ORIGINAL ARTICLE

Modelling COVID-19 in Senegal and China with count autoregressive models

Lucien Diégane Gning[1] · Aba Diop[2] · Mamadou Lamine Diagne[3] · Jean Tchuenche[4,5]

Abstract

COVID-19 is a global health burden. We propose to model the dynamics of COVID-19 in Senegal and in China by count time series following generalized linear models. One of the main properties of these models is that they can detect potentials trends on the contagion dynamics within a given country. In particular, we fit the daily new infections in both countries by a Poisson autoregressive model and a negative binomial autoregressive model. In the case of Senegal, we include covariates in the models contrary to the Chinese case where the fitted models are without covariates. The short-term predictions of the daily new cases in both countries from both models are graphically illustrated. The results show that the predictions given by the negative binomial autoregressive model are more accurate than those given by the Poisson autoregressive model.

Keywords Poisson auto-regressive models · Count time series · Coronavirus · Negative binomial

Aba Diop, Mamadou Lamine Diagne and Jean Tchuenche have contributed equally to this work.

✉ Lucien Diégane Gning
luciengning@gmail.com

Aba Diop
aba.diop@uadb.edu.sn

Mamadou Lamine Diagne
mlamine.diagne@univ-thies.sn

Jean Tchuenche
jmtchuenche@gmail.com

1 LERSTAD, UFR Sciences appliquées et Technologie, Université Gaston BERGER, Saint-Louis, BP 234, Senegal

2 Equipe de Recherche en Statistique et Modèles Aléatoires(ERESMA), Universitê Alioune Diop, Bambey, Senegal

3 Département de Mathématiques, UFR Sciences et Technologie, Universite de Thiès, Thiès, Senegal

4 School of Computer Science and Applied Mathematics, University of the Witwatersrand, Private Bag 3, Wits 2050, Johannesburg, South Africa

5 School of Computational and Communication Sciences and Engineering, Nelson Mandela African Institution of Science and Technology, P.O. Box 447, Arusha, Tanzania

Introduction

A cluster of pneumonia caused by the coronavirus (COVID-19) emerged in Wuhan, China (WHO 2019; Paules et al. 2020). Since its emergence in December 2019, COVID-19, a deadly respiratory diseases has been a global public health burden. With a high human transmission rate, COVID-19 lead to a rapid global spread of the disease, and the World Health Organization (WHO) declared it a pandemic (WHO 2019, 2021). Contingency measures to mitigate the spread of the virus included physical distancing, face masks, or the most stringent lock down (WHO 2019).

COVID-19 has spread throughout the world, intruding on almost every aspect of daily life and negatively affecting the world economy (Li et al. 2020). As many other countries, the Senegalese government adopted the recommendations of the WHO, namely the strict respect of non-therapeutic prevention measures. In addition to these hygienic measures, the Senegalese government also put in place a short lock-down period to reduce the rapid spread of the disease in the country.

Several mathematical models have been developed in the literature including but are not limited to epidemiological models (López and Rodó 2020; Saikia et al. 2021; Tripathi et al. 2021), statistical models (Chan et al. 2021; Aidoo et al. 2022; Ganiny and Nisar 2021; Roy et al. 2020) and neural networks (Niazkar and Niazkar 2020; Shad et al. 2021).

Herein, we propose statistical models that could help our understanding of the disease contagion dynamics, as well as when the peak of contagion is reached, so that preventive measures (such as mobility restrictions) can be applied and/or relaxed. Indeed, we use a Poisson auto-regressive model (Fokianos and Tjøstheim 2011, 2012; Fokianos et al. 2009; Agosto et al. 2016), and a negative binomial autoregressive model (Christou and Fokianos 2014; Lange 2001; Liboschik et al. 2017) to fit and predict the daily new infections in China and Senegal. Both models used in this paper are special cases of the count times series following generalized linear models family, see, for example, Liboschik et al. (2017); Christou and Fokianos (2015). Note that our work extends some of the results obtained by (Agosto and Giudici (2020)). Indeed, they applied the Poisson auto-regressive model (without covariates) to the available China, Iran, South Korea and Italy data to investigate whether the spread of the disease has a trend. We propose to model the COVID-19 data available in China and Senegal by the Poisson autoregressive model and the negative binomial autoregressive model, including covariates only in the Senegalese case. For each country, we fit the models on the data and provide short-term of sample predictions of the daily new cases for both models. The models' fits and the predictions are compared for each country. Finally, note that for the case of China, we do not have the required data to incorporate covariates into the models.

The rest of the paper is organized as follows: Sect. 2 describes the data used in our analysis. In Sect. 3, the theoretical tools (the models) employed to analyze the trend of daily positive cases of COVID-19 in both countries are described. In Sect. 4, we present a practical application using real data, outline the results of the modelling. Finally, we discuss the results obtained and present the limitations and perspectives for futures directions in Sect. 5.

