Crime and rain: is there a correlation?

Tauã Vital and Eduardo Almeida

July 2020

Abstract

In this paper, time series models are used to address the relation between property crime and rain for the city of Rio de Janeiro. Daily data for the years 2013 to 2017 were used to make short-term forecasting of property crime to the first month of 2017. Results showed that there is no statistically significant relation between the two variables. Artificial Neural Networks showed a better accuracy perfomance when compared to traditional time series models, such as ARIMA and SARIMA.

Keywords: crime; rain; time series; ARIMA; ANN.

1. Introduction

As crime rate increases in Brazil, the topic has risen to one of the major concerns in the domestic policy agenda, so has the need to relevant empirical discoverings that can explain the phenomenon. Since the 1950's, and specially after Becker (1968) proposes his theory of rational criminal behaviour, we saw a profusion of researchers dedicating their time to explore the incentives and the disincentives that leads an individual to commit a crime. The majority of those studies are focused on exploring the relation between crime and socio-demographic variables, such as income, gender, and inequality (Kelly (2000); Entorf and Spengler (2000); Levitt (2004)). While a few others addresses the impacts of public policies created to fight crime (Donohue and Levitt (2001); Levitt (1998); Pereira Filho and De Sousa (2018); Vital (2018)). However, as enphasized by Cohn (1990), these traditional criminological variables changes slowly over time and can not explain short term variations in crime rate.

Traditionally, the economic crime theory neglects the effect of situational contexts in the decision of an individual to commit a crime. Vital, De Souza, Faciroli, et al. (2020) proposes a new approach in which the decision to commit a crime is based on two factors. First everyone has inside yourself a tendency to do some evil, which explains the behaviour of psycopaths for exemple. Second, the individuals are influenced for the environment in which they live in. This last one can explain better short term variations of crime, such as the possibility of a change in the weather influence the decision to one to commit a fellony.

Weather is one variable that is mostly neglected in behaviour criminlogical theories. The effects of temperature in human behaviour has been a subject of interested of researchers for a long time (Cohn (1990); Horrocks and Menclova (2011); Field (1992); Anderson, Anderson, Dorr, DeNeve, and Flanagan (2000); Stec and Klabjan (2018). Findings reveal that the higher temperatures leads to more activities, and more criminality (Michel, Wang, Selvarajah, Canner, Murrill, Chi, Efron, and Schneider (2016)). However the relation between crime and rain is not well stablished in the literature (Cohn (1990)). So aiming to fill this gap, this paper raise a simple question: is there a correlation between crime and rain?

Not wanting to sound unacademic, but we believe that is a common sentiment of lethargy among people in rainy days. And a more formal hypothesis, based on the rationality of individual, incorporated in crime theory by Becker (1968), rainy days raise the opportunity cost of criminals by two means. First, during periods of rain, fewer people are walking on the streets, so the chance of encounter between the offender and its victim are smaller. Second, usually in big cities, rainy days are associated with an increase in traffic, so offenders that go from suburban areas to distant regions of the city to commit crimes must take that into account.

We sought to adress the relation between property crime and rain using daily data of Rio de Janeiro city in Brazil. We collected microdata from police reports and compiled a daily dataset of robbery and theft of cars from January 2013 to January 2017. Then, combining with meteorological data from the National Institute of Spatial Research - a Brazilian government agency, we explored the exogeneity of rain, employing an ARIMA model with rain as a exogenous regressor to access the relation between the variables. Chen, Yuan, and Shu (2008) showed that an ARIMA model has a great accuracy forecasting the short term property crime, namely weekly, in China.

