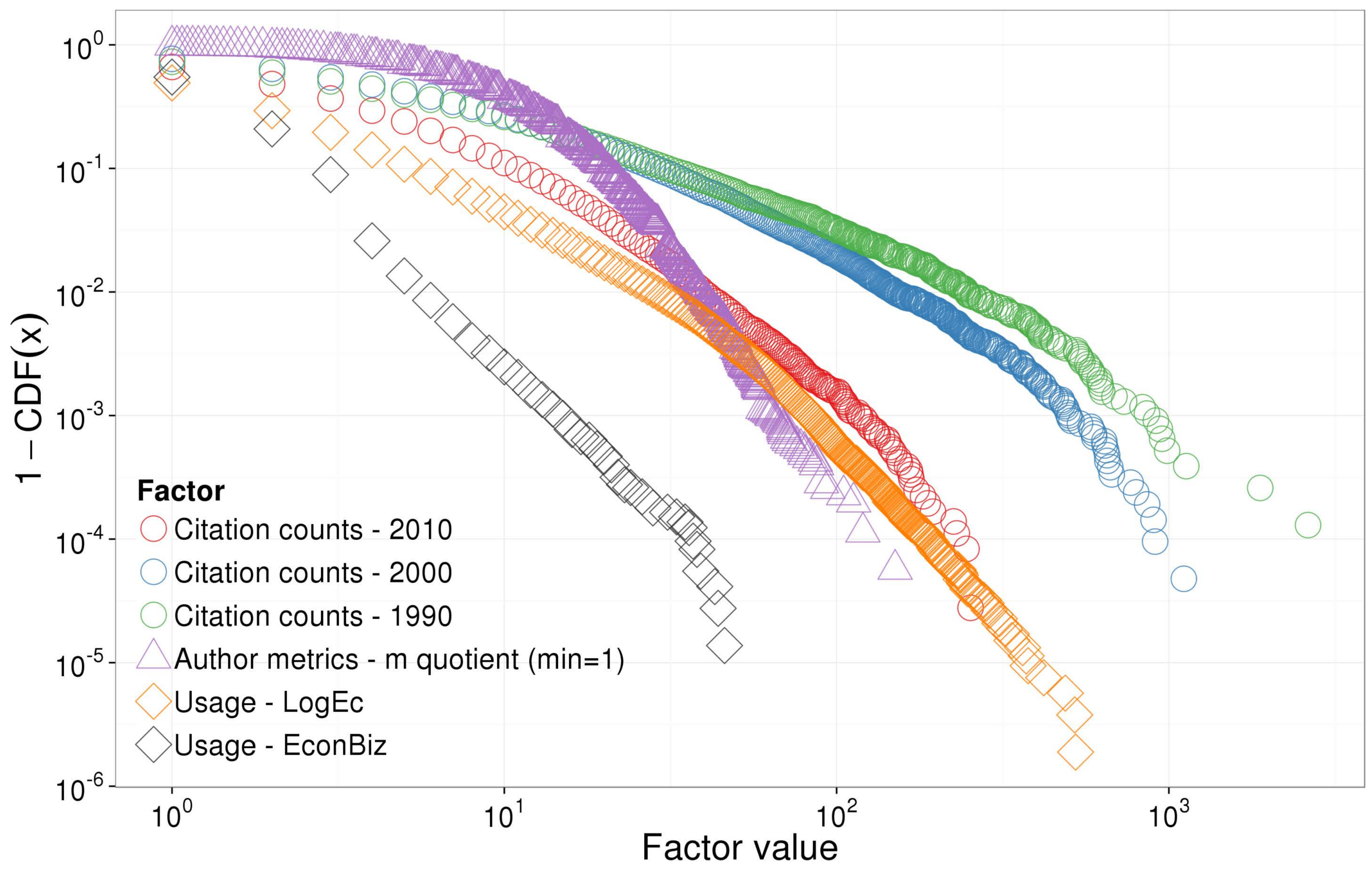
GOALS

- Enable weighing of factors against each other (in a linear model); i.e., establish a common utility scale as in multi-attribute utility theory.
- Remove biases from individual factors.
- Make weights in the linear model less sensitive to changes in underlying data (when data for factors are updated).

