Appendix for "When BLUE Is Not Best"[*]

Daniel K. Baissa[†]

Carlisle Rainey[‡]

Replication of Clark and Golder (2006)

Clark and Golder (2006) attempt to "rehabilitate" Duverger's (1963) classic explanation for the number of political parties in a system. They write:

> According to Duverger, the mechanical effect of electoral institutions favoring large parties creates incentives for strategic entry and strategic voting. Parties without a chance of winning are encouraged to withdraw. If these parties fail to withdraw, then voters will have an incentive to vote strategically in favor of better placed parties. Thus disproportional systems with low district magnitudes are likely to reduce the demand for political parties created by social heterogeneity (p. 694).

For our replication, we focus specifically on their hypothesis:

> HYPOTHESIS: Social heterogeneity increases the number of electoral parties only when the district magnitude is sufficiently large.

This suggests that the marginal effect of social heterogeneity should be positive and statistically significant under permissive electoral rules (i.e., large district magnitude) and about zero and statistically insignificant (though see Rainey 2014) under restrictive electoral rules (i.e., district magnitude near one).

To measure their key outcome variable, the number of political parties, Clark and Golder use the *effective* number of political parties (ENEP), where $\text{ENEP} = \sum_{i=1}^{n} \frac{1}{v_i^2}$, for v_i represents the vote share of party i in the election. Similarly, for their measure of social heterogeneity,

[*]We thank Bill Clark and Matt Golder for making their data available to us. The analyses presented here were conducted with `R` 3.2.2. All data and computer code necessary for replication are available at github.com/carlislerainey/heavy-tails.

[†]Daniel K. Baissa is a Ph.D. student in the Department of Government, Harvard University, 1737 Cambridge St., Cambridge, MA, 02138 (dbaissa@g.harvard.edu).

[‡]Carlisle Rainey is Associate Professor of Political Science, Florida State University, Room 531B Bellamy Building, 113 Collegiate Loop, Tallahassee, FL, 32306 (crainey@fsu.edu).

Clark and Golder use the effective number of ethnic groups (ENEG), taken from Fearon's (2003) measure of ethnic fragmentation. To measure the restrictiveness of the electoral rules, Clark and Golder use the average district magnitude in the election.

Figure 1 plots the key explanatory variables, district magnitude and the effective number of ethnic groups, along the horizontal and vertical axes, respectively. The size of the points indicates the effective number of political parties for each election. The hypothesis predicts that the large points should lie in the upper-right section of the plot and small points should lie near the horizontal and vertical axes.
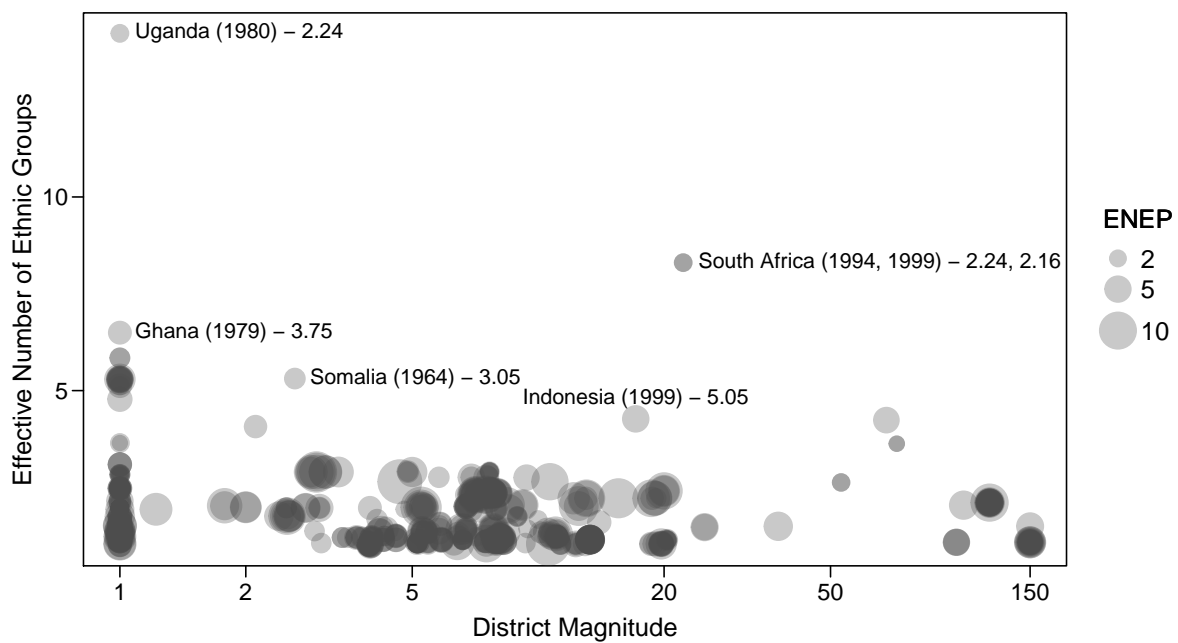
Figure 1: The distribution of district magnitude (on the log scale) and ethnic heterogeneity. Note: the point sizes indicate the number of political parties. According to Clark and Golder's hypothesis, large points should lie in the upper-right portion of the plot and small points should lie along the axes.

