Analysis of the factors that contribute to the industrial conflict (A changer)

Clara le Draoulec, Maeva Kameni

March 22, 2022

Introduction:

In the United States, in order to legislate the memberships in labor union, right-to-work laws have been passed, known as the Taft–Hartley Act. However, this law is not imposed in all states. Therefore, in this analysis, we propose to address if there is a difference in terms of industrial conflict between unionized and non-unionized states and determine to what extent certain factors might contribute to this conflict.

Four factors are selected as explanatory variables. The first variable is the degree of unionization of the manufacturing labor force. The second variable is a dummy variable and describes the status of the state, where 0 represents right-to-work states and 1 is assigned to union-shop states. The third variable is a measure of the variation in the industrial structure between states. The latter aims at describing how the union strength is distributed throughout the industries of a state. Finally, the fourth variable represents the size agricultural labor force

Exploratory Data Analysis:

Before performing any analysis on the data set, the first step is to further look at the individual variables, examine their distribution, search for any outliers that can influence and bias our model and eventually our conclusions.

Let's now have a look at the output variable, namely the degree of conflict.

Table 1: Summary of the variable factors

	State postal abbreviation	Degree of conflict	Degree of unionization	Union shop dummy	Sector	Ag force
	Length:48	Min. :0.08	Min. : 8.3	Min. :0.00	Min. : 3.2	Min. : 1.3
	Class :character	1st Qu.:0.74	1st Qu.:18.8	1st Qu.:0.00	1st Qu.: 7.1	1st Qu.: 4.6
	Mode :character	Median :1.11	Median :25.0	Median :1.00	Median :11.7	Median : 8.0
		Mean :1.32	Mean :27.3	Mean :0.56	Mean :15.1	Mean : 9.9
		3rd Qu.:1.85	3rd Qu.:36.3	3rd Qu.:1.00	3rd Qu.:18.0	3rd Qu.:13.1
		Max. :3.82	Max. :53.3	Max. :1.00	Max. :50.0	Max. :32.8

