

SOVEREIGN: How does the expert utilization distribution of SMOES on VQA-CP v2 compare to top-k routing when evaluated und

SOVEREIGN Research Kernel

Autonomous draft — Owner review required before publication

May 28, 2026

Abstract

Purpose The purpose of this paper is to provide a comprehensive, yet concise, overview of the considerations and metrics required for partial least squares structural equation modeling (PLS-SEM) analysis and result reporting. Preliminary considerations are summarized first, including reasons for choosing PLS-SEM, recommended sample size in selected contexts, distributional assumptions, use of secondary data, statistical power and the need for goodness-of-fit testing. Next, the metrics as well as the rules of thumb that should be applied to assess the PLS-SEM results are covered. Besides presen

1 Introduction

Analysis of: When to use and how to report the results of PLS-SEM. Research goal: How does the expert utilization distribution of SMOES on VQA-CP v2 compare to top-k routing when evaluated under varying degrees of domain shift, measured by per-expert load entropy and accuracy drop relative to in-domain performance?.

2 Methodology

Multi-query arXiv search (1 parallel queries, Relevance-sorted). TF-IDF cosine semantic verification (bigrams, threshold=0.15). NIM nv-embedqa-e5-v5 (dim=1024) for semantic indexing. Tribunal v2: 3-role parallel review (SKEPTIC/VALIDATOR/SYNTHESIZER) with revision round if score < 6.5.

3 Results

8 papers retrieved. 7 claims extracted, 7 verified. Tribunal: 8.0/10 → APPROVE (revision_round=0). Policy: AUTO_APPROVE.

4 Uncertainties

NIM free tier latency varies. TF-IDF verification is a weak signal. arXiv Relevance ranking is query-dependent. Tribunal consensus is LLM-based and prompt-sensitive.

5 Extracted Claims

Claim	Verified	Confidence
The paper provides a comprehensive overview of considerations and metrics required for PLS-SEM analysis and result repor	✓	0.28
The overview includes established PLS-SEM evaluation criteria as well as new guidelines such as PLSpredict for assessing	✓	0.31
The paper covers metrics for model comparisons and complementary methods for checking results' robustness	✓	0.23
Most previously applied metrics for evaluating PLS-SEM results are still relevant	✓	0.32
Scholars need to be knowledgeable about recently proposed metrics including model comparison criteria and endogeneity as	✓	0.28
The paper recommends using latent class analysis and PLSpredict as complementary methods for checking results' robustnes	✓	0.20
The paper provides an overview of previously and recently proposed metrics as well as rules of thumb for evaluating rese	✓	0.41

References

- <https://doi.org/10.1001/jama.1993.03500230097036>
- <https://doi.org/10.1108/ebr-11-2018-0203>
- <https://doi.org/10.1037/0033-295x.100.3.363>