To test this hypothesis, Clark and Golder fit the following linear model using least squares:

$$\begin{aligned}
\text{ENEP}_i = \beta_0 &+ \beta_1 \text{ENEG}_i + \beta_2 \log(\text{Magnitude}_i) + \beta_3 \text{Upper-Tier Seats}_i \\
&+ \beta_4 \text{Presidential Candidates}_i + \beta_5 \text{Proximity}_i \\
&+ \beta_6 \text{Ethnic}_i \times \log(\text{Magnitude}_i) + \beta_7 \text{Ethnic}_i \times \text{Upper-Tier Seats}_i \\
&+ \beta_8 \text{Presidential Candidates}_i \times \text{Proximity}_i + \epsilon_i \; ,
\end{aligned} \tag{1}$$

The first key coefficient in this analysis is β_1, which captures the effect of social heterogeneity when district magnitude is one (i.e., the log of district magnitude is zero) and there are no upper-tier seats. According to the hypothesis, β_1 should be about zero. The second key coefficient is β_6, which captures how the effect of social heterogeneity changes with the electoral rules. According to the hypothesis, β_6 should be positive, so that the effect of social heterogeneity becomes (perhaps more) positive as the district magnitude increases.

Clark and Golder use least squares to obtain their estimates of the model coefficients, but worry about their estimates of the standard errors. They write that "[t]he crucial thing to remember is that although OLS is consistent with longitudinal data, the standard errors may be incorrect" (p. 690). They discuss several options and ultimately settle on robust standard errors clustered by country, but show that their conclusions are robust to alternative approaches to estimating standard errors. However, they do not address the possibility of a non-normal error distribution or its potential impact on the coefficient estimates. This is especially concerning given that the effective number of electoral parties is bounded below by zero, perhaps creating an error distribution with a skew to the right.

To get an initial sense of how the results might change using an alternative (perhaps more efficient) estimator, we replicated the estimates from four models in their Table 2 using least squares and the biweight estimator. For these initial estimates, we make no attempt to account for the clustered nature of the data in calculating the standard errors, but do supply the usual 90% confidence intervals to serve as a lower-bound on the uncertainty. Figure 2 presents these estimates and confidence intervals.

The crucial estimate $\hat{\beta}_6$ changes substantially depending on the choice of estimator. This key estimate, which the theory suggests should be positive, remains negative in the 1990s sample including new democracies, shrinks substantially toward zero in the 1990s sample that includes only established democracies, and *becomes negative* in the large sample of countries from 1946–2000 that includes new democracies *and* the large sample that only includes established democracies. When results depend on the choice of estimator, it is especially

