

The Heavy Tails of Electoral Data

The Importance of Robust Estimators^{*}

Dan Baissa[†]

Carlisle Rainey[‡]

Abstract

Researchers studying the consequences of comparative electoral institutions, as well as other areas of political and social science, often estimate linear regression models on continuous outcomes of interest using least squares. These outcomes include measures of the number of political parties, proportionality, and vote share, among others. While it is well known that least-squares estimates are often sensitive to single, influential data point, this knowledge has not led to appropriate practices when using least-squares estimators. We highlight the importance of using more robust estimators (at least as a robustness check) and discuss several approaches to detect, summarize, and communicate the influence of particular data points. We conclude with a reanalysis of Clark and Golder (2006) and show that their conclusions depend on several influential data points. Removing these data or using a robust estimator substantially weakens their key conclusions about the conditional relationship between social heterogeneity and electoral rules in influencing the number of political parties.

^{*}We thank Bill Clark and Matt Golder for making their data available to us. The analyses presented here were conducted with R 3.1.0. All data and computer code necessary for replication are available at github.com/carlislerainey/meaningful-inferences.

[†]Dan Baissa is an M.A. student in the Department of Political Science, University at Buffalo, SUNY, 520 Park Hall, Buffalo, NY 14260 (kellymcc@buffalo.edu).

[‡]Carlisle Rainey is Assistant Professor of Political Science, University at Buffalo, SUNY, 520 Park Hall, Buffalo, NY 14260 (rcrainey@buffalo.edu).

Introduction

Appendix

The Heavy Tails of Electoral Data

A