Modeling the Spread of COVID-19 in Lebanon: A Bayesian Perspective

Samer A. Kharroubi*

Department of Nutrition and Food Sciences, Faculty of Agricultural and Food Sciences, American University of Beirut, Beirut, Lebanon

This article investigates the problem of modeling the trend of the current Coronavirus disease 2019 pandemic in Lebanon along time. Two different models were developed using Bayesian Markov chain Monte Carlo simulation methods. The models fitted included Poisson autoregressive as a function of a short-term dependence only and Poisson autoregressive as a function of both a short-term dependence and a long-term dependence. The two models are compared in terms of their predictive ability using mean predictions, root mean squared error, and deviance information criterion. The Poisson autoregressive model that allows capturing both short-term and long-term components performs best under all criterions. The use of such a model can greatly improve the estimation of number of new infections, and can indicate whether disease has an upward/downward trend, and where about every country is on that trend, so that containment measures can be applied and/or relaxed. The Bayesian model is flexible in characterizing the uncertainty in the model outputs. The model is also applicable to other countries and more time periods as data becomes available. Further research is encouraged.

Keywords: Bayesian statistic, statistical modeling, Poisson autoregressive model, prediction, COVID-19

INTRODUCTION

As the Coronavirus Disease 2019 (COVID-19) pandemic progresses, countries around the world, including Lebanon, are increasingly implementing a range of responses that are intended to help prevent the transmission of this disease. Until a COVID-19 vaccine becomes available, strict measures from closing schools and universities to locking down entire cities and countries were enforced to suppress the virus transmission, thereby, slowing down the growth rate of cases and rapidly reducing case incidence.

Structured mathematical and statistical techniques can be potentially powerful tools in the fight against the COVID-19 pandemic. These techniques allow the COVID-19 transmission to be modeled, so the resulting models can be used to predict and explain COVID-19 infections. This may be of great usefulness for health care decision-makers, as it gives them the time to intervene on the local public health systems, thereby take the appropriate actions to contain the spreading of the virus to the degree possible. A statistical model that describes the spread of the disease over time is, therefore, essential to this endeavor. Since the outbreak of the pandemic, there has been a scramble to use and explore various statistical techniques, and other data analytic tools, for these purposes.

Advancement in statistical modeling, such as Bayesian inference methods facilitates the analysis of contagion occurrence through time. It is well-documented that infectious diseases grow exponentially and are usually driven by the basic reproduction number $R$ (see for example [1]) for a given population. The value of $R$ is defined as the ratio between consecutive new occurrences
of infection. This describes a short-term dependence. However, according to [2] for the case of COVID-19, incubation time varies substantially among individuals and incidence and so measurement may not be uniform across different populations, thereby a long-term dependence should be induced. This implies that, a model that describes the contagion dynamics of COVID-19 should ideally contain both short-term and long-term components as determining factors of newly infected counts. Such a model is hugely important to understand whether the contagion of the virus has a trend (upward or downward) and where exactly each country stands on that trend. This can provide support to decision-makers involved in contrasting the spread of the COVID-19 to perceive the effectiveness of their policy measures against the virus and what their future steps should be.

We aim in the presenting paper to model and predict the number of COVID-19 infections in Lebanon using Bayesian methods. To this aim, we propose two different models of increasing complexity using Bayesian Markov chain Monte Carlo (MCMC) simulation methods. These models include Poisson autoregressive as a function of a short-term dependence only and Poisson autoregressive as a function of both a short-term dependence and a long-term dependence. The Poisson autoregressive model that includes both short-term and long-term memory components performs best in terms of mean predictions, root mean squared error (RMSE), and deviance information criterion (DIC). A model of this kind, while mathematically expressing the current practices in the modeling on the global spread of COVID-19, produces findings that could be beneficial for policy decision-makers. Our contribution in the present study is a Bayesian statistical model for the spread of COVID-19 which, by accounting for dependence between infection counts, can better detect the contagion curve dynamics, so can shed some light on the understanding of its possible future path. The Bayesian model is also flexible in characterizing inputs to regression models and more comprehensive in characterizing the uncertainty in the model outputs.