3
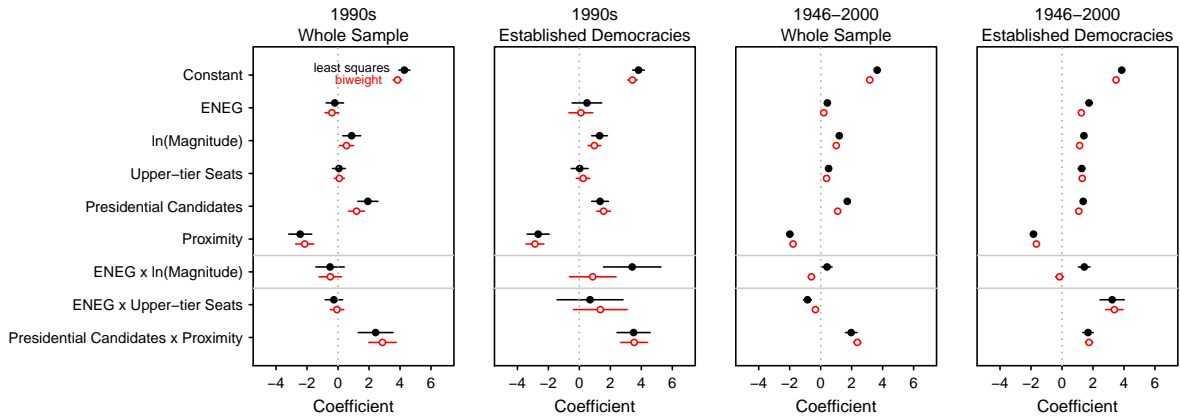
Figure 2: Coefficient estimates of Clark and Golder's (2006) linear model using a least-squares and biweight estimator.
Note: the explanatory variables are standardized to have mean zero and standard deviation one-half. The black lines and points show the least squares estimates and 90% confidence intervals and the red lines and points show the biweight 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.

important to carefully examine the residuals.

For the remainder of our analysis, we focus on the estimates from large sample of countries from 1946–2000 that includes only established democracies. Figure 3 presents the histogram of the residuals from the least-squares estimates and Figure 4 presents the QQ plot for these residuals. Both figures indicate a substantial skew to the right. While this does not necessarily lead to biased estimates, it does, in our view, suggest that the linear model would be more appropriate for a transformed outcome variable. If the transformation makes the errors more closely approximate a normal distribution, then the least squares estimators will be more efficient. The model for the transformed outcome can capture potentially interesting substantive effects as well.

The maximum likelihood estimate of the Box-Cox transformation parameter λ is about $-\frac{1}{3}$ and the confidence interval does not include zero, which suggests that a log-transformation does not quite eliminate the skew. We re-estimated the model using both a log-transformation and Box-Cox transformation with $\lambda = -\frac{1}{3}$. Figure 5 presents the histograms of the residuals from these two regression models. Notice that log-transforming the effective number of electoral parties does not quite eliminate the skew in the residuals. However, the Box-Cox transformation with $\lambda = -\frac{1}{3}$ produces highly symmetric residuals.

Figure 6 shows the QQ plot for the residuals. The left panel confirms that a right-skew
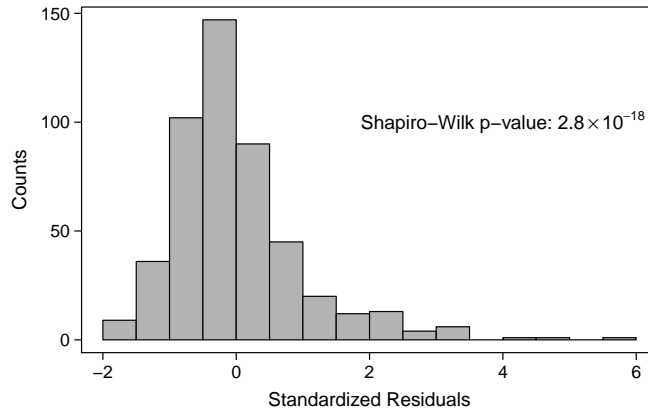
Figure 3: A histogram of the residuals from Clark and Golder's (2006) main model.
Note: the residuals are not approximately normal, but have a strong skew to the right. For example, one would rarely expect to observe residuals more that three standard deviations from zero if the assumption of normality holds. In these data, we have several residuals more that three standard deviations away and one nearly six standard deviations away, suggesting that some transformation of the outcome variable might be useful.
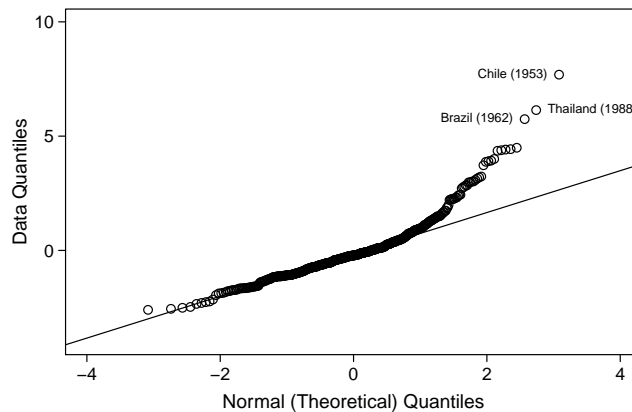

