Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v15n1p110

Forecasting the number of vehicles thefts in Campinas/Brazil using a Generalized Linear Autoregressive Moving Average model

By Pala, Carvalho, Sáfadi

Published: 20 May 2022

This work is copyrighted by Università del Salento, and is licensed under a Creative Commons Attribuzione - Non commerciale - Non opere derivate 3.0 Italia License.
For more information see:
http://creativecommons.org/licenses/by-nc-nd/3.0/it/

Electronic Journal of Applied Statistical Analysis
Vol. 15, Issue 01, May 2022, 110-122
DOI: 10.1285/i20705948v15n1p110

Forecasting the number of vehicles thefts in Campinas/Brazil using a Generalized Linear Autoregressive Moving Average model

Luiz Otávio de Oliveira Pala*, Marcela de Marillac Carvalho, and Thelma Sáfadi

Federal University of Lavras, Brazil.
Department of Statistics

Published: 20 May 2022

By definition, thefts are considered the act of taking away other people's mobile possessions for personal use or for others, affecting crime rates, economic indicators and enabling recent studies to create risk zones in society, contributing to insurance pricing in actuarial methods. This paper analyzes the number of vehicle thefts of 38 locations near *Campinas/São Paulo*, Brazil, using a GLARMA(p,q) model with Poisson and Negative Binomial response. The main feature of GLARMA(p,q) is to consider the peculiarities of counting data as high dispersion. As a result, it was possible to verify the adequacy and usefulness of the model for counting data. With specific techniques for estimating time series related to the public security area, patterns can be better understood, revealing relevant information that can be added to decision-making processes to direct public policies.

keywords: Actuary, Brazil, glarma model, thefts, vehicles.

1 Introduction

By definition, thefts are considered the act of taking away other people's mobile possessions for personal use or for others, contrasting with robbery, which is characterized

*Corresponding author: luiz.pala@estudante.ufla.br

©Università del Salento
ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

by the act of taking other people's mobile possessions for personal use or for others, through major threat or violence against the owner (Brasil, 1940). In the general sphere of crime, themes involving crime and violence have been the object of study of many countries, mainly relating the involved costs to society and individuals (Silva, 2015).

In the social context, carjacking and thefts of vehicles are among the primary crimes identified by the *Secretária Nacional de Segurança Pública* (Filho and Guedes, 2018), affecting crime rates and enabling studies concerning these offenses in society (Freitas and Gonçalves, 2017). Studies with such focus can be seen in Nunley and Zietz (2015), in the elaboration of risk zones for crimes such as carjacking and theft of vehicles in 83 census regions of San Francisco and in the analysis between inflationary process and crime rate by Chung and Kim (2018).

In the actuarial background, specifically in pricing, variables as vehicle traffic location, robberies, thefts, collision and fire impact in premium costs (Teixeira and Scalon, 2016). Consequently, individual attributes and risk levels can lead to different insurance and coverage costs, allowing the development of risk groups (Schmeiser and Wagner, 2014). For further detailing about pricing and actuarial analysis, see Denuit and Walhin (2007).

New insurance products have been marketed based on use, which are called Usage-based insurance (UBI). According to Grzadkowska (2018), these are products that the premium can be weighted through the driver's behavior, location and usage of the vehicle. In this case, vehicle data can be provided via telematics, feeding the insurer's database. This manner of contracting has shown an increasing trend in recent years (Zhou and Chen (2017), Bian and Liang (2018)). However, according to Bian and Liang (2018), studies are still needed to measure how much and how the driver's behavior can affect the insurance premium and risk, for example.

Statistical methods have contributed to pricing and crime rate analysis in several locations (Denuit and Walhin (2007), Freitas and Gonçalves (2017), Chung and Kim (2018)). Time series models have enabled the analysis of vehicle thefts and robberes in Filho and Guedes (2018) and Mao and Ye (2018), and also criminality aspects in Vujić and Koopman (2016) using intervention models. On the other hand, linear generalized models have allowed pricing, portfolio analysis and allocation of policyholders in accordance with risk level.

Morettin and Toloi (2018) define a time series (Y_t) as a set of observations of a phenomenon ordered in time. One of the main objectives of a time series analysis is to build stochastic models that can describe the phenomenon (Shumway and Stoffer, 2006). As stated in the literature, assuming the additive effect, Y_t can be decomposed by three latent components, denoting: trend (T_t), seasonality (S_t) and random error (e_t) under the assumption that $e_t \sim$ WN(0, σ^2). That is, $Y_t = T_t + S_t + e_t$, t = 1, ..., N (Morettin and Toloi, 2018).

