Master Thesis Proposal

Leonard Kosanke Valentin Kriegmair

Humboldt University Berlin

Master Thesis Proposal

Introduction

Monte Carlo simulations are an extensively utilized tool for assessing and comparing statistical methods in quantitative empirical science. Their key advantage lies in the ability to evaluate the performance of estimation and inference methods by analyzing them in light of known simulated population models and values. Despite the clear insight into the explicitly specified underlying data structures that simulations provide, they are not immune to common pitfalls that thus far have been predominantly associated with the replication crisis in empirical quantitative research (Lohmann et al., 2022). In simulation studies, a key challenge is ensuring that results can be generalized to real-world applications. These studies must ensure that the chosen performance metrics, experimental factors, and inference models accurately reflect and test the complexities and variability of real-world scenarios. Given the impracticality of simulating and analyzing every possible inference model and use case, these studies inherently involve a multitude of decision-making elements potentially prone to bias. Further, the decisions of selecting evaluation criteria, target models and simulation specifications can have substantive impact on the results. One approach that addresses these issues at the level of population model selection empirically grounds this decision by sampling from model configurations in the scientific literature (Bollmann et al., 2015).

Here we propose a novel approach to address generalizability as well as biases on all decision-making elements of simulation studies. We aim to introduce and test an adversarial collaboration framework for simulation studies. The approach is tailored for comparison studies in particular, which are critical in the validation of new statistical methods. This aligns with the recognized need for neutral and rigorous assessment in such studies (Boulesteix et al., 2013). Adversarial collaboration, piloted by (Mellers et al., 2001) and popularized by (Kahneman, 2011), is a research method increasingly recognized within the quantitative empirical research community for enhancing scientific rigor. Praised as

"The next science reform" (Clark & Tetlock, 2021) adversarial collaboration is the process of disagreeing scholars working jointly to resolve scientific disputes. This method entails identifying points of empirical disagreement, designing mutually agreed upon methods to test competing hypotheses, and jointly publishing results, irrespective of the outcome. The authors highlight its role in promoting fair tests and truthful representation of opposing views, thereby enhancing epistemic accountability and reducing research ambiguity. Additionally, juxtaposing and debating competing positions in this way can improve generalizability of results.

Our goal is to transfer these benefits of adversarial collaboration to simulation studies. To do so, we want to conduct a focused case study: comparing a newly proposed iterative structural after measurement (SAM) estimation approach to structural equation model (SEM) estimation with traditional non-iterative SEM estimation (Dhaene & Rosseel, 2023; Robitzsch, 2022; Rosseel & Loh, 2022). In their respective simulations, the authors results differed up to the point of contradiction, providing us with ideal grounds for conducting adversarial collaboration. While Dhaene and Rosseel (2023) concluded that SAM estimation generally outperforms non-iterative SEM in small samples, Robitzsch (2022) did not find the methods to differ. Similarly, whereas the former found SAM to be more robust against model misspecification, the latter argued the opposite to be true. The idea is that we, the authors of this paper, represent these competing positions, with each of us advocating for one of the different set of findings, as detailed later. This allows us to examine the practical feasibility of adversarial collaboration in Monte Carlo simulation studies as well its potential to enhance methodological rigor and generalizability in this domain of research. Hence, such an investigation will lead us to address the following research question:

By applying our adversarial collaboration framework to the comparison of iterative structural after measurement (SAM) and traditional non-iterative estimation of structural equation models (SEM), can we demonstrate its practical applicability and technical

feasibility for conducting Monte Carlo simulation studies?

<div align="center">**Method**</div>

To answer this research question, we created the following procedure to be conducted for the adversarial collaboration process:

Outline

In general, the procedure consists of two parts. In the first part, each researcher conducts an independent preliminary study as a replication of their side of the argument. Leonard Kosanke will be replicating relevant parts of the study by Robitzsch (2022), and Valentin Kriegmair will replicate the study by Dhaene and Rosseel (2023). The replications include the generation of individual simulation protocols, as suggested by Morris et al. (2019).

In the second part, a joint study, including a joint simulation protocol, is pursued. Here, the main part of the adversarial collaboration takes place. Each step of the individual simulation studies (from the substantive research question up until the interpretation of the results) is scrutinized and debated between collaborators, using adversarial collaboration techniques as listed below. Decisions are made and documented based on the most convincing argument presented, if possible. This process starts based on the results of Part 1, which could include contradictory findings, subjective interpretations of ambiguous results, or strong opinions on expected outcomes as a consequence of the preliminary results.

To facilitate a structured comparison and the integration of our studies through the collaboration process, we have established a common framework for conducting our individual simulation studies. This framework allows for the expected divergences in results while providing a basis for systematic comparison and synthesis. Additionally, we have identified two substantive research questions to co-align the individual simulation studies:

1. How do SAM and traditional SEM methods (including ML and ULS) compare in terms of bias, Mean Squared Error (MSE), and convergence rates in small to

moderate samples?

2. What is the impact of model misspecifications, such as residual correlations and cross-loadings, on the performance of SAM compared to traditional SEM methods?

In the end, each collaborator will report all results in their own thesis.

We are aware that adversarial collaboration has its limitations when it comes to decision making and has lead to unresolvable disagreements in previous studies Cowan et al. (2020) and Mellers et al. (2001). In order to preemptively reduce this possibility, we want to conduct this study in a structured and formalized manner. To this end, we propose to use the tools presented in the next section. These are also meant to facilitate the evaluation of the adversarial collaboration with regards to our research question.