Figure 4: QQ plot of the residuals in Clark and Golder's (2006) main model.
Note: if the residuals were approximately normal, then the points in the QQ plot would approximately follow the plotted line. However, notice that the positive residuals deviate sharply from the theoretical expectations. This also suggests that some transformation of the outcome variable might be useful.

remains after the log-transformation, as suggested by the histogram in the left panel of Figure 5. The right panel of Figure 6 confirms that the Box-Cox transformation removes much or all of this skew, as suggested by the right panel of Figure 5.

Although the Box-Cox transformation removes much or all of the skew, the residuals retain
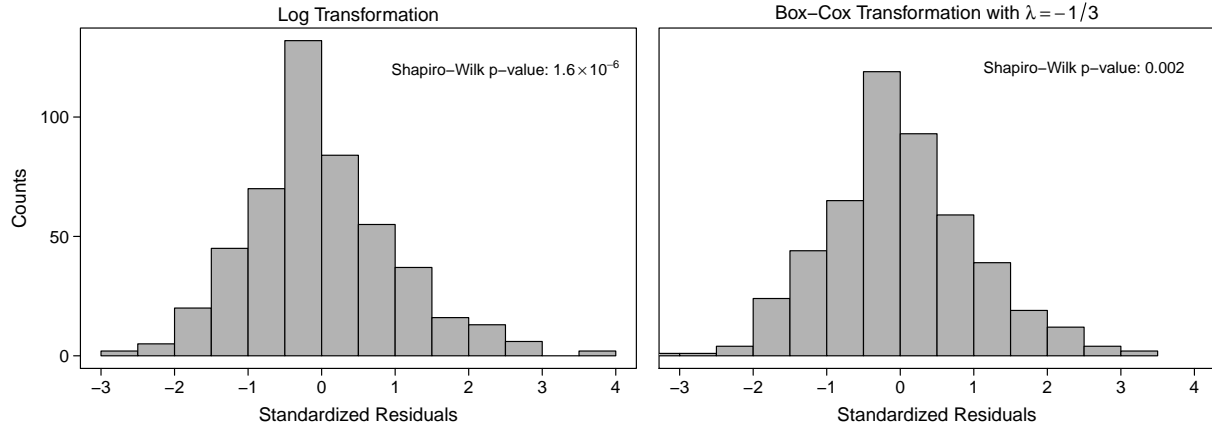
Figure 5: Histograms of the residuals after transforming the outcome variable.
Note: the left panel shows that the log transformation does not quite remove all of the skew, but the right panel shows that the Box-Cox transformation with $\lambda = -\frac{1}{3}$ creates an approximately symmetric error distribution.
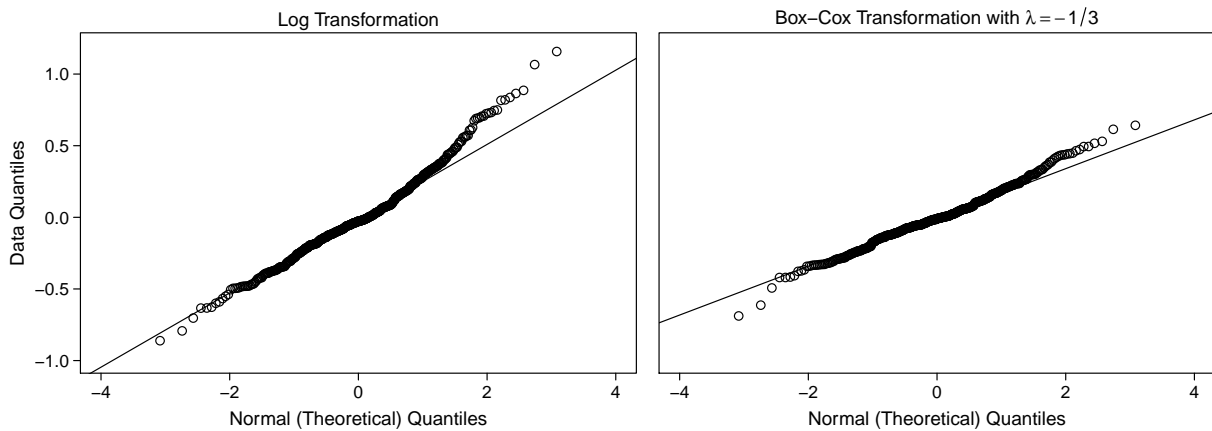
Figure 6: QQ plots of the residuals after transforming the outcome variable.
Note: the left panel shows that the log transformation does not quite remove all of the skew, but the right panel shows that the Box-Cox transformation with $\lambda = -\frac{1}{3}$ creates an approximately symmetric error distribution, though it has heavier tails than a normal distribution.