Stec and Klabjan (2018) showed that artificial neural networks have a better accuracy in a classification setting of crime prediction. The authors employed a mix of recurrent neural networks with convulation neural networks to predict if the number of crimes belongs to certain ranges 0-5, 6-10, etc. The results presented that artificial neural networks have a great accuracy in predicting crime. Also, Stec and Klabjan (2018) showed that the inclusion of variables related to weather and public transportation improved the accuracy of the models. The goal of this paper is not classification but regression, aiming to forecast the number of crimes in a certain day in Rio de Janeiro. Given the results presented by the authors we decided to include in our analysis the use of artificial neural networks to predict the number of crimes in Rio de Janeiro. Thorugh the best of our knowledge, there is no other study doing the same.

Besides this introduction section, this work is divided as follow: section 2 presents the dataset, as well as the empirical methodology; section 3 presents the results and the discussion; section 4 are dedicated to robustness check; finally section 5 concludes the paper.

2. Dataset and Methods

We compiled daily data of crime from the police reports of Rio de Janeiro for the entire years of 2013 to 2016, defining these period as our training set, while January 2017 data was used as the test set. Much of the literature of weather on crime are focused on violent crimes (Cohn and Rotton (1997); Michel et al. (2016). However it is expected that rain can influence more in the offenders decision when that is a property crime. That being said, we used car theft and robbery as a proxy for criminal activity. Also, using another types of minor crimes as dependent variable can lead to poor results due to under reporting of such felonies. Specially in days of heavily rain, we conjecture that is less likely that an individual will report a property crime, if the stolen good was of small value. Rainy data is in milimeters of precipitation throughout the day. Exhibits 1 and 2 shows the plots for the time series of crime and rain, while Table 1 present the description of the time series.

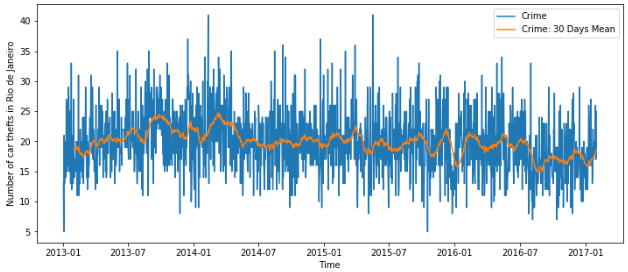

Fig. 1. Crime series

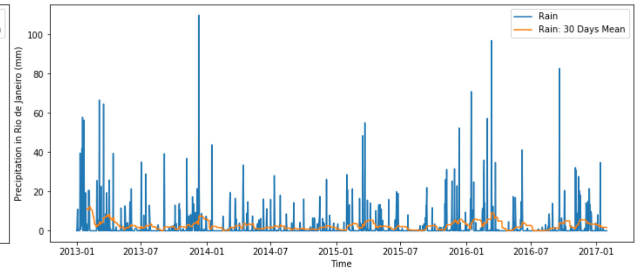

Fig. 2. Rain series

Table 1: Description of the time series in Rio de Janeiro

Time serie	Description	Source	Mean	Std. Dev.	Max.	Min.
crime	Number of car thefts	ISP-RJ	19.796	5.300	41.0	4.000
rain	Precipitation in milimiters	INPE	2.555	8.130	109.8	0.000

The choice of Rio de Janeiro as our alaysis unit is that temperature in the city is usualy high with low variance along the year (the average temperature was 23.2°C in the last ten years) , so it is more likely to isolate the effect of rain.

Our empiral estrategy consists of two approaches. First, we consider a linear approach through an ARIMA(p,d,q) model with rain as a exogenous regressor.

$$crime_t = \alpha_0 + A(L)crime_{t-1} + c_0 rain_t + B(L)\epsilon_t \tag{1}$$

We assume that the input process $rain_t$ and the white noise process ϵ_t are boht stationary and mutually independent. *A(L)* and *B(L)* are polynomials in the lag operator assigning weights to past values of $crime_t$ and ϵ_t.

