U.S. Homicides Analysis

1. Abstract

One of the main functions of any government is to guarantee the safety and peace of its citizens. The analysis of criminal behavior is one of many tools the government has available to fight criminal activity. In this study we focus in understanding the overall behavior of homicides in the Unite States and its relationship with different variables. By understanding the crimal behavior law enforcement agencies can make decisions such as where and when personnel has to be allocated.

2. Introduction

The goal of the study is to answer five main questions:

1. Is the behavior of homicides by victim's gender different?

2. Is there a difference in the behavior of homicide rates across regions in the United States?

3. Is there a relationship between homcides and unemployment rate?

4. Is there a relationship between homcides and the temperature in New York state?

5. What kind of relationship exist between the clearance rate of homicides and homicides rate?

The required data for the project was gathered from different sources:

1. Homicides data[1]: FBI's Suplemental Homicide Report (SHR). The SHR is one of many reports of the Uniform Crime Reporting (UCR) program, which is part of a nationwide cooperative statistical effort from different agencies with the goal to meet the need for reliable crime statistics conceived in 1929. The dataframe we worked on is a transformation of the original SHR:
    a. Each record represents a victim in a murder or non-negligent manslaughter case.
    b. There is a total of 638,454 records and 26 variables.
    c. The main variables of interest are:
        i. "City".
        ii. "State": (50 states and D.C.).
        iii. "Year": 1980-2014.
        iv. "Month": Jan-Dec.
        v. "Crime Solved": Yes/No.
        vi. "Victim's age": 0-99 or unkown.
        vii. "Victim's sex": Female, Male, or Unknown.
        viii. "Victim's race": Asian/Pacific Islander, Black, Native American/Alaska Native, White or Unkown.
        ix. "Perpetrator's age": 0-99 or unkown.
        x. "Perpetrator's sex": Female, Male, or Unknown.
        xi. "Perpetrator's race": Asian/Pacific Islander, Black, Native American/Alaska Native, White or Unkown.
        xii. "Relationship": victim to offender (28 values).
        xiii. "Weapon": weapon used by perpetrator (16 values).

[1] The vicims as a result from the 1995 Oklahoma City bombing are included (168). The victims as a result of the events of September 11, 2001 are not included.

2. U.S. Population: U.S. Census Bureau.
    a. Yearly data from 1980-2014.
    b. We assumed a linear growth between years.

3. Unemployment rate[2]: Bureau of Labor Statistics.
    a. Definition: percentage of labor force without employment.
    b. Monthly data from 1980-2014.
    c. Seasonally adjusted.

4. New York weather: National Climatic Data Center (NCDC).
    a. Monthly data form 1980-2014.
    b. Average air temperature.

3. Methods

The first step of the analysis was to do some exploratory analysis of the homicide data. We created a time series by grouping the original data by year and month. The homicides time series can be seen in Figure 1.

We can see there are high peaks, specially some months around 1990 and 1992. In order to understand what is the cause of these extreme observations we decided to go one level down in the data, that is, look at the homicides by state.

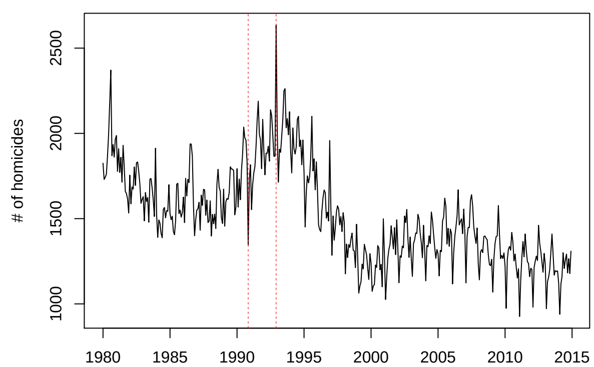

Figure 1: Number of homicides by month in the United States 1980-2014.

After looking at the monthly homicides by state we found out that we had some missing data in some states. This problem was found mainly in 6 states which can be seen in Figure 2.

3.1 Missing data

After discovering the general problem of missing data we identified 4 types of problems:

1. Complete year data missing, that is 12 months of a given year.
2. Part of year data missing.
3. All data of a given year reported in December. An example can be seen in the state of Alabama in the early 90's in Figure 2.
4. Extremely low/high values. An example can be seen in the state of Illinois or New York in the first 10 years of data in Figure 2.

[2]Series ID LNS14000000.

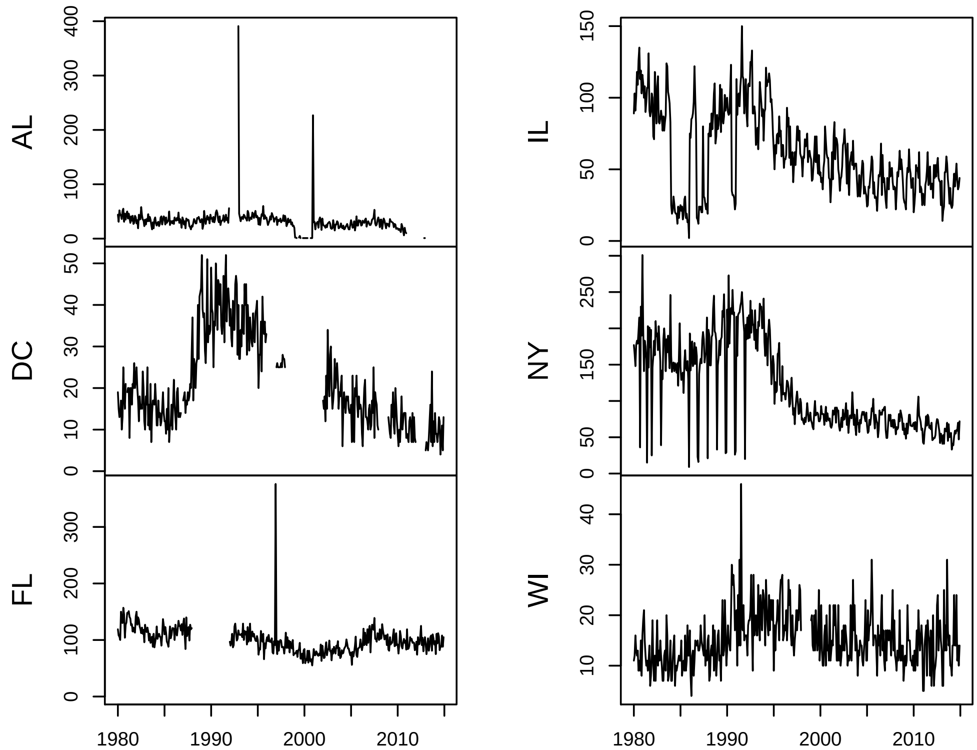

Figure 2: States with missing data.

In order to solve this problem, we proposed two solutions:
1. Yearly estimates with temporal disaggregation.
2. State Space Models (SSM).

3.1.1 Yearly estimates with temporal disaggregation

Temporal disaggregation methods are used to disaggregate low frequency time series to higher frequency time series (e.g., from year to month), where either the sum or other statistic of the high frequency series is consistent with the low frequency series. The disaggregation can be done with or without one or multiple high frequency inicator series.

3.1.1.1 Theory

In this study we used the most common method known as Chow-Lin, the basic idea is as follows:

For simplicity, lets assume the low frequency is years and the high frequency are months. First, a preliminary monthly series $\vec{p}$ is determined and the differences between the annualized perliminary series and the observed anual series are distributed among the preliminary series. That is,

$$\hat{\vec{m}} = \vec{p} + D\vec{e} \qquad \vec{e} = \vec{y} - C\vec{p}$$

Where $\vec{y}$ is a $n_y \times 1$ vector of the yearly series, C is the $n_y \times n_m$ conversion matrix that performs the annualization, where n_y and n_m denote the yearly and monthly observations respectively, D is the $n_m \times n_y$ distribution matrix, and $\hat{\vec{m}}$ is $n \times 1$ vector with the final estimation of the monthly series.

The preliminary monthly series $\vec{p}$ is defined by performing the GLS regression $\vec{y} = C\vec{m} = CX\vec{\beta} + C\vec{v}$, where X is a $n_m \times m$ matrix and m denotes the number of indicator series, and $\vec{v}$ is the error vector with covariance matrix Σ. The main assumptions is that the linear relationship between the anual series CX and $\vec{y}$ holds between the monthly series X and $\vec{m}$, that is, $\vec{m} = X\vec{\beta} + \vec{v}$. So the preliminary monthly series is $\vec{p} = X\hat{\beta}$ and β is estimated in the usual way.

$$\hat{\vec{\beta}} = [X'C'(C\Sigma C')^{-1}CX]^{-1}X'C'(C\Sigma C')^{-1}\vec{y}$$

After having defined the preliminary series, the next step is benchmarking, making sure the anual totals are consistent with the original series. Which is done by minimizing $(\vec{m} - X\hat{\vec{\beta}})'\Sigma^{-1}(\vec{m} - X\hat{\vec{\beta}})$ with respect to the constraint $\vec{y} = C\vec{m}$. The solution is given by:

$$\vec{m} = X\hat{\vec{\beta}} + \Sigma C'(C\Sigma C')^{-1}(\vec{y} - CX\hat{\vec{\beta}}) = X\hat{\vec{\beta}} + D\vec{e}$$

So, the distribution matrix D is defined as $\Sigma C'(C\Sigma C')^{-1}$. The monthly residuals are assumed to have an AR(1) structure and $e_t = \rho e_{t-1} + \epsilon_t$, where $\epsilon_t \sim WN(0, \sigma_\epsilon)$, so Σ has the following structure and is found by estimating the autoregresive parameter ρ:

$$\Sigma = \frac{\sigma_\epsilon^2}{1-\rho^2}\begin{bmatrix} 1 & \rho & \dots & \rho^{n-1} \\ \rho & 1 & \dots & \rho^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \dots & 1 \end{bmatrix}$$

3.1.1.2 Application

The temporal disaggregation methodology is implemented in R with a package called `tempdisagg`. We will now describe the process of how we used this methodology in the states which had missing data. We will use the state of Florida as an example.