tails that are slightly heavier than the tails of the normal distribution. The residuals have larger positive and negative quantiles than one would expect under a normal distribution. This suggests the residuals follow a heavy-tailed distribution, perhaps resembling a t distribution with degrees of freedom in the range of six to twelve. Even after the Box-Cox transformation, the Shapiro-Wilk test rejects the null hypothesis of normality with $p = 0.002$. Assuming that these residuals follow a t-distribution, we can estimate the degrees of freedom using

maximum likelihood, which points toward a t distribution with about ten degrees of freedom. Recall that the biweight estimator is a (slightly) more efficient estimator for t_{10} distributed errors. However, because the least squares estimator eliminates large errors at the expense of many small and moderate errors, the error distribution might have *even heavier* tails than the least squares estimates suggest.

Most importantly, after the Box-Cox transformation, the residuals are well-behaved. They are highly symmetric and only slightly heavy-tailed. In spite of a nearly ideal situation, the simulations in Table 1 and Figure 2 in the main text suggest that the biweight estimate is slightly more efficient in this application. But these differences are slight so it can be valuable to carefully examine a least squares estimator and a more robust estimator.

Quantities of Interest

To estimate the quantity of interest—the effect of social heterogeneity as the permissiveness of the electoral rules varies—we re-estimate Clark and Golder's model with and without the Box-Cox transformation using both the least squares and biweight estimators. To obtain standard errors, we use the cluster bootstrap suggested by Harden (2012) to calculate the confidence intervals for each model.

Figure 7 shows the effect of increasing social heterogeneity from ENEG = 1.06 (10th percentile) to ENEG = 2.48 (90th percentile) as the district magnitude varies. The upper-left figure replicates Clark and Golder's approach (except for the cluster-bootstrap confidence intervals) and replicates their finding, which they summarize:

> [These results] clearly illustrate that in established democracies, ethnic hetero-
> geneity significantly increases the number of parties once the electoral system
> is sufficiently permissive. This is exactly what Duverger's theory predicts. To
> be more specific, Figure 1a [our upper-left panel of Figure 7], based on the
> pooled model with established democracies, indicates that ethnic heterogeneity
> will increase the number of electoral parties once we move beyond nonpermissive
> electoral systems with single-member districts when [Magnitude = 1].