Data

The data we shall analyse consist of the confirmed daily new cases of COVID-19 in Senegal and in China. The data source is the daily World Health Organization reports, available at https://ourworldindata.org/coronavirus. The time range of these observations begins on 2nd of March 2020 for Senegal (5th of March 2020 for China) and ends on 10th of October 2021 for both countries. The choices of the two countries are based on the fact that China was the first country where the disease emerged, and Senegal is among the top twenty most affected countries on the African continent. The daily new cases used to fit the models are plotted in Figs. 1 and 2. Because we wish to make short-term forecast from September to October 2021, we use 95% of the data for the parameters estimation and the subsequent 5% of the data for the models' assessment. It is important to note that 5% of the data represent roughly one month for both countries. The reason for not considering long-term predictions of COVID-19 is because the influence of external factors that have not been taken into account in the models could be large, for example, a political decision, a vaccine, a drug, the temperature, new variants of the virus and other uncontrolled factors that make medium- and long-term forecasts uncertain. However, the more data one has for the estimation, the better the precision of the estimated parameters, even if we lose accuracy on the forecast error since we have less test data.

Furthermore, for Senegal, we have additional variables that will be used as covariates. These variables are described below.
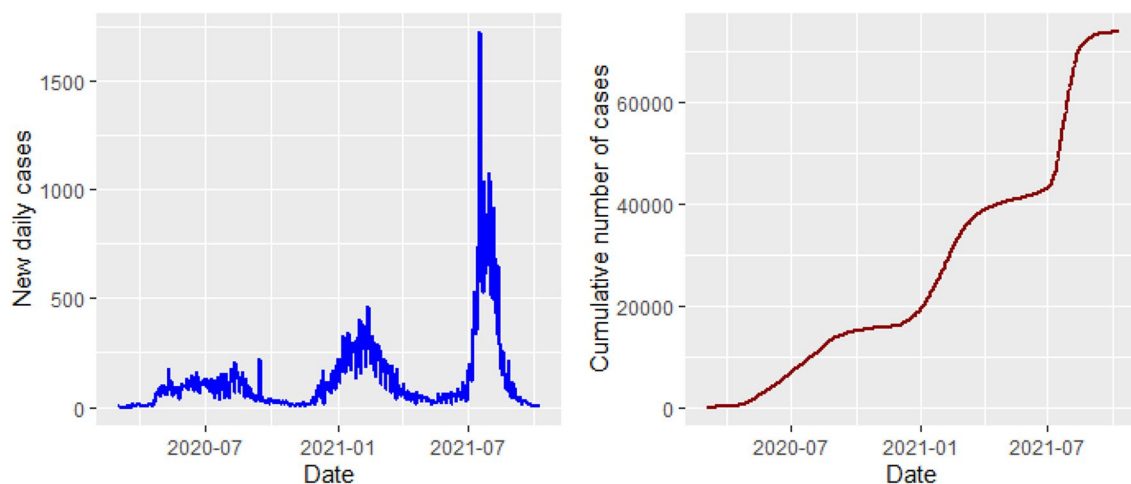
Fig. 1 Daily reported COVID-19 cases in Senegal

- x_1 : the number of daily COVID-19 community cases (i.e., cases in which infected do not know where the acquire the infection).
- x_2 : the number of daily COVID-19 imported cases (i.e., cases from infective immigrants).
- x_3 : the number of daily COVID-19 contact cases (i.e., infected have an idea of where/from whom s/he acquired the infection).

Finally, we summarize COVID-19 data from both countries, respectively, in Tables 1 and 2.

Methodology

The models

In what follows, Y_t denotes the number of newly confirmed cases on day t, $t = 1, \ldots, n$. We consider the time series of counts $\{Y_t\}$, and we denote by $\boldsymbol{x}_t = (x_{t,1}, \ldots, x_{t,r})^T$ the time-varying vector of the r covariates. Let $\mathcal{F}_{t-1}$ be the σ-field generated by $\{y_0, \ldots, y_{t-1}, \boldsymbol{x}_t\}$ i.e., the process from time $t = 0$ to time $t - 1$, and the potential covariates at time t, and λ_t the conditional mean $\mathbb{E}(Y_t \mid \mathcal{F}_{t-1})$. Further details on this point can be found in Douc et al. (2013); Jung et al. (2006); Kedem and Fokianos (2002); Woodard et al. (2011). We are interested in the models of the form:

$$\log(\lambda_t) = \beta_0 + \beta_1 \log(1 + y_{t-1}) + \alpha_1 \log(\lambda_{t-1}) + \boldsymbol{\eta}^T \boldsymbol{x}_t, \quad (1)$$

where:

$$\max \left(\mid \beta_1 \mid, \mid \alpha_1 \mid, \mid \beta_1 + \alpha_1 \mid \right) < 1. \quad (2)$$

This last condition on the parameters β_1, α_1 guarantees the existence of a stationary and ergodic solution of the process $\{Y_t\}$ for the fitted model, see (Fokianos and Tjøstheim 2011; Liboschik et al. 2017) for more details.