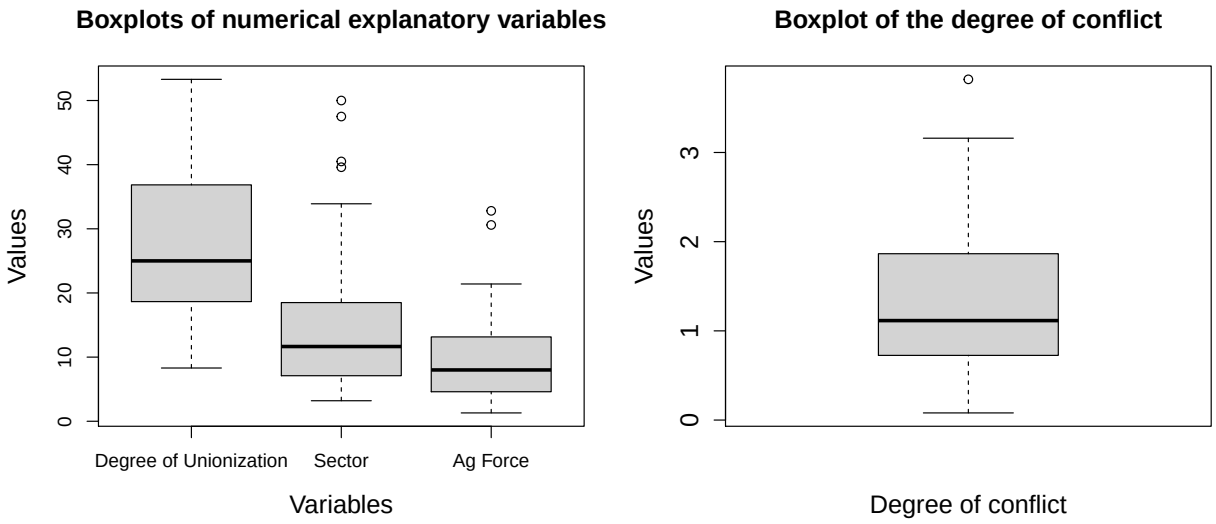

Figure 1: Boxplots of the numerical variables

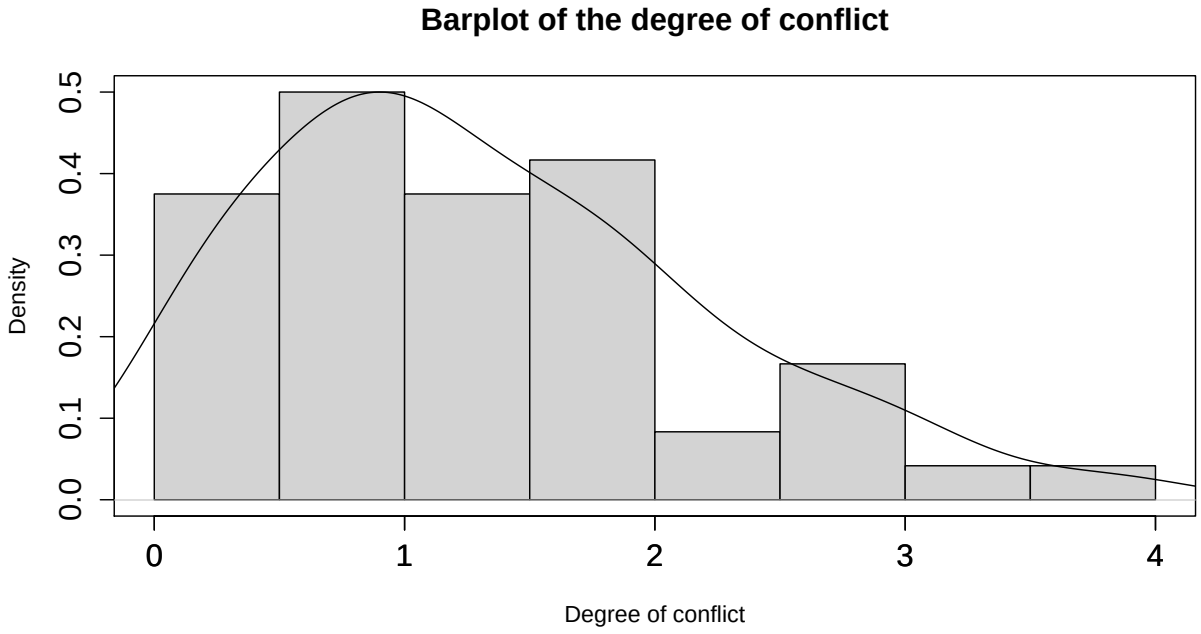

Figure 2: Distribution of the degree of conflict

Given this figure, we can now affirm that our output variable follows a gaussian distribution.

Finally, concerning the dummy variable, a barplot is preferred to assess the distribution

Figure 3: Distribution of the states with respect to their union status

This latter figure ensures that there is fairly the same of number of observation in the two groups that we will compare in the rest of the analysis.

Now we will analysis the distribution of the differents variable factors according to our groups namely Right-to-work states and Union-shop States (ca on maintien que c'est trop) PAS FORCEMENT NECESSAIRE JE TROUVE QUE C?EST UN PEU OVERKILL

Let's see how the variables are linked to each other before beginning the model fitting. It could help us make sure that the model fits properly afterwards.

Statisitics and model fitting:

ECRIRE UN TRUC BIEN:

The main goal first step in the analysis is a direct comparison of the observed levels of industrial conflict in the right-to-work stated and union-shop states for the period 1957-1962

- the measure of industrial conflict is the proportion of time lost due to work stoppages (man hours relative to total estimated working time)

The independent t-test between the unionized and non-unionized states suggests that there is no significant difference in terms of degree of conflict.