FACTORS AND DATA SOURCES

DATA TYPE	POPULARITY FACTOR	DATA SOURCE
A) Citation counts	No. of citations for item	CitEc (external)
	Citation impact for journal	SCImago Journal Rank, CitEc (external)
B) Author metrics	Citation impact for author (m quotient: h-index divided by scientific age; see Hirsch, 2005)	CitEc (external)
C) Usage data	No. of record views	Web analytics tool (internal), LogEc (external)
	No. of clicks on full text	Web analytics tool (internal), LogEc (external)
	No. of loans at local library	Library's local system (internal)

CUMULATIVE DISTRIBUTIONS OF THE RAW DATA



PROBLEMS

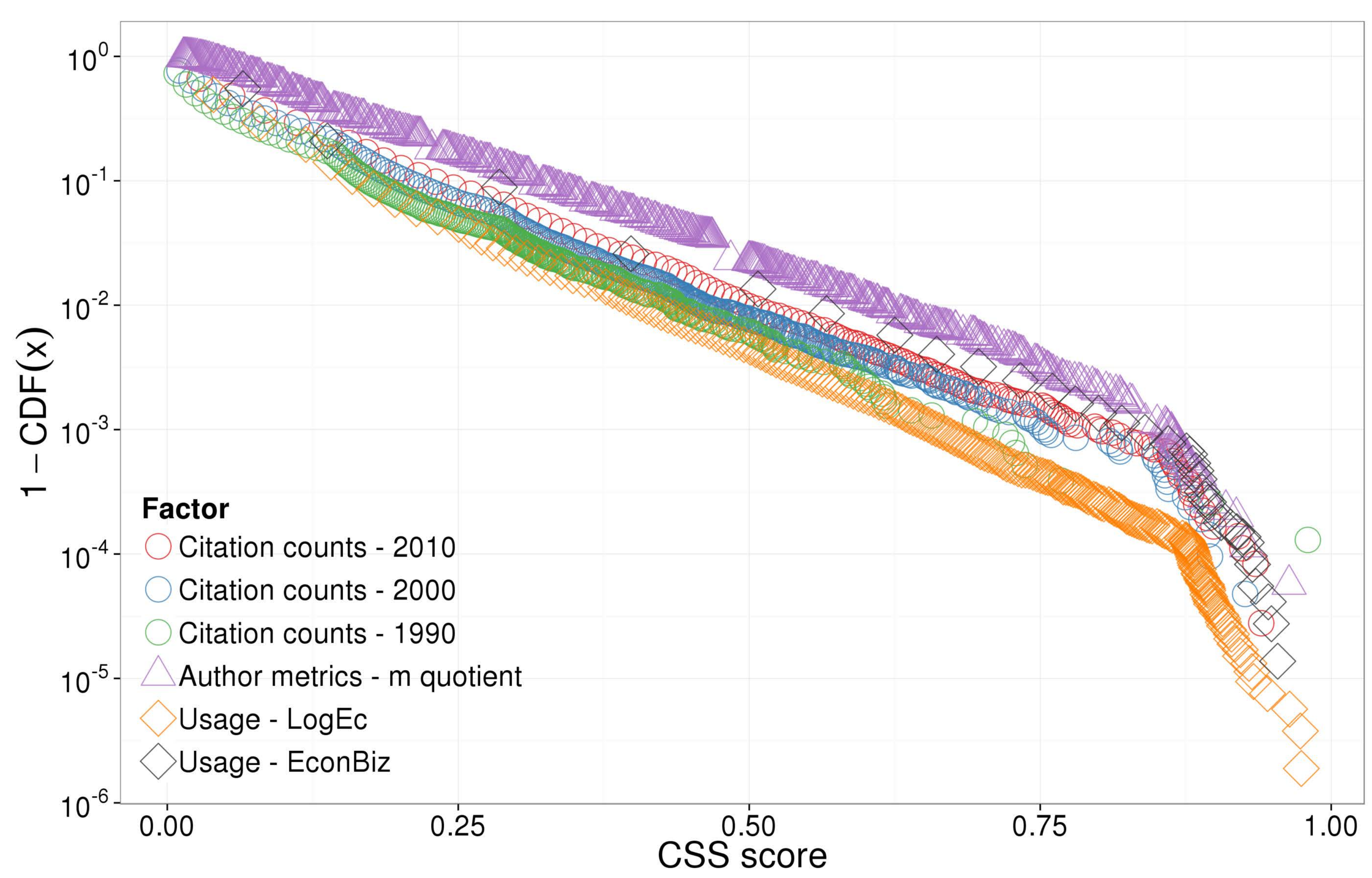
- Citation counts for documents of different age are biased due to citation obsolescence.
- Usage data are biased due to source (different usage per document ratio).
- The different factors are incommensurable with each other a priori.

CHARACTERISTIC SCORES AND SCALES (CSS)

- Method proposed by Glänzel & Schubert (1988) to find characteristic classes in citation distributions (e.g., papers that are "poorly cited", "fairly cited", "remarkably cited", or "outstandingly cited").
- The classes are found by iteratively calculating truncated moments: The first class boundary is set to the mean of the distribution, $\beta_1 = \mu$, the second boundary is given by the mean of the distribution truncated at the first boundary, $\beta_2 = \mathrm{mean}(\{x_i | x_i \geq \beta_1\})$. Finally, the k-th class boundary is given by