Second, we relax the linear hypothesis employing a one hidden layer feed forward neural network with the precipitation as an exogneous regressor. Hyndman and Athanasopoulos (2018) provides a nice function *nnetar* in the *forecast* package in R. Basically the function trains 20 NNs by adopting random starting values and then obtains the mean of the resulting predictions to compute the forecasts. The neural net assumes in this scenario the following form:

$$\hat{crime}_t = \hat{\beta}_0 + \sum_{j=1}^{t} \hat{\beta}_j \psi(\mathbf{X}^T \bullet \hat{\delta}_j) \tag{2}$$

where $\mathbf{X}^T$ consists of the transpose matrix lags of $crime_t$ and the exogenous variable.

Then, the function $\psi\ (\mathbf{X}^T \bullet \hat{\delta}_j)$ has the logistic form:

$$\psi(\mathbf{X}^T \bullet \hat{\delta}_j) = \left[1 + exp\left(- \hat{\delta}_{j0} + \sum_{i=1}^{p} \hat{\delta}_{j1} \bullet cr\hat{i}me_{t-1}\right)\right]^{-1}, j = 1, ..., K \tag{3}$$

The nonlinearity arises through the lagged $crime_{t-1}$, entering in a flexible way through the logistic functions of equation 3. The number of logistic functions (K) included is known as the number of hidden nodes.

The relation between crime and rain will be evaluated in two manners. First, through the statistical significance of the coefficient in the ARIMA model. Second, comparing the forecasting performance of the models including or not the exogenous variable (Stec and Klabjan (2018)). In case, the inclusion of the precipitation variable in the model improves the forecasting accuracy of crime, we can assume that there is a correlation between the variables.

3. Results and Discussion

3.1. ARIMA results

Determining a parsimonious model involving a simple form for equations 1 is the main task in the ARIMA methodology (Shumway and Stoffer (2017)). The first step to identify the best fitted model is determining the if the series are stationary (dos Santos and Kassouf (2012). For that we employed augmented Dickey-Fuller (ADF), Phillips-Perron (PP) and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test statistics. Table 1 shows the results.

It is important to emphasize that the sthocastic trend of the time seris can be linked to the variance of the data. Taking that into account before testing for the presence of unit roots in the crime series we employed the Box-Cox test (Box and Cox (1964)) to evaluate the necessity of transforming the data. The maximum likelihood estimation procedure showed that no transformation was necessary.

The Augmented Dickey-Fuller (ADF) (Dickey and Fuller (1979)) test was used to test the null hypothesis of the presence of a unit root in the time series sample. Presence of unit root implies that the series is non-stationary. The alternative hypothesis is that the series under test is stationarity or trend-stationarity. The ADF test involves the following regression:

$$\Delta x_t = \mu + \gamma t + \alpha x_{t-1} + \sum_{j=1}^{k-1} \beta_j \Delta x_{t-j} + u_t \tag{4}$$

where Δ is the operator difference and u_t is a white noise. The α parameter is analyzed based on its negativity and significance. Table 2 shows the results of the ADF test for both time series, crime and rain.

Table 2: Unit root test for the time series

Time serie	ADF	KPSS	PP	Conclusion
crime	-6.382***	3.645	-35.709***	Data has no unit root and is stationary
rain	-34.410***	0.2954	-34.436***	Data has no unit root and is stationary

As it can be seen both series are stationaries, therefore no differentiation is neccessary prior to the estimations. The next step is to determine the lag autorregressive order AR(p) and the moving average order MA(q). Autocorrelation function and partial autocorrelation function are used to help determining those orders, but also a variety of models were estimated and the decision of the best order for the ARIMA(p,d,p) are based on the AIC and BIC criterias. Lemon, Partridge, et al. (2017) showed that there is a relation between the number of assaults and days of the weekend, emphasizing the importance of taking into account the week seasonality in the crime pattern. The ETF decomposition does not show a seasonal pattern in the time series of homicides. We came into conclusion that the best fit was the ARIMA(1,0,1) model in which we include rain as an exogenous variable. Table 3 presents the results for the model, and figure 3 shows the regression fit in the out-of-sample period set (January 2017).