The model parameters are interpreted as follow: β_0 is the intercept, whereas β_1 and α_1 represent respectively the short-term dependence on the previous time, and the long-term dependence on all past values of the observed process. Finally, the parameter η_i, $i = 1, \ldots, r$, represents the effect of
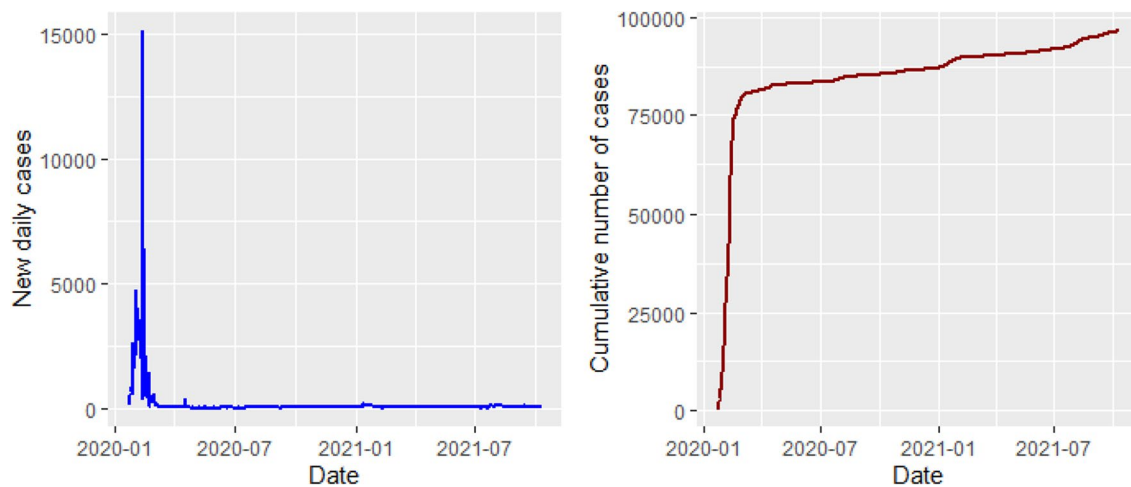

Fig. 2 Daily reported COVID-19 cases in China

Table 1 Descriptive Statistics of COVID-19 data in Senegal

Variable	Observations	Total	Mean	Std.Dev.	Min.	Max.
Confirmed cases	588	73837	125.57	182.77	0	1722
Contact cases	588	27902	47.45	54.37	0	744
Community cases	588	47845	81.37	145.97	0	978
Imported cases	588	413	0.70	2.09	0	27

Table 2 Descriptive Statistics of COVID-19 data in China

Variable	Observations	Total	Mean	Std.Dev.	Min.	Max.
Confirmed cases	625	95956	153.53	806.82	0	15133

the covariate x_i on the logarithm of the expected number of new infections at date t.

The models of the form (1) are log-linear models and allow negative dependence contrary to the INGARCH model (Fokianos and Tjøstheim 2011; Liboschik et al. 2017).

Poisson autoregressive model

The Poisson autoregressive (PAR) model is specified by adding to the model (1) the following distributional assumption for Y_t given $\mathcal{F}_{t-1}$:

$$Y_t \mid \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t). \tag{3}$$

This implies that:

$$\mathbb{P}(Y_t = y \mid \mathcal{F}_{t-1}) = \frac{\lambda_t^y \exp(-\lambda_t)}{y!}, \quad y = 0, 1, \ldots, \tag{4}$$

and:

$$\text{Var}(Y_t \mid \mathcal{F}_{t-1}) = E(Y_t \mid \mathcal{F}_{t-1}) = \lambda_t. \tag{5}$$

Hence, because the conditional mean is identical to the conditional variance, this is a major drawback related to its application. An alternative count time series model resolving this particular issue is presented in the following subsection.

Negative binomial autoregressive model

For the negative binomial autoregressive (NBAR) model, we assume in addition to (1) that:

$$Y_t \mid \mathcal{F}_{t-1} \sim \text{NegativeBinomial}(\lambda_t, \phi), \quad \phi \in]0, +\infty[. \tag{6}$$

This implies that:

$$\mathbb{P}(Y_t = y \mid \mathcal{F}_{t-1}) = \frac{\Gamma(\phi + y)}{\Gamma(y + 1)\Gamma(\phi)} \left(\frac{\phi}{\phi + \lambda_t} \right)^\phi \\ \left(\frac{\lambda_t}{\phi + \lambda_t} \right)^y, \quad y = 0, 1, \ldots \tag{7}$$

where Γ is the gamma function defined by $\Gamma(x) = \int_0^\infty u^{x-1} \exp(-x) \, dx$, for $x > 0$.

In this case, $\text{Var}(Y_t \mid \mathcal{F}_{t-1}) = \lambda_t + \lambda_t^2 / \phi$, ϕ is referred to as dispersion parameter. Note that when $\phi \to +\infty$, the Negative Binomial distribution converges to the Poisson distribution.

Both models presented in this work fall within the purview of count time series following generalized linear models family (Liboschik et al. 2017; Box and Jenkins 1970; Fuller 1976).

Maximum likelihood estimation

Both models considered in this work were fitted by using the maximum likelihood method, see, for example, Akaike (1998) for more details. That is, by maximizing:

$$\ell_1(\boldsymbol{\theta}) = \sum_{t=1}^n (y_t \log(\lambda_t) - \lambda_t - log(y_t!)), \tag{8}$$

for the Poisson autoregressive model and:

$$\ell_2(\boldsymbol{\theta}, \phi) = \sum_{t=1}^n \left(y_t \log(\lambda_t) - (\phi + y_t) \log(\phi + \lambda_t) \right. \\ \left. + \phi \log(\phi) + \log \left(\frac{\Gamma(\phi + y_t)}{y_t! \Gamma(\phi)} \right) \right), \tag{9}$$

for the negative binomial autoregressive model, with respect to $\boldsymbol{\theta}$ and ϕ, where $\boldsymbol{\theta} = (\beta_0, \beta_1, \ldots, \beta_p, \alpha_1, \ldots, \alpha_q, \eta_1, \ldots, \eta_r)^T \in \mathbb{R}^{p+q+r+1}$ and satisfying both conditions (1) and (2).

