The Effect of Income on Amount to Live Normally in Russian Longitudinal Monitoring Survey Data using OLS, GEE, and Linear Mixed Models

We were interested in the relationship between a person's income on the amount of money they believe they need to live normally. In theory, the amount of money a person needs to live should be stable since products are the same price for every person. The Russian Longitudinal Monitoring Survey data from UNC Dataverse, though, had a range from 200 to 3,000,000 of unknown monetary units. These extremes suggested that factors either external or internal to the individual were affecting the believed amount of money to live normally. Specifically we considered the year and landscape status as external factors and age, gender, education, and occupation as internal factors along with income as the variable of interest. Since the data was longitudinal, we employed models that would group observations so as to account for dependencies within individuals over time. All models suggested a positive association between the squared rooted income and square rooted amount to live normally around 0.20-0.30 with no significant difference between the two documented genders once interactions between gender and other variables were considered.

We replaced income and amount to live normally with their respective square root to improve model performance. We started with OLS modeling as a baseline analysis to determine whether grouping the data was necessary. Without any grouping or interactions, the OLS model found that all the variables included to be significant at the 0.05 level and that income had an estimate of 0.287 with a significance at the 0.01 level. GEE, which allowed for clustering longitudinally, resulted in nearly identical estimates. A mixed linear model with a random effect longitudinally and random slopes for year gave different results such as to income with an estimate of 0.225. The model used did not provide significance measures but the t-values suggested that many of the variables could be significant, especially the intercept and income. Without any interactions, the models so far indicated gender to be significant in determining the amount to live normally.

We ran all three models again, this time introducing interactions between gender with age and education. The mixed linear models could not converge with so many interactions, so we limited the number of variables and interactions included to the ones that showed the most extreme t-values. Landscape status and occupation were served similar purposes as control variables and had similar t-values, so we only included landscape status. Education had the least extreme t-values, so we also dropped education. The resulting mixed linear model resulted in income having an estimate of 0.238 and a t-value that was ten times that of most of the other variables. The same inputs for OLS and GEE yielded an estimate of 0.307 with 0.01

significance for both and a significant difference between genders. We ran all the variables with the desired interactions for OLS and GEE since there was a greater chance of convergence, which resulted similarly again with an estimate of 0.289 for income with 0.01 significance and no significance for gender.

All models resulted in an estimate around 0.20-0.30 for income with 0.01 significance as well as an intercept around 9-10 with 0.01 significance. Income likely has a positive association with the amount to live normally, and there is a starting value of 100 even with no income. We did not include any observations that did not include income since there was no way to tell whether no response automatically indicated no income. The mixed model with gender interactions did not converge, so we cannot conclude whether there were gender differences in the amount to live normally. Only 14,000 of the 370,000 observations remained after data cleaning, so there were not enough observations to properly analyze all interactions. The OLS and GEE gender interaction models, though, suggested that there were no gender differences, only slight differences in the interaction of gender with other variables such as education. Interestingly, there were few differences between the OLS and GEE models, suggesting that there might have been too many dependencies between the longitudinal clusters not accounted for. The mixed model grouping both longitudinally and by year would have accounted for more of those dependencies.