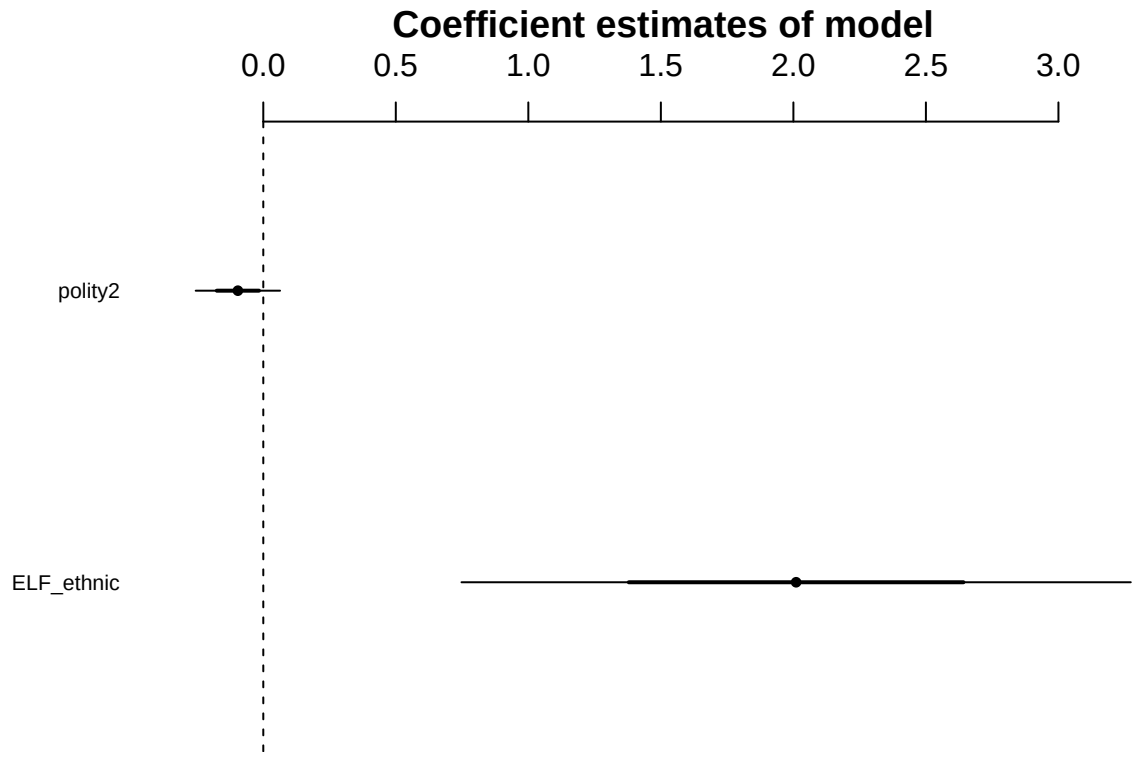
Midterm

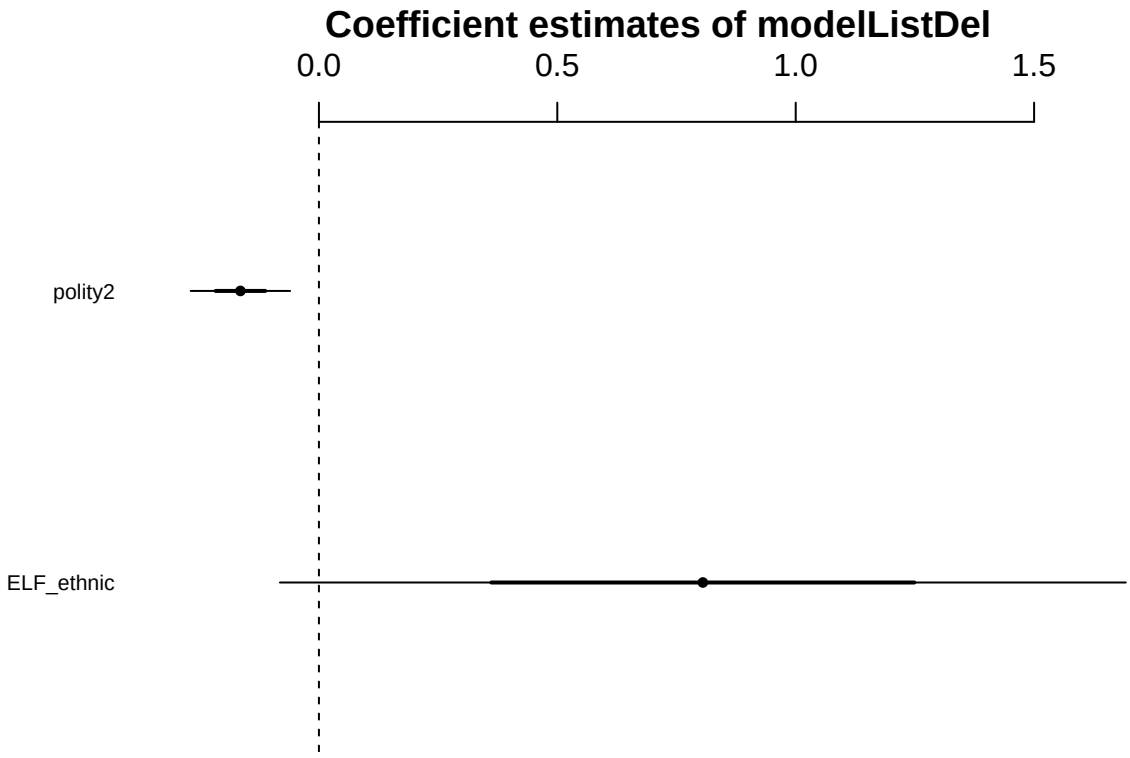
Devin McCarthy

Tuesday, March 17, 2015

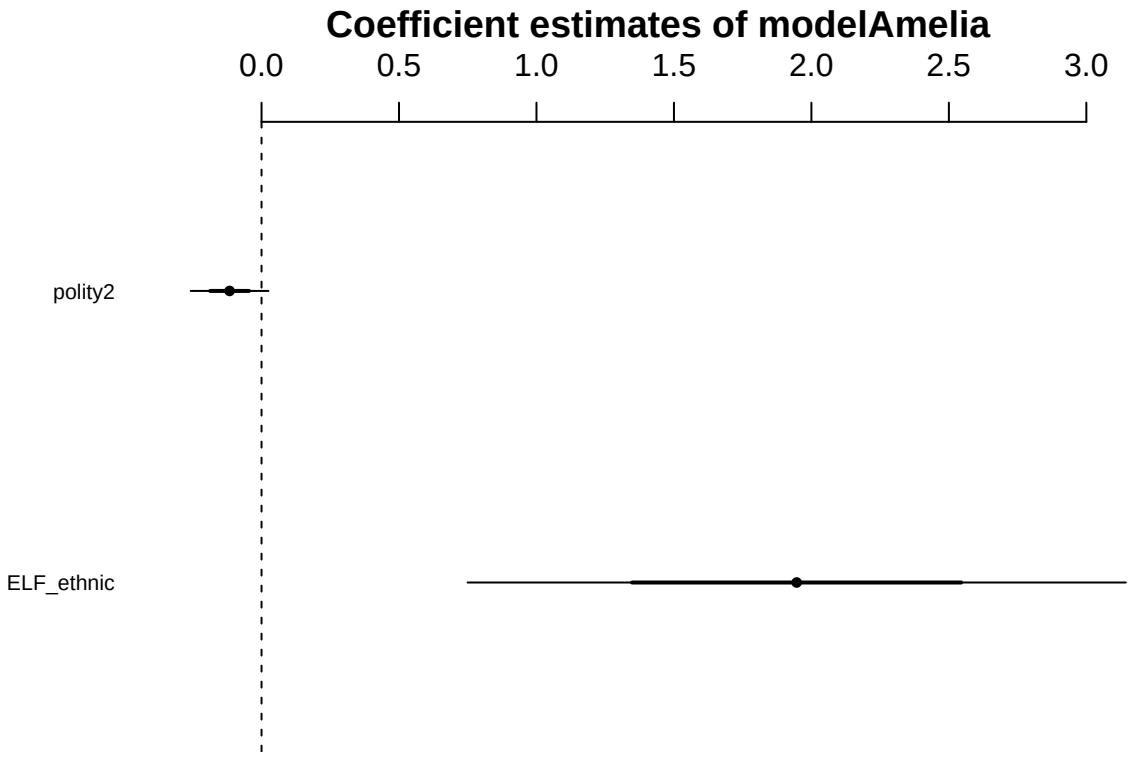
Three models are used to analyze this dataset: "model", "modelListDel", and "ModelAmelia." The first "model" is a straightforward OLS regression on the complete data set. Because no data is missing in this model, its coefficient estimates can be considered to be the "true" values of the coefficients. The ELF_ethnic variable in this model has a coefficient of 2, while the polity2 variable has a coefficient of -.1. Only the ethnic fractionalization variable is significant at a p-value of $< .05$. The plot below displays these coefficients as well as confidence intervals for the estimates.



The second model, "modelListDel", uses a data set identical to the first except that six observations in the polity2 variable have been eliminated. The model uses a listwise deletion approach to dealing with the NA values. This has a considerable effect on the model's estimates: the coefficient of ELF_ethnic drops to .8 and the coefficient of polity2 decreases to -.16. Furthermore, polity2's effect is now significant at a p-value of $< .05$.



The third model, "modelAmelia," uses a multiple imputation approach, Amelia, to dealing with the NA values in the polity2 variable. The imputed data produces coefficients much closer to the "true" data than the model using listwise deletion. The coefficient on ethnic fractionalization is 1.95, and the coefficient on the polity score is -.12. ELF_ethnic is significant at a p-value of $< .05$, and polity2 is no longer significant. These results suggest that the Amelia method is far superior to listwise deletion at approaching accurate coefficient estimates when data is missing.


Coefficient estimates of modelAmelia
0.0
0.5
1.0
1.5
2.0
2.5
3.0
polity2
ELF_ethnic