To our knowledge, this study would be the first in the Middle East to analyze and predict the spread of COVID-19 using Bayesian methods and, therefore, neighboring countries would benefit from this study along with the model until similar studies are conducted in the region.

**METHODS**

**Data Source**
The Ministry of Public Health has started to release a daily bulletin about COVID-19 infections in Lebanon since 23 February 2020. Data are available from the website of the Ministry of Public Health (MoPH) [3] and worldometer website [4]. The overall temporal distribution of daily counts of COVID-19 cases (blue line) is presented in Figure 1. The data covers the period from February 23 to April 18, 2020. The plot indicates that COVID-19 contagion in Lebanon has achieved a complete cycle. Specifically, Figure 1 shows an upward trend until a peak is reached on March 23 and after this date, a decreasing trend is then observed.

**Bayesian Methods**
Bayesian methods allow the use of information other than the study data, into the analysis [5]. Such information is represented as a prior distribution and is combined with the likelihood function to give a posterior distribution on which inferences are subsequently made. In Bayesian analysis, the use of subjective/informative a priori beliefs is not a necessity as “vague” priors can be utilized. Given the focus in the present study is not the incorporation of prior information, all prior distributions that will be used for model parameters are considered to be “vague.” However, in order to examine the robustness of our results under different prior probability distributions, sensitivity analysis is performed. We consider this finding in more detail in the discussion section.

**MCMC Methods**
Bayesian methods rest upon the computation of the posterior distributions for model parameters. MCMC methods [6] are computer-intensive methods that would allow user to draw samples from the posterior distribution, without the need to explicitly compute the posterior distribution. Each model in this study will be fit using the software package WinBUGS [7] and the relevant WinBUGS code is provided in the Appendix.

**Model Development**
Two different models were fitted to the data as follows:

1. Poisson autoregressive as a function of a short-term dependence only
2. Poisson autoregressive as a function of both a short-term dependence and a long-term dependence

Following Agosto and Giudici [2], the number of new cases $y_t$ reported at time (day) $t$ is assumed to follow a Poisson distribution i.e.,

$$ y_t \sim \text{Poisson}(\lambda_t), $$

with a log-linear autoregressive intensity specification, as follows:

$$ \log(\lambda_t) = \alpha + \beta \log(1 + y_{t-1}) \quad (\text{Model 1}) $$

$$ \log(\lambda_t) = \alpha + \beta \log(1 + y_{t-1}) + \gamma \log(\lambda_{t-1}) \quad (\text{Model 2}) $$

In each model, the inclusion of 1 in $\log(1 + y_{t-1})$ allows to address the problem generated by zero values, $\alpha$ represents the intercept term and $\beta$ expresses the short-term dependence of the expected number of cases reported at time $t$, $\lambda_t$, on these observed in the previous day (time $t-1$). The $\gamma$ component in model 2 corresponds to a trend component and, more specifically, it represents the long-term dependence of $\lambda_t$ on all past counts of the observed process. Note that the use of a log-linear autoregressive intensity specification, rather than linear, allows for negative dependence. Inference for this model was conducted via the Maximum Likelihood estimation method in Agosto and Giudici [2].
Model Estimation

Both models were implemented from a Bayesian perspective using Gibbs sampling MCMC simulation methods using WinBUGS software [7]. The relevant code to undertake the Bayesian models is given in the Appendix. For every model, an initial 10,000 iterations were run as a “burn-in” to reach convergence. To assess convergence, two parallel chains were started from different initial values, and the ratio of the within-chain to between-chain variance was then monitored and converged at about one, indicating convergence had been reached [8]. The initial run was then followed by an additional 50,000 iterations for parameter estimation purposes. To this end, the prior distributions for all the regression parameters (\(\alpha\), \(\beta\), and \(\gamma\)) were specified as

\[
\alpha, \beta, \gamma \sim N(0, 10^6).
\]

That is, centered at zero with a large variance so as to be relatively non-informative.