But this evidence breaks down once we adjust for the non-normality of the errors. The lower-left panel of Figure 7 shows that using the biweight estimator as a robust alternative to least squares produces estimated interaction in the opposite direction as the hypothesis predicts, though small and not statistically significant. The upper-right panel of Figure 7 shows that transforming the outcome variable to make the data more consistent with the
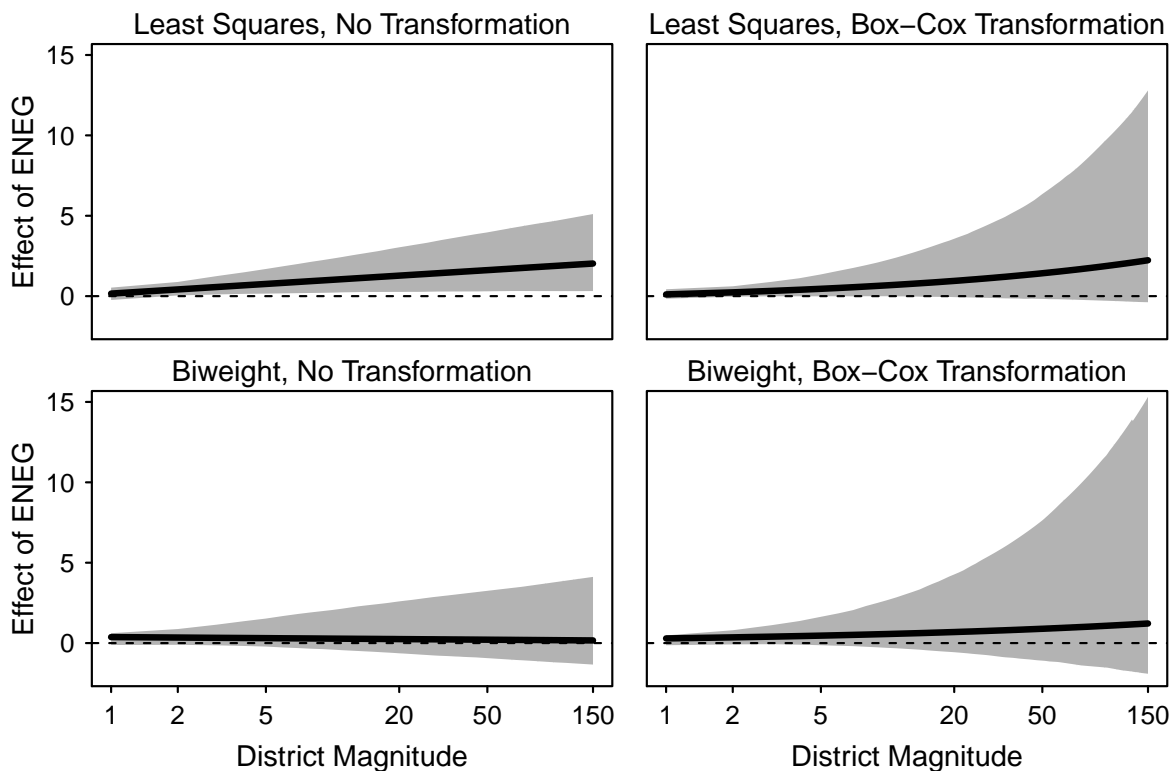
Figure 7: The estimated effects and 90% confidence intervals of increasing social heterogeneity from ENEP = 1.06 (10th percentile) to ENEP = 2.48 (90th percentile) on the effective number of electoral parties as the district magnitude varies.
Note: only the model that does not use robust estimators or transform the outcome variable provides evidence in support of the hypothesis.

assumed normal-linear model substantially increases the uncertainty across the range of district magnitude, so that small negative effects are now plausible, as well as much larger positive effects.

However, we argue that the best approach for these data is to transform the outcome variable to obtain a roughly symmetric error distribution *and* use the biweight estimator to address the remaining heavy tails shown in the left panel of Figure 6. This approach reduces effect of social heterogeneity somewhat across the range of district magnitude.