There are several time series models, being the autoregressive integrated moving average ARIMA(p, d, q) a class of these. These models are suitable for modeling series with non-explosive behavior and that become stationary when the difference operator is used (Morettin and Toloi, 2018). Besides that, it is possible to specify seasonal models within the class SARIMA(p, d, q)(P, D, Q)s for series with a cyclic and periodic behavior. However, these classic models for time series from Box and Jenkins (1976) are specified assuming a Gaussian distribution, like the ARMA(p, q) process.

Some time series models may have particular characteristics, such as discrete and financial series. In discrete phenomena, generalized linear models (GLM), from Nelder and Wedderburn (1972), and its extensions are used. However, these methodologies rarely consider the serial correlation between observations. Thus, mathematical approaches for the analysis of a discrete time series derived from regression methods and from time series models were proposed in (Cox (1981), Zeger (1988), Mckenzie (1988), Li (1994), Shephard (1995), Davis and Wang (1999), Davis and Wang (2000), Davis and Streett (2003) and Benjamin and Stasinopoulos (2003)).

According to Dunsmuir and Scott (2015), these models use GLM in conjunction with ARMA(p, q) to model a wide variety of time series, such as the GLARMA(p, q) (Generalized linear autoregressive moving average), which allows the assumption of counting distributions, with the advantage of modeling autocorrelation, a common problem in time series. See, for example, Dunsmuir and Scott (2015) reviewing the theory of model applications; Petukhova and Z. (2018), comparing the performance of ARIMA(p,d,q), GLARMA(p,q) and Random forests to predict the number of Influenza virus (IAV-S) cases in Canada; and Mukhopadhyay and Thatte (2019) in the study of malaria cases in India, using models with a Binomial Negative and Poisson response.

Dunsmuir (2016) defines the class GLARMA(p, q) as a non-Gaussian state space process, in which the state of the process is linearly dependent on covariates and where the distribution of observations follows to the exponential family. In this model, $y_1, ..., y_n$ denotes the series, $\mathbf{x}_t$ denotes the covariables and $F_t = \{Y_s : s < t, \mathbf{x}_s : s \leq t\}$ denotes the past information of y_t and the past and present information of $\mathbf{x}_t$. Then, $f(y_t|W_t) = \exp\{y_t W_t - a_t b(W_t) + c_t\}$, where a_t and c_t are sequences of constants, c_t is a function of y_t, and W_t is a function of the elements in F_t. Consequently, the conditional average of the response is $u_t = E(Y_t|W_t)$, being u_t related to Wt through a link function. As seen in Dunsmuir and Scott (2015), W_t is specified as $W_t = \mathbf{x}_t^{'}\beta + Z_t$, being, $Z_t = \sum_{i=1}^{p} \phi_i(Z_{t-i} + e_{t-i}) + \sum_{i=1}^{q} \theta_i e_{t-i}$, where ϕ_i and θ_i are, respectively, autoregressive and moving average parameters.

The Poisson and Negative Binomial distributions are commonly used in GLARMA(p,q) model. According to Dunsmuir and Scott (2015), it is used for Poisson:

$$f(y_t, W_t) = \frac{u_t^{y_t} e^{-ut}}{y_t!}.$$

And for the Negative Binomial:

$$f(y_t, W_t) = \frac{\Gamma(\alpha + y_t)}{\Gamma(\alpha)\Gamma(y_t + 1)} \left[\frac{\alpha}{\alpha + u_t}\right]^{\alpha} \left[\frac{u_t}{\alpha + u_t}\right]^{y_t}.$$

In the above cases, the link function used is logarithmic, therefore, $u_t = exp(W_t)$. In this manner, the parameters of the process can be estimated and predictions and analyzes can be carried out next. For more properties about these models, see Dunsmuir and Scott (2015) and Dunsmuir (2016).

Forecasts can be of interest to the researcher, such as crime rates in different regions feeding a pricing model, for instance. This paper compiles the number of vehicle thefts of 38 locations near *Campinas/SP* enabling the management and direction of public policies such as in Filho (2004) and Napoleão (2005).

When building models for the time series of police occurrences for vehicle thefts, it is necessary to assess peculiarities of counting data, considering discrete probability distributions for non-negative integers such as the Poisson or Negative Binomial distributions. Thus, the aim of this work is to study the time series of the number of vehicle thefts in 38 locations near Campinas between January 2010 and March 2020, analyzing the adequacy and usefulness of the GLARMA(p, q) models with Poisson and Binomial Negative responses to describe the phenomenon.

