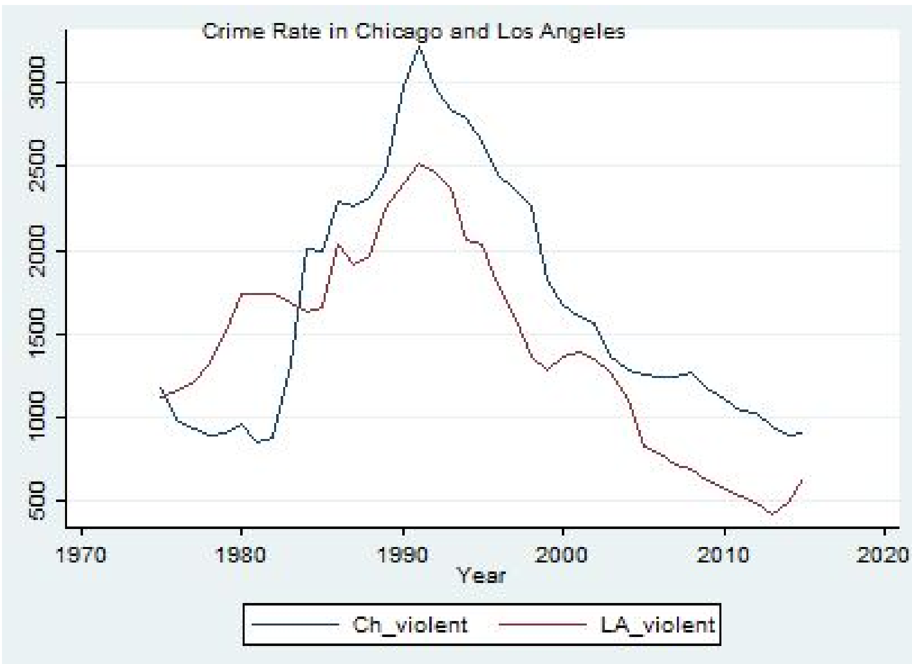
Violent Crime Rate in Chicago and Los Angeles

Sona Avagyan

February 2023

1 Introduction

This paper is aimed to analyze the relationship of violence crime rate of two large cities of the United States. Historical analysis shows that the crime rates have a decreasing pattern for these two cities. When we look at the simple plot of the historical crime rate per capita figures, we can observe that these two series are seemed to move together. In this paper, we would like to analyze whether there indeed exists a relationship between the crime rates in two cities, and whether these series are cointegrated or not.



Data Description

Data is extracted from Kaggle, it contains the annual crime rates per capita of two US major cities: Chicago and Los Angeles. It covers the period from 1975 to 2015 and only the number of violent crimes is counted in the dataset. The term "violent crime" refers to cases when a victim is harmed by or threatened with violence.

It is also important to consider that our dataset is small, the data is collected on an annual basis for 41 years and the total number of observations in the dataset is exactly 41. This is a major limitation for the analysis, however, still, we can see whether there is some relationship between the crime rates of those cities.

2 Test for stationary

Firstly we need to conduct a test for stationary and see In time series analysis first of all it is crucial to understand what kind of data set we use and whether we deal with stationary variables or non stationary ones. Hence the first thing that we did was implement the Augmented Dicky Fuller test and see whether the initial variables of violent crime per capita are stationary or not.

As we can see from the tables the Dicky-Fuller test confirmed our initial suggestion that the variable violent crime per capita is a non stationary variable since we fail to reject the null hypothesis (T statistics -0.948 $<$ Critical value of -1.688). These are the results for the test with a drift. We also conducted another test including the trend in order to understand whether there is some trend around which we can claim that the variable is stationary. The t statistics is again less than the critical value, hence we fail to reject the null hypothesis and we deal with nonstationary variable of LA violent crime per capita

```
. dfuller la_violent, drift lags(1)

Augmented Dickey-Fuller test for unit root         Number of obs   =        39

                              ———————————— Z(t) has t-distribution ————————————
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
──────────────────────────────────────────────────────────────────────────────
 Z(t)             -0.948            -2.434            -1.688            -1.306
──────────────────────────────────────────────────────────────────────────────
p-value for Z(t) = 0.1747
```

```
. dfuller ch_violent, drift lags(1)

Augmented Dickey-Fuller test for unit root         Number of obs   =        39

                              ———————————— Z(t) has t-distribution ————————————
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
──────────────────────────────────────────────────────────────────────────────
 Z(t)             -1.441            -2.434            -1.688            -1.306
──────────────────────────────────────────────────────────────────────────────
p-value for Z(t) = 0.0790
```

We also conducted similar tests for Chicago which showed the same result as for LA. Therefore, we can state that we have two nonstationary variables in the data set for our analysis.

For further purposes, we also test the integration order of the processes and from the below results we can see that we have first-order integrated series I(1), as the variables are stationary in their first lag differences. So, we can conclude that we have two random walk processes in the dataset.