Because the scale of the estimates and uncertain differ drastically across the choice to transform the outcome variable or not, as well as the range of district magnitude, we present the relevant quantities of interest in Table 1. In this case, we are interested in the effect of substantially increasing social heterogeneity from ENEG = 1.06 (10th percentile) to ENEG =

2.48 (90th percentile) on the number of political parties when district magnitude is one (10th percentile) and also when district magnitude is 14 (90th percentile). Since the hypothesis suggests that this effect should be larger when the district magnitude is larger, we are also interested in the difference between these two effects. For simplicity, we focus our discussion on the differences between the typical approach, least squares with no transformation, and the approach we recommend, transformation *and* robust estimation.

The first row of Table 1 suggests that in countries with single-member districts, a substantial increase in ENEG (10th to 90th percentile) increases the ENEP by about 0.16 [-0.24; 0.53] parties. On the other hand, in large, multimember districts (magnitude of 14), the same increase in social heterogeneity increases the ENEP by about 1.14 [0.26; 2.67] parties. This is just as the hypothesis predicts. Further, this increase of 0.98 [0.06; 2.65] is large and statistically significant.[1]

	First-Difference When $M = 1$		First-Difference When $M = 14$		Second-Difference	
	Est.	90% CI	Est.	90% CI	Est.	90% CI
No Transformation						
Least Squares	0.16	[-0.24; 0.53]	1.14	[0.26; 2.67]	0.98	[0.06; 2.65]
Biweight	0.37	[-0.12; 0.62]	0.26	[-0.51; 2.33]	-0.11	[-0.96; 2.05]
Box-Cox Transformation						
Least Squares	0.11	[-0.16; 0.43]	0.80	[-0.03; 2.84]	0.69	[-0.19; 2.73]
Biweight	0.29	[-0.13; 0.48]	0.62	[-0.41; 3.34]	0.33	[-0.78; 3.14]

Table 1: The quantities of interest from least squares and biweight estimates, with and without the Box-Cox transformation of the outcome variable.
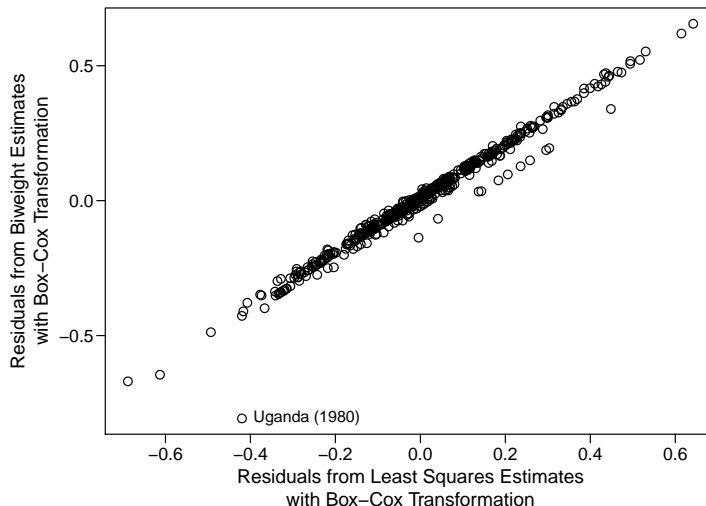Note: the least squares estimates without transforming the outcome variable are consistent with Clark and Golder's hypothesis. However, transforming the outcome variable, using the robust biweight estimator, or both substantially reduces the amount of evidence that these data offer in favor of the hypothesis.

However, once we make an effort to account for the non-normality of the residuals by transforming the outcome variable *and* using the robust biweight estimator, this evidence for interaction weakens slightly. This model suggests that, in single-member districts, a substantial increase in social heterogeneity increases the ENEP by about 0.29 [-0.13; 0.48] parties. This is about twice Clark and Golder's initial estimate. In large, multimember districts (magnitude of 14), the estimate shrinks to 0.62 [-0.41; 3.34], which is about half of Clark and Golder's estimate. This leads to an estimated increase of 0.33 [-0.78; 3.14] in the effect of social heterogeneity as we move from single-member districts to large, multimember districts—about one-third of Clark and Golder's initial estimate.