2 Materials and Methods

The data analyzed corresponds to the number of occurrences of vehicle thefts recorded in the region of *Campinas, São Paulo*, Brazil. Data are available in the form of monthly time series from January 2010 to March 2020 by the *Secretaria de Segurança Pública do Estado de São Paulo* (SSP, 2020). The analyzed region includes 38 locations, shown in Table 1.

Table 1: Cities allocated by the *Secretaria de Segurança Pública do Estado de São Paulo* to the region of *Campinas, São Paulo*, Brazil.

Cities			
Águas de Lindoia	Indaiatuba	Mogi Guaçu	Santo Ant. de Posse
Amparo	Itapira	Mogi Mirim	Serra Negra
Atibaia	Itatiba	Monte Al. do Sul	Socorro
Bom J. dos Perdões	Itupeva	Morungaba	Tuiuti
Bragança Paulista	Jaguariúna	Nazaré Paulista	Valinhos
Cabreúva	Jarinu	Paulínia	Vargem
Campinas	Joanópolis	Pedra Bela	Várzea Paulista
Campo L. Paulista	Jundiaí	Pedreira	Vinhedo
Estiva Gerbi	Lindóia	Pinhalzinho	
Holambra	Louveira	Piracaia	

In Figure 1, we have the geographic location of each city cited in Table 1 under a radius of 81.8 kilometers around Campinas, which is one of the most populous cities in Brazil. Considering straight line distance, *Joanópolis* is the furthest city from *Campinas*, with an approximate distance of 80.5 kilometers. *Valinhos* is the closest, with an approximate distance of 12 kilometers.

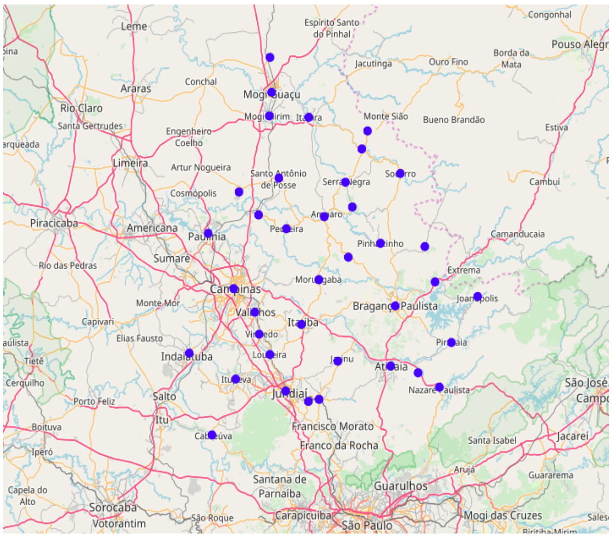

Figure 1: Geographic location of the localities included in the Campinas region, according to SSP (2020)

In order to analyze the trend component in the series, we performed the Cox-Stuart test. In this test, the series is grouped in pairs as follows: (y_1, y_{1+c}), (y_2, y_{2+c}),..., (y_{N-c}, y_N), being c = N/2 if N is even, or c = (N+1)/2 if N is odd. In each pair (y_i, y_{i+c}), positive $(y_i < y_{i+c})$ or negative $(y_i > y_{i+c})$ signs are associated and ties are eliminated. Denoting N as the number of pairs where $y_i \neq y_{i+c}$, the test confronts the null hypothesis of no trend against the presence of trend (Morettin and Toloi, 2006). As a result, a trend variable T = 1, ..., N was specified. Regarding the seasonal component (S_t), periodic covariates were defined in order to control the possible deterministic seasonal effect as follows: $\cos(2\pi t/12)$, $\sin(2\pi t/12)$, $\cos(2\pi t/6)$ and $\sin(2\pi t/6)$, for t = 1, ..., N. In addition to these, a binary variable was created to evaluate the effect of the months of December, that is:

$$D = \begin{cases} 1, & \text{if } \ t = 12\gamma \\ 0, & \text{otherwise} \end{cases},$$

for $\gamma = \{1, \ldots, 10\}$.

We considered models with Poisson and Negative Binomial response. The latter was used to control high data dispersion. Fisher's scoring method with 50 iterations was used to estimate the parameters of the model, available in the Glarma package of Dunsmuir and Scott (2018). To analyze the assigned distribution, we used the histogram from probability integral transformation (PIT). PIT compares the predictive distribution values with the data, that is, if the data was collected from the predictive distribution, PIT follows a uniform distribution (Dunsmuir, 2016). In this paper, we used the PIT extension for discrete distributions from Czado and Held (2009). As a form of interpretation, deviations from the histogram pattern, such as a U shape, indicate high dispersion. For further details, see Czado and Held (2009), Dunsmuir and Scott (2015), and Dunsmuir (2016).