The calculation of the parameter estimates of both models is then a non-linear optimization problem with linear inequality constraints. Many algorithms solving this kind of problems are developed in the literature (see for example Lange 2001). The *R* package *tscount* developed by Liboschik et al. (2017) fitted models of the form (1) using a quasi-conditional maximum likelihood estimation (more details on this method can be found in Martin et al. (2012)). Using this package in the case of the Poisson auto-regressive model, we obtain an ordinary maximum likelihood estimator, whereas in the case of the negative binomial auto-regressive model, we obtain a quasi-maximum likelihood estimator. Then, for the PAR model, we calculate the maximum likelihood estimator parameters using the *R* package *tscount*, while for the NBAR model these maximum likelihood estimator parameters are calculated by direct constrained optimization of the log-likelihood function (the source codes are available on this github repository https://github.com/LUGNI2/Count-Times-Series-for-COVID-19). The constraints on the parameters are defined by equation (2), and the initial values are obtained by fitting a generalized linear model as in Fokianos and Tjøstheim (2011). It should be noted that the calculation of the models parameter estimators is based on the non-linear minimization with linear inequality constraints using an adaptive barrier algorithm. For more details on this, see the *R* package *constrOptim* (https://search.r-project.org/CRAN/refmans/alabama/html/constrOptim.nl.html). Finally, note that the source code of the *tscount* package is available on the Comprehensive *R* Archive Network (CRAN) repository (https://cran.r-project.org/package=tscount).

Model goodness of fit

Many tools to evaluate the goodness of fit developed for generalized linear models as well as for times series can be utilized to evaluate our proposed models' fits and its predictive performance. Some of them are recalled below (and for further details on this approach, see Liboschik et al. 2017).

Akaike Information Criterion

The Akaike Information Criterion (AIC) introduced in Akaike (1998) is defined as:

$$AIC = 2K - 2\hat{\ell}, \tag{10}$$

where $\hat{\ell}$ is the maximized value of the log-likelihood function of the model and K is the number of model parameters. The idea is to select the model that have the lowest AIC.

Bayesian Information Criterion

Formally, the Bayesian Information Criterion (BIC) is defined as:

$$BIC = 2K \log(n) - 2\hat{\ell}, \tag{11}$$

where K represents the number parameters, n the sample size, and $\hat{\ell}$ the maximized value of the log-likelihood function. The model with the lower BIC is preferred in a model selection procedure.

Residuals

Various types of residuals are defined in the literature, see for example (Hilbe 2011). Response residuals are given by:

$$r_t = y_t - \hat{\lambda}_t, \tag{12}$$

while Pearson residuals are given by:

$$r_t^P = \frac{(y_t - \hat{\lambda}_t)}{\sqrt{\hat{\lambda}_t + \hat{\lambda}_t^2/\hat{\phi}}}, \tag{13}$$

for $t = 1, \ldots, T$, where $\hat{\lambda}_t$ denotes the fitted value of the conditional mean λ_t, and $\hat{\phi}$ the estimate of the dispersion parameter ϕ.

The empirical auto-correlation function (Liboschik et al. 2017) of these residuals is used as a tool for detecting a potential serial dependence unexplained by the fitted model. Furthermore, a graph of squared residuals r_t^2 against the corresponding fitted values $\hat{\lambda}_t$ shows the relation between the conditional mean and the conditional variance. Indeed, if the Poisson assumption on the conditional distribution is appropriate, the points scatter around the identity function, while if the Negative binomial assumption is appropriate, the points scatter around a quadratic function (Ver Hoef and Boveng 2007).

Probability integral transform

Gneiting et al. (2007) proposed the probability integral transform (PIT) as a tool for assessing the probabilistic calibration of the predictive distribution of a model. We will use a non-randomized version of the PIT for count data proposed by Czado et al. (2009). They show that if the predictive distribution is correct, the PIT will have a standard uniform distribution. According to Czado et al. (2009), the PIT is defined by:

$$F_t(u \mid y) = \begin{cases} 0, & u \leq P_t(y-1) \\ \frac{u - P_t(y-1)}{P_t(y) - P_t(y-1)}, & P_t(y-1) \leq u \leq P_t(y), \\ 1, & u \geq P_t(y). \end{cases} \tag{14}$$

where $P_t(y) = \mathbb{P}(Y_t \leq y \mid \mathcal{F}_{t-1})$ is the predictive distribution of the model. For the observed values $y_t, (t = 1, \ldots, n)$, they proposed to compare the mean PIT defined by:

$$\bar{F}(u) = \frac{1}{n} \sum_{t=1}^{n} F_t(u \mid y_t), \quad 0 \leq u \leq 1, \tag{15}$$

to the cumulative distribution function of the standard uniform distribution, that is, the identity function. This comparison is performed by plotting an histogram for the PIT which is interpreted as follows: a U-shape indicates underdispersion of the predictive distribution, while an upside-down U shape points to overdispersion.

