Dale Coleman

15 March 2017

INFX 572 A

Final Paper: Regression Analysis: 1907 Romanian Peasants’ Revolt

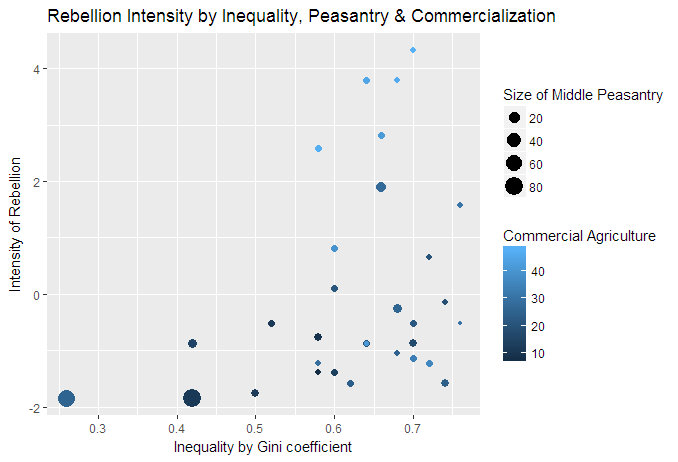
The dataset I selected for this paper contains five variables related to the 1907 Romanian Peasants’ Revolt. The “Chirot” dataset comes from the “car” package. The data originates from a study by Chirot and Ragin, published in *American Sociological* Review, examining the effects of the market and traditionalism on this famous peasant rebellion (1975). I selected this dataset primarily because of my interest in socioeconomic systems, and the root causes of organized and spontaneous mass resistance action. On a practical level, I found this dataset to be clean and complete, lending itself nicely to regression work (which was its initial intended purpose).

Each observation in this dataset represents a county in Romania in 1907. Individual counties are not specified, but that level of specificity is not required to answer the question I will be asking (although it would have been nice to be able to create a geographical visualization of the data). All five variables are continuous numeric variables, representing ordinal indices of factors relating to the revolt. They are as follows (437):

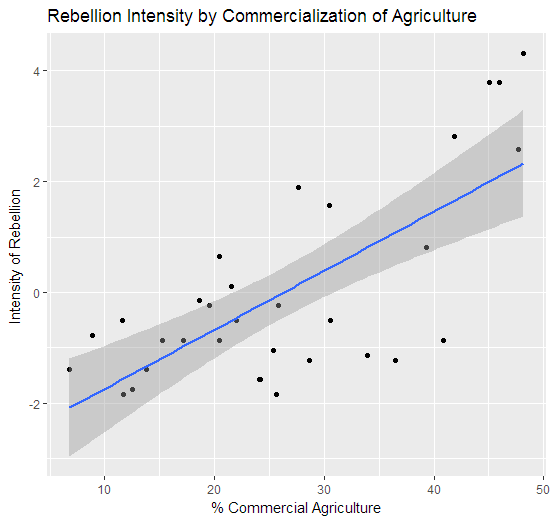
|  |  |
| --- | --- |
| intensity | Measure rebellion intensity, measured by combining the standardized spread of rebellion by percentage of villages and the standardized violence of the rebellion by number of deaths. |
| commerce | Commercialization of agriculture, measured by percentage of arable land devoted to wheat |
| tradition | Traditionalism, measured by percentage of illiteracy in rural population |
| midpeasant | Middle peasantry strength, measured by percentage of land owned in units of 7 to 50 hectares |
| inequality | Gini coefficient of inequality |

Of these variables “intensity” and “tradition” stand out as potentially problematic. In the paper the authors point out the difficulty of measuring the intensity of a rebellion (433). In the end they create a metric based on percent of villages participating and number of deaths during the revolt. As we will see later, this increases the ambiguity of our regression result, but the method underlying this metric is solid. The “tradition” metric could more aptly be described as the rural illiteracy percentage. The original study includes a detailed explanation of why rural literacy rates are reliable proxy statistics for traditionalist worldviews, which I will forego here (435). The logic behind the remaining statistics is pretty self-explanatory.

Like the original study, I am interested in demonstrating and visualizing the factors contributing to the peasant revolt. Specifically, I would like to know **what effect, commercialization, traditionalism and inequality had on the intensity of revolt by county.** I arrived at this question after performing exploratory analysis on this dataset.



