

Peer Review

Review of: "Statistically Significant Linear Regression Coefficients Solely Driven by Outliers in Finite-Sample Inference"

Muhammad Amin¹

1. University of Auckland, New Zealand

Dear Felix Reichel,

Firstly, I would like to commend you on the thoroughness and clarity of your research article titled "Statistically Significant Linear Regression Coefficients Solely Driven by Outliers in Finite-Sample Inference." Your investigation into the impact of outliers on linear regression coefficients is both timely and relevant, especially given the widespread use of linear regression in various fields of research.

Strengths of the Paper

- 1. Clarity and Structure:** The paper is well-organized and easy to follow. The introduction effectively sets the stage for the research, clearly outlining the problem and the importance of addressing it. The subsequent sections logically build on this foundation, providing a comprehensive exploration of the topic.
- 2. Methodological Rigor:** Your use of numerical simulations in R to demonstrate the impact of outliers on regression coefficients is a strong point. The comparison with robust Huber regression adds depth to your analysis, clearly illustrating the limitations of classical OLS regression and the benefits of robust methods.
- 3. Relevance and Contribution:** The topic is highly relevant to the fields of econometrics and statistics. Your findings highlight a critical issue that can lead to misleading inferences in regression analysis, which is a valuable contribution to the academic community.
- 4. Diagnostic Tools:** The detailed discussion on diagnostic statistics and tools for detecting influential observations is particularly useful. This section provides practical guidance for researchers on how

to identify and mitigate the impact of outliers in their own work.

Areas for Improvement

- 1. Robustness Checks:** While you have demonstrated the impact of a single outlier through simulation, it would be beneficial to include additional robustness checks. For instance, you could explore the effects of multiple outliers or different types of outliers (e.g., outliers in the dependent variable vs. outliers in the independent variable) to further validate your findings.
- 2. Real-World Data Application:** The paper would be strengthened by applying your methodology to a real-world dataset. This would provide empirical evidence of the practical implications of your findings and demonstrate how robust regression techniques can improve the reliability of statistical inferences in real-world scenarios.
- 3. Literature Integration:** Although you have cited relevant literature, a more comprehensive review of existing work on outlier detection and robust regression methods would enhance the paper's credibility. This could include a discussion of alternative robust regression techniques and their relative strengths and weaknesses.
- 4. Visual Aids:** The inclusion of more visual aids, such as additional residual plots or graphical representations of the impact of outliers, would help readers better understand the concepts and results. Visual aids can often convey complex ideas more effectively than text alone.
- 5. Discussion on Practical Implications:** While you have discussed the dangers of relying solely on p-values, a more detailed discussion on the practical implications of your findings for researchers and practitioners would be valuable. This could include specific recommendations on how to approach regression analysis in the presence of outliers and the potential consequences of ignoring these issues.

Specific Comments

Section 3: The example you provide to illustrate the impact of a single outlier is compelling. However, it would be helpful to include a brief explanation of how the outlier was generated and why it was chosen. This would provide context and make the example more relatable to readers.

Section 4: The discussion on single outlier tests is thorough, but it might be useful to include a brief comparison of these tests with other outlier detection methods. This would give readers a better understanding of the range of tools available and their respective advantages and disadvantages.

Section 5: The discussion on the normality assumption is clear, but it would be beneficial to include a brief overview of how violations of this assumption can be detected and addressed in practice. This could include a mention of specific statistical tests or diagnostic plots that researchers can use.

Conclusion

Overall, your paper provides a valuable contribution to the field by highlighting the critical issue of outliers in linear regression analysis. With a few additional enhancements, particularly in terms of robustness checks, real-world applications, and a more comprehensive literature review, the paper could be even stronger. I look forward to seeing the final version of your work and its impact on the academic community.

Declarations

Potential competing interests: No potential competing interests to declare.