Marginal Calibration

The marginal calibration is the difference of the average predictive cumulative distribution function and the empirical cumulative distribution function of the observations:

$$\frac{1}{n} \sum_{t=1}^{n} P_t(y) - \frac{1}{n} \sum_{t=1}^{n} \mathbb{1}(y_t \leq y), \quad y \in \mathbb{R}. \tag{16}$$

Practically, we will plot the marginal calibration diagram for the data. The more equation (16) is close to zero, the more the predictions from the model are appropriate.

Practical Application

We applied the models presented in the previous sections and fit them to data on the number of new daily cases of individuals infected with of COVID-19 infection in Senegal and China.

Note that, for Senegal, we applied the models including covariates which satisfy

$$\begin{aligned} \log(\lambda_t) = &\beta_0 + \beta_1 \log(1 + y_{t-1}) + \alpha_1 \log(\lambda_{t-1}) \\ &+ \eta_1 x_{1,t} + \eta_2 x_{2,t} + \eta_3 x_{3,t}, \end{aligned} \tag{17}$$

Table 3 PAR model estimates for Senegal

Parameter	Estimate	Std. Error	p-Value
β_0	0.9664	0.0323	$< 2 \, 10^{-16}$
β_1	0.5058	0.0116	$< 2 \, 10^{-16}$
α_1	0.2710	0.0162	$< 2 \, 10^{-16}$
η_1	0.0004	0.00002	$< 2 \, 10^{-16}$
η_2	0.0218	0.0024	$< 2 \, 10^{-16}$
η_3	0.0012	0.00003	$< 2 \, 10^{-16}$

Table 4 NAR model estimates for Senegal

Parameter	Estimate	Std. Error	p-Value
β_0	1.0342	0.1426	$4.13 \, 10^{-13}$
β_1	0.7152	0.0470	$< 2 \, 10^{-16}$
α_1	0.0050	0.0664	0.9400
η_1	0.0008	0.0002	0.0002
η_2	-0.0026	0.0093	0.7814
η_3	0.0035	0.0006	$7.4 \, 10^{-10}$
ϕ	4.7488	0.4091	$< 2 \, 10^{-16}$

Table 5 PAR model estimates for China

Parameter	Estimate	Std. Error	p-Value
β_0	0.1783	0.0055	$< 2 \, 10^{-16}$
β_1	0.3978	0.0021	$< 2 \, 10^{-16}$
α_1	0.5888	0.0021	$< 2 \, 10^{-16}$

Table 6 NAR model estimates for China

Parameter	Estimate	Std. Error	p-Value
β_0	0.15963	0.0421	0.0002
β_1	0.48811	0.0221	$< 2 \, 10^{-16}$
α_1	0.49261	0.0232	$< 2 \, 10^{-16}$
ϕ	1.95641	0.1228	$< 2 \, 10^{-16}$

where x_1, x_2 and x_3 are the covariates defined in Sect. 2. In contrast, for China, we apply the models without covariates, satisfying

$$\log(\lambda_t) = \beta_0 + \beta_1 \log(1 + y_{t-1}) + \alpha_1 \log(\lambda_{t-1}). \quad (18)$$

Results and discussion

The maximum likelihood estimates of the parameters, their standard errors and their *p*-values for statistical significance tests for the fitted models are provided in Tables 3, 4, 5, 6.

Tables 3 and 4 show that, for Senegal, both models include all covariates at 5% significance level except the daily imported COVID-19 cases from the NAR model. For China, we observe the presence of short-term dependence and long-term trend. Indeed using the Chinese data, the estimated parameters of the auto-regressive coefficients (β_1 and α_1) for both models are all significant. For Senegal, we have the same conclusions excepted for the NAR model that does not include the parameter representing long-term dependence (α_1) as significant. Note that, for Senegal and for the NAR model, although all covariates are statistically significant, their influence on the daily new cases are very weak because their regression coefficients are very small. For example, according to the estimated value of η_1, if the expectation of new cases for yesterday, the number of contact cases for today and the number of imported cases for today are all close to 0, 100 new cases observed yesterday and 1 community case observed today generate about 5 new expected cases today.

The numerical calculations are carried out using the open source software *R* (R Development Core Team 2021). To investigate the models goodness of fit, we summarized the value of AIC, BIC and their Root Mean Square Error (RMSE) in Table 7. Note that with all considered criteria, the NAR model fits better the data than the PAR model for both countries. In Figs. 3, 4, 5, 6, 7, 8, 9, 10, we plot the marginal calibration and the PIT histogram of the models for both countries. Thus, for the data from both countries, the PIT histogram corresponding to the PAR model is approximately *U*-shaped, an indication that the PAR is not adequate to fit the data. As opposed to this, the PIT histogram corresponding to NAR model appears to approach better the standard uniform distribution for both countries. It follows that the NAR model is more satisfactory than the PAR model to fit the COVID-19 data in Senegal and China.