Adversarial Collaboration (AC) Techniques:

At each step of decision making in the Joint Study, any of the following techniques can be used, depending on their applicability:

- Core Disagreements: Arrive at clearly defined core disagreements in the research question or objectives that might be the origin for conflicts in successive steps. (Clark et al., 2022)

- Assumption check: List, question, and categorize assumptions (Kardos & Dexter, 2017)

- Red-Teaming: what if scenarios (Kardos & Dexter, 2017)

  - e.g. Generate and test alternative estimands with the goal of invalidating the arbiters decisions.

- Quality of information (Kardos & Dexter, 2017): checking the adversaries quality of evidence on a paper-level

- Third neutral arbiter: In case of fundamentally unresolvable disagreements, a third neutral arbiter (Aaron Peikert) will be consulted to try to resolve them.

Documentation - Decision log

The results of the *Joint Study* will be presented in the main papers in a concise way, focussing on the most crucial decisions.

In addition, a separate decision log will serve as a detailed and complete documentation of all decisions made. Here, we summarize the AC-techniques used, as well as their consequence for the decision-making process.

In our mind, decision making can be based on four distinct grounds: Evidence-based, pragmatic reasons, arbitrary reasons or other reasons (e.g. personal values, political issues). In some cases, more than one of these are present. In such cases, we want to rank their relevance to a decision, if possible. An example could be the following log entry:

```
Decision element: factor of sample sizes
Result: c(50, 100, 200)
Grounds: Primary - Evidence-based, as per Peikert et al. (2020);
         Secondary - Consensus based on the average of individual suggestions.
```

Evaluation - Diary entries

The adversarial collaboration in Part 2 (Joint study), is documented and evaluated from each collaborators perspective using semi-structured diary entries based on Shah and Leeder (2016), after each step. This has two purposes: Firstly, accumulating evidence for the evaluation of the AC and the collaboration procedure we propose. Secondly, it can provide valuable information that helps us improve this procedure. Thus, the resulting diaries will be used to evaluate the applicability of the adversarial collaboration and answer the research question.

Timetable

Upon receiving approval, we aim to conduct the individual simulation studies until March 15, 2024. The joint study is scheduled to commence immediately afterward, from

March 16 to April 15, 2024. Finally, the period from April 16 to April 30, 2024, will be dedicated to writing our respective theses. We acknowledge that these dates serve as a preliminary timeline and may be subject to adjustments as the project progresses.

References

Bollmann, S., Heene, M., Küchenhoff, H., & Bühner, M. (2015). What can the Real World do for simulation studies? A comparison of exploratory methods [00000 Publisher: Universitätsbibliothek der Ludwig-Maximilians-Universität München]. https://doi.org/10.5282/UBM/EPUB.24518

Boulesteix, A.-L., Lauer, S., & Eugster, M. J. A. (2013). A Plea for Neutral Comparison Studies in Computational Sciences (M. Gasparini, Ed.). *PLoS ONE*, *8*(4), e61562. https://doi.org/10.1371/journal.pone.0061562

Clark, C., & Tetlock, P. (2021, June). Adversarial Collaboration: The Next Science Reform. https://doi.org/10.1007/978-3-031-29148-7_32

Cowan, N., Belletier, C., Doherty, J. M., Jaroslawska, A. J., Rhodes, S., Forsberg, A., Naveh-Benjamin, M., Barrouillet, P., Camos, V., & Logie, R. H. (2020). How Do Scientific Views Change? Notes From an Extended Adversarial Collaboration. *Perspectives on Psychological Science*, *15*(4), 1011–1025. https://doi.org/10.1177/1745691620906415

Dhaene, S., & Rosseel, Y. (2023). An Evaluation of Non-Iterative Estimators in the Structural after Measurement (SAM) Approach to Structural Equation Modeling (SEM). *Structural Equation Modeling: A Multidisciplinary Journal*, *30*(6), 926–940. https://doi.org/10.1080/10705511.2023.2220135

Kahneman, D. (2011). *Thinking, fast and slow* [Pages: 499]. Farrar, Straus; Giroux.

Kardos, M., & Dexter, P. (2017). A Simple Handbook for Non-Traditional Red Teaming.

Lohmann, A., Astivia, O. L. O., Morris, T. P., & Groenwold, R. H. H. (2022). It's time! Ten reasons to start replicating simulation studies. *Frontiers in Epidemiology*, *2*, 973470. https://doi.org/10.3389/fepid.2022.973470

Mellers, B., Hertwig, R., & Kahneman, D. (2001). Do Frequency Representations Eliminate Conjunction Effects? An Exercise in Adversarial Collaboration. *Psychological Science*, *12*(4), 269–275. https://doi.org/10.1111/1467-9280.00350

Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate
   statistical methods [_eprint:
   https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.8086]. *Statistics in Medicine*,
   38(11), 2074–2102. https://doi.org/10.1002/sim.8086

Robitzsch, A. (2022). Comparing the Robustness of the Structural after Measurement
   (SAM) Approach to Structural Equation Modeling (SEM) against Local Model
   Misspecifications with Alternative Estimation Approaches. *Stats*, *5*(3), 631–672.
   https://doi.org/10.3390/stats5030039

Rosseel, Y., & Loh, W. W. (2022). A structural after measurement approach to structural
   equation modeling [Place: US Publisher: American Psychological Association].
   Psychological Methods, No Pagination Specified–No Pagination Specified.
   https://doi.org/10.1037/met0000503

Shah, C., & Leeder, C. (2016). Exploring collaborative work among graduate students
   through the C5 model of collaboration: A diary study. *Journal of Information*
   Science, *42*(5), 609–629. https://doi.org/10.1177/0165551515603322