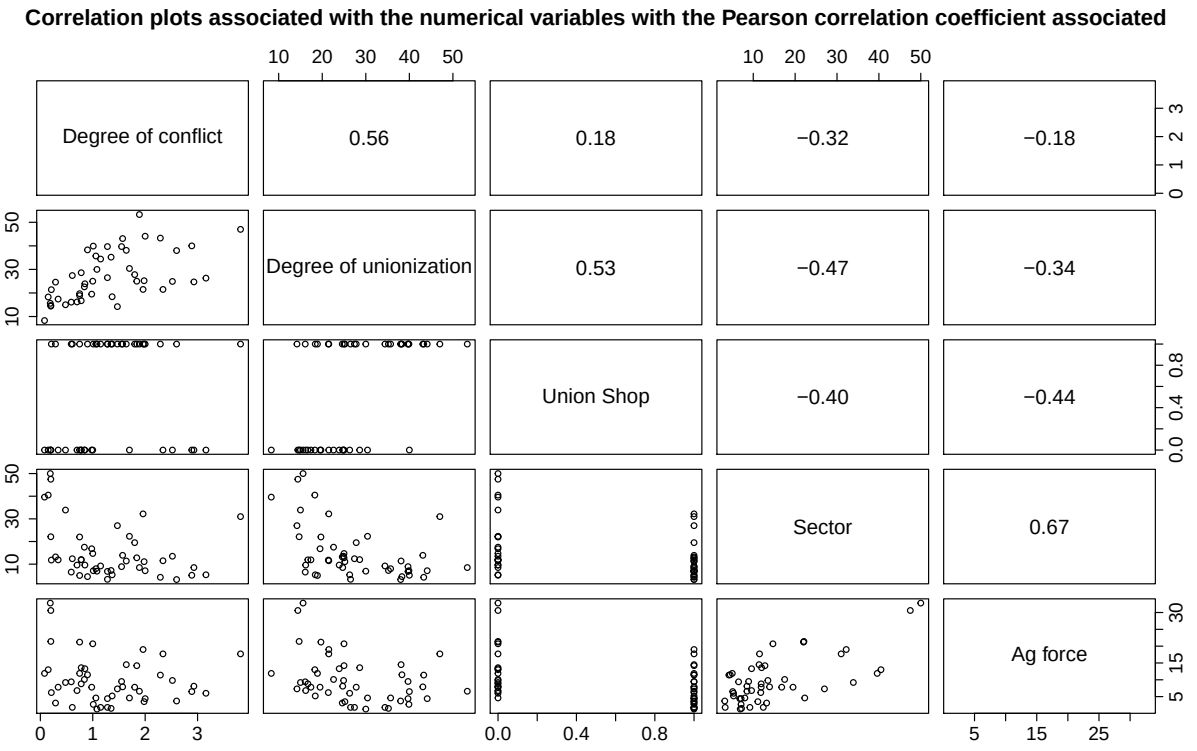

Figure 4: Correlation plots between the explanatory variables

PAS FORCEMENT NECESSAIRE: (RESULT PAPER: AVERAGE PER YEAR: right-to-work : 0.19% // non-right to work:0.24% –> significative at 5%

AVERAGE TOTAL MAN HOURS LOST: right to work: 267,151 hours per year // union-shop : 712,603 hours per year –> significative –> causative factors: conceivably be favtor other than the existaence or non existence of a right to work law)

–> test the statements (better test t-test independant) : - Agriculture > industry in right to work whereas industry > agriculture.

" Where the relatively non-unionized agricultural sector is of major importance in the economy of a state, the theoretical level of industrial conflict should be less than where this is not the case".

- Try to explain the outcome variable using the variable factors (variance explained R^2 maybe adj-R^2) –> multiple linear regression (2 digits and ^ on estimates) : state how model fitted (ie, LS) CLEARLY describe how model selected define all terms

–> assumptions of the multiple linear regression: 1. A linear relationship between the dependent and independent variables 2. The independent variables are not highly correlated with each other (function VIF pour check si c'est bien bon) 3. The variance of the residuals is constant 4. Independence of observation

–> state the assumptions: (ON THE ERROR) 1. errors have mean 0 2. errors are homoscedastic (same variance) 3. errors are uncorrelated 4. errors are normally distributed

stepwise regression Since the EDA allowed us to see that there are some outliers, we will perform a Cook's plot to detect if there is any influential points.