The Florida homicides series, as seen in Figure 3 had 4 complete years of missing data and a extremly high value at December 1996.

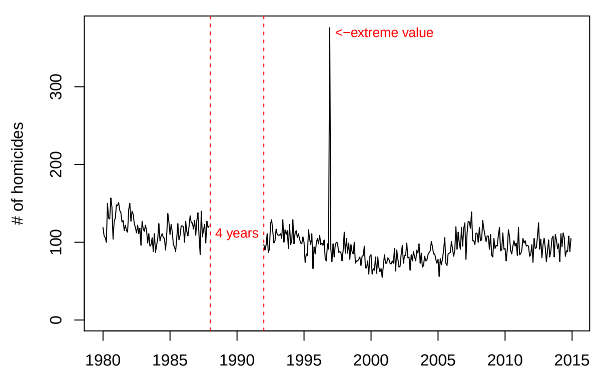

Figure 3: Florida's homicide time series.

The first step was to remove the extreme value and estimate it using by interpolation. Then we estimate the total number of homicides for each of the 4 years with missing data. The estimates can be seen in the top of Figure 4.

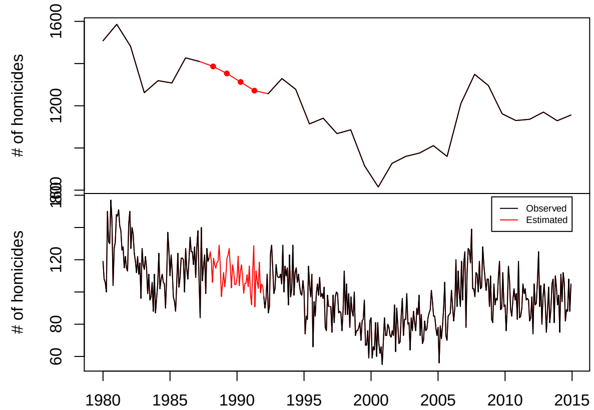

Figure 4: (Top) Florida's homicides by year. (Bottom) Florida's Homicides by month.

Now that we have a complete yearly series we apply the temporal disaggregation. In this case, we want to disaggregate the 4 yearly estimates into a monthly series, and we used the previous 4 years of the monthly series as our indicator series (January 1984 - December 1987). The resulting series can be seen in the bottom of Figure 4.

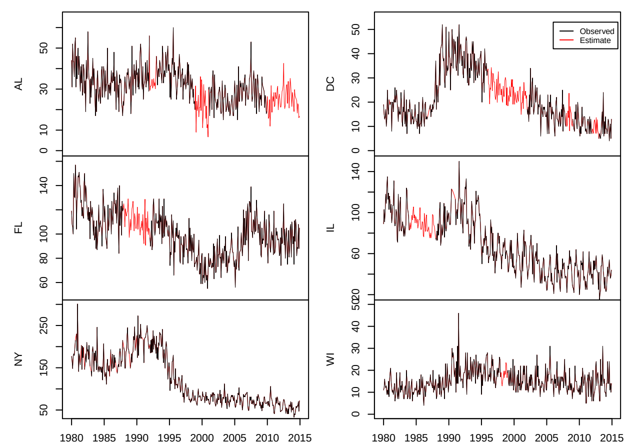

Figure 5: Temporal disaggregation results.

We repeated the process through all 6 states that presented similar problems. The overall result can be seen in Figure 5.

3.1.2 State Space Models (SSM)

In this section we solve the problem of missing data using the State Space Models framework. Where $\vec{x}_t = \Phi\vec{x}_{t-1} + \vec{w}_t$ is the state equation and $\vec{y}_t = A_t\vec{x}_t + \vec{v}_t$ is the observation equation, and

$$\vec{y}_t = \begin{pmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \\ y_{4,t} \\ y_{5,t} \\ y_{6,t} \end{pmatrix} = \begin{pmatrix} Alabama_t \\ DC_t \\ Florida_t \\ Illinois_t \\ NewYork_t \\ Wisconsin_t \end{pmatrix}$$

$$A_t = \begin{pmatrix} I_{\{y_{1,t} \neq NA\}} & 0 & 0 & 0 & 0 & 0 \\ 0 & I_{\{y_{2,t} \neq NA\}} & 0 & 0 & 0 & 0 \\ 0 & 0 & I_{\{y_{3,t} \neq NA\}} & 0 & 0 & 0 \\ 0 & 0 & 0 & I_{\{y_{4,t} \neq NA\}} & 0 & 0 \\ 0 & 0 & 0 & 0 & I_{\{y_{5,t} \neq NA\}} & 0 \\ 0 & 0 & 0 & 0 & 0 & I_{\{y_{6,t} \neq NA\}} \end{pmatrix}$$

The results of the SSM method can be seen in Figure 6.

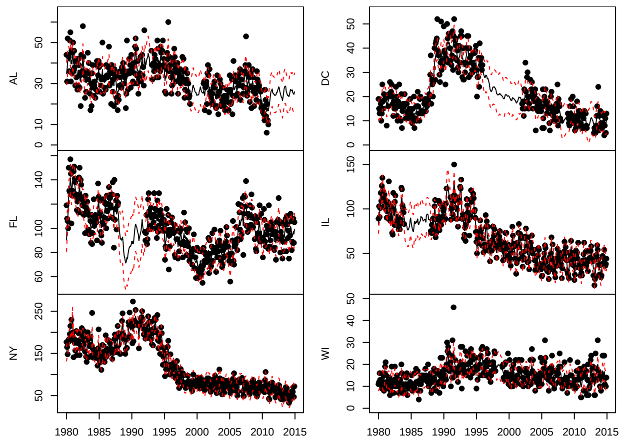

Figure 6: State Space Model for missing data.

To compare both methods, we aggregate the data to get the national time series. We can see in Figure 7 that both methods yield similar results and extreme observations were reduced.

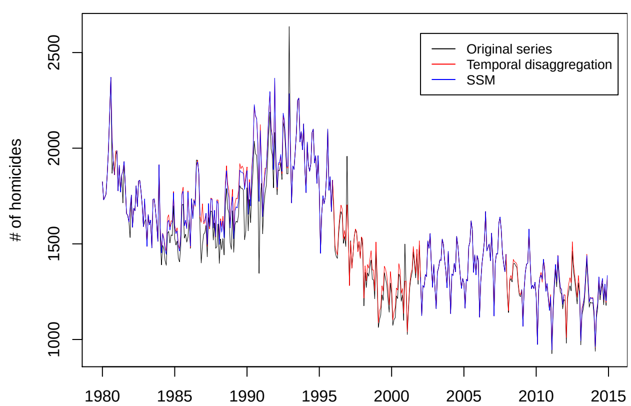

Figure 7: Missing data national results.

3.2 U.S. Homicide Rate

To analize the homicides we decided to work with the series corrected by temporal disaggregation and we also devided the number of homicides by the population and multiplied by 100,000 to get the homicide rate time series that can be seen in Figure 8.

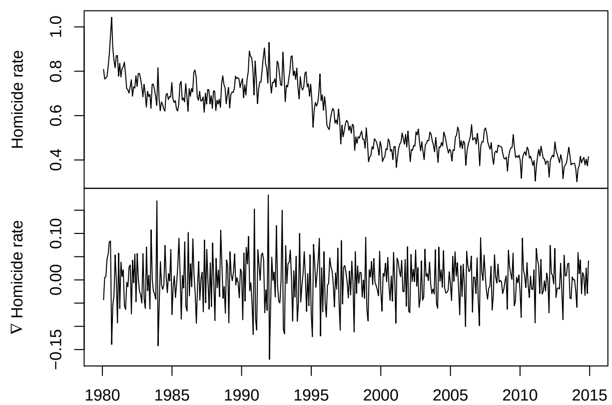

Figure 8: U.S. homicide rate per 100,000 inhabitants and the differenced homicide rate .

To check if the homicide rate series is stationary we performed the Augmented Dickey-Fuller test where H_0 is that the series has unit root and H_1 is that the series is stationary. The resulting p-value of the test is 0.2556, so we can not reject the null hypothesis. In order to have a stationary series we used the first difference of the original series and repeated the Augmented Dickey-Fuller test. This time the resulting p-value was <0.01, so we reject the null hypothesis and conclude the series is now stationary. The differenciated series can be seen in Figure 8.

Now with a stationary series we take a look at the ACF[3] and PACF. We can see in the left part of Figure 9 that there is a peak every 12 lags which suggest there is a seasonal component. The right of Figure 9 shows the ACF and PACF of the first differenced and then seasonally differenced series. First, we take a look at the behavior of the seasonal part in Figure 9, in the ACF we can identify a high peak at lag 1 (12 months) and then there is a cutoff, in the PACF we can see also the high peak at lag 1 and what we think is a decay in the folowwing lags (24 and 36 months). Next, inspecting the ACF and PACF within the seasonal lag (1,. . . ,11), it appears that the ACF cuts off after lag 1 and the PACF decays. Based on this results, we decided to fit an $ARIMA(0, 1, 1) \times (0, 1, 1)_{12}$. The diagnostics for the fit are displayed in the left of Figure 10. We can see some outliers in the Q-Q plot and still some autocorrelations in the residuals and a poor performance in the Ljung-Box test which for each aggregated lag H, the null hypothesis H_0 is that the residuals are indepentent and H_1 is they are not independent.

We decided to use the `auto.arima` function in the `forecast` package to compare resulting models with the previous model. The `auto.arima` function with constraints d=1 and D=1 chose an $ARIMA(1, 1, 1)\times(2, 1, 1)_{12}$. The diagnostics for the fit are displayed in the right of Figure 10. We can see that the autocorrelation in the residuals is almost gone and the Ljung-Box test does better.

[3]The ACF value for lag 0 is removed to avoid scale distortions in the plot.