Table 3: ARIMA(1,0,1) with rain as an exogenous regressor

	Coefficient	Standard-error	p-value
AR(1)	0.991	0.007	0.001
MA(1)	-0.955	0.016	0.001
Intercept	19.529	0.681	0.001
Rain	0.018	0.016	0.255
Number of observations		1461	
AIC		8935.356	
BIC		8961.790	
Jarque Bera		44.78	
Ljung-Box		96.22	

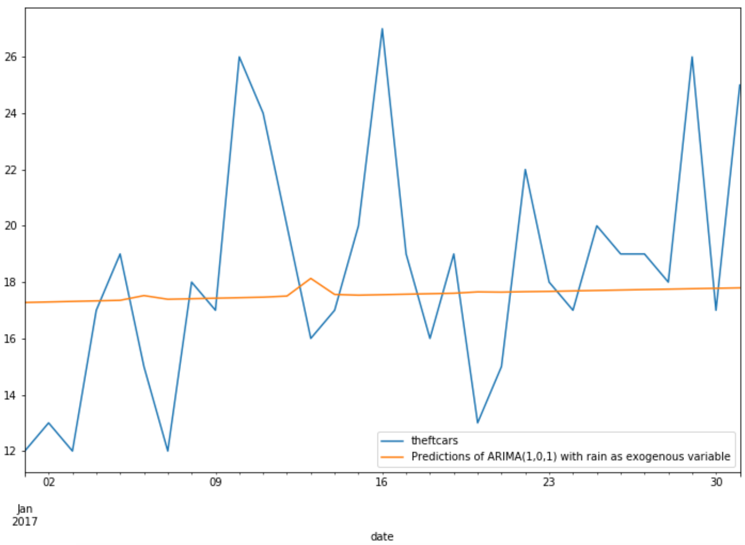

Fig. 3. Out of sample predictions of the ARIMA(1,0,1) with rain as exogenous variable model

The number of car thefts is statistically significant linked to the number of car the thefts in the day before (t - 1) as well as with the lagged error term. The coefficient of the precipitation is positive but not statiscally significant at conventional levels. In this scenario we do not found evidence that the millimeters of rain precipitation have an impact on the number of car thefts on Rio de Janeiro in the period of 2013 to 2016. Figure 3 shows the fit of the model in the out of sample estimations for the period of January 2017. As it can be seen the ARIMA model produces a smooth estimation of the series.

3.2. Accuracy evaluation

Another way to evaluate if rain is important to explain the number of car thefts in Rio de Janeiro is to test if the inclusion of this variable in the models improve the out-of-sample forecasting accuracy. Be that as it may, in this subsection we estimated a bigger number of models, in which some we included the rain variable as an exogenous whilst in others we did not. The out-of-sample period was stablished as January 2017.

We used three measures to evaluate the accuracy of the models:

(i) Mean Squared Error (MSE):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(crime_i - cr\hat{i}me_i\right) \tag{5}$$

(ii) Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(crime_i - cr\hat{i}me_i\right)}$$

(iii) Mean Absolute Error (MAE):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|crime_i - cr\hat{i}me_i\right| \tag{7}$$

As mentioned in the previous section we employed classic univariate time series models with rain as an exogenous variable, such as the ARIMA(p,d,q) and taking into account that crime can present also seasonality, we tested for the presence for weekly seasonality throgh SARIMA(p,d,q)(P,D,Q,m) models, despite that the ETF Decomposition of the homicide time series does not show a seasonal pattern. As Stec and Klabjan (2018) showed, artificial neural networks can have a great accuracy perfomance in a classification setting of crime prediction. That being said, we employed an aritificial neural network to predict the number of crimes.

In order to test the hypothesis that only days with heavy rain have an impact on criminal behaviour, we tested the inclusion of a dummy variable in the models, assuming value equal to one if it rained more than 40 millimeters in that day, and zero otherwise. Table 4 presents the results of the estimations.