Table 7 Goodness of fits

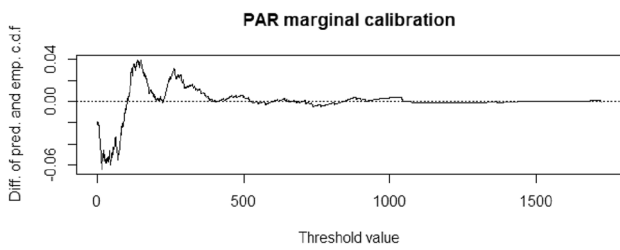| Model	AIC	Senegal			China		
		BIC	RMSE	AIC	BIC	RMSE
PAR	13523.11	13587.03	13.55788	85897.5	85929.8	25.4619
NAR	5499.788	5574.354	11.08462	5208.864	5251.96	20.3465

Fig. 3 Marginal calibration for the models fitting data from Senegal
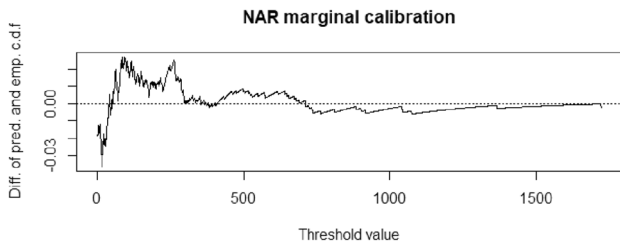

Fig. 4 Marginal calibration for the models fitting data from Senegal
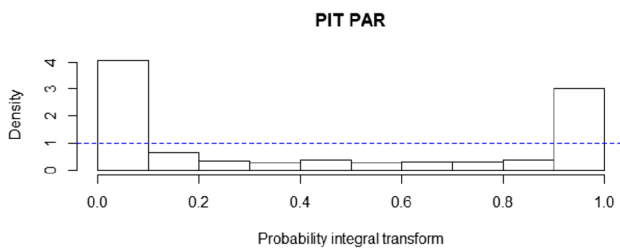

Fig. 5 PIT for the PAR model fitting data from Senegal
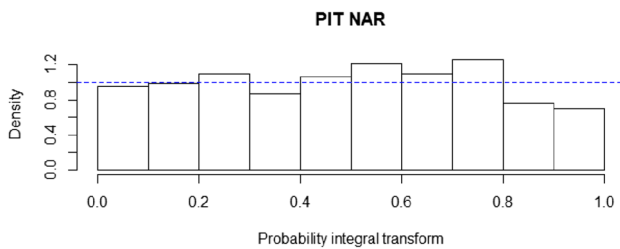

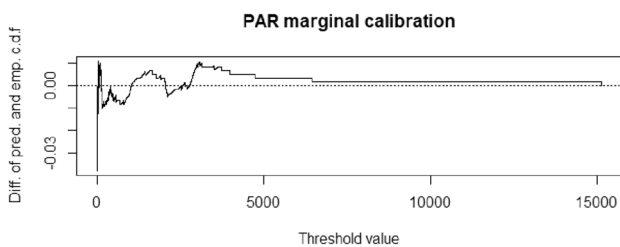**Fig. 6** PIT for the NAR model fitting data from Senegal

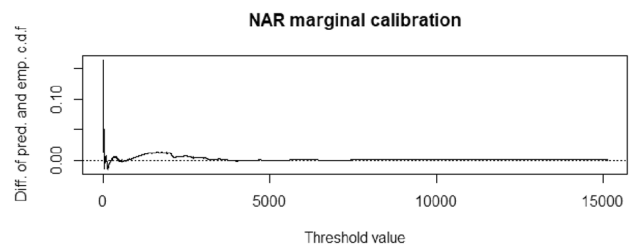**Fig. 7** Marginal calibration for the models fitting data from China

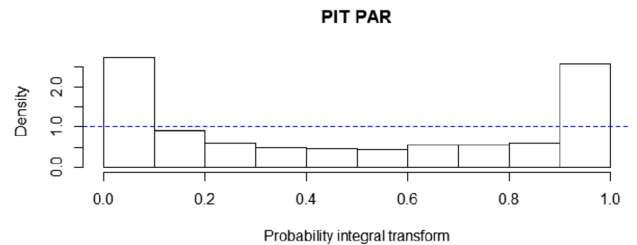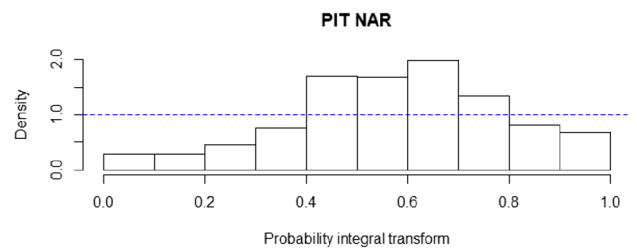**Fig. 8** Marginal calibration for the models fitting data from China

Fig. 9 PIT for the PAR model fitting data from China

**Fig. 10** PIT for the NAR model fitting data from China

Finally, the predictions of the daily new cases using the models for both countries are plotted in Figs. 11 and 12. As excepted, the predictions given by the NAR model appear more accurate than those given by the PAR model.

For both countries, it appears that the predicted values and the real values of the daily reported cases are relatively closed, although the models slightly overestimate the number of new cases. A possible explanation of this fact is that there exists many factors that can influence the outbreak, such as the individuals behaviour, political decisions, among others that are not accounted for in the models. Therefore, the obtained results should be analyzed and interpreted with caution.

Conclusion

The COVID-19 pandemic has major socio-economic challenges globally. With multiple strains of the virus co-circulating, mitigation strategies to curb the spreads of the various strains is a daunting task for health authorities.

