Thank you Sara for the introduction of the products. In previous meetings with client, we found that studying the causes of crime may help us to learn more about the consequences of crime patterns. Then, jump to the next part: modeling. We decide to use regression model to find how the variables affect crime rate. (滚动鼠标)

We refer to FBI reports, some factors are known to affect the number and types of crimes. Here, we select some of the most important factors on the left(滚动鼠标). And we found the corresponding variables on the right to measure these factors(滚动鼠标). When choosing these variables, consider the difficulty of data acquisition.

We collected these data from 1995 to 2018 and analyzed it based on FBI data from 1995 to 2018. We focus on violent crime and property crime. The data looks like this. (滚动鼠标)

The modeling includes two parts: time series regression model and panel data regression model. (滚动鼠标)\*2

The regression function is in the R file, and the path can be found in the readme file on github. We also documented the method in a pdf file. (滚动鼠标)

The R file contains two functions: plm\_inference\_state\_level for time series regression, and plm\_inference\_overall for panel data regression.

Let's start with the first one, the time series regression model. (滚动鼠标)The model aims to develop how the selected variables affect each state year by year. Therefore, the analysis is based on each state.

The model looks like this(滚动鼠标)

* Where i from …t… .x … i from 1 to 51, represents states (50) and District of Columbia
* t from 1995 to 2018, represents the year number
* to represents 5 predictors after standardization, including unemployment rate (unemp\_rate), population density (pop\_density), poverty rate (pov\_rate), police officer rate (pol\_officer\_rate) and medium age(med\_age)
* y represents our response-the logarithm of the crime rate (the number of crimes divided by population). We made the model twice. y represents the violent crime rate for the first time, and y represents the property crime rate for the second time.
* is error of the model

Here, since we are performing time series regression analysis, the variance of the residuals may not be a constant and may have autocorrelation that violate the assumption of independent errors. Therefore, we need PACF to test whether the residuals have autocorrelation. If so, we use the Newey-west method to correct the estimation of variance(滚动鼠标). Then, we calculate the t statistic and its corresponding p-value to test whether the coefficient is significant at the 5% level.

Our function only needs to input the data file showed before, and then it will give the result of state name, regression coefficient, t statistic, p value and importance (to show whether the estimated coefficient is significantly not zero).

Let us look at the results and explain. (滚动鼠标)

The result is in an excel file called regression\_output. Here, I will only select one state of violent crime rate to explain the results.

Here, I show the time series regression results of violent crime in North Dakota. We can see that the population density is positive and significant. (3滚动鼠标) The positive coefficient tells us that as population density increases, violent crimes increase. This is not a good sign. Because the population density is calculated by dividing the population by the area of ​​the state. In the long run, as the population grows, the population density is rising, and the crime rate is also rising. This means that the police and the government are not doing a good job of controlling crime. For police officers, (滚动鼠标) this is negative and significant. This means that, for example, as the police officer rate increases, the crime rate decreases. Then we can conclude that for this state, increasing the ratio of police officers is a good way to control crime. The poverty rate is not significant, (滚动鼠标) which means that we do not have enough evidence to reject the null hypothesis that the coefficient is equal to zero. Therefore, at least from 1995 to 2018, the poverty rate did not play an important role in North Dakota, or maybe it does not directly affect violent crime, but indirectly affects other factors, which are then reflected in the violent crime rate. Similarly, we can see that the unemployment rate (滚动鼠标)and median age (滚动鼠标)are both significant and positive. Then the state should pay more attention to those unemployed. From these significant coefficients, we can compare the absolute values ​​and conclude that for this state, the sequence of important variables are population density>police officer ratio>medium age>unemployment rate. Of course, there may be other factors that influence the rate of violent crime in North Dakota.

This is an example of how to interpret the results of a time series regression model. So can we show something for all states? Let's take a look at this plot(滚动鼠标). The graph uses coefficients obtained from the time series regression model of all states. If the variable coefficient is significant, we will use the original number. If not, then set the value of the coefficient as 0. Then, use these five coefficient numbers to calculate the Euclidean distance between states and draw this graph. If the color is orange, it means that the distance between the two states is greater. If the color is light blue, it means that the distance between the two states is very small. Here, we can see that North Dakota(滚动鼠标) has the largest distance with other states, followed by South Dakota. (滚动鼠标)

If we use these coefficient data for clustering, then we will get this graph. (滚动鼠标)We can see that there are three clusters on the map. States with the same color also have similar causes that affect crime. for example, compared with the coefficient, cluster 3 (blue) indicates that the population density is not significant or positive but with small value. Cluster 1 (red) indicates that the population density is negatively significant. Cluster 2 (green) indicates that the population density is positive and the coefficient is large (greater than 1). The results of the cluster can reveal to a certain extent that these states have similar reasons for crime.

(滚动鼠标)For some states, such as Nevada and Pennsylvania, we found that their police officer rate (滚动鼠标)coefficient is positively correlated with the violent crime rate. That's interesting. In order to find out the direction of the causal relationship between the police officer rate and crime rate, we conducted a Granger causality test. (滚动鼠标) As a result, violent crimes in Arizona and Illinois affected the rate of police officers. The ratio of police officers in Texas and New Hampshire affects crime. for Idaho, they affect each other. the states in grey means we can’t determine the direction because the sequence it not stationary, to analyze it we need more years data.

Since time constraints and the violent crime often have a serious impact, we only discuss violent crimes in detail and deal with property crimes quickly.

for property crimes, the analysis in each state is very similar. Here, I want to skip it. let’s look directly at the plots and maps. (滚动鼠标)This is different from violent crime at the state level. You can see the orange gathers in these states. (滚动鼠标(滚动鼠标. and this is the cluster map of property crime and granger causality test for police officer rate versus property crime. (滚动鼠标(滚动鼠标

(滚动鼠标The second panel data regression model analyzes the relationship between predictors and crime rates nationwide. The model is like this, (滚动鼠标 where t represents 1995-2018. X is the design matrix and betas are the coefficient of the 5 predictor variables. Y is the logarithmic crime rate. We assume that β is affected by time. Then we want to split the time effect and only want to know the effect from predictor variables (we call it fixed effect). Hence, we build the model for β:(滚动鼠标Text

Description automatically generated

newey west (滚动鼠标

Here’s the result of violent crime. (滚动鼠标

We can see that all of our predictors are significant. (滚动鼠标 The population density and median age are negative(滚动鼠标(滚动鼠标(滚动鼠标. The police officer rate(滚动鼠标, poverty rate (滚动鼠标and unemployment rate(滚动鼠标 are all positive. Please note that these coefficients have different meanings from previous time series regressions. We focus on the annual differences between states. In states with high population density and median age, the violent crime rate tends to be low. States with a higher police rate, poverty rate, and unemployment rate have a lower tendency to violent crime.

(滚动鼠标(滚动鼠标For property crime, population density is not significant, and other predictors have the same coefficient signs which means they can be interpret the same as the violent crime.

Then let’s review of the product. (滚动鼠标….that’s all, thank you! if you have any questions, feel free to ask us!