We used Pearson residuals, as also used by Mukhopadhyay and Thatte (2019), although there are other variations, such as Score residuals. Stationarity condition was analyzed with autocorrelation function and Ljung–Box test, supported in the condition of asymptotic normality stated in (Dunsmuir and Scott, 2015). Denoting $\mathbf{\Psi} = (\phi^{'}, \theta^{'})^{'}$, we applied the Wald test in $\mathbf{\Psi}$, under H0: $\mathbf{\Psi} = 0$, that is, indicating if the Z_t structure is significant. See, for instance, the Wald statistic in (Dunsmuir and Scott, 2015). For our study, we used the package Glarma of Dunsmuir and Scott (2018) available in R (Team, 2020).

3 Results and Discussion

During the investigated period, there was a monthly average of 737.61 vehicles thefts in the region. Considering that every month has 30 days, it corresponds to 24.6 vehicles per day. We observed a possible seasonal pattern every December (Figure 2a), where the dashed horizontal line represents the monthly average of the period. Note that in December there is a possible reduction in the number of vehicle thefts with an average of 635.5 vehicles, which is a reduction of 102.11 vehicles when compared with the monthly average.

To verify this result, we applied the t-student test to compare the means. According to the test at a 5% significance level, we can affirm that the December mean is different from the general mean (p-value = 0.0005), under a null hypothesis of equality of means. This result is opposite to the study of Gonçalves (2008), showing that in December there is a greater number of thefts in *Belo Horizonte, Minas Gerais*, justified by greater capital movement, higher number of vehicles and lack of parking. This raises the hypothesis that these patterns may change between locations, being necessary to attribute these specific conditions in pricing of insurance, for instance.

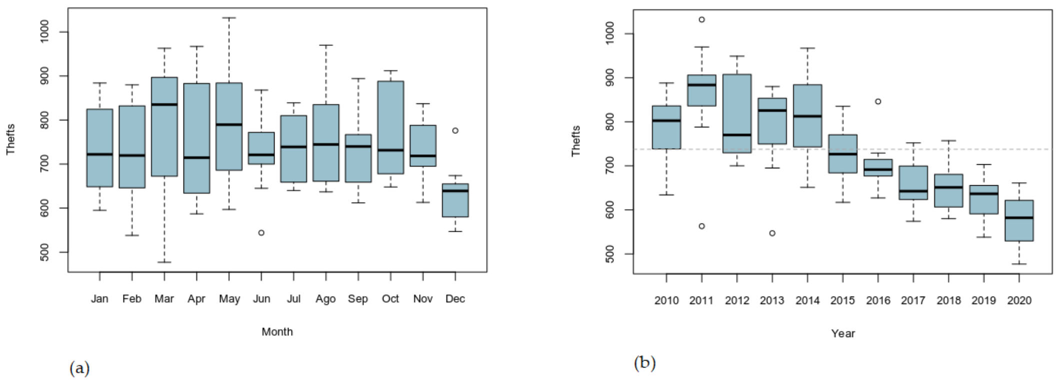

Figure 2: Box-plot of the number of vehicle thefts in relation to the months between January 2010 and March 2020 (a) and years 2010 to 2020 (b)

Returning to the Campinas region, we analyzed the monthly and daily frequency distributions of thefts, showed in Figure 3(a) and 3(b), respectively. Daily frequency was estimated considering months have 30 days. In the same Figure, we presented the normal distribution estimates for the data. According to Shapiro-Wilk test, the hypothesis of normal distribution for the data was not rejected. This can be justified because, even in the case of counting data, the Poisson and Negative Binomial distributions are related to the normal distribution, as seen in Casella and Berger (2014).

The 95% quantiles were estimated assuming normal distribution, making risk measures possible in the region, and represented by the vertical lines in gray (Figure 3 (a) and (b)). Concerning the monthly distribution, we found that the probability of the number of vehicle thefts in a month being greater than 921 vehicles is 5%. Similarly, the probability of the number of vehicle thefts in any given day being more than 31 is 5% in the region studied. This measure can be used for this locality's sub-regions, contributing to pure premium estimates. In contrast, we estimated the 95% quantiles to *Ribeirão Preto* region, resulting in 684 thefts (monthly data) and 23 thefts (daily data).