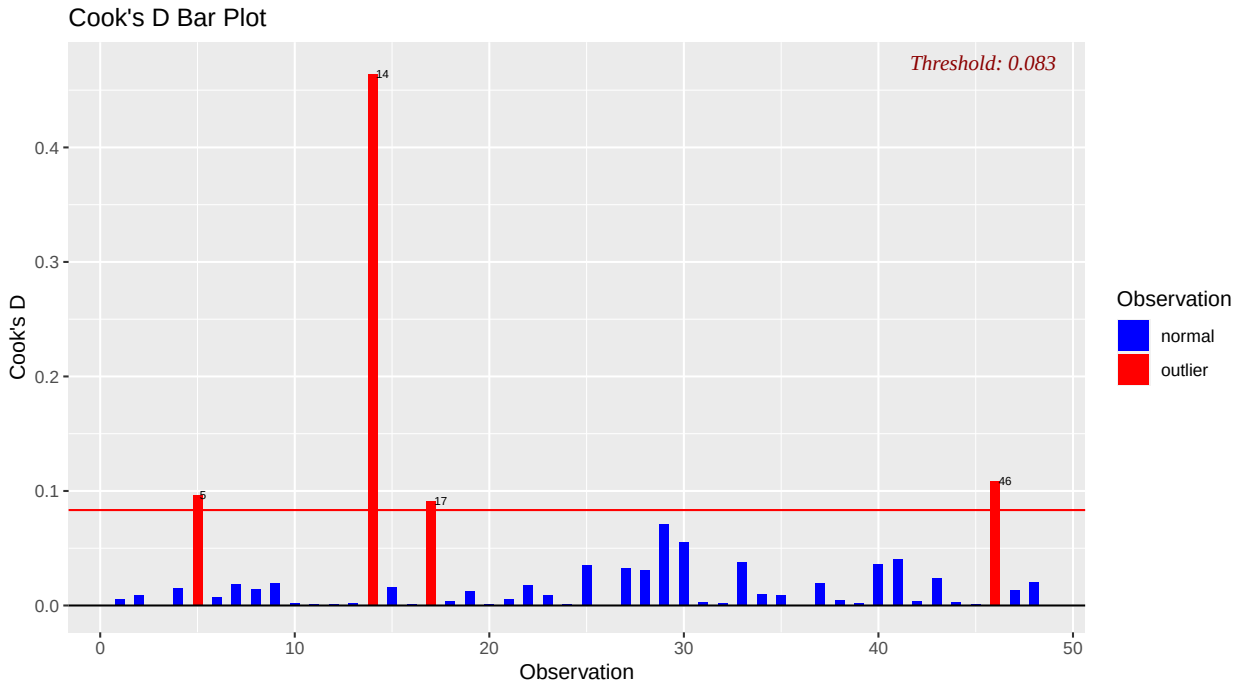

Figure 5: Cook's Distance Bar Plot for our model

Given that results we sill remove those points to see if these indeed influence the results of our multiple linear regression

Model Assessment:

- State how the model fitted (Ordinary Least square) degree_of_conflict ~ remaining variable factors
- check the assumptions –> qqplot of the residuals and residuals vs. fitted –> function VIF pour check la mutlicollinearité (les variables indé sont elles vraiment inépendantes entre elles)