Table 4: Accuracy measures of the esimations

Model specification	MSE	RMSE	MAE
ARIMA(1,0,1)	16.508	4.063	3.072
ARIMA(1,0,1) with rain	16.743	4.092	3.108
ARIMA(1,0,1) with dummy of heavy rain	16.553	4.069	3.076
SARIMA(1,0,0)(2,0,2,7)	17.130	4.139	3.057
SARIMA(1,0,2)(1,0,2,7) with rain	17.901	4.230	3.162
SARIMA(1,0,0)(1,0,2,7) with dummy of heavy rain	16.072	4.009	2.925
ANN(18,1,10)	5.375	2.893	4.224
ANN(18,1,10) with rain	5.375	2.893	4.224
ANN(18,1,10) with dummy of heavy rain	5.375	2.893	4.224

As it is shown in Table 4, the inclusion of variables related to rain does not improve the accuracy results in the out of sample estimations. These results are in consonance with the results of Table 3.

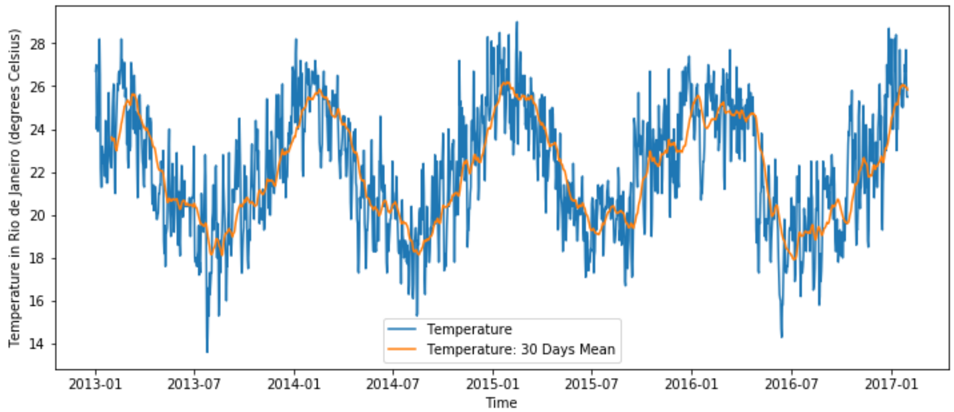

Fig. 4. Temperature series

4. Robustness check

Cohn (1990) identified a relation between short time variations in the crime rate and temperature. Despite that we do not find any statistically significant relation between the number of car thefts and precipitation, maybe that is a relation between the latter and the avarage daily temperature.

Therefore we estimates a modification version of equation 1:

$$crime_t = \alpha_0 + A(L)crime_{t-1} + c_0 temperature_t + B(L)\epsilon_t \tag{8}$$

where temperature is the avarage daily temperature in the city of Rio de Janeiro in degrees Celsius. Figure 3 shows the time series of the latter.

It is clear the seasonality pattern in the time series. Before estimations we used the Hodrick–Prescott filter to deseasonalize the data. We followed the same procedure described in subsection 3.1, we arrived at a ARIMA(6,0,1) as the best fit. Table 5 shows the results of the estimations.

Table 5: ARIMA(6,0,1) with temperature as an exogenous regressor

	Coefficient	Standard-error	p-value
AR(1)	0.888	0.048	0.001
AR(2)	-0.026	0.035	0.452
AR(3)	0.031	0.036	0.403
AR(4)	0.021	0.038	0.577
AR(5)	-0.012	0.036	0.734
AR(6)	0.049	0.030	0.098
MA(1)	-0.852	0.041	0.001
Intercept	1.032	0.457	0.024
Temperature	-0.049	0.070	0.485
Number of observations		1461	
AIC		8942.687	
BIC		8995.556	
Jarque Bera		42.53	
Ljung-Box		71.31	

The p-value associated with the exogenous regressor temperature is not statistically significant at conventional levels. Evaluating the accuracy of the prediction using this model, the mean squared error is 15.361, the root mean square error is 3.919, and the mean absolute error is 2.976. When compared to the model estimated in subsection 3.1, ARIMA(1,0,1) with rain as an exogenous variable, we found an improvement in accuracy prediction. This can be interpreted as changes in temperature having a stronger relation with the number of homicides than rain.

