

MLE Midterm

Katie Webster

March 16, 2015

1 Analysis

I calculated three linear models: model, modelListDel, and modelAmelia. Figure 1 displays the coefficient estimates and confidence intervals for each model: the black, baseline estimates correspond to model, red estimates correspond to modelAmelia, and blue estimates correspond to modelListDel.

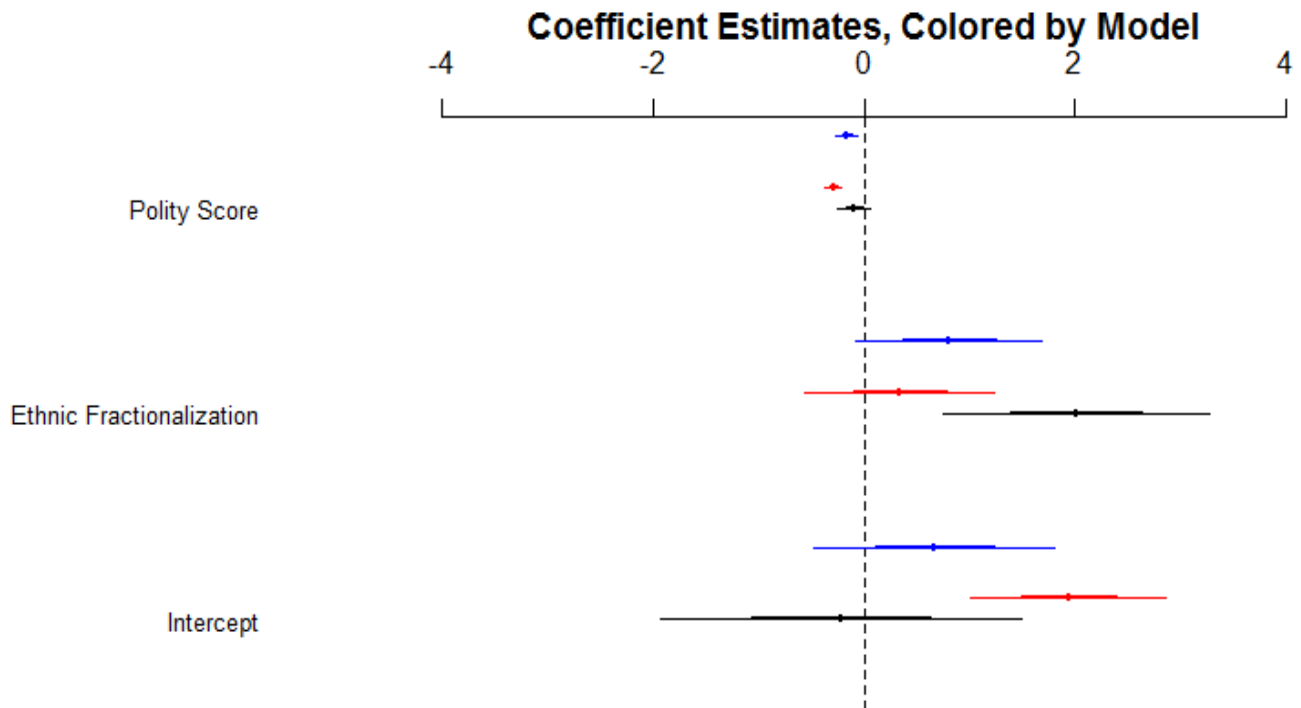


Figure 1 shows that there are some noticeable differences between each model's coefficient estimates. The modelListDel (blue) estimates are closer than the modelAmelia (red) coefficient estimates to the original model's estimates. Table 1 shows that the Amelia model has the lowest standard errors and p-values out of all three models, such that the estimates have a much higher significance than those in the base model (or the Listwise Deletion model). The Listwise Deletion model's standard errors are also smaller than the baseline model's, underscoring the fact that, when data is missing, the statistical significance of coefficient

estimates are likely to be biased upwards.

Table 1: Standard Errors of Each Coefficient, by Model

Model	<i>Baseline</i>	<i>ListwiseDeletion</i>	<i>Amelia</i>
Polity	0.07959	0.05215	0.0383
ELF	0.63132	0.44353	0.45215
Intercept	0.85367	0.56938	0.45975

That the Amelia model performs worse than the Listwise Deletion model might seem surprising at first glance, because the Amelia model is supposed to impute the missing data from `modelListDel`, so one might expect more accurate coefficients. However, imputation works best if the missing data are randomly missing; here, the missing data is not missing at random. Polity scores are missing only for the six highest values of the dependent variable (the Gini coefficient). This would bias the imputation results and explain why the Listwise Deletion model actually produces coefficient estimates that are closer to the original model's estimates.