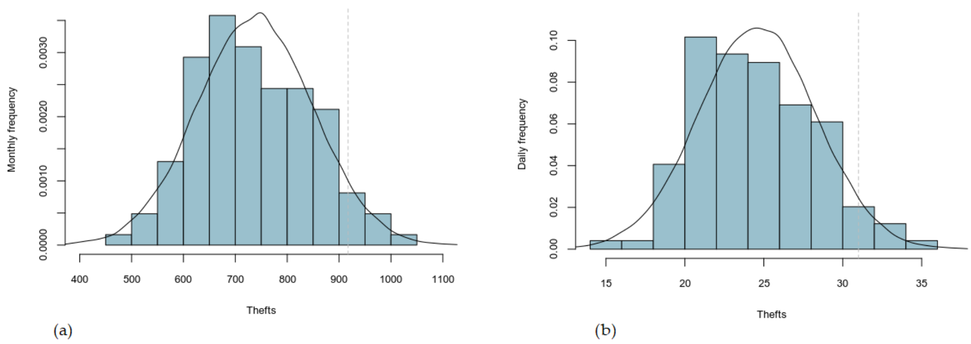

Figure 3: Distribution of vehicle thefts in *Campinas* region in the period analyzed: monthly (a) and daily (b) estimated averages

Considering *Campinas* population in 2010 being 3,306,396 inhabitants, we have a rate of 28.48 thefts per 10,000 inhabitants in 2010. In the same year, countries as Colombia, Peru and Chile had rates of theft or attempted theft of 47.1; 51.5 and 132.6 per 10,000 inhabitants, respectively (OECD, 2019). Note that after 2015, there was a reduction in the median when compared to previous years (Figure 2(b)). We tested and confirmed that the monthly average between 2015 and 2019 is smaller than the previous years (p-value $\sim$ 0).

Figure 4 shows the time series of thefts in the region. Both points represent the extreme values in the period. May 2011 was the month with the highest number of thefts, corresponding to 1,031 vehicles. March 2020 was the month with the lowest number of thefts, totaling 411 vehicles. This reduction in March 2020 may be an impact of the quarantine policies adopted in the State of *São Paulo* due to Covid-19, which reduced the circulation of people and vehicles.

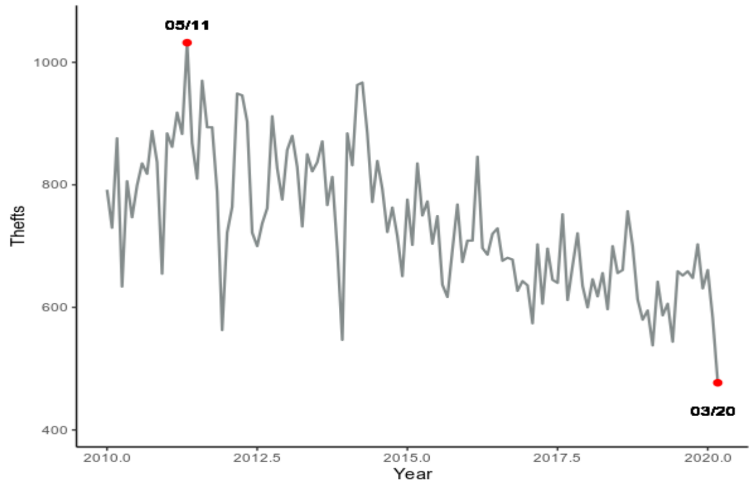

Figure 4: Time series of the monthly number of vehicle thefts in Campinas region during the months of January 2010 and March 2020

Graphically, it can be noted the presence of the trend (T_t) component in the time series, a fact also pointed out by the Cox-Stuart test (p-value ~ 0). Furthermore, the sample variance (S^2) estimated is approximately 17 times higher than the sample mean, indicating high dispersion. For the model, the autoregressive and moving average orders were identified with the autocorrelation (ACF) and partial autocorrelation (PACF) functions. Significant autocorrelations were verified in the first series lags in both functions. For parsimony reasons, models with lower lags were built, as shown in Table 2. In these models, only the significant covariates were maintained, considering a significance level of 5%.

Table 2: GLARMA(p, q) models considered and the respective information criteria AIC, BIC and HQC

Model	**Distribution**	**AIC**	**BIC**	**HQC**	**Wald**
GLARMA(1,0)	Poisson	1180.31	1897.19	1887.17	0.000
GLARMA(2,0)	Poisson	1841.19	1860.87	1849.18	0.000
GLARMA(1,0)	Neg. Bin.	1403.59	1414.84	1408.15	0.005
GLARMA(2,0)	Neg. Bin.	1400.92	1414.98	1406.64	0.002

Regarding the models presented in table 2, the GLARMA(2,0) model presented the best AIC and HQC information criteria. Moreover, analysis of the PIT histogram validated the adoption of the Negative Binomial distribution when compared to the Poisson models, as seen in Figure 5. Models with Poisson distribution 5 (a) and 5 (b) are penalized in the tails due to high dispersion, which did not occur in the models with Negative Binomial distribution 5 (c) and 5 (d).