Fig. 11 Observed and predicted values of daily new cases in Senegal
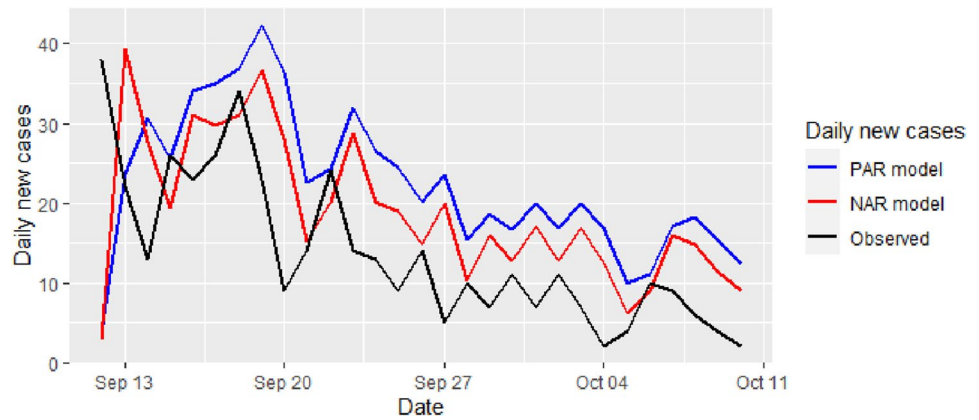
Fig. 12 Observed and predicted values of daily new cases in China
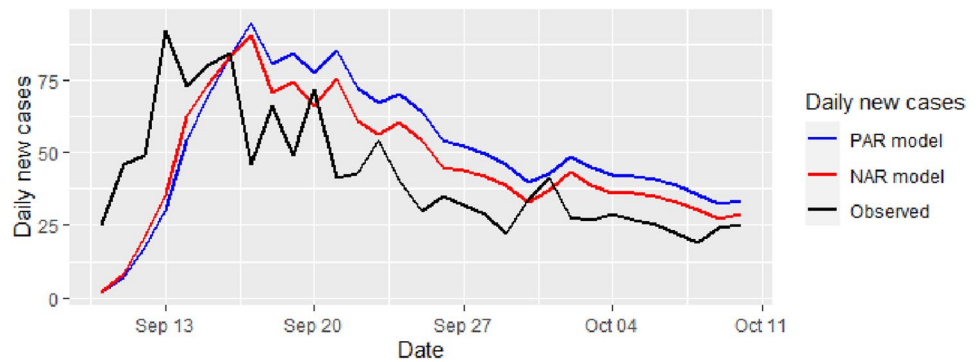
We fitted the daily reported new COVID-19 cases in China and Senegal using two types of count time series following generalized linear models with covariates in the case of Senegal: the negative binomial auto-regressive model (NAR) and the Poisson auto-regressive model (PAR). Results obtained from our analysis reveal that the NAR model fits better the daily new COVID-19 cases than the PAR model for both countries. Furthermore, we find a short-term trend and a long-term trend for China, whereas for Senegal, only a short-term trend is founded. In conclusion, the NAR model could be useful to health policy decision-makers in China and Senegal as the estimated values of the trend parameters can be monitored in any time period during the outbreak. They reveal the contagion dynamics in a country during a given period.

However, our proposed study has some limitations. For example, in the case of Senegal, one needs covariates that influence the number of daily new infections more than the covariates used herein to better understand the dynamics of the outbreak. Also, it could be interesting to incorporate significant covariates in the case of China, but as of the time of writing this paper, such data were not available to us. The models we investigated use a time series process that cannot capture potentials seasonality on the data. Finally, the conditional distribution used does not capture the disease multiple waves.

This study, which is not exhaustive, could be extended in various ways, by considering a count time series with the conditional distribution following a univariate generalized Waring distribution to understand better the COVID-19 dynamics, and to capture the disease multiple waves, or by accounting for the virus incubation period in the models. One could also use a more general time series such as Auto-Regressive Integrated Moving Average (ARIMA). Another direction that extends this study is to use more general time series such as Auto-Regressive Integrated Moving Average (ARIMA) in the models.

Declarations

References

Agosto A, Cavaliere G, Kristensen D, Rahbek A (2016) Modelling corporate defaults: Poisson autoregressions with exogenous

covariates (parx). J Empir Financ 38(6):40–63. https://doi.org/10.1016/j.jempfin.2016.02.007

Agosto A, Giudici P (2020) A Poisson autoregressive model to understand Covid-19 contagion dynamics. Risk 8(3):1–8. https://doi.org/10.3390/risks8030077

Aidoo EN, Ampofo RT, Awashie GE, Appiah SK, Adebanji AO (2022) Modelling Covid-19 incidence in the African sub-region using smooth transition autoregressive model. Model. Earth Syst. Environ. 8:961–966. https://doi.org/10.1007/s40808-021-01136-1

Akaike H (1998) Information theory and an extension of the maximum likelihood principle, In *Parzen, E., Tanabe, K., Kitagawa, G. (eds) Selected Papers of Hirotugu Akaike*, ed. in Statistics, S.S., 978–1–4612–7248–9. New York: Springer

Box GE, Jenkins GM (1970) Time series analysis: forecasting and control. Holden Day, San Francisco