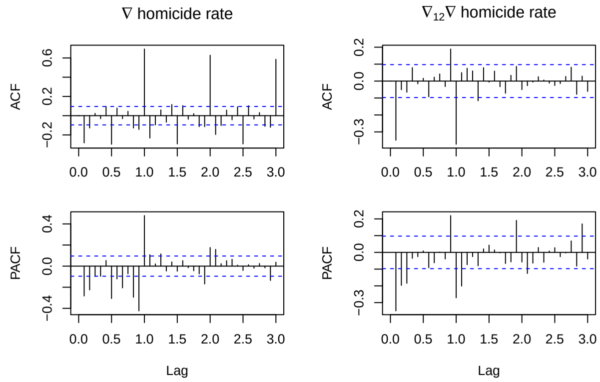

Figure 9: ACF and PACF of the first differenced and seasonal differenced U.S. homicide rate.

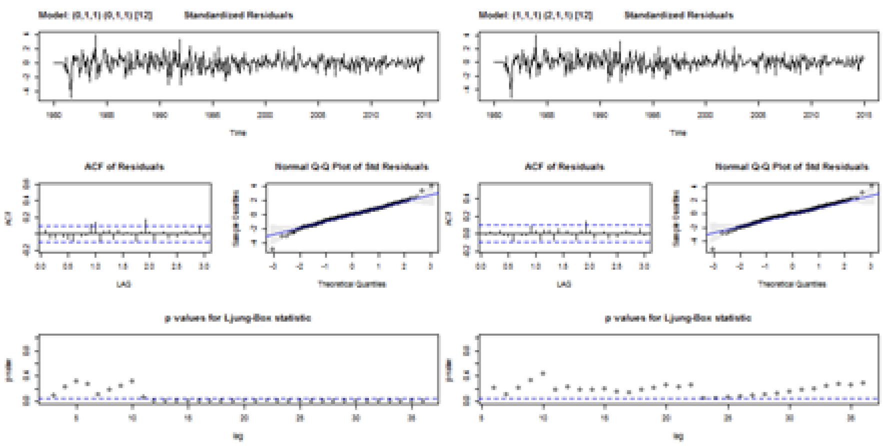

Figure 10: Diagnostics for the ARIMA(0,1,1)x(0,1,1) s=12 model fit (left) and the ARIMA(1,1,1)x(2,1,1) s=12 model fit (right) on the U.S. homicide rate.

3.3 Victim's Gender

The first thing we did was to look at the homicides series by gender, which can be seen in Figure 11. It is clear that the male homicides are approximately 3.5 times the female homicides[4], and at a first glance the behavior between series appears to be similar. To make a more precise observation of the similarity of the behavior we looked at their respective first differenced and then seasonal differenced ACF and PACF.

Results, which are displayed in Figure 12, show that both ACFs have a peak at the first seasonal lag and then a cutoff, and in the case of the female's PACF there is a decay in the seasonal lags and the male's PACF shows a cutoff after lag 1, this could suggest the seasonal component is different between series. The within seasonal behavior for both series is almost the same, a cutoff after the first lag and a decay in the PACF.

Even though we identified a somewhat different behavior in the seasonal components between series, we fitted an $ARIMA(0,1,1) \times (0,1,1)_{12}$ to both and took a look at the diagnostics. The left plot in Figure 13 shows the female victims series diagnostics and int the middle plot the male victims series diagnostics. We can see that the model appears to fit well to the female series with some outliers identified in the Q-Q plot. In the case of the male series the diagnostics show that some autocorrelation is left in the residuals.

[4]Since the population distribution between gender is aproximately 50-50% we did not use the homocide rate to make the comparison.

We tried a different models in the male victim series and the model that had better diagnostics was an $ARIMA(0,1,1) \times (2,1,1)_{12}$. Exploring the diagnostics, in the right plot of Figure 13, we can see there is still some autocorrelation in the residuals.

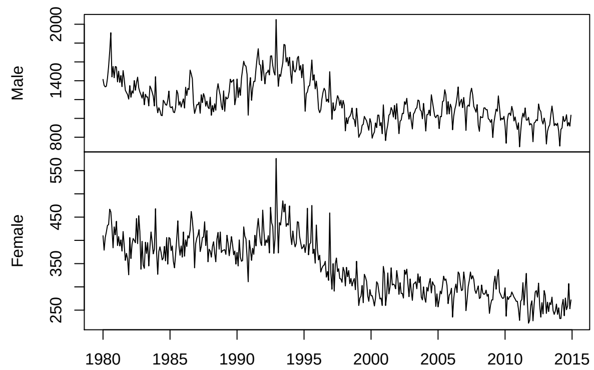

Figure 11: U.S. homicides by victim's gender.

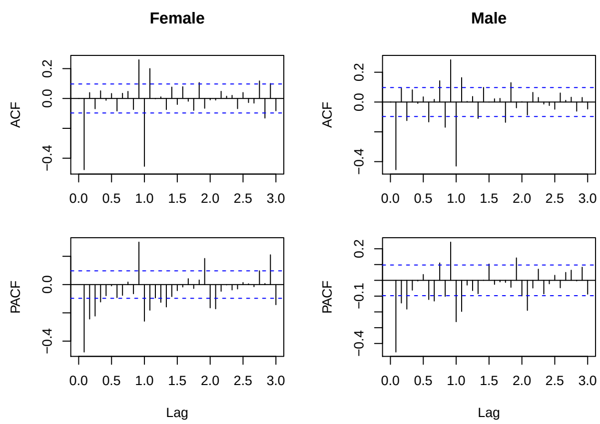

Figure 12: ACF and PACF of the first differenced and the seasonally differenced series by victim's gender.

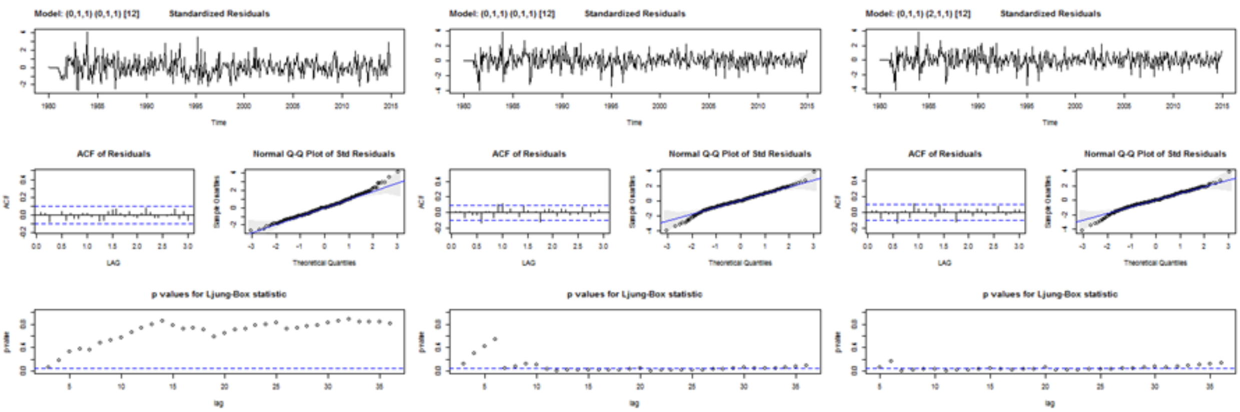

Figure 13: Diagnostics of the ARIMA(0,1,1)x(0,1,1) s=12 model fit to both the female (left) and male (middle) series and the ARIMA(0,1,1)x(2,1,1) s=12 model fit to the male series (right).

3.4 U.S. Regions

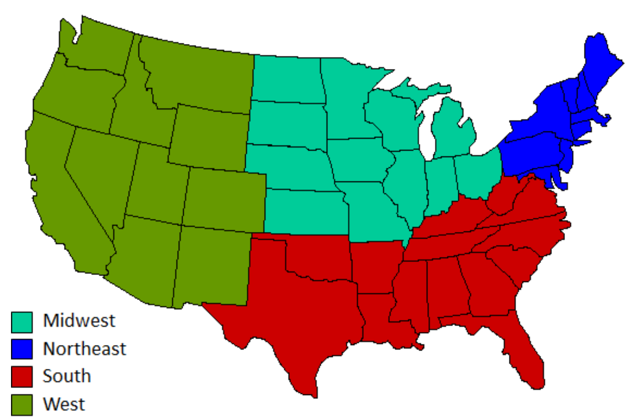

Figure 14: Census Regions and Divisions of the United States.

Using the U.S. Census Bureau definition the homicides data was grouped into four regions, Midwest, Northeast, South, and West. Figure 14[5] displays how the regions are formed, and the top plot of Figure 15 shows the number of homcides series for each region. It is clear the number of homicides in the south region is bigger than all other regions although they all apear to have similar trends and behavior.

This comparison is not fairly done since the population of each region has to be taken into consideration. The normalized series, or in other words the homicide rate, is shown the bottom plot of Figure 15. The normalization closes the gap between regions but the south region remains on top.

To understand their behavior we compare the ACF and PACF of the first differenced and then seasonally differenced homicide rate series of each region. Results are displayed in Figure 16.

In the Midwest region ACF we can see peaks in the fist two seasonal lags and then a cutoff, in the PACF we identify a decay. Within the seasonal part of the ACF we see peaks in the first two lags and a decay in the PACF. This led us to fit an $ARIMA(0,1,2) \times (0,1,2)_{12}$, which diagnostics look good with a couple of outliers in the Q-Q plot. The diagnostics are displayed in the left plot of Figure 17.

[5]Alaska and Hawaii are not shown. They are part of the West region.

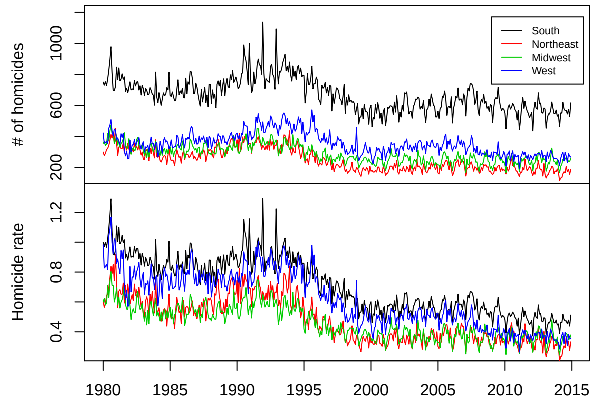

Figure 15: U.S. homicides (top) and homicide rate (bottom) by region.