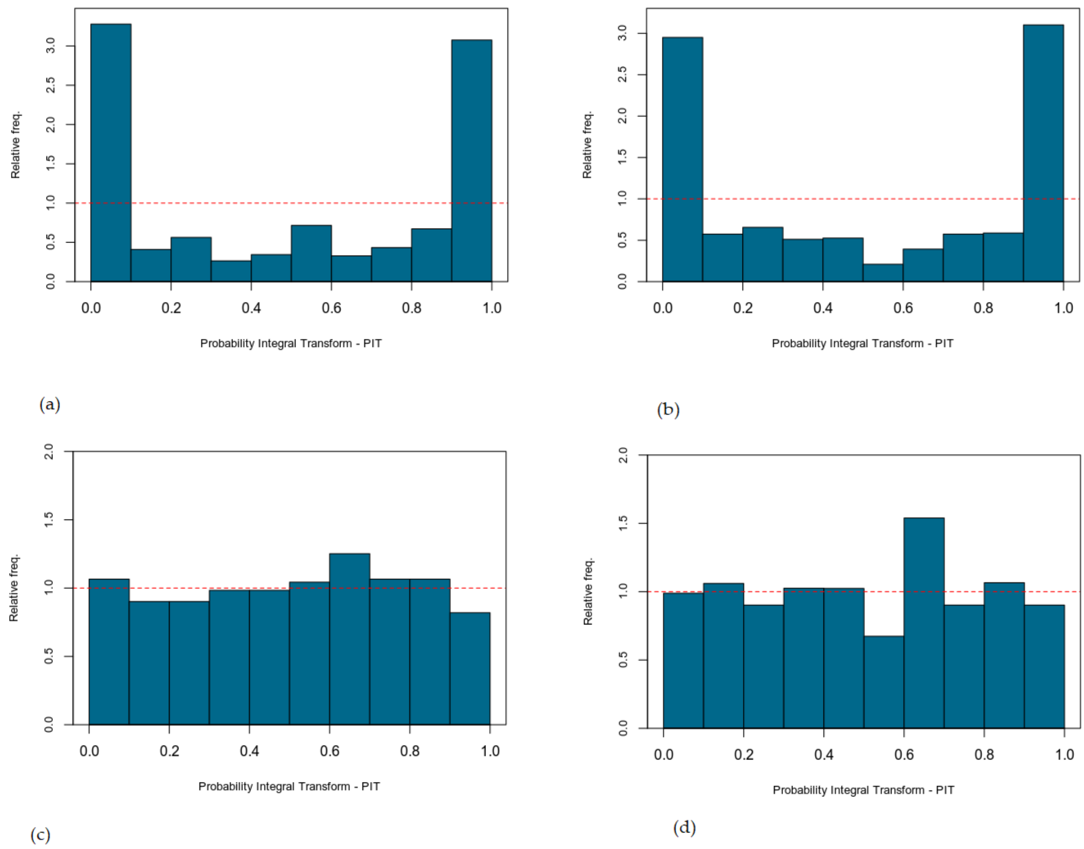

Figure 5: PIT histogram of models with Poisson and Negative Binomial distribution

Pearson residuals of GLARMA(2,0) with Negative Binomial distribution were analyzed and presented a stationary pattern (Figure 7) with no significant autocorrelations until the twentieth lag. The estimated coefficients are shown in Table 3, note that the deterministic seasonal control covariates $\cos(2\pi t/12)$, $\sin(2\pi t/12)$, $\cos(2\pi t/6)$ and $\sin(2\pi t/6)$ were removed because they were not significant.

We also observe the trend and December effect in the phenomenon. Concerning the trend, the effect was statistically significant, indicating a reduction in the average of thefts as time passed. The same occurred to December months, indicating a reduction in the average of thefts. In relation to the autoregressive parameters, we found that both lags are significant as well as the α parameter of the Negative Binomial distribution. The Wald test indicated that $\mathbf{\Psi}$ is significant (p-value $= 0.0022$) and the Ljung–Box test indicated that the residuals are independent (p-value $= 0.8002$).