My initial inclination was to measure land inequality against intensity of rebellion. The exponential growth of both the Gini coefficient and rebellion intensity and the outlier status of two large middle peasantry counties, led me to suspect that the midpeasant variable was a mediator and should be removed from the regression model I was constructing. Counties have a large middle peasanty as a result of having low inequality. To put it another way, low inequality and a large middle class are effectively synonymous. Comparing the intensity of the rebellion with the commercialization of agriculture yielded a much more obvious linear relationship.



Armed with these preliminary plots, I decided to structure my regression equation as follows:

intensity = β0 + β1commerce + β2inequality + β3tradition

Running this regression yielded some interesting results. An intercept value of -13.34744 indicates that in a theoretical county with no commercial agriculture, no income inequality and no traditionalism (or no rural illiteracy) the “intensity” of the rebellion would be -13.3. As mentioned previously, this is an ambiguous metric that covers loss of life and spread of revolt in terms of percentages of villages. This is a weakness in the data and the original study. Because this is an aggregate value we have no real way to quantify this metric outside of relative terms.

Inequality in this regression was a non-significant variable, with a p-value of 0.4777. Commercialization of agriculture was highly significant with a p-value of 8.92e-05 and traditionalism barely significant with a p-value of 0.0524. I decided to rerun the regression using just commerce and tradition as my variables.

intensity = β0 + β1commerce + β2tradition

This raised the intercept value to -12.61149, and slightly increased the significance of tradition, with a p-value of 0.483:

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -12.61149 4.78390 -2.636 0.0133 \*

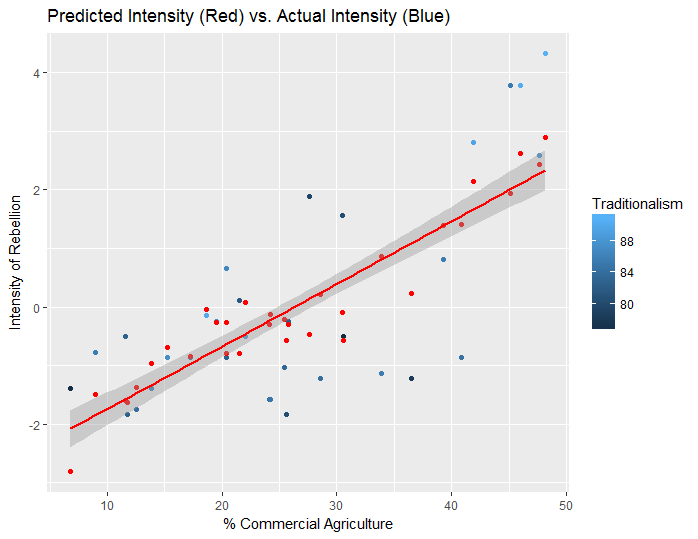
commerce 0.09522 0.01885 5.052 2.19e-05 \*\*\*

tradition 0.11992 0.05816 2.062 0.0483 \*

In a country with no commercial agriculture and no traditionalism, the intensity of the rebellion would be -12.61. For every percentage increase in commercial agriculture the intensity of the rebellion increases by .09522 (with a standard error of 0.01885 ) and for every percentage increase in traditionalism (or every percentage decrease in rural literacy) the intensity of the rebellion increases by 0.11992 (with a weak correlation and a standard error of 0.05816). To reiterate, we know that this shows that there is a standardized, relative increase in violence and spread of revolt, but neither the dataset, nor the study authors provide us with a way to disaggregate these figures into specific predictions of injuries or property destruction or loss of life.

There are too many potential lurking variables to draw firm conclusions. We don’t know the nature of the violence being calculated in the intensity statistic. The original study authors don’t know how to distinguish violence committed by peasants vs. violence committed by police. As mentioned previously, there are almost certainly geographic implications to which we are blind. Regarding my methodology, I think this dataset is well suited for this type of regression work.

Keeping in mind the aforementioned caveats, we can still draw some general implications from this analysis. It appears that the strongest variable correlating with rebellion intensity is commercialization of agriculture. At the risk of going full Karl Marx, this appears to demonstrate an increasingly volatile peasant reaction according to the degree to which they are alienated from the product of their labor. Additionally, there is a light correlation between rural illiteracy and rebellion intensity. This corresponds with what we know about the tumultuous social and economic climate surrounding industrialization in Europe.



References:

Chirot, D., & Ragin, C. (1975). The Market, Tradition and Peasant Rebellion: The Case of Romania in 1907. *American Sociological Review,* *40*(4), 428-444.