---

[1]Using cluster-robust standard errors, Clark and Golder find that the product term is *not quite* significant. We use cluster-bootstrap standard errors and find the coefficient is *barely* significant.

Differences Between the Estimates

One major advantage of robust estimators is that these estimators allow unusual cases to stand out. Figure 8 compares the residuals from the least squares fit and the biweight fit after the Box-Cox transformation. Notice that the residuals tend to agree, except for the 1980 election in Uganda, which is the largest residual in the biweight fit, but does not stand out among the least squares residuals.

Figure 8: The relationship between the least squares and biweight estimates after the Box-Cox transformation.
Note: the biweight estimate allows the unusual case of Uganda to stand out from the others.

Similarly, Figure 9 presents the 35 smallest weights from the biweight fit. Recall that as cases become increasingly inconsistent with most of the data, the biweight estimator increasingly downweights these cases, potentially to zero. As we might expect, given the residuals in Figure 8, the 1980 election in Uganda receives zero weight. This election is consistent with the hypothesis because it features single member districts, an extremely large number of ethnic groups, and only two major political parties. Unfortunately for the hypothesis, though, this case is inconsistent with most of the data.
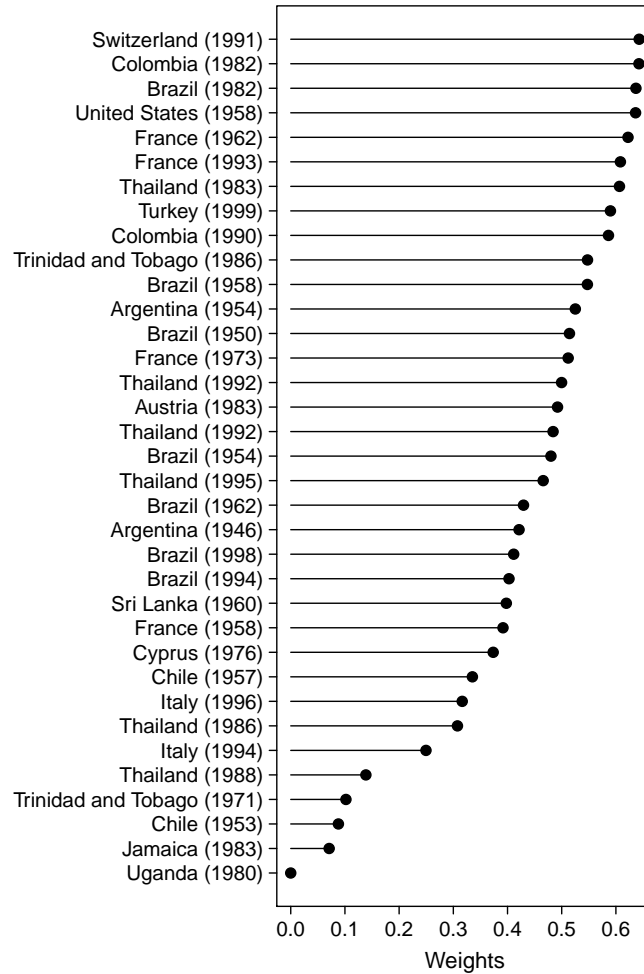
Figure 9: The final weights implied by the biweight estimator.
Note: because the 1980 election in Uganda is quite different from the remainder of the data, the biweight estimator downweights this case all the way to zero.

References

Clark, William Roberts, and Matt Golder. 2006. "Rehabilitating Duverger's Theory: Testing the Mechanical and Strategic Modifying Effects of Electoral Laws." *Comparative Political Studies* 39(6):679–708.

Duverger, Maurice. 1963. *Political Parties: Their Organization and Activity in the Modern State.* New York: Wiley.

Harden, Jeffrey J. 2012. "Improving Statistical Inference with Clustered Data." *Statistics, Politics, and Policy* 3(1):1–27.