

POLSCI 733: Midterm (Assignment 7)

Moohyung Cho

March 17, 2015

In this analysis, I analyze the impact of *ethnic fractionalization* and the level of democracy (measured by *POLITY score*) on *income inequality*. In particular, I compare three models with different data manipulations: (1) original model with full data, (2) the model with ruling out missing data (listwise deletion), and (3) the model with imputation by using Amelia. Table 1 presents the regression results of three models. Figure 1 is the coefficient plots of three models.

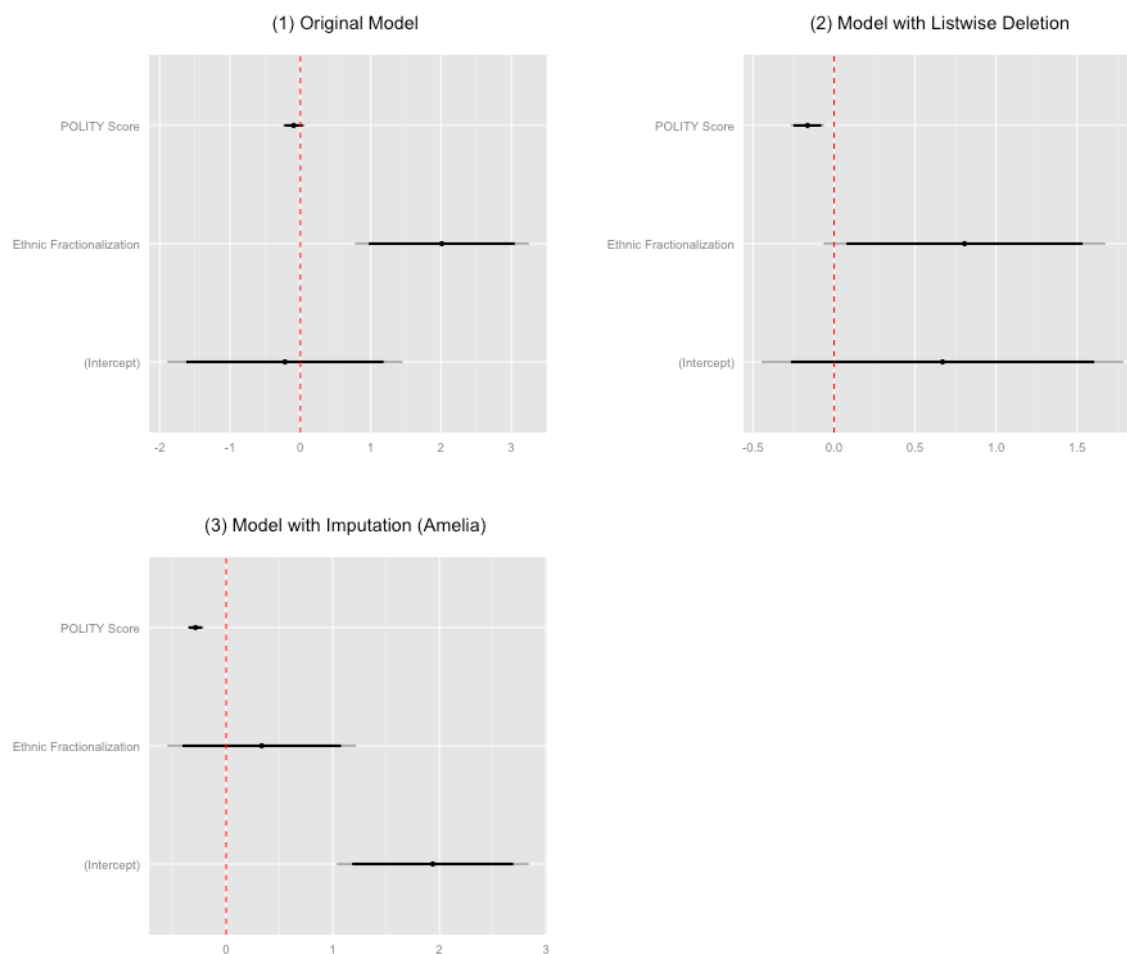
Table 1: Comparison of Three Models

	Dependent Variable: Income Inequality		
	(1) Original Model	(2) Listwise Deletion	(3) Imputation (Amelia)
Intercept	-0.220 (0.854)	0.669 (0.569)	1.938*** (0.460)
Ethnic Fractionalization	2.010*** (0.631)	0.805* (0.444)	0.334 (0.452)
POLITY Score	-0.096 (0.080)	-0.165*** (0.052)	-0.287*** (0.038)
Observations	46	40	46
R ²	0.340	0.448	0.704
Adjusted R ²	0.309	0.418	0.690
Residual Std. Error	0.831 (df = 43)	0.514 (df = 37)	0.557 (df = 43)
F Statistic	11.057*** (df = 2; 43)	20.891*** (df = 2; 37)	51.113*** (df = 2; 43)

Note:

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Figure 1: Coefficient Plots of Three Models



In original Model (1), only ethnic fractionalization attains statistical significance at $p < 0.01$, and other variables are not significant. The positive coefficient of ethnic fractionalization indicates that one unit increase of the level of ethnic fractionalization increases income inequality by approximately 2. In Model (2), 6 observations are missing and excluded in regression analysis. While the statistical significance of ethnic fractionalization is weak ($p < 0.1$), POLITY score is negatively associated with income inequality and statistically significant at $p < 0.01$, indicating that income inequality decreases as countries are more democratic.

In Model (3), missing data are imputed by Amelia. The regression result shows that POLITY score is still statistically significant at $p < 0.01$, whereas ethnic fractionalization does not reach statistical significance. The model also suggests that the negative impact of POLITY score on income inequality is stronger than in the model with listwise deletion.

(The coefficient of POLITY score decreases from -0.165 to -0.287.) Furthermore, it should be noted that R^2 in Model (3) is much larger than that in other two models: it jumps to over 0.7, indicating that the model with imputation better explains the original data compared to other two models. In the same vein, the value of F -statistic also jumps to over 51.

In comparison with the model with listwise deletion, the imputed model makes a significant explanatory variable (POLITY score) more powerful. In addition, standard errors of variables in Model (3) is generally smaller than those in Model (2) (except standard error of ethnic fractionalization which is slightly larger in Model (3)). Such reduction of standard errors also shortens the length of confidence intervals in Model (3), which contributes to better estimation. It also increases R^2 and F -statistic, implying a better explanatory power than Model (2). In this respect, my analysis is consistent with the general claim that imputation is a better method for estimation than listwise deletion when dealing with missing data. However, I find a sharp contrast between original model and imputed model. Specifically, the statistical significances of each independent variable are reversed in Models (1) and (3). Based solely on this analysis, it is hard to figure out why such difference occurs and which model is more appropriate for estimation.