Chan S, Chu J, Zhang Y, Nadarajah S (2021) Count regression models for Covid-19. Phys A 563:0378–4371. https://doi.org/10.1016/j.physa.2020.125460

Christou V, Fokianos K (2014) Quasi-likelihood inference for negative binomial time series models. J Time Ser Anal 35(1):55–78. https://doi.org/10.1111/jtsa.12277

Christou V, Fokianos K (2015) On count time series prediction. J Stat Comput Simul 85(2):357–373. https://doi.org/10.1080/00949655.2013.823612

Czado C, Gneiting T, Held L (2009) Predictive model assessment for count data. Biometrics 65(4):1254–1261. https://doi.org/10.1111/j.1541-0420.2009.01191.x

Douc R, Doukhan P, Moulines E (2013) Ergodicity of observation-driven time series models and consistency of the maximum likelihood estimator. Stochastic Process Appl 123(7):2620–2647. https://doi.org/10.1016/j.spa.2013.04.010

Fokianos K, Rahbek A, Tjøstheim D (2009) Poisson autoregression. Comput Stat Data Anal 51(2):350–364. https://doi.org/10.1198/jasa.2009.tm08270

Fokianos K, Tjøstheim D (2011) Log-linear Poisson autoregression. J Multivar Anal 102:563–578. https://doi.org/10.1016/j.jmva.2010.11.002

Fokianos K, Tjøstheim D (2012) Nonlinear Poisson autoregression. Ann Inst Stat Math 64(1):205–225. https://doi.org/10.1007/s10463-012-0351-3

Fuller WA (1976) Introduction to statistical time series. Wiley, New York

Ganiny S, Nisar O (2021) Mathematical modeling and a month ahead forecast of the coronavirus disease 2019 (Covid-19) pandemic: an Indian scenario. Model Earth Syst Environ 7:29–40. https://doi.org/10.1007/s40808-020-01080-6

Gneiting T, Balabdaoui F, Raftery AE (2007) Probabilistic forecasts, calibration and sharpness. J R Stat Soc B 69(2):243–268. https://doi.org/10.1111/j.1467-9868.2007.00587.x

Hilbe JM (2011) Negative binomial regression, 2nd edn. Cambridge University Press, Cambridge

Jung RC, Kukuk M, Liesenfeld R (2006) Time series of count data: modelling, estimation and diagnostics. J Am Stat Assoc 104:1430–1439. https://doi.org/10.1016/j.csda.2006.08.001

Kedem B, Fokianos K (2002) Regression models for time series analysis. Wiley Series in Probability and Statistics, Hoboken, NJ

Lange K (2001) Numerical analysis for statisticians. Springer, New York

Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y (2020) Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med 382:1199–1207

Liboschik T, Fokianos K, Fried R (2017) tscount: An R package for analysis of count time series following generalized linear models. J Stat Softw 82(5):1–51. https://doi.org/10.18637/jss.v082.i05

López L, Rodó X (2020) A modified SEIR model to predict the Covid-19 outbreak in Spain and Italy: simulating control scenarios and multi-scale epidemics. Results Phys 21:103746. https://doi.org/10.1016/j.rinp.2020.103746

Martin V, Hurn S, Harris D (2012) *Quasi-Maximum Likelihood Estimation. In Econometric Modelling with Time Series: Specification, Estimation and Testing (Themes in Modern Econometrics, pp. 307–351).* Cambridge: Cambridge University Press

Niazkar HR, Niazkar M (2020) Application of artificial neural networks to predict the Covid-19 outbreak. Glob Health Res Policy 5(50). https://doi.org/10.1186/s41256-020-00175-y

Paules CI, Marston HD, Fauci AS (2020) Coronavirus infections-more than just the common cold. JAMA 323(8):707–708. https://doi.org/10.1007/82_2017_25

R Development Core Team (2021) R: A language and environment for statistical computing. http://www.R-project.org

Roy S, Bhunia GS, Shit PK (2020) Spatial prediction of Covid-19 epidemic using arima techniques in India. Model Earth Syst Environ 7:1385–1391. https://doi.org/10.1007/s40808-020-00890-y

Saikia D, Bora K, Bora MP (2021) Covid-19 outbreak in India: an SEIR model-based analysis. Nonlinear Dyn 104(4):4727–4751. https://doi.org/10.1007/s11071-021-06536-7

Shad M, Sharma YD, Singh A (2021) Forecasting of monthly relative humidity in Delhi, India, using SARIMA and ANN models. Earth Syst Environ Model. https://doi.org/10.1007/s40808-022-01385-8

Tripathi A, Tripathi RN, Sharma D (2021) A mathematical model to study the Covid-19 pandemic in India. Earth Syst Environ Model. https://doi.org/10.1007/s40808-021-01280-8

Ver Hoef JM, Boveng PL (2007) Quasi-Poisson vs. negative binomial regression: How should we model overdispersed count data? Ecology 88(11):2766–2772. https://doi.org/10.1890/07-0043.1

WHO. 2019. World Covid-19 report (december 2019). https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (accessed on 4 December 2021)

WHO (2021) Who coronavirus (Covid-19) dashboard. https://covid19.who.int/

Woodard DB, Matteson DS, Henderson SG (2011) Stationarity of generalized autoregressive moving average models. Electron J Stat 5:800–828. https://doi.org/10.1198/016214503388619238