The Northeast ACF shows a peak at the first seasonal and the first within seasonal lag. In the PACF we can see a decay in both cases. We fitted an $ARIMA(0,1,1) \times (0,1,1)_{12}$ but the diagnostics shown in the middle plot of Figure 17 show some correlation in the residuals. We then fitted an $ARIMA(2,1,1) \times (2,1,1)$, which was the result of the `auto.arima` function and diagnostics can be seen in the right plot of Figure 17.

The South region shows similar behavior as the Northeast in the ACF and PACF so we fitted an $ARIMA(0,1,1) \times (0,1,1)_{12}$ but the diagnostics shown in the left plot of Figure 18 show some correlation in the residuals. We then fitted an $ARIMA(1,1,2) \times (0,1,2)$, which was the result of the `auto.arima` function and diagnostics can be seen in the middle plot of Figure 18.

The West region was the most problematic one, the behavior of the ACF and PACF suggested an $ARIMA(0,1,2) \times (0,1,1)_{12}$ but the diagnostics, shown in the right plot of Figure 18, are not good. We tried a different models and none showed any improvement in the diagnostics.

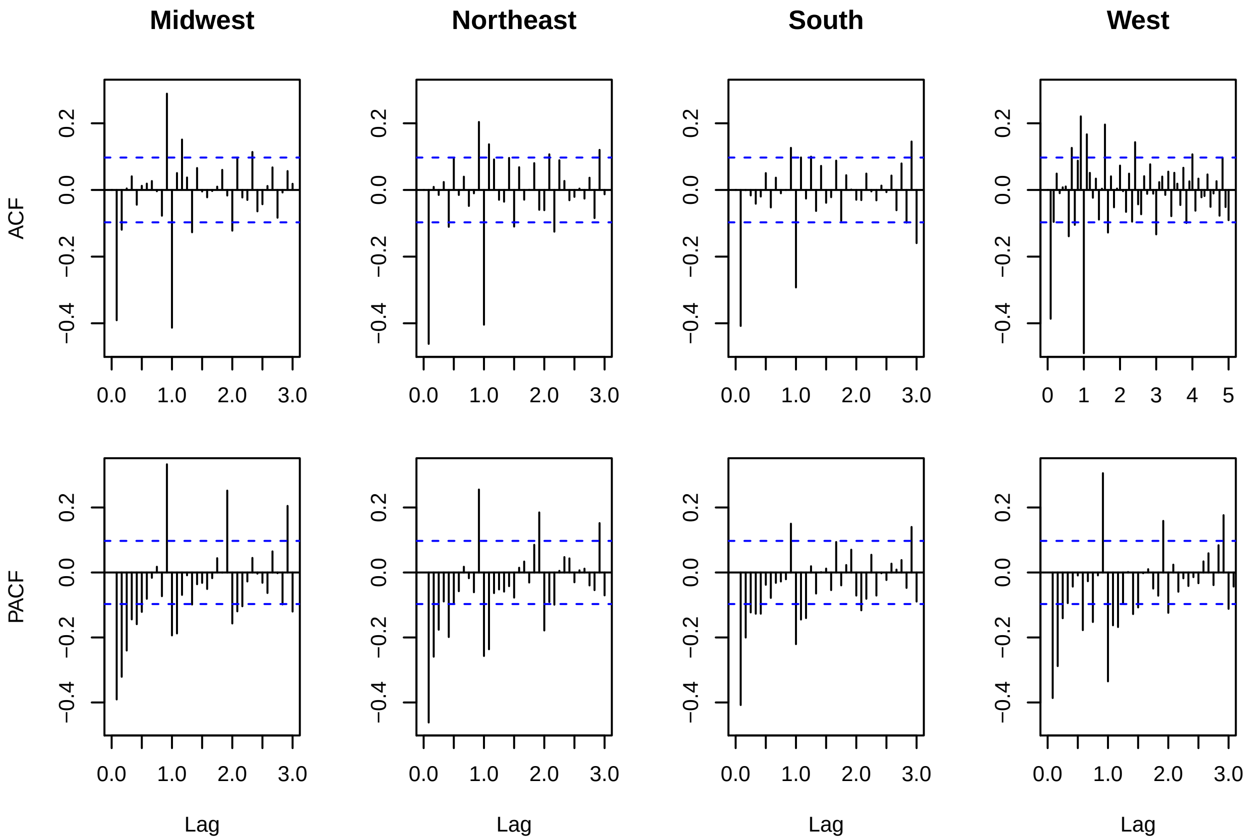

Figure 16: ACF and PACF of the first differenced and then seasonally differenced homicide rate by region.

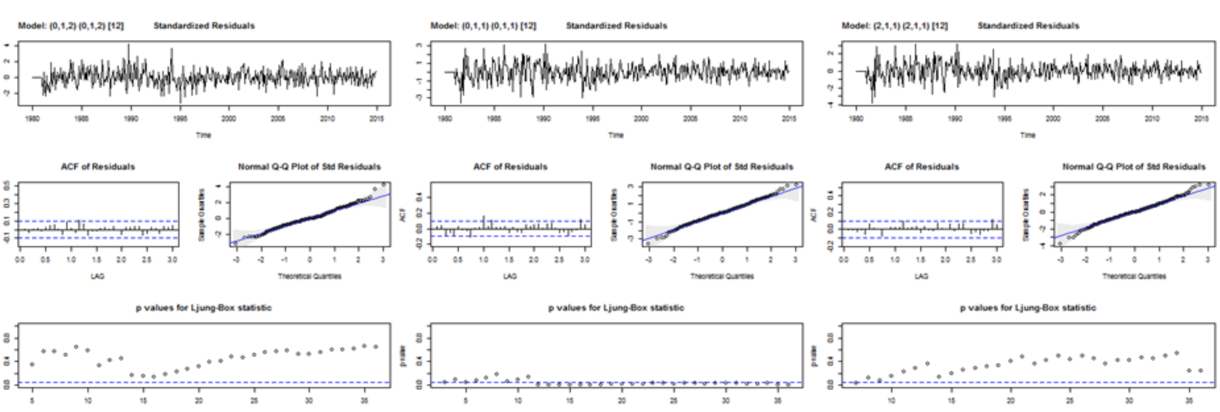

Figure 17: Diagnostics for the ARIMA(0,1,2)x(0,1,2) s=12 model fit on the Midwest homicide rate series (left), the ARIMA(0,1,1)x(0,1,1) s=12 model fit (middle) on the Northeast homicide rate, and the ARIMA(2,1,1)x(2,1,1) s=12 model fit (right) on the Northeast homicide rate series.

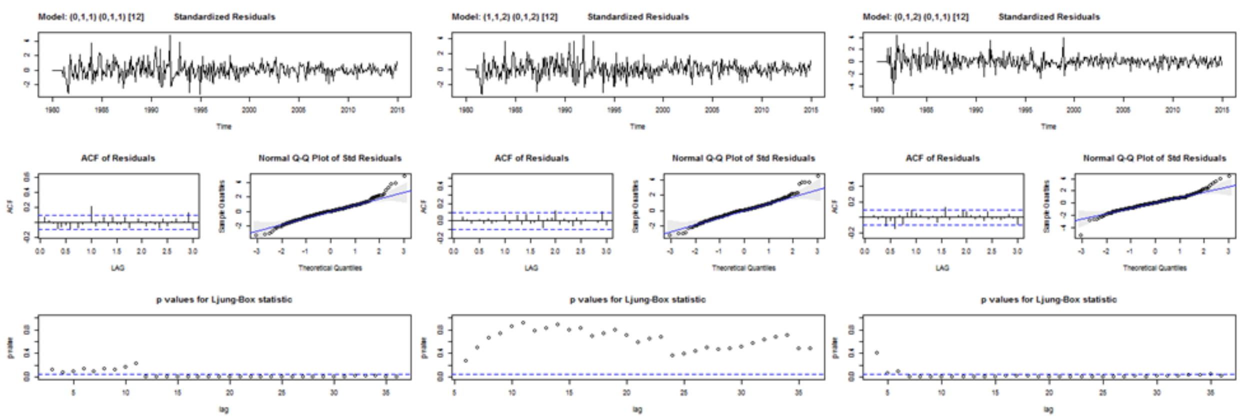

Figure 18: Diagnostics for the ARIMA(0,1,1)x(0,1,1) s=12 model fit (left) and the ARIMA(1,1,2)x(0,1,2) s=12 model fit (middle) on the South homicide rate series, and the ARIMA(0,1,2)x(0,1,1) s=12 model fit (right) on the West homicide rate series.

3.5 Unemployment Rate

Unemployment rate might influence homicide rate, because we assume that when people are unemployed they look for different ways to provide for themselves and/or their families to survive. Unemployment means no fixed income, and no fixed income means unstable life styles, when people really need money to survive, they may turn into illicit activities. Figure 19 displays the unemployment rate[6] and homicide rate in the same graph.

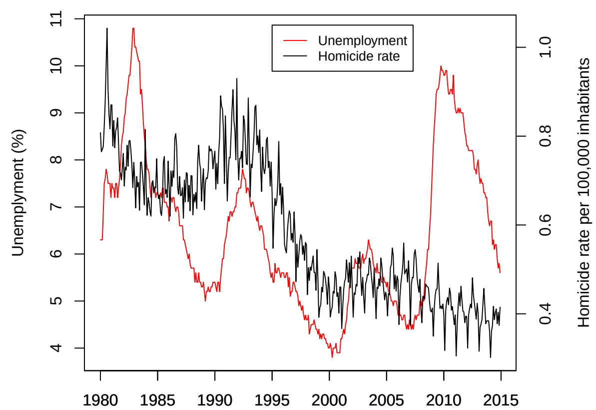

Figure 19: U.S. homicide and unemployment rate.

Unemployment rate and homicide rate appear to be I(1) with the p-value equal 0.1507 and 0.2556 respectively for the Augmented Dickey-Fuller test. Since both variabiles have intergration order of one, we use Engle Granger methodology to test the two variables cointergration. The Engle Granger Cointegration test (with H_0: the series have no cointegration) resulted in not rejecting the null hypothesis, so these two variables do not have cointergration relationship, which means we have to work with the differenced version of both variables. Differenced unemployment rate and Homicides rate appear to be stationary with a p-value less than 0.01 for the Augmented Dickey-Fuller test. The next step was to fit a VAR model.