$$\beta_k = \mathrm{mean}(\{x_i | x_i \geq \beta_{k-1}\})$$

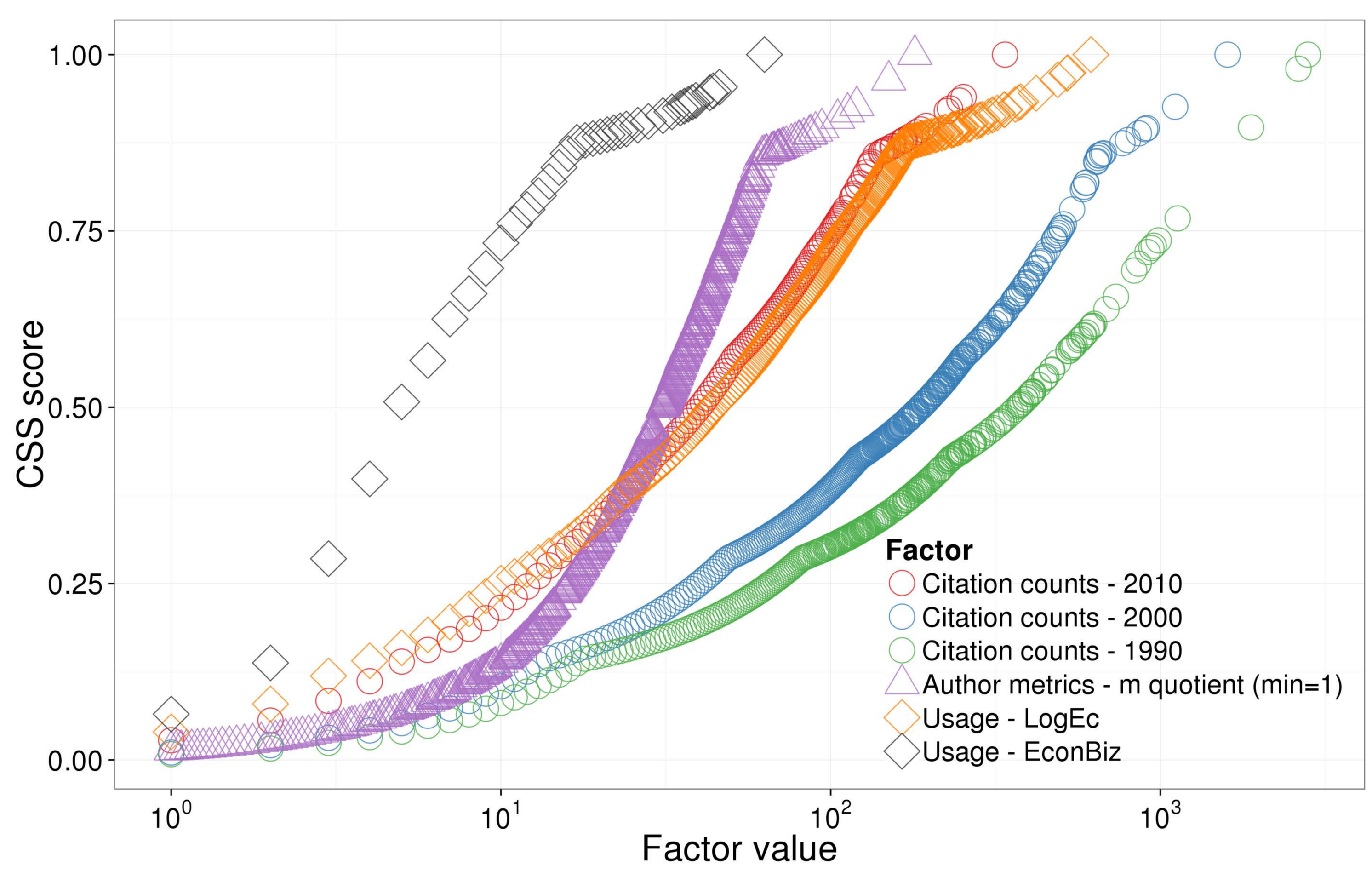
CUMULATIVE DISTRIBUTIONS AFTER CSS TRANSFORMATION



CONCLUSION

- CSS method works well to remove citation obsolescence bias from citation counts.
- CSS method works reasonably well
  a) to normalize usage data from different sources,
  b) to normalize and align the different factors.
- However, the CSS method cannot fully compensate for all artifacts in the raw distributions.
- Since the method is quasi parameter-free, it might be especially interesting for LIS, if no training data are available.
- Effectiveness of CSS scores as utilities in an overall relevance model must still be evaluated in retrieval performance studies.

NON-LINEAR EQUIVALENCES INDUCED BY CSS TRANSFORMATION



REFERENCES

Glänzel, W., & Schubert, A. (1988). Characteristic scores and scales in assessing citation impact. Journal of Studies in International Education, 14(2), 123–127.
Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. In Proceedings of the National Academy of Sciences of the United States of America (Vol. 102, pp. 16569–16572).

Project Partners
ZBW Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics
Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences
Funded by
DFG Deutsche Forschungsgemeinschaft
LibRank
www.librank.info