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 important 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) an show that their conclusions depend on several influential data points. Removing these data or using a robust estimator substantially weaken 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

The linear regression model can be written as $E(y|X) = X\beta + \epsilon$, where y is an outcome variable of interest (usually roughly continuous), X is a $n \times (k+1)$ matrix containing a single column of ones and k columns holding k explanatory variables, β is a $(k+1) \times 1$ matrix of model coefficients, and ϵ is an $n \times 1$ matrix of errors. Researchers in political science commonly estimate this model with ordinary least squares (OLS) by minimizing the squared residuals, $\hat{\beta}^{OLS} = \arg\min S(b)$, where $S(b) = \sum_{i=1}^{n}(y_i - X_i b)^2$. That is, OLS estimators choose the estimate $\hat{\beta}$ that minimizes the sum of the squared residuals. Under the assumption that the errors ϵ_i follow independent and identical normal distributions with mean zero and unknown variance, the OLS estimator is the minimum variance unbiased estimator (MVUE).

Even if the errors do not follow independent and identical normal distributions, the Gauss-Markov Theorem guarantees the least-squares estimator is the best (i.e., minimum variance) *linear* unbiased estimator if the errors have mean zero and constant (and finite) variance. However, this should provide little comfort to researchers because their is little statistical or substantive reason to restrict themselves to *linear* estimators.

At first glance, one might take the linearity restriction under Gauss-Markov to refer to the structure of the model, such that $E(y|X) = X\beta$ falls into the class of "linear" regression models, but $E(y|X) = e^{X\beta}$ does not. Indeed, this is the sense in which we use "linear" in the term "linear regression." However, the "linear" restriction in the Gauss-Markov Theorem refers to a highly technical and obscure statistical criterion that requires that the estimates be a linear function of the outcome variables, so that $\hat{\beta}_j = \lambda_1 y_1 + \lambda_2 y_2 + \ldots \lambda_n y_n$, so that the weights λ_i are allowed to depend on X, but not on y.[1] In other words, Gauss-Markov does not require a linear *model* of the form $E(y|X) = X\beta$, but it does require a linear estimator of the form $\hat{\beta}_j = \lambda_1 y_1 + \lambda_2 y_2 + \ldots \lambda_n y_n$.

[1] Formally, linearity requires that $\hat{\beta} = My$, where M depends on the matrix X. For the case of least-squares, $M = (X'X)^{-1}X'$.

However, we argue that restricting ourselves to linear estimators is unnecessary and unproductive. There is no statistical reason to restrict ourselves to linear estimators, except for mathematical convenience, and there are substantive reasons to reject this restriction. For example, if the researcher is aware that one case has a unusually large outcome variable (conditional on the explanatory variables), then the researcher might wish to weight that case less than the other, more typical cases so that one atypical case does not exert several times more impact on the estimates than other, typical cases. Indeed, substantive researchers might wish to attach zero weight to extremely unusual cases because these cases might be due to a different substantive process.

It is not often appreciated that if the errors do not follow independent and identical normal distributions, then the OLS is no longer the MVUE–other estimators might outperform OLS.

Replication of Clark and Golder (2006)

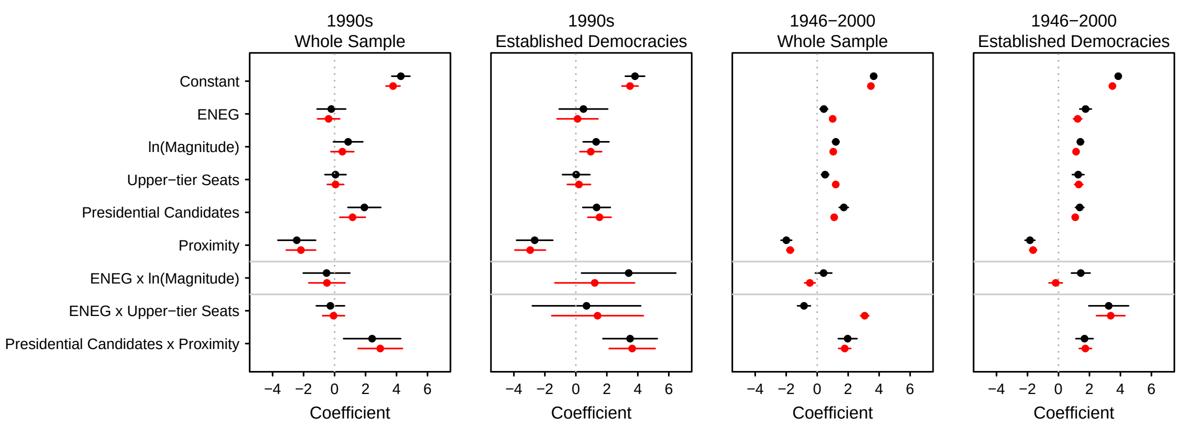

FIGURE 1: Replication of Clark and Golder (2006) using MM-estimation with explanatory variables standardized to have mean zero and standard deviation one-half. The black lines and points show the OLS estimates and 90% confidence intervals and the red lines and points show the MM estimates and confidence intervals. Notice that the coefficient for the product of the effective number of ethnic groups and the district magnitude changes drastically with the choice of estimator.

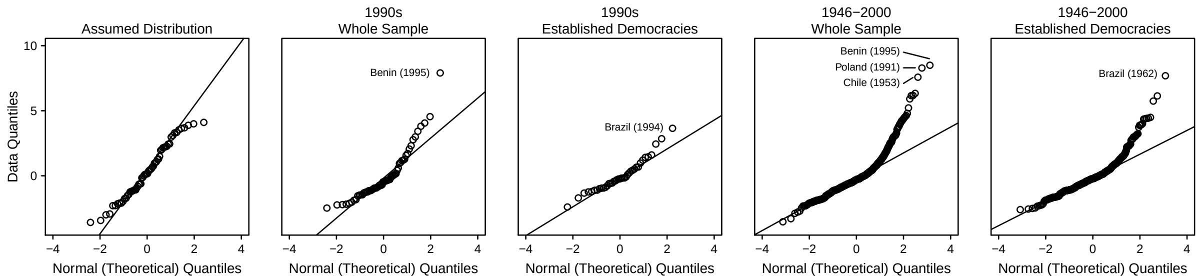

FIGURE 2: caption here.

Appendix

The Heavy Tails of Electoral Data

A