We can see in Figure 20 that after fitting a VAR(12) model, there appears to be no relationship between the differenced homicide rate and unemplyment rate. The Granger Causality test is used to check wether there is a causality relationship between two series. For example, the H_0 is that differenced unemployment series do not Granger-cause the differenced homicide rate. We applied this test twice, once in each direction, and the p-values (0.98, 0.97) indicate that these two varibles have no Granger Causes between each other.

[6]Percentage of of labor force without employment. Seasonally adjusted.

$\nabla Unemployment_t$

	Estimate	Std. Error	t value	Pr(>|t|)	
dif_U.l1	-1.083e-02	5.054e-02	-0.214	0.830410	
dif_H.l1	-3.058e-01	2.132e-01	-1.434	0.152398	
dif_U.l2	1.133e-01	5.029e-02	2.252	0.024877	*
dif_H.l2	-3.413e-01	2.324e-01	-1.469	0.142717	
dif_U.l3	2.075e-01	5.067e-02	4.096	5.14e-05	***
dif_H.l3	-1.806e-01	2.397e-01	-0.754	0.451521	
dif_U.l4	1.211e-01	5.162e-02	2.345	0.019538	*
dif_H.l4	-2.152e-01	2.409e-01	-0.894	0.372137	
dif_U.l5	1.893e-01	5.178e-02	3.655	0.000293	***
dif_H.l5	-2.399e-01	2.388e-01	-1.004	0.315830	
dif_U.l6	1.050e-01	5.260e-02	1.996	0.046633	*
dif_H.l6	-1.459e-01	2.329e-01	-0.627	0.531281	
dif_U.l7	6.091e-02	5.265e-02	1.157	0.248097	
dif_H.l7	-1.523e-01	2.312e-01	-0.659	0.510451	
dif_U.l8	-1.561e-02	5.218e-02	-0.299	0.764974	
dif_H.l8	-1.651e-01	2.345e-01	-0.704	0.481890	
dif_U.l9	-3.020e-02	5.151e-02	-0.586	0.558039	
dif_H.l9	-1.058e-01	2.352e-01	-0.450	0.652955	
dif_U.l10	-4.062e-02	4.975e-02	-0.816	0.414756	
dif_H.l10	-2.737e-01	2.321e-01	-1.179	0.239107	
dif_U.l11	4.988e-02	4.945e-02	1.009	0.313783	
dif_H.l11	-2.757e-01	2.257e-01	-1.222	0.222582	
dif_U.l12	-1.084e-01	4.939e-02	-2.195	0.028747	*
dif_H.l12	-2.302e-01	2.091e-01	-1.101	0.271729	
const	-3.528e-03	1.657e-02	-0.213	0.831524	
trend	-5.374e-06	6.604e-05	-0.081	0.935185	

$\nabla homicide_t$

	Estimate	Std. Error	t value	Pr(>|t|)	
dif_U.l1	1.025e-02	1.057e-02	0.969	0.332920	
dif_H.l1	-4.137e-01	4.459e-02	-9.278	< 2e-16	***
dif_U.l2	1.142e-02	1.052e-02	1.086	0.278286	
dif_H.l2	-3.358e-01	4.860e-02	-6.911	2.04e-11	***
dif_U.l3	-5.093e-03	1.060e-02	-0.481	0.631004	
dif_H.l3	-2.479e-01	5.011e-02	-4.946	1.14e-06	***
dif_U.l4	-7.023e-03	1.079e-02	-0.651	0.515676	
dif_H.l4	-2.457e-01	5.037e-02	-4.878	1.58e-06	***
dif_U.l5	1.818e-03	1.083e-02	0.168	0.866757	
dif_H.l5	-1.970e-01	4.994e-02	-3.945	9.49e-05	***
dif_U.l6	3.026e-03	1.100e-02	0.275	0.783418	
dif_H.l6	-3.060e-01	4.870e-02	-6.284	9.05e-10	***
dif_U.l7	-8.008e-04	1.101e-02	-0.073	0.942061	
dif_H.l7	-2.760e-01	4.835e-02	-5.708	2.30e-08	***
dif_U.l8	-3.417e-03	1.091e-02	-0.313	0.754389	
dif_H.l8	-2.704e-01	4.903e-02	-5.515	6.46e-08	***
dif_U.l9	-6.026e-03	1.077e-02	-0.559	0.576196	
dif_H.l9	-2.066e-01	4.918e-02	-4.200	3.33e-05	***
dif_U.l10	5.857e-03	1.040e-02	0.563	0.573766	
dif_H.l10	-2.363e-01	4.853e-02	-4.870	1.64e-06	***
dif_U.l11	1.124e-02	1.034e-02	1.087	0.277751	
dif_H.l11	-1.616e-01	4.719e-02	-3.425	0.000681	***
dif_U.l12	-1.222e-02	1.033e-02	-1.183	0.237564	
dif_H.l12	4.792e-01	4.373e-02	10.959	< 2e-16	***
const	-6.758e-03	3.466e-03	-1.950	0.051915	.
trend	1.276e-05	1.381e-05	0.924	0.355895	

Figure 20: Differenced unemployment and homcide rate VAR(12) model parameters.

3.6 Temperature

In this analysis, we are plan to run multivariate analysis to figure out is there any relationship between the temperture and the number of homicides. Since weather varies significantly nationwide, we decided to use homicide cases and temperture records in New York state. We decide to use New York City's temperature to represent New York state becuase : 1. New York state shares the same weather cross whole state. 2 New York state crime report shows that about 2/3 murders happen inside New York City, so there is almost no implication in using New York City's temperature. Figure 21 shows the number of homicides cases in New York State and temperture in New York City.

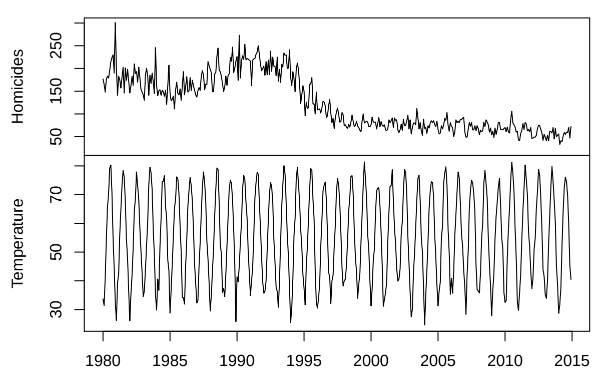

Figure 21: Homicides cases and Temperature in New York.

For Tempertuare we can easy to say that it's stationary and we can confirm this from p-value less than 0.01 for the Augmented Dickey-Fuller test. In order to make homicides cases in New York stationary, we differentiate the data. The differential of number of New York homicides cases is stationary since the p-value of Augmented Dickey-Fuller test is less than 0.01.

Now that both factors are stationary, we fit a VAR model to these two series. Using the BIC criteria for the `VARselect` function we chose to use lag = 6. We can see in Figure 22 that there are some relevant parameters.

$\nabla homicide_t$

	Estimate	Std. Error	t value	Pr(>|t|)	
dif_NYH.l1	-0.7095612	0.0497455	-14.264	< 2e-16	***
NYWeather_ts..1..l1	0.0759372	0.2874561	0.264	0.79179	
dif_NYH.l2	-0.5336245	0.0590249	-9.041	< 2e-16	***
NYWeather_ts..1..l2	0.0430580	0.3064826	0.140	0.88834	
dif_NYH.l3	-0.4134635	0.0626515	-6.599	1.32e-10	***
NYWeather_ts..1..l3	-0.4427275	0.3047401	-1.453	0.14706	
dif_NYH.l4	-0.3414711	0.0626328	-5.452	8.75e-08	***
NYWeather_ts..1..l4	0.4619418	0.3047411	1.516	0.13035	
dif_NYH.l5	-0.3034459	0.0589233	-5.150	4.10e-07	***
NYWeather_ts..1..l5	0.0921349	0.3045360	0.303	0.76240	
dif_NYH.l6	-0.0814824	0.0497725	-1.637	0.10240	
NYWeather_ts..1..l6	-0.7537645	0.2854131	-2.641	0.00859	**
const	28.0134997	19.3066460	1.451	0.14757	
trend	-0.0001582	0.0079802	-0.020	0.98419	

$Temperature_t$

	Estimate	Std. Error	t value	Pr(>|t|)	
dif_NYH.l1	-0.006955	0.008312	-0.837	0.40323	
NYWeather_ts..1..l1	0.424038	0.048030	8.829	< 2e-16	***
dif_NYH.l2	0.006413	0.009862	0.650	0.51591	
NYWeather_ts..1..l2	0.131431	0.051209	2.567	0.01064	*
dif_NYH.l3	-0.005938	0.010468	-0.567	0.57089	
NYWeather_ts..1..l3	-0.091136	0.050918	-1.790	0.07423	.
dif_NYH.l4	0.005105	0.010465	0.488	0.62596	
NYWeather_ts..1..l4	-0.133623	0.050918	-2.624	0.00902	**
dif_NYH.l5	-0.007529	0.009845	-0.765	0.44486	
NYWeather_ts..1..l5	-0.225513	0.050884	-4.432	1.21e-05	***
dif_NYH.l6	-0.001876	0.008316	-0.226	0.82162	
NYWeather_ts..1..l6	-0.293257	0.047689	-6.149	1.89e-09	***
const	65.524613	3.225875	20.312	< 2e-16	***
trend	0.001190	0.001333	0.892	0.37288	

Figure 22: Differenced homicide and temperature VAR(6) model parameters

3.7 Clearance rate

The last analysis we propose to do was to see what kind of relationship does the homicide rate has with the clearance rate. The clearance rate is defined as the percentage of homicides in a given month that were solved. In Figure 23 we can see both series.