Commenter le plot

commenter le plot

Conclusion

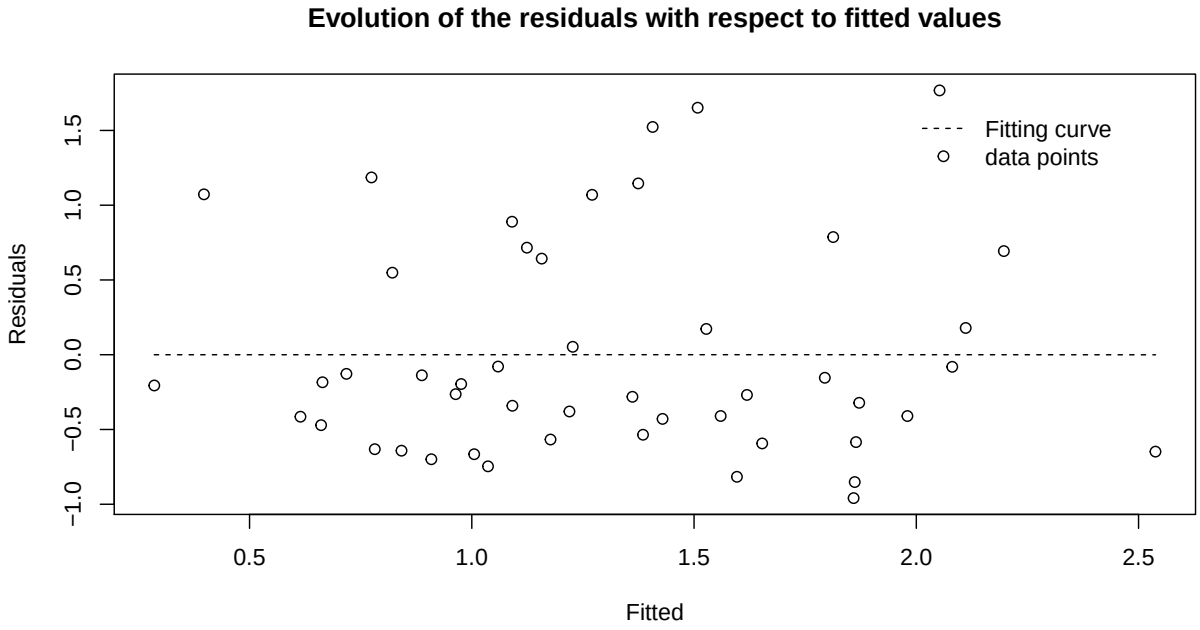

Figure 6: Residuals vs. fitted plot

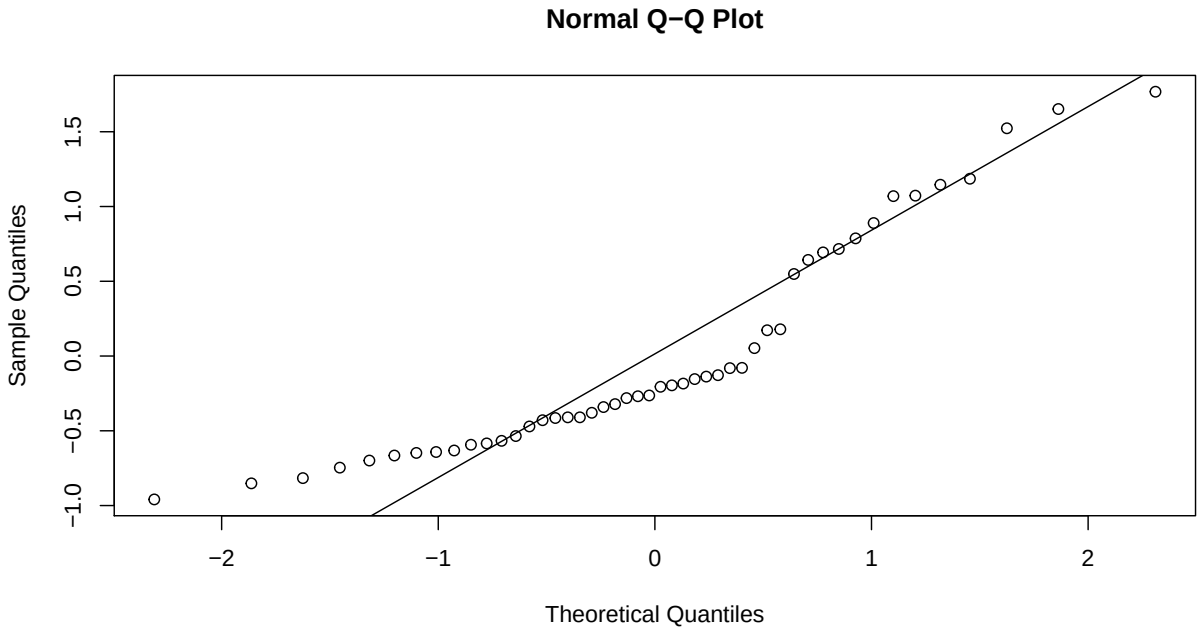

Figure 7: Q-Q plot

Table 2: Analysis of the multicollinearity of the numerical explanatory variables

	Variance Inflation Factor
Degree_of_unionization	1.562638
Union_shop_dummy	1.541870
Sector	2.043418
Ag_force	1.933582