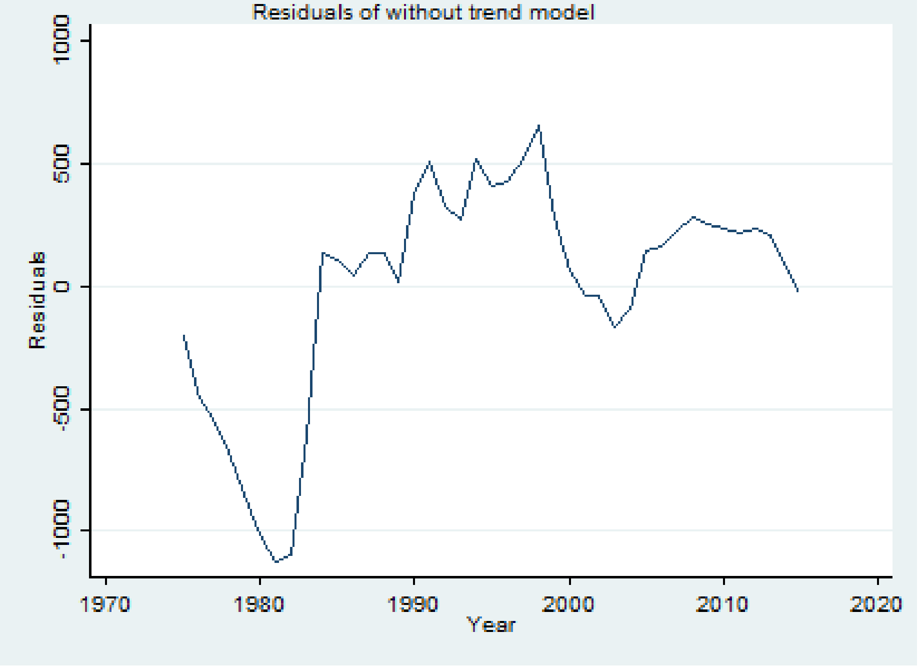
```
. dfuller d.ch_violent

Dickey-Fuller test for unit root                   Number of obs   =        39

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -3.873            -3.655            -2.961            -2.613
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0022
```

```
. dfuller d.la_violent

Dickey-Fuller test for unit root                   Number of obs   =        39

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -3.808            -3.655            -2.961            -2.613
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0028
```

3 Conintegration

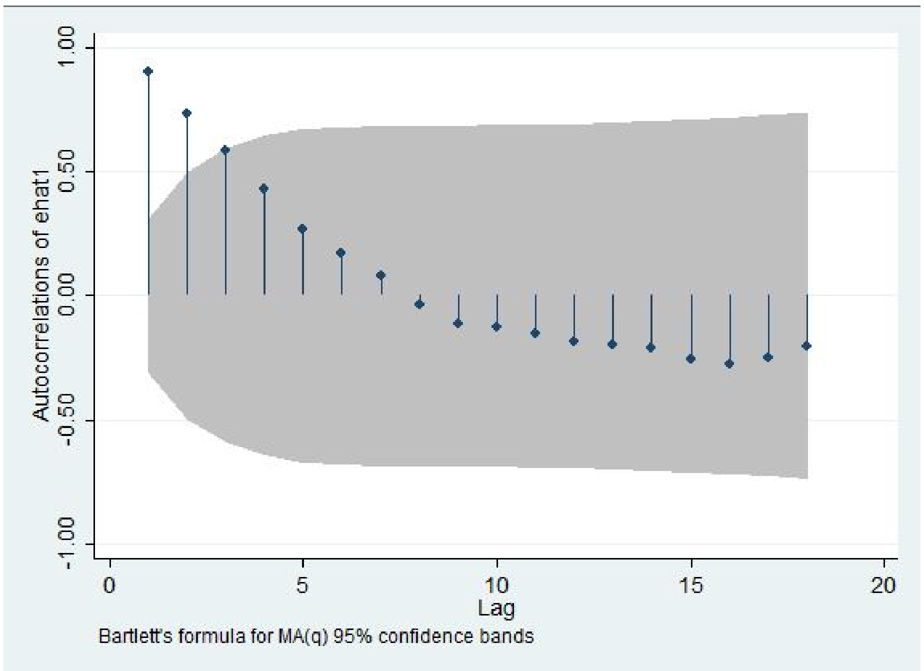
As we have already checked for stationarity, now we can proceed on and check whether there is cointegration between our two series, meaning that our goal is to see whether the crime rates of Chicago and Los Angeles diverge drastically over time, or despite some short-term deviations, they still move together in the long run. For this purpose, we create two regressions: the first one only includes crime rates for Chicago and LA while the second one also includes the time variable to capture the trend effect in the regression. From the regression results, we get the predicted residuals and when we plot them. At the first sight, it is hard to identify whether the residuals follow the stationary process or not. Thus, to understand, we run another Dicky-Fuller test for both regressions and get that both error terms are also nonstationary processes. As we see that the error term is also nonstationary which indicates that we don't have cointegration for crime rates in Chicago and in LA. (T stat of -2.731 $<$ critical value of -3.404 at 95% confidence interval)

Residuals Plot



4 Autocorrelation

So far, we conducted stationarity tests, and cointegration test. Now let's understand whether there is autocorrelation in our series. For that, we can plot a correlogram. From the bellow correlogram, we can observe that we have second-degree autocorrelation of the error term, at two lags of the error term is outside of the 95% confidence interval.