Table 3: Estimated coefficients of the GLARMA model (2, 0) with Negative Binomial distribution

	Estimated	**Standard Error**	Pr—>(Z)—
Intercept	6.7855	0.0241	0.0000
T	-0.0029	0.0003	0.0000
D	-0.1322	0.0271	0.0000
ϕ_1	0.0267	0.0097	0.0062
ϕ_2	0.0214	0.0090	0.0179
α	132.27	20.530	0.0000

Considering the parameters estimated by the process, the model can be described as follows:

$$Y_t|F_t \sim \text{Neg Bin}(u_t, 132.27) \quad (1)$$

$$\log(u_t) = 6.7855 - 0.0029T - 0.1322D + Z_t \quad (2)$$

$$Z_t = 0.0267(Z_{t-1} + e_{t-1}) + 0.0214(Z_{t-2} + e_{t-2}). \quad (3)$$

Finally, Figure 6 shows the adjusted and observed values for the time series.

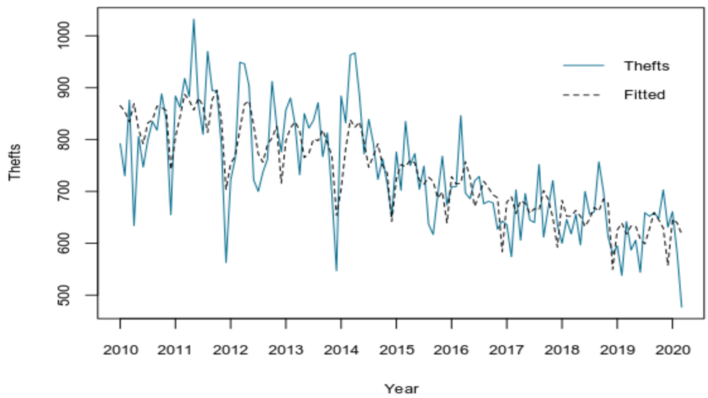

Figure 6: Adjusted and observed values of the time series in the analyzed period

4 Conclusion

Aiming to study the time series of vehicle thefts in 38 locations near Campinas between January 2010 and March 2020, this work investigated the adequacy and usefulness of

GLARMA(p,q) models with Poisson and Binomial Negative response, suitable for the peculiarities of counting time series. The models were adjusted by Fisher score method with 50 iterations, with the insertion of covariables for the deterministic seasonal control given by $\cos(2\pi t/12)$, $\sin(2\pi t/12)$, $\cos(2\pi t/6)$ and $\sin(2\pi t/6)$.

The series showed a downward trend and high dispersion in the period studied. Among the adjusted models, the GLARMA(2,0) process with Negative Binomial distribution was the most indicated in this case. Besides that, the seasonal control covariates were removed from the model because they are not significant. The values adjusted by the model showed a good predictive performance to describe the series' behavior over the period, as well as for forecasting purposes, under the conditions considered in the study.

It was possible to verify the adequacy and usefulness of the model with GLARMA(p, q) class considering specific distributions for counting data. With specific techniques for estimating time series related to the public security area, patterns can be better understood, revealing relevant information that can be added to decision-making processes in order to direct public policies. As future suggestions, we intend to expand the GLARMA(p, q) class to the other regions of São Paulo, enabling comparisons between risks of theft.

Acknowledgement

This study was financed in part by *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil* (CAPES) – Finance Code 001.

References

Benjamin, A., R. A. and Stasinopoulos, M. (2003). Generalized autoregressive moving average models. *Journal of the American Statistical Association*, 461(98):214–223.

Bian, Y., Y. C. Z. J. and Liang, L. (2018). Good drivers pay less: A study of usage-based vehicle insurance models. *Transportation Research Part A*, 107:20–34.

Box, G. and Jenkins, G. (1976). *Time series analysis, forecasting and control.* Holden-Day.

Brasil, B. (1940). Decreto lei n$^{\underline{o}}$ 2.848, de 07 de dezembro de 1940.

Casella, G. and Berger, R. (2014). *Inferência Estatística.* Cengage Learning.

Chung, J. and Kim, H. (2018). Crime risk maps: A multivariate spatial analysis of crime data. *Geographical Analysis*, 51(4):1–25.

Cox, D. (1981). Statistical analysis of time series: Some recent developments. *Scandinavian Journal of Statistics*, 8:93–115.

Czado, C., G. T. and Held, L. (2009). Predictive model assessment for count data. *Journal of the International Biometric Societ*, 65(4):–.

Davis, R., D. W. and Streett, S. (2003). Observation-driven model for poisson counts. *Biometrika*, 90(4):777–790.

Davis, R., D. W. and Wang, Y. (1999). *Modelling time series of count data. In S. Ghosh, editor, asymptotics, nonparametric & Time Series*. Marcel Dekker.