Model Validation

The two models were compared in terms of their coefficients with their associated 95% credible intervals (CI), as well as their predictive performance using plots of predicted to actual values, calculations of the mean predictions, RMSE and DIC. The RMSE criterion for the mean is defined as:

\[
RMSE = \left(\frac{\sum_{t=1}^{T} (y_t - \hat{y}_t)^2}{T}\right)^{1/2}
\]

where \(y_t\) is the observed value of new infections reported at time (day) \(t\), \(\hat{y}_t\) is the fitted value and \(T\) is the number of time points in the sample. In addition, the DIC, which combines measures of both model fit and model complexity, is defined by

\[
DIC = D + PD
\]

where \(D\) represents the posterior mean deviance and \(PD\) is the effective number of parameters representing model complexity. The minimum DIC denotes the model best fitting the data [9].

| Parameter | Model 1 | Model 2 | China [2] |
|-----------|---------|---------|-----------|
| \(\alpha\) | 0.743 (0.476, 1.003) | 0.169 (0.038, 0.301) | 0.402 |
| \(\beta\) | 0.704 (0.611, 0.798) | 0.608 (0.514, 0.693) | 0.815 |
| \(\gamma\) | NA | 0.332 (0.241, 0.429) | 0.131 |
| RMSE | 8.56 | 7.68 | NA |
| DIC | 517.5 | 444.7 | NA |
RESULTS

Model Estimation and Validation

Table 1 shows the estimated autoregressive coefficients for both models together with their associated 95% CI. We notice that all coefficients had the expected positive sign as well as their credible intervals excluding zero, indicating the presence of a short-term dependence for model 1 and both a short-term dependence and a long-term trend for model 2. A testing of the models’ performance is also shown in Table 1, where model 2 was found to perform best by scoring the best RMSE and DIC with 7.68 and 444.7, respectively, in comparison to model 1 (RMSE = 8.56, DIC = 517.5).

Model Predictions

Model 2 has been tested in terms of its predictive ability where the resulting fitted mean occurrences (orange line) have been plotted along with the actual occurrences (blue line) in Figure 1. As can be seen from the plot, the Poisson autoregressive model as a function of both a short-term dependence and a long-term dependence predicts the data quite well. With aim of better interpreting the short-term time series and long-term time series of model 2, we notice from Table 1 that the estimated $\gamma$ parameter is lower than $\beta$, confirming that a downward trend data is accumulated. Additionally, we split the Lebanese data into two time periods and we separately fit the model for each data set. More specifically, we first fit the model on the data, which covers the period from February 23 to March 23, 2020 and then on the data from March 24 onwards. Analysis from the first data set revealed that $\gamma$ was larger than $\beta$, confirming the presence of an upward trend (the $\gamma$ component) which absorbs the short-term component. After this date, the estimated $\gamma$ parameter became lower, so a downward trend data is accumulated.

Agosto and Giudici [2] drew a similar conclusion from their analysis of the Chinese data, which covers the period from January 20 to March 15, 2020. Their estimated $\beta$ and $\gamma$ parameters (final column of Table 1) revealed that the contagion cycle was in a downward trend ($\gamma < \beta$). On the other side, their analysis for South Korea revealed non-significant estimate for the estimated $\gamma$ parameter, confirming absence of a trend effect on the daily cases. However, for Italy, their results showed that the estimated $\beta$ parameter was smaller than $\gamma$, suggesting that the trend of the contagion has not peaked yet.

Uncertainty in Model Predictions

A key potential advantage of the Bayesian approach is that it produces estimates of the uncertainty in the number of new infections predictions from the model. The classical models, like [2], produce data on the uncertainty in the model parameters, however, they do not produce estimates of the uncertainty in the number of new infections predictions from the model. Figure 2 shows the probability distributions for the last four daily count predictions ($\lambda_{55}$–$\lambda_{58}$) from the model. From these distributions, the mean, median, standard deviation, and corresponding 95% credible intervals along with Monte Carlo (MC) error can be computed. The distributional statistics for all daily counts predictions from the model are reported in Table 2. These results show that Bayesian method is more flexible in characterizing inputs to regression models and more comprehensive in characterizing the uncertainty in the model outputs.

