Rebellion in Eastern Ukraine

In March and April 2014, rebels took up arms in Donbas (Donetsk and Luhansk). While the Ukrainian government took a decisive turn toward pro-Europe, these protesters waved Russian flags and proclaimed the establishment of Donetsk and Luhansk People's Republics. Over time, the rebels extended their territorial control, will 61% of the municipalities in Donetsk and Luhansk regions falling under rebel's control after one year. But surprisingly, 80% of municipalities did not experience any rebel violence. What explains the difference between municipalities that experienced rebel violence and those which did not?

Partly because Donbas has traditionally been dominated by Russian-speaking people, the Donbas uprising is often attributed to ethnolinguistic differences. However, the Euromaidan revolution was in part triggered by economic concerns: the pro-Russian president, Viktor Yanukovych, decided not to sign a free trade agreement with the EU, whereas the Maidan protesters insisted on establishing free trade with the EU. In the Ukrainian economy, Donbas plays an important role as the industrial heartland with 'the big three' industries: metals, mining, and machine. Reducing tariffs with the EU, as the Euromaidan protesters demanded, might expose the industry to international competition. For Ukraine, metals are internationally competitive while coal mining is subsidized by the state. Machine building industry is, on the other hand, vulnerable to trade shocks because they are highly dependent on Russia.

In this exercise, you will analyze the factors that differentiated municipalities that underwent rebel violence and those that didn't. How does the variation in industry affect the likelihood of rebel violence? Are economic factors equally important as the ethnolinguistic factors, when explaining the rebellion in Eastern Ukraine? This exercise is based on the following study:

Zhukov, Yuri. 2016. "Trading hard hats for combat helmets: The economics of rebellion in eastern Ukraine *Journal of Comparative Economics*, 44(1), pp. 1-15.

The data set contains 3037 observations, each of which is a municipality (i.e. cities, towns, villages) in Donetsk and Luhansk provinces. Variables in this data set are described below:

Name	Description
`municipality`	municipality name
`district`	district to which the municipality belongs
`region`	region to which the municipality belongs
`latitude`	latitude of the municipality
`longitude`	longitude of the municipality
`ind_machine`	proportion of the local labor force employed in machine-building industry
`ind_mining`	proportion of the local labor force employed in mining industry
`ind_metals`	proportion of the local labor force employed in metals industry
`russian_lang`	proportion Russian-speaking people in the 2001 Ukrainian Census
`forest`	forest cover
`elev`	elevation
`dist_road`	distance to the nearest road
`pop_density`	population density
`dist_russia`	distance to the Russian border

Name	Description
`yanukovych`	percent of the popular vote received by Viktor Yanukovych in the 2010 presidential election
`rebel_violence`	incidence of any rebel violence in the municipality
`until_control`	number of days until rebel control

Question 1

We begin by familiarizing ourselves with the region's geography, using `latitude` and `longitude` to visualize the geographical location of municipalities.

a) Create a scatterplot of municipalities, where each region (`Donetsk` and `Luhansk` in `region`) is marked with different colors. Find four districts (`district`) with the highest population density (`pop_density`), and mark municipalities in each of the four districts in the plot, using different colors and labeling with their district names.

b) Create a new variable `industry` that indicates which industry (`metals`, `mining`, or `machine`) had the highest proportion of local labor force. Create a new scatterplot of municipalities, where each of the three dominant industry is marked with different colors. Add crosses for municipalities that experienced any rebel violence (`rebel_violence`). What relationships does the scatterplot reveal between the dominant industry and the incidence of rebel violence?

c) Create a new variable `russian_quintile` that divides the municipalities into five groups, based on the quintile of the `russian_lang` variable. Repeat the same analysis as (b). Create a new scatterplot of municipalities, where each of the five groups are marked with different colors. Add crosses to municipalities that experienced any rebel violence (`rebel_violence`). What relationships does the scatterplot reveal between the dominant language and the incidence of rebel violence?

Question 2

We examine if there exists any difference in the experience of rebel violence, between municipalities with strengths in different industries. For the sake of simplicity, we assume independence across observations throughout this exercise.

a) Use the new variable `industry` that you created in the previous question. For the group of municipalities with strength in machine, mining, and metal industries, estimate the probability of experiencing any rebel violence. For each industry, report both the point estimate and its 95% confidence intervals.

b) Perform two-sided hypothesis tests to examine if the municipalities with strength in each of three industries experienced different degrees of rebel violence. Conduct two tests comparing between machine and mining, as well as between machine and metals. For each test, state both the null and alternative hypotheses explicitly and interpret the result.

c) Repeat the same analysis with the language factor. Create a new variable `russian_maj` that represents whether Russian language is dominant (whether the proportion of Russian speakers is above 0.5 or not), and perform a two-sided hypothesis test for examining if there's difference between these two groups in the degree of rebel violence. For each test, state both the null and alternative hypotheses explicitly and interpret the result.

Question 3

The incidence of rebel violence, which we examined in the previous question, does not necessarily mean that the municipalities fell under the rebel control. In this question, we analyze the determinants of territorial control, using linear regression. As the outcome variable, create a new binary variable `rebel_control`, which takes 1 if the municipality fell under rebel control and 0 otherwise. Note that those municipalities without any experience of rebel control (i.e., `nocontrol`) are marked as `1650` in the variable `until_control`.

a) Fit a linear regression by regressing the variable `rebel_control` on the proportion of each industry (metals, mining, and machine) as well as the proportion of Russian-speaking people. Interpret the coefficients including an intercept. State the underlying assumption for valid inference under the linear model. Be sure to also comment on the goodness-of-fit of the model.

b) Fit the same linear regression model as in Question 3(a), but add all six covariates (`forest`, `elev`, `dist_road`, `pop_density`, `dist_russia`, and `yanukovych`). Interpret the result. How do the covariates affect the results and the plausibility of the assumption you stated in 3(a)?

c) Fit the same linear regression model as in Question 3(b), but add the interaction terms between each industry and the proportion of Russian-speakers. Interpret the result. How do the interaction terms affect the results and the plausibility of the assumption you stated in 3(a)?

Question 4

Donbas is comprised of two regions: Donetsk and Luhansk. Some cities in Donetsk region have strong industrial sectors. For instance, Mariupol is well-known for the metal industry, as represented by Azov Steel Plant.

a) Conduct the two-sided hypothesis tests to examine if there is any significant difference between the municipalities in `Donetsk` and `Luhansk` regions (using the `region` variable). Start by comparing the outcome variable (`rebel_control`) between the two regions. Also, compare the explanatory variables (`ind_machine`, `ind_mining`, `ind_metals`, `russian_lang`) beteen the two regions. Interpret the result.

b) Restrict the sample to the municipalities in Luhansk, in order to rule out the possibility that the analysis in the previous questions might be influenced by the large industrial towns in Donetsk. With this new subset of the data, conduct the regression analysis from Q3 (a), (b), and (c) again. Interpret the results. What are the important factors that could explain rebel control in Luhansk based on this analysis? Comment on the prediction error as well: how well do the models predict rebel control of the municipalities? How do the ability to predict changes with different models?