Davis, R., D. W. and Wang, Y. (2000). On autocorrelation in a poisson regression model. *Biometrika*, 87:491–505.

Denuit, M., M. X. P. S. and Walhin, J. (2007). *Actuarial modelling of claim counts: risk classification, credibility and bonus-malus systems*. Wiley.

Dunsmuir, W., L. C. and Scott, D. (2018). *Glarma: Generalized Linear Autoregressive Moving Average Models*.

Dunsmuir, W. (2016). *Generalized Linear Autoregressive Moving Average Models. In: RICHARD, Davis (ed.). Handbook of Discrete-Valued Time Series*. Chapman & Hall.

Dunsmuir, W. and Scott, D. (2015). The glarma package for observation-driven time series regression of counts. *Journal of Statistical Software*, 67(7):1–36.

Filho, L. (2004). *Distribuição espacial da violência em campinas: uma análise por geoprocessamento*. PhD thesis, UFRJ.

Filho, A., Z. G. and Guedes, E. (2018). Análise temporal das subtrações de veículos em salvador (ba). *Conjuntura & Planejamento*, 193:47–61.

Freitas, T., C. A. and Gonçalves, G. (2017). A estimação do Índice geral de criminalidade (igcrime) para os municípios do rio grande do sul. *Ensaios FEE*, 38(3):499–520.

Gonçalves, A. (2008). Geografia dos furtos de veículos em belo horizonte/minas gerais. Master's thesis, Pontifícia Universidade Católica de Minas Gerais.

Grzadkowska, A. (2018). *What is usage-based insurance?*

Li, K. (1994). Time series models based on generalized linear models: Some further results. *Biometrics*, 50:506–511.

Mao, Y., D. S. D. J. Z. W. W. C. and Ye, X. (2018). Space–time analysis of vehicle theft patterns in shanghai, china. *International Journal of Information*, 7(9).

Mckenzie, E. (1988). Some arma models for dependent sequences of poisson counts. *Advances in Applied Probability*, 20:822–835.

Morettin, P. and Toloi, C. (2006). *Análise de séries temporais*. Blucher.

Morettin, P. and Toloi, C. (2018). *Análise de séries temporais: Modelos lineares univariados*. Blucher.

Mukhopadhyay, A., T. R. S. P. G. N. and Thatte, U. (2019). Modeling and forecasting indian malaria incidence using generalized time series models. *Communications in Statistics: case studies, data analysis and applications*, 5(2).

Napoleão, P. (2005). Criminalidade urbana e condições de vida na região administrativa de campinas (sp) no ano de 2000: Uma análise espacial. Master's thesis, Universidade Estadual Paulistas.

Nelder, J. and Wedderburn, R. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A*, 135(3):370–384.

Nunley, J., S. M. S. R. and Zietz, J. (2015). The impact of inflation on property crime. *Contemporary Economic Policy*, 34(2):483–499.

OECD, O. f. E. C. a. D. (2019). *Regional social and environmental indicators: safety in regions.*

Petukhova, T., O. D. M. B. D. R. and Z., P. (2018). Assessment of autoregressive integrated moving average (arima), generalized linear autoregressive moving average (glarma), and random forest (rf) time series regression models for predicting influenza a virus frequency in swine in ontario, canada. *PLoS ONE*, 13(6).

Schmeiser, H., S. T. and Wagner, J. (2014). Unisex insurance pricing: Consumers perception and market implications. *The Geneva Papers*, 39:322–350.

Shephard, N. (1995). Generalized linear autoregressions. *Economics Papers*, 8.

Shumway, R. and Stoffer, D. (2006). *Time series analysis and its applications with R examples.* Springer.

Silva, C. (2015). Determinantes da vitimização do brasil. *Revista cadernos de economia*, 19:30–46.

SSP, S. d. S. P. d. E. d. S. P. (2020). *Estatísticas.*

Team, R. C. (2020). *R: A Language and Environment for Statistical Computing.*

Teixeira, F. and Scalon, J. (2016). A dependência espacial do valor do prêmio de automóvel. *Revista Brasileira de risco e seguro*, 20(11):20–54.

Vujić, S., C. J. and Koopman, S. (2016). Intervention time series analysis of crime rates: The case of sentence reform in virginia. *Economic Modelling*, 57:311–323.

Zeger, S. (1988). A regression model for time series of counts. *Biometrika*, 75(4):621–629.

Zhou, L., C. Q. L. Z. Z. H. and Chen, C. (2017). Speed-based location tracking in usage-based automotive insurance. *IEEE 37th International Conference on Distributed Computing Systems*, pages 2252–2257.