DISCUSSION

In this article, we have analyzed, by modeling, the trend of the current COVID-19 pandemic in Lebanon along time. We have developed two different models of increasing complexity using Bayesian MCMC simulation methods, and found that the Poisson autoregressive as a function of both a short-term dependence and a long-term dependence provides the best fit to the data. The use of Poisson autoregressive model that allows to
| $\lambda$ | Mean prediction | Standard deviation | Median prediction | Lower bound of the 95% CI | Upper bound of the 95% CI | MC error |
|---------|----------------|--------------------|-------------------|---------------------------|--------------------------|---------|
| $\lambda_1$ | 0.9032 | 0.7494 | 0.6856 | 0.117 | 2.861 | 0.01872 |
| $\lambda_2$ | 1.297 | 0.6414 | 1.194 | 0.3645 | 2.807 | 0.01829 |
| $\lambda_3$ | 1.337 | 0.4112 | 1.317 | 0.6114 | 2.186 | 0.01005 |
| $\lambda_4$ | 1.934 | 0.2957 | 1.389 | 0.8306 | 1.988 | 0.007653 |
| $\lambda_5$ | 1.445 | 0.2453 | 1.434 | 0.9905 | 1.968 | 0.007352 |
| $\lambda_6$ | 1.867 | 0.2763 | 1.86 | 1.35 | 2.465 | 0.009348 |
| $\lambda_7$ | 1.735 | 0.2524 | 1.72 | 1.276 | 2.298 | 0.009061 |
| $\lambda_8$ | 2.091 | 0.2872 | 2.076 | 1.566 | 2.719 | 0.01054 |
| $\lambda_9$ | 2.342 | 0.3093 | 2.325 | 1.773 | 3.011 | 0.01148 |
| $\lambda_{10}$ | 3.156 | 0.3729 | 3.141 | 2.448 | 3.946 | 0.01397 |
| $\lambda_{11}$ | 3.779 | 0.4076 | 3.766 | 2.997 | 4.624 | 0.01525 |
| $\lambda_{12}$ | 4.217 | 0.4196 | 4.208 | 3.409 | 5.085 | 0.01562 |
| $\lambda_{13}$ | 5.587 | 0.3912 | 5.361 | 2.883 | 4.432 | 0.01447 |
| $\lambda_{14}$ | 3.723 | 0.3983 | 3.696 | 2.987 | 4.571 | 0.01473 |
| $\lambda_{15}$ | 3.333 | 0.3954 | 3.308 | 2.622 | 4.202 | 0.01462 |
| $\lambda_{16}$ | 4.478 | 0.4239 | 4.467 | 3.667 | 5.369 | 0.01569 |
| $\lambda_{17}$ | 5.834 | 0.4705 | 5.633 | 4.715 | 6.581 | 0.01733 |
| $\lambda_{18}$ | 5.79 | 0.4558 | 5.784 | 4.903 | 6.726 | 0.0166 |
| $\lambda_{19}$ | 8.293 | 0.6026 | 8.297 | 7.116 | 9.451 | 0.02176 |
| $\lambda_{20}$ | 10.99 | 0.797 | 11.01 | 9.45 | 12.55 | 0.02831 |
| $\lambda_{21}$ | 6.409 | 0.5255 | 6.39 | 5.442 | 7.503 | 0.01885 |
| $\lambda_{22}$ | 7.628 | 0.4764 | 7.617 | 6.723 | 8.624 | 0.01653 |
| $\lambda_{23}$ | 10.67 | 0.5928 | 10.67 | 9.506 | 11.81 | 0.01989 |
| $\lambda_{24}$ | 10.37 | 0.4995 | 10.37 | 9.407 | 11.35 | 0.0158 |
| $\lambda_{25}$ | 11.82 | 0.5463 | 11.83 | 10.75 | 12.87 | 0.01664 |
| $\lambda_{26}$ | 12.48 | 0.542 | 12.48 | 11.44 | 13.53 | 0.01563 |
| $\lambda_{27}$ | 14.11 | 0.6219 | 14.11 | 12.9 | 15.35 | 0.0176 |
| $\lambda_{28}$ | 15.49 | 0.6858 | 15.49 | 14.17 | 16.85 | 0.01894 |
| $\lambda_{29}$ | 16.98 | 0.7716 | 16.97 | 15.51 | 18.52 | 0.02119 |
| $\lambda_{30}$ | 25.03 | 1.998 | 24.97 | 21.3 | 29.1 | 0.06745 |
| $\lambda_{31}$ | 24.1 | 1.455 | 24.08 | 21.28 | 26.95 | 0.04464 |
| $\lambda_{32}$ | 22.2 | 1.101 | 22.18 | 20.12 | 24.38 | 0.03019 |
| $\lambda_{33}$ | 26.2 | 1.6 | 26.17 | 23.24 | 29.38 | 0.04873 |
| $\lambda_{34}$ | 26.75 | 1.598 | 26.71 | 23.71 | 29.93 | 0.04796 |
| $\lambda_{35}$ | 28.8 | 1.845 | 28.75 | 25.27 | 32.48 | 0.05678 |
| $\lambda_{36}$ | 26.31 | 1.496 | 26.27 | 23.48 | 29.32 | 0.04382 |
| $\lambda_{37}$ | 24.22 | 1.311 | 24.19 | 21.72 | 26.85 | 0.03771 |
| $\lambda_{38}$ | 24.66 | 1.326 | 24.64 | 22.17 | 27.3 | 0.03774 |
| $\lambda_{39}$ | 17.32 | 1.126 | 17.32 | 15.1 | 19.54 | 0.03682 |
| $\lambda_{40}$ | 17.6 | 0.9141 | 17.58 | 15.85 | 19.43 | 0.02879 |
| $\lambda_{41}$ | 17.42 | 0.8237 | 17.4 | 15.86 | 19.09 | 0.02255 |
| $\lambda_{42}$ | 16.97 | 0.7732 | 16.94 | 15.51 | 18.54 | 0.02063 |
| $\lambda_{43}$ | 16.34 | 0.7365 | 16.32 | 14.95 | 17.83 | 0.01966 |
| $\lambda_{44}$ | 15.23 | 0.7129 | 15.21 | 13.87 | 16.67 | 0.01998 |
| $\lambda_{45}$ | 12.43 | 0.7635 | 12.4 | 10.98 | 13.99 | 0.025 |
| $\lambda_{46}$ | 13.53 | 0.6141 | 13.51 | 12.39 | 14.79 | 0.01722 |
| $\lambda_{47}$ | 11.56 | 0.6646 | 11.53 | 10.33 | 12.93 | 0.02145 |
| $\lambda_{48}$ | 15.92 | 0.6674 | 15.91 | 14.68 | 17.29 | 0.01711 |
| $\lambda_{49}$ | 12.74 | 0.6063 | 12.72 | 11.6 | 13.96 | 0.0179 |

