LLM-Assisted Replication for Statistical Research in Social Science

So Kubota[1], Hiromu Yakura[2], Samuel Coavoux[3],
Sho Yamada[4], and Yuki Nakamura[5]

[1]Graduate School of Economics, Tohoku University

[2]Max Planck Institute for Human Development

[3]CREST, ENSAE, Institut Polytechnique de Paris

[4]Graduate School of Public Policy, The University of Tokyo

[5]Faculty of Engineering, The University of Tokyo

January 30, 2026

Abstract

The replication crisis, the failure of scientific claims to be validated by further research, is one of the most pressing issues for empirical research. This is partly an incentive problem: replication is costly and less rewarded than original research. Large language models (LLMs) have accelerated scientific production by streamlining writing, coding, and reviewing, yet this acceleration risks outpacing verification. To address this, we present an LLM-based system that replicates statistical analyses from social science papers and flags potential problems. Quantitative social science is particularly well-suited to automation because it relies on standard statistical models, shared public datasets, and uniform reporting formats such as regression tables and summary statistics. We present a prototype that iterates LLM-based text interpretation, code generation, execution, and discrepancy analysis, demonstrating its capabilities by reproducing key results from a seminal sociology paper. We also outline application scenarios including pre-submission checks, peer-review support, and meta-scientific audits, positioning AI verification as assistive infrastructure that strengthens research integrity.[1]