We also conducted a Breusch Godfrey test for serial correlation which also showed the same results. Thus, this indicates that if we proceed with our analysis using only a simple regression model, we will end up with false standard errors. Therefore we decided to use an ARDL model.

5 ARDL Model

As we observed above, we have an I(1) process with no cointegration and serial correlation, which means that we can design an ARDL model with the first differences to estimate the relationship between violent crime rates in Chicago and Los Angeles. However, before designing the model we need to understand the lag level of our model so that we can come up with the best model. We used the Akaike information criterion (AIC) to identify the lag level of the model. For both series, the AIC test showed that we should take the first lag difference in the model.

AIC Test For Chicago

```
. varsoc D.ch_violent

   Selection-order criteria
   Sample:  1980 - 2015                        Number of obs      =        36

  +---------------------------------------------------------------------------+
  |lag |    LL      LR      df    p      FPE       AIC      HQIC      SBIC    |
  |----+----------------------------------------------------------------------|
  |  0 | -243.535                      46510.1   13.5853   13.6006   13.6293  |
  |  1 | -239.645  7.7796*   1  0.005  39615.7*  13.4247*  13.4554*  13.5127* |
  |  2 | -239.637   .0166    1  0.897    41871   13.4798   13.5259   13.6118  |
  |  3 | -239.463  .34883    1  0.555  43859.5   13.5257   13.5871   13.7016  |
  |  4 |   -238.8   1.326    1  0.250    44728   13.5444   13.6212   13.7644  |
  +---------------------------------------------------------------------------+
   Endogenous:  D.ch_violent
    Exogenous:  _cons
```

AIC Test For Los Angeles

```
varsoc D.la_violent

 Selection-order criteria
 Sample:  1980 - 2015                        Number of obs      =        36

+---------------------------------------------------------------------------+
|lag |    LL      LR      df    p      FPE       AIC      HQIC      SBIC    |
|----+----------------------------------------------------------------------|
|  0 | -230.268                      22255.8   12.8482   12.8636   12.8922  |
|  1 | -227.273    5.99*   1  0.014  19922.8*  12.7374*  12.7681*  12.8254* |
|  2 | -227.054  .43845    1  0.508  20811.7   12.7808   12.8268   12.9127  |
|  3 | -227.053  .00077    1  0.978  22011.9   12.8363   12.8977   13.0122  |
|  4 | -227.046  .01525    1  0.902  23280.2   12.8914   12.9682   13.1114  |
+---------------------------------------------------------------------------+
 Endogenous:  D.la_violent
  Exogenous:  _cons
```

ARDL Regression Output

```
. reg D.ch_violent L.D.ch_violent D.la_violent L.D.la_violent

      Source |       SS           df       MS      Number of obs   =        39
-------------+----------------------------------   F(3, 35)        =      4.69
       Model |  455391.445         3  151797.148   Prob > F        =    0.0074
    Residual |  1131711.29        35  32334.6083   R-squared       =    0.2869
-------------+----------------------------------   Adj R-squared   =    0.2258
       Total |  1587102.74        38  41765.8615   Root MSE        =    179.82

------------------------------------------------------------------------------
D.ch_violent |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  ch_violent |
        LD. |   .3653754   .1510888     2.42   0.021     .0586489    .6721019
             |
  la_violent |
         D1. |   .3927614   .2182156     1.80   0.081    -.0502397    .8357626
         LD. |   .0845622   .2330801     0.36   0.719    -.3886156    .5577399
             |
       _cons |   7.571902   29.01074     0.26   0.796    -51.32304    66.46684
------------------------------------------------------------------------------
```

As we can observe from the above regression output the change in the crime rate of Los Angeles does not have a significant impact on the change in crime rates in Chicago at 95% confidence level, as the values are higher than 0.05. However, we can see that the first lag difference of the Chicago crime rate has a positive significant impact on the difference of Chicago crime rate, which is in fact expected (increase by a factor of 0.365). The crime rate in a specific city now is expected to be dependent on its lag values. This result shows us that there is no relationship between the crime rate figures of these two large cities in the United States. Although, from the initial point, it seemed that these two series are tightly moving together over time.

To conclude, we conducted multiple tests in order to confirm the non stationarity of our series and then we also check for a cointegreation, to understand whether there is some relationship between the crime rate figures over 41 years of Chicago and Los Angeles. We have come up with one lag differenced ARDL model to estimate the series. Eventually, the results show no significant relationship between the crime rates in LA and Chicago.