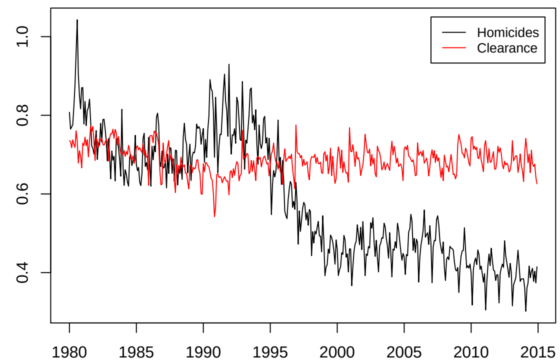

Figure 23: U.S. homicide rate per 100,000 inhabitants and clearance rate as percentage.

The clearance rate appears to be stationary with a p-value less than 0.01 for the Augmented Dickey-Fuller test. We proceeded to fit a VAR model for the clearance rate and the first differenced homicide rate to see how the two series interact with each other. Using the `VARselect` function and the BIC criteria we determined that the number of lags should be 12.

The estimated parameters of the VAR(12) model are shown in Figure 24, which show there is some dinamic relationship between the series. Instead of using all 12 lags we set lags 7 to 10 parameters to zero. After this we used the `refVAR` function from the `MTS` package that sets insignifficant parameters to zero. Looking at the CCF of the residuals (off-diagonal) of Figure 25 we can see that there is almost no correlation in the residuals.

Because the clearance rate is always observed after the homicides, even months/years after the homicides, a more useful model would be to use the the homicide rate as a predictor for the clearance rate.

We fitted a linear model using the clearance rate as the response and lags of the differenced homicide rate as the predictors, assuming the errors were independent, and we analyzed the residuals of the fitted model to determine the structural relationship between them. The residuals' ACF and PACF can be seen in Figure 26. It seems that an $ARIMA(0,1,1) \times (0,1,1)_{12}$ would be a good fit for the errors. With this result, we fited a linear model with errors having the stated structure and the residuals of this model appear to be white(Figure 27).

$\nabla Homicides_t$

	Estimate	Std. Error	t value	Pr(>|t|)	
hom.l1	-4.821e-01	4.597e-02	-10.488	< 2e-16	***
cl.l1	-9.222e-02	6.607e-02	-1.396	0.163631	
hom.l2	-3.564e-01	5.147e-02	-6.924	1.87e-11	***
cl.l2	5.125e-02	6.955e-02	0.737	0.461632	
hom.l3	-2.257e-01	5.375e-02	-4.199	3.35e-05	***
cl.l3	2.086e-01	6.886e-02	3.029	0.002622	**
hom.l4	-2.082e-01	5.493e-02	-3.789	0.000176	***
cl.l4	2.729e-02	6.924e-02	0.394	0.693710	
hom.l5	-1.298e-01	5.549e-02	-2.338	0.019892	*
cl.l5	-2.553e-02	6.898e-02	-0.370	0.711490	
hom.l6	-2.389e-01	5.408e-02	-4.417	1.30e-05	***
cl.l6	1.199e-01	6.943e-02	1.727	0.084947	.
hom.l7	-2.134e-01	5.284e-02	-4.039	6.49e-05	***
cl.l7	-2.393e-01	6.875e-02	-3.481	0.000557	***
hom.l8	-1.944e-01	5.239e-02	-3.709	0.000239	***
cl.l8	-1.348e-01	6.958e-02	-1.938	0.053355	.
hom.l9	-1.344e-01	5.150e-02	-2.610	0.009400	**
cl.l9	-1.042e-01	6.993e-02	-1.490	0.136934	
hom.l10	-1.794e-01	5.014e-02	-3.579	0.000389	***
cl.l10	4.778e-02	7.013e-02	0.681	0.496078	
hom.l11	-1.501e-01	4.779e-02	-3.142	0.001811	**
cl.l11	1.022e-01	6.998e-02	1.460	0.145113	
hom.l12	4.224e-01	4.374e-02	9.656	< 2e-16	***
cl.l12	-4.929e-02	6.709e-02	-0.735	0.462935	
const	5.573e-02	4.933e-02	1.130	0.259342	
trend	7.283e-06	1.340e-05	0.544	0.587086	

$Clearance_t$

	Estimate	Std. Error	t value	Pr(>|t|)	
hom.l1	-3.698e-02	3.490e-02	-1.060	0.289948	
cl.l1	3.386e-01	5.016e-02	6.750	5.51e-11	***
hom.l2	-9.161e-02	3.907e-02	-2.345	0.019564	*
cl.l2	7.704e-02	5.280e-02	1.459	0.145371	
hom.l3	-1.422e-01	4.080e-02	-3.484	0.000551	***
cl.l3	3.001e-02	5.228e-02	0.574	0.566301	
hom.l4	-1.475e-01	4.170e-02	-3.538	0.000453	***
cl.l4	5.501e-02	5.257e-02	1.046	0.296055	
hom.l5	-1.870e-01	4.213e-02	-4.438	1.19e-05	***
cl.l5	9.933e-02	5.237e-02	1.897	0.058620	.
hom.l6	-1.065e-01	4.106e-02	-2.594	0.009866	**
cl.l6	-9.303e-03	5.271e-02	-0.177	0.859993	
hom.l7	-2.338e-02	4.012e-02	-0.583	0.560394	
cl.l7	2.051e-02	5.220e-02	0.393	0.694574	
hom.l8	-2.393e-03	3.978e-02	-0.060	0.952058	
cl.l8	3.538e-02	5.282e-02	0.670	0.503367	
hom.l9	-6.949e-03	3.910e-02	-0.178	0.859034	
cl.l9	-7.994e-02	5.309e-02	-1.506	0.132938	
hom.l10	4.845e-02	3.806e-02	1.273	0.203804	
cl.l10	6.181e-04	5.324e-02	0.012	0.990743	
hom.l11	-9.544e-02	3.628e-02	-2.631	0.008866	**
cl.l11	1.306e-01	5.312e-02	2.458	0.014410	*
hom.l12	-1.093e-01	3.321e-02	-3.291	0.001091	**
cl.l12	1.593e-01	5.093e-02	3.129	0.001891	**
const	9.638e-02	3.745e-02	2.573	0.010449	*
trend	6.663e-07	1.017e-05	0.066	0.947806	

Figure 24: Parameters estimates for the VAR(12) model.

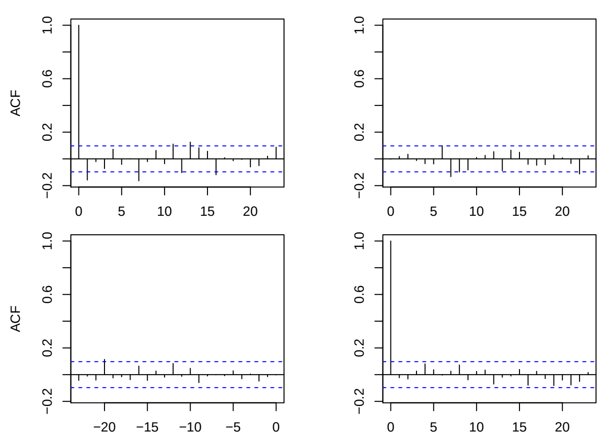

Figure 25: ACF and CCF (off-diagonal) for the VAR(12) model with parameters set to zero.

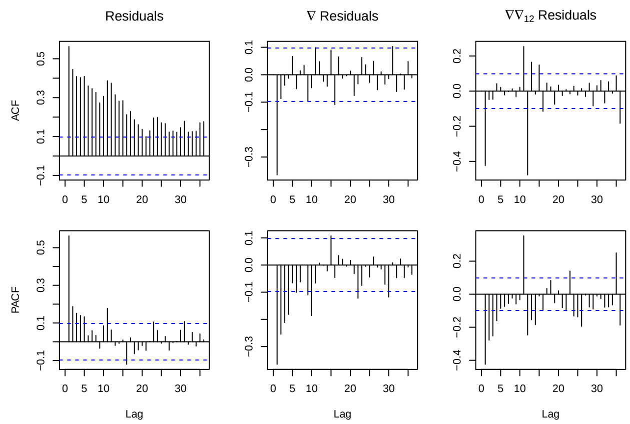

Figure 26: ACF and PACF of the linear model residuals.

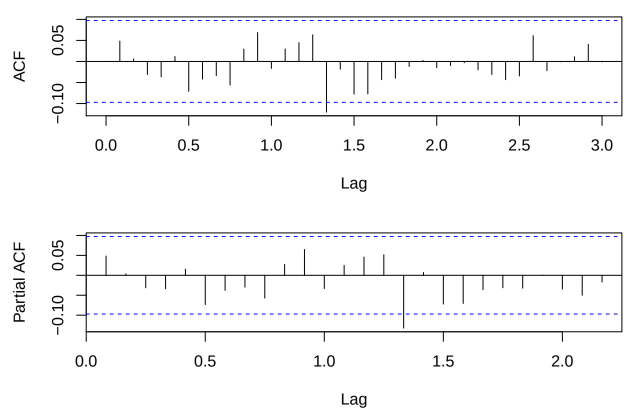

Figure 27: ACF and PACF of the residuals of linear model with correlated errors.

4. Results

4.1 U.S. Homicides rate behavior

The model we chose after examining the ACFs and PACFs of the national homicide rate was an $ARIMA(0,1,1) \times (0,1,1)_{12}$, which equation form is:

$$\nabla_{12}\nabla h_t = \theta_1(B)\Theta_1(B^{12})w_t$$

$$h_t = h_{t-1} + h_{t-12} - h_{t-13} + \theta w_{t-1} + \Theta w_{t-12} + \theta\Theta w_{t-13} + w_t$$

$$\hat{\theta} = -0.5106 \qquad \hat{\Theta} = -0.8042 \qquad \sigma_w^2 = 0.000865 \qquad AIC = -6.0433$$

The `auto.arima` function selected an $ARIMA(2,1,1) \times (1,1,1)_{12}$ which equation form is:

$$\phi_1(B)\Phi_2(B^{12})\nabla_{12}\nabla h_t = \theta_1(B)\Theta_1(B^{12})w_t$$

$$(1 - \phi B - \Phi_1 B^{12} - \Phi_2 B^{24} + \phi\Phi_1 B^{13} + \phi\Phi_2 B^{25})\nabla_{12}\nabla h_t = (1 + \theta B + \Theta B^{12} + \theta\Theta B^{13})w_t$$

$$\hat{\phi} = 0.2426 \quad \hat{\Phi}_1 = 0.1767 \quad \hat{\Phi}_2 = 0.0196 \quad \hat{\theta} = -0.7033 \quad \hat{\Theta} = -0.8793 \quad \sigma_w^2 = 0.000838 \quad AIC = -6.0608$$

We chose the second model as the final model since it had better diagnostics and the AIC is smaller. We used both models to forecast the next three years. The predictions are shown in Figure 28. Although the detailed data for 2015 has not been released, the total number of homicides reported for 2015 is 15,696. Our model predicted a total of 14,844 and a prediction interval of (13,283 ; 16,404) for 2015. The observed data was a bit higher than our prediction (~5.7%), but the real value was inside the 95% prediction interval.