5. Conclusion

Considering the temporal autocorrelation in the criminal activity we assessed the relation between the number of homicides and rain in two ways. First, through the statistical significance of the variable in a ARIMA model. Second, relaxing the linearity assumption and adopting a artificial neural network to evaluate if the inclusion of variables linked to rain improved the accuracy of the models. In both scenarios, there is not a statistical correlation between crime and rain.

Time series model of ARIMA and Aritificial Neural Networks were used to make short-term forecasting of property crime for the city of Rio de Janeiro in Brazil. The fitting and forecasting results were compared between the methods as well with the inclusion of the exogenous regressor of rain. The result shows that Artificial Neural Networks model fits the data well and makes higher accurate forecasting than ARIMA model. This work can be helpful to the local police stations and municipal governments in showing the efficiency of ANN in forecasting crime, therefore this tool can be used to improve the process of decision-making and emergency management.

References

Anderson, C. A., Anderson, K. B., Dorr, N., DeNeve, K. M., Flanagan, M., 2000. Temperature and aggression. In: *Advances in experimental social psychology*, Elsevier, vol. 32, pp. 63–133.

Becker, G. S., 1968. Crime and punishment: An economic approach. In: *The economic dimensions of crime*, Springer, pp. 13–68.

Box, G. E., Cox, D. R., 1964. An analysis of transformations. Journal of the Royal Statistical Society: Series B (Methodological) 26, 211–243.

Chen, P., Yuan, H., Shu, X., 2008. Forecasting crime using the arima model. In: *2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery*, IEEE, vol. 5, pp. 627–630.

Cohn, E. G., 1990. Weather and crime. The British Journal of Criminology 30, 51–64.

Cohn, E. G., Rotton, J., 1997. Assault as a function of time and temperature: A moderator-variable time-series analysis. Journal of Personality and Social Psychology 72, 1322.

Dickey, D. A., Fuller, W. A., 1979. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American statistical association 74, 427–431.

Donohue, J. J., Levitt, S. D., 2001. The impact of legalized abortion on crime. Quarterly Journal of Economics .

dos Santos, M. J., Kassouf, A. L., 2012. Avaliação de impacto do estatuto do desarmamento na criminalidade: Uma abordagem de séries temporais aplicada à cidade de são paulo/assessing the disarmament statute impact on crime rates: a time series approach applied to são paulo city. Economic Analysis of Law Review 3, 307.

Entorf, H., Spengler, H., 2000. Socioeconomic and demographic factors of crime in germany: Evidence from panel data of the german states. International review of law and economics 20, 75–106.

Field, S., 1992. The effect of temperature on crime. The British Journal of Criminology 32, 340–351.

Horrocks, J., Menclova, A. K., 2011. The effects of weather on crime. New Zealand Economic Papers 45, 231–254.

Hyndman, R. J., Athanasopoulos, G., 2018. Forecasting: principles and practice. OTexts.

Kelly, M., 2000. Inequality and crime. Review of economics and Statistics 82, 530–539.

Lemon, D., Partridge, R., et al., 2017. Is weather related to the number of assaults seen at emergency departments? Injury 48, 2438–2442.

Levitt, S. D., 1998. The Relationship between Crime Reporting and Police: Implications for the Use of Uniform Crime Reports. Journal of Quantitative Criminology .

Levitt, S. D., 2004. Understanding why crime fell in the 1990s: Four factors that explain the decline and six that do not.

Michel, S. J., Wang, H., Selvarajah, S., Canner, J. K., Murrill, M., Chi, A., Efron, D. T., Schneider, E. B., 2016. Investigating the relationship between weather and violence in baltimore, maryland, usa. Injury 47, 272–276.

