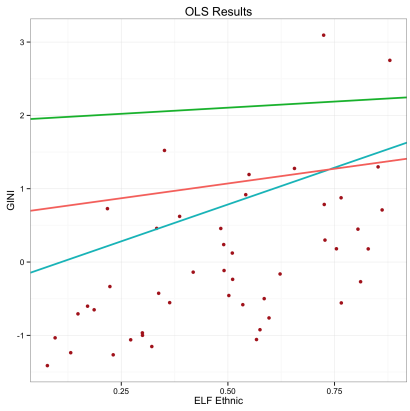
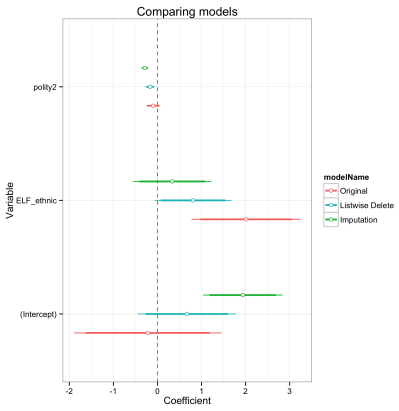
Midterm

Spencer Dorsey

March 16, 2015





Presented in the above figures are the visual regression results and coefficient plot comparing the various models. Upon initial analysis it is clear that, compared to the regression run on the complete data set, the listwise deletion regression and the imputation regression each have advantages and disadvantages. As seen in the OLS Results figure (which show the relationship between the independent variable of interest and the dependent variable), the listwise deletion method produced estimates that are generally closer to the original (this is also show clearly in the Comparing models figure) but the slope of the estimates is significantly different. The imputed model has a much closer slope to the original but a significantly different intercept.

Despite the apparent success of the listwise deletion method, I would argue the the imputation method is still more reliable as the characteristics of the relationship between the variables is maintained to a higher degree (with the obvious exception of the intercept).

It also should be noted that these results reflect only a single imputation while the generally accepted minimum is five imputations. Additionally, imputation will generally be more successful when the missing data is more-or-less randomly distributed. If the data that was excluded in this particular case is the result of a specific pattern then the imputation may be biased (explaining the divergent intercepts between the original and imputed models).