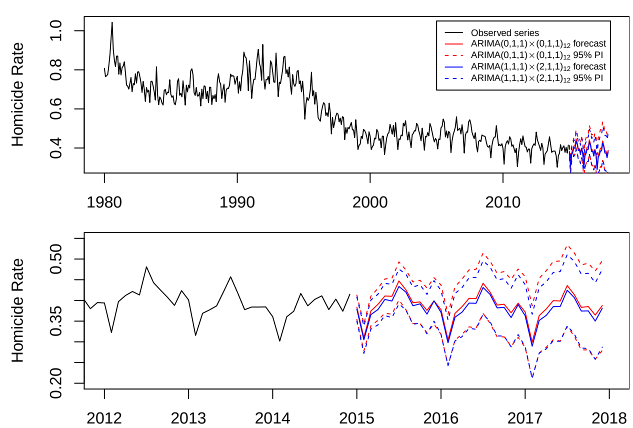

Figure 28: Three year forecast of the national homicide rate.

4.2 Victim's gender

The final models chosen for each gender were an $ARIMA(0,1,1) \times (0,1,1)_{12}$ for the female victims series and a $ARIMA(0,1,1) \times (2,1,1)_{12}$ for the male victims series. Their equational form and estimates are:

$$\nabla_{12}\nabla f_t = \theta_1(B)\Theta_1(B^{12})w_t$$

$$f_t = f_{t-1} + f_{t-12} - f_{t-13} + \theta w_{t-1} + \Theta w_{t-12} + \theta\Theta w_{t-13} + w_t$$

$$\hat{\theta} = -0.7439 \qquad \hat{\Theta} = -0.9602 \qquad \sigma_w^2 = 520$$

$$\Phi_2(B^{12})\nabla_{12}\nabla m_t = \theta_1(B)\Theta_1(B^{12})w_t$$

$$(1 - \Phi_1 B^{12} - \Phi_2 B^{24})\nabla_{12}\nabla m_t = \theta w_{t-1} + \Theta w_{t-12} + \theta\Theta w_{t-13} + v_t$$

$$\hat{\Phi}_1 = 0.1212 \qquad \hat{\Phi}_2 = 0.0239 \qquad \hat{\theta} = -0.5382 \qquad \hat{\Theta} = -0.8743 \qquad \sigma_v^2 = 4,779$$

We used this models to forecast the following three years, which are shown in Figure 29. The total prediction for the 2015 year was 3,160 for female victims and 11,585 for male victims for a total of 14,744 homicides. The sum of the estimates is very close to the estimate using the national homicide rate in the previous section (14,844).

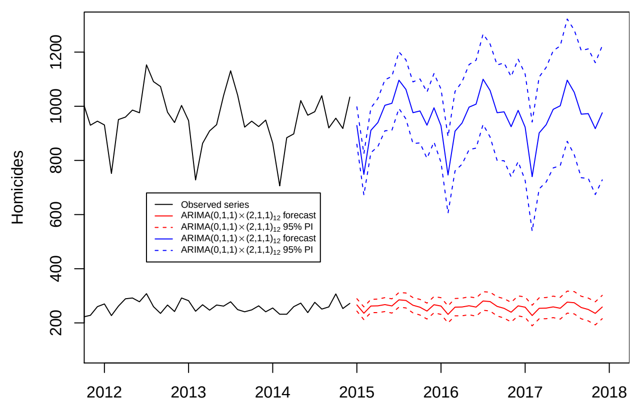

Figure 29: Three year homicide forecast by victim's gender (top male victim, bottom female victim).

4.3 U.S. Regions

The final models for each region are:

Midwest: $ARIMA(0,1,2) \times (0,1,2)_{12}$

$$\nabla_{12}\nabla M_t = \theta_2(B)\Theta_2(B^{12})w_t$$

$$\nabla_{12}\nabla M_t = (1 + \theta_1 B + \theta_2 B^2)(1 + \Theta_1 B^{12} + \Theta_2 B^{24})w_t$$

$$\hat{\theta}_1 = -0.6875 \qquad \hat{\theta}_2 = -0.1039 \qquad \hat{\Theta}_1 = -0.8319 \qquad \hat{\Theta}_2 = -0.0580 \qquad \sigma_w^2 = 0.001473$$

Northeast: $ARIMA(2,1,1) \times (2,1,1)_{12}$

$$\phi_2(B)\Phi_2(B^{12})\nabla_{12}\nabla N_t = \theta_1(B)\Theta_1(B^{12})w_t$$

$$(1 - \phi_1 B - \phi_2 B^2)(1 - \Phi_1 B^{12} - \Phi_2 B^{24})\nabla_{12}\nabla M_t = (1 + \theta B)(1 + \Theta_1 B^{12})w_t$$

$$\hat{\phi}_1 = 0.1491 \qquad \hat{\phi}_2 = 0.1102 \qquad \hat{\Phi}_1 = 0.1689 \qquad \hat{\Phi}_2 = 0.0497 \qquad \hat{\theta} = -0.8301 \qquad \hat{\Theta} = -1.0000 \qquad \sigma_w^2 = 0.002291$$

South: $ARIMA(1,1,2) \times (0,1,2)$

$$\phi_1(B)\nabla_{12}\nabla S_t = \theta_2(B)\Theta_2(B^{12})w_t$$

$$(1 - \phi B)\nabla_{12}\nabla S_t = (1 + \theta_1 B + \theta_2 B^2)(1 + \Theta_1 B^{12} + \Theta_2 B^{24})w_t$$

$$\hat{\phi} = 0.5684 \qquad \hat{\theta}_1 = -1.1751 \qquad \hat{\theta}_2 = 0.2709 \qquad \hat{\Theta}_1 = -0.619 \qquad \hat{\Theta}_2 = -0.2304 \qquad \sigma_w^2 = 0.002659$$

West: $ARIMA(0,1,2) \times (0,1,1)$

$$\nabla_{12}\nabla W_t = \theta_2(B)\Theta_1(B^{12})v_t$$
$$\nabla_{12}\nabla S_t = (1 + \theta_1 B + \theta_2 B^2)(1 + \Theta B^{12})v_t$$
$$\hat{\theta}_1 = -0.5348 \quad \hat{\theta}_2 = -0.1402 \quad \hat{\Theta} = -0.9236 \quad \sigma_v^2 = 0.003003$$

We used each one of this model to make a three year forecast for the homcide rate. The results can be seen in Figure 30.

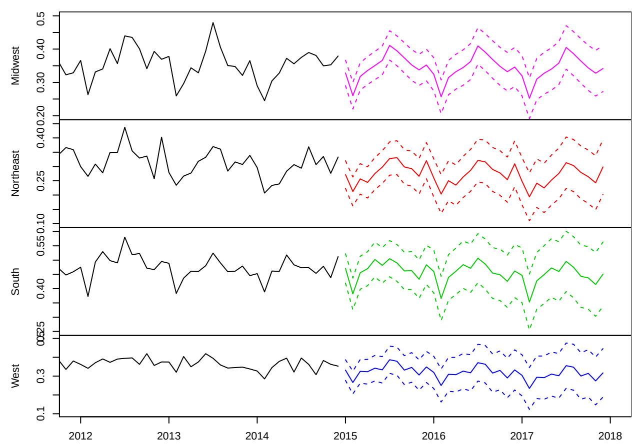

Figure 30: Three year homicide forecast by region.

4.4 Unemployment

Through VAR model and Granger Causilty test, we find out that there is no significant relation between Unemployment rate and homicides rate in the USA at all, which is we did not expect. Figure 31 shows the distribution of homicides by type relationship between victims and offenders. This graph show that 42% of murders happen between people who know each other and only 15% between strangers. Acquaitance murders often count as crime of passion and crime of plan. Ecomic issues will have small influence in crime of passion.

There are a lot of factors that play a role in crimes of passion, for example, the persons' character, education level, poverty, just to name a few. Maybe this can explain why unemployment rate has almost no influnce on the number of homicides. Other reports also point out, with mature social relief, unemployed citizens have some basic guarantee for their daily life, so they will not go further than propety crimes, unless they are inherently criminals.

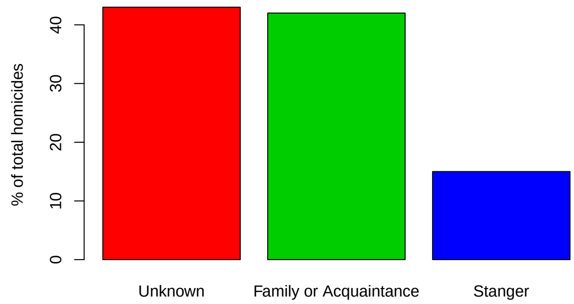

Figure 31: Distribution of homicides by perpetrator-victim relationship.