Pereira Filho, O. A., De Sousa, M. D. C. S., 2018. Avaliação De Impacto Das Guardas Municipais Com O Uso De Tratamentos Binários, Multivalorados E Contínuos. In: *Anais do XLIV Encontro Nacional de Economia [Proceedings of the 44th Brazilian Economics Meeting]*, ANPEC-AssociaÃ § Ã£ o Nacional dos Centros de PÃs-GraduaÃ § Ã£ o em . . . , no. 194.

Shumway, R. H., Stoffer, D. S., 2017. Time series analysis and its applications: with R examples. Springer.

Stec, A., Klabjan, D., 2018. Forecasting crime with deep learning. arXiv preprint arXiv:1806.01486 .

Vital, T., De Souza, D., Faciroli, J., et al., 2020. Unemployment, poverty and police performance: an ardl analysis of crime in sã£ o paulo. Economics Bulletin 40, 128–139.

Vital, T. M., 2018. Uma análise de impacto da guarda municipal no Brasil .

Appendices

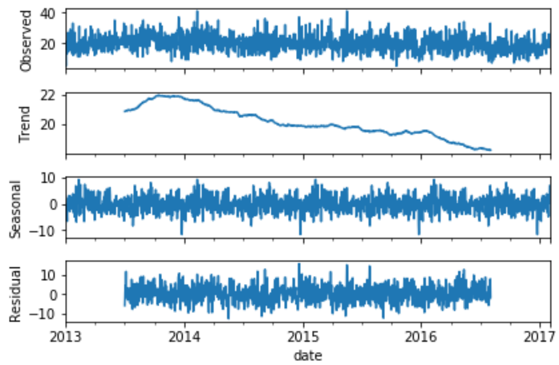

Fig. 5. ETF Decomposition of crime series

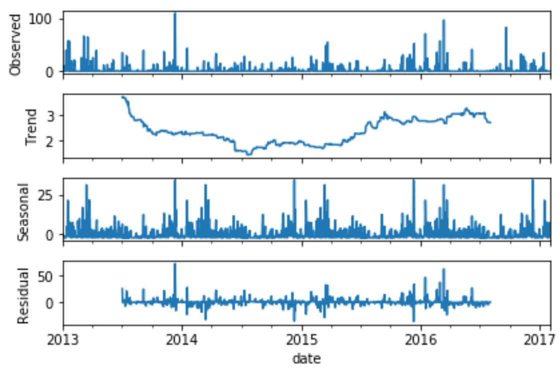

Fig. 6. ETF Decomposition of rain series

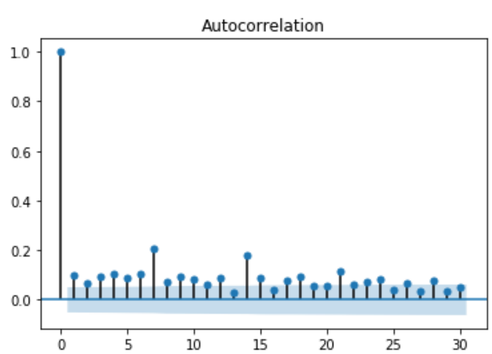

Fig. 7. ACF of crime series

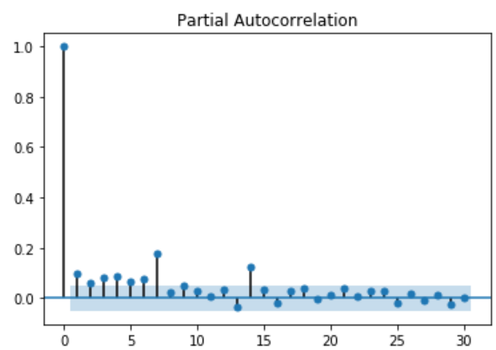

Fig. 8. PACF of crime series