4.5 Temperature

The final VAR(6) model we fitted has the following parameters:

$$\hat{\boldsymbol{\phi}}_1 = \begin{pmatrix} -0.7096 & 0.0758 \\ -0.0071 & 0.4253 \end{pmatrix} \quad \hat{\boldsymbol{\phi}}_2 = \begin{pmatrix} -0.5336 & 0.043 \\ -0.0063 & 0.132 \end{pmatrix} \quad \hat{\boldsymbol{\phi}}_3 = \begin{pmatrix} -0.4134 & -0.4427 \\ 0.0061 & -0.0914 \end{pmatrix}$$

$$\hat{\boldsymbol{\phi}}_4 = \begin{pmatrix} -0.3414 & 0.462 \\ 0.0049 & -0.134 \end{pmatrix} \quad \hat{\boldsymbol{\phi}}_5 = \begin{pmatrix} -0.3034 & 0.0921 \\ -0.0077 & -0.2255 \end{pmatrix} \quad \hat{\boldsymbol{\phi}}_6 = \begin{pmatrix} -0.0815 & -0.754 \\ -0.002 & 0.0476 \end{pmatrix}$$

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{w}} = \begin{pmatrix} 360.3 & 3.46 \\ 3.46 & 10.08 \end{pmatrix}$$

We fitted an $ARIMA(1,1,1) \times (2,1,0)_{12}$ to the homicides in New York state using a similar procedure as in previous chapters, to compare predictions with the multivariate model. The prediction can be seen in Figure 32.

Given that the 2015 data has not been released we found a different source of estimates in the New York State Crime Report. A total of 610 homicides were reported for 2015, comparing with 616 in 2014, there was a 1% decrease. The ARIMA model predicts 634 homicides, which comparing it with 642 in 2014 (using our data) represents a 1.2% decrease. The VAR(6) model predicits 664 for 2015, comparing with 642 in 2014 represents an increase of 3.4%.

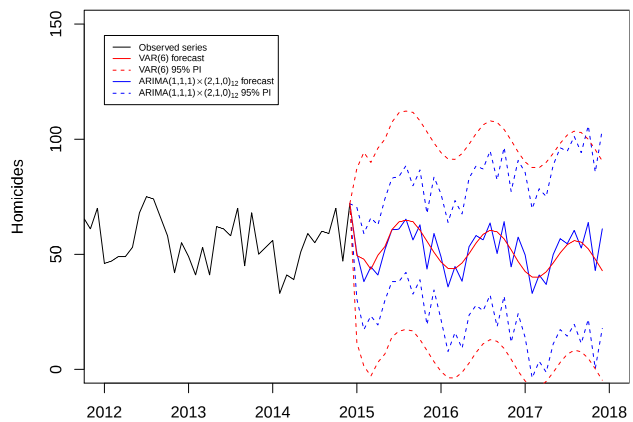

Figure 32: Three year forecast of the New York state number of homicides cases

4.6 Clearance rate

The final VAR model chosen was:

$$\hat{\phi}_1 = \begin{pmatrix} -0.295 & -0.0711 \\ 0 & 0.3542 \end{pmatrix} \quad \hat{\phi}_2 = \begin{pmatrix} -0.181 & 0 \\ -0.0608 & 0.116 \end{pmatrix} \quad \hat{\phi}_3 = \begin{pmatrix} -0.105 & 0.241 \\ -0.119 & 0 \end{pmatrix}$$

$$\hat{\phi}_4 = \begin{pmatrix} -0.107 & 0 \\ -0.143 & 0 \end{pmatrix} \quad \hat{\phi}_5 = \begin{pmatrix} -0.0577 & 0 \\ -0.1614 & 0.0925 \end{pmatrix} \quad \hat{\phi}_6 = \begin{pmatrix} -0.1309 & 0 \\ -0.0907 & 0 \end{pmatrix}$$

$$\hat{\phi}_7 = \hat{\phi}_8 = \hat{\phi}_9 = \hat{\phi}_{10} = \mathbf{0}$$

$$\hat{\phi}_{11} = \begin{pmatrix} 0 & 0 \\ -0.107 & 0.118 \end{pmatrix} \quad \hat{\phi}_{12} = \begin{pmatrix} 0.593 & -0.115 \\ -0.115 & 0.175 \end{pmatrix}$$

$$\hat{\mathbf{\Sigma}} = \begin{pmatrix} 1.067454e^{-03} & -1.383492e^{-06} \\ -1.383492e^{-06} & 5.323408e^{-04} \end{pmatrix}$$

We used this model to forecast 2015 homicide rate and clearance rate, shown in Figure 33. The total estimate for 2015 was 14,580 homicides and a clearance rate of 67%.

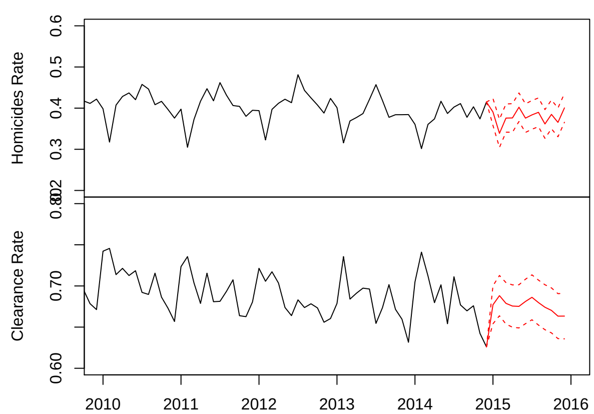

Figure 33: Three year forecast of the national homicide rate and clearance rate.

The model for the linear regression with correlated errors is:

$$
\begin{aligned}
cl_t = {} & 0.0756\nabla h_t - 0.0042\nabla h_{t-1} - 0.0550\nabla h_{t-2} - 0.0906\nabla h_{t-3} - 0.0756\nabla h_{t-4} - 0.0589\nabla h_{t-5} - 0.0884\nabla h_{t-6} \\
& - 0.0191\nabla h_{t-7} - 0.0194\nabla h_{t-8} - 0.1143\nabla h_{t-9} - 0.0943\nabla h_{t-10} - 0.0951\nabla h_{t-11} - 0.0439\nabla h_{t-12} + v_t
\end{aligned}
$$

where v_t is an $ARIMA(0,1,1) \times (0,1,1)_{12}$ with $\hat{\theta} = -0.7527$ and $\Theta = -0.8699$.

5. Conclusions

The final model for the national homicide rate is an $ARIMA(2, 1, 1) \times (1, 1, 1)_{12}$, which predictions for 2015 was a total of 14,844 homicides. The observed value for 2015 was aproximately 5.7% higher than our prediction and it was contained by the prediction interval. There was an increase of 9.4% in the national number of homicides with respect to 2014, which was kind of unexpected because the trend show a slow decay in the previous years.

1. Is the behavior of homicides by victim's gender different?

In the victim's gender analysis, it was clear from the begining that the number of homicides of males is much higher than of females, aproximately 3.5 times, but the trend of both series is very similar. The behavior of the homicides by gender it is clearly different, as we show by fitting distinct models to each series.

The forecast for 2015 was 3,160 female victims and 11,585 victims, for a total of 14,744, which is close to the national estimate. It is worth noting that there is an average of 28 unidentified gender murder victims and that this were not included in the analysis.

2. Is there a difference in the behavior of homicide rates across regions in the United States?

Our suspicion about the South region of the country was confirmed by the data. The South region has the highest homicide rate and all other three regions have very similar homicides rates at least in the last 4 years.

We found out that the homicide rate's behavior across the four regions is different. Each region had his specific model fitted and even though we had problems with the West region, it was clear its behavior was different from all other regions.

The forecast of homicides for each region was:

Region	Homicides
Midwest	2,835
Northeast	1,906
South	6,774
West	3,047
Total	**14,563**

3. Is there a relationship between homcides and unemployment rate?

The results of anaylsis show that homicides are unique, that they are not economically motivated. "Murder and its junior partner, assault, for instance, are mainly precipitated by such things as anger, sexual jealousy, perceived insults or threats, and long-standing personal quarrels, and they are frequently facilitated by disinhibitors such as alcohol or drugs. These factors, it appears, are unaffected by economic downturns or upswings."(Los Angeles Times). This citation explains excatly why there is a limited relationship between homicides and unemployment rate.

4. Is there a relationship between homcides and the weather in New York state?

Our anaylsis shows us that there is a relationship between temperature and homocides in New York. Homicidies happen more during summer time and this is maybe just because people do more outside activities and interact more with each other. We do not recommend to use weather escpecially the tempeture as an influence factor in future analysis, because the multivariate model appears to have a bad predictive performance, at least in 2015, compared to the univariate model.

5. What kind of relationship exist between the clearance rate of homicides and homicides rate?

We determined there is an interaction between the two variables by fitting a multivariate model an obtaining significant parameters. We then fitted a linear model with correlated errors using the clearance rate as the response variable, because it is observed after the homicides, and the differenced lagged homicide rate as the explanatory variable. We saw that all laged explanatory variables had a negative relationship with the clearance rate, which makes sense since the higher their value means an increase in the homicide rate that translates into more worload for law enforcement officers.

An interesting analysis would be to add as an explanatory variable the government expenditure in law enforcement aggencies or the historical headcount of these agencies, to see wether a significant higher clearance rate could be achieved by hiring more personnel.

Bibliography

C. Sax and P.Steiner, "Temporal Disaggregation of Time Series", The R Journal Vol. 5/2, December 2013, ISSN 2073-4859.

C. Sax and P. Steiner. "tempdisagg: Methods for Temporal Disaggregation and Interpolation of Time Series", 2013. URL http://CRAN.R-project.org/package=tempdisagg. R package version 0.25.

S. Wójcik, "Temporal Disaggregation of Time Series with Regularization Term", Barometr Regionalny Tom 14 NR 3, 2016.

G. Chamberlin, "Temporal Disaggregation", Economic & Labour Market Review, Nov. 2010.

R. Shumway and D. Stoffer, "Time Series Analysis and Its Applications with R Examples", Third Edition, Springer, 2011.

"Cointegration: The Engle and Granger approach" URL http://www2.warwick.ac.uk/fac/soc/economics/staff/gboero/personal/hand2_cointeg.pdf G. S. Maddala, In-Moo Kim, "Unit Roots, Cointegration, and Structural Change", Cambridge University Press, 1998

Sucheta Pradhan, "Does Unemployment Increase Crime Rate?", 2013. URL http://www.buzzle.com/articles/does-unemployment-increase-crime-rate.html

Barry Latzer, "Do hard times spark more crime?", 2014, URL http://www.latimes.com/opinion/op-ed/la-oe-latzer-crime-economy-20140124-story.html

Andrew M. Cuomo, "New York State Crime Report", May 2016, URL http://www.criminaljustice.ny.gov/crimnet/ojsa/NYSCrimeReport2015-prelim.pdf