(Continued)
distributions. More discussions regarding the choice of prior distributions are available in Spiegelhalter et al. [10].

**DATA AVAILABILITY STATEMENT**

Publicly available datasets were analyzed in this study. Data are available from the website of the Ministry of Public Health in Lebanon on https://www.moph.gov.lb/en/Pages/2/24870/novel-coronavirus-2019.

**AUTHOR CONTRIBUTIONS**

SK: conceptualization, formal analysis, methodology, software, validation, writing—original draft, writing—review, and editing.

**REFERENCES**

1. Biggerstaff M, Cauchemez S, Reed C, Gambhir M, Finelli L. Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature. *BMC Infect Dis.* (2014) 14:480. doi: 10.1186/1471-2334-14-480
2. Agosto A, Giudici P. A poisson autoregressive model to understand COVID-19 contagion dynamics. *Risks.* (2020) 8:77. doi: 10.2139/ssrn.3551626
3. Ministry of Public Health. *Novel Coronavirus 2019.* (2020). Available online at: https://www.moph.gov.lb/en/Pages/2/24870/novel-coronavirus-2019 (accessed April 19, 2020).
4. Worldometer. *Coronavirus Cases.* (2020). Available online at: https://doi.org/10.1101/2020.01.23.20018549V2 (accessed April 19, 2020).
5. Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian Data Analysis.* London: Chapman and Hall (1995). doi: 10.1201/9780429295841
6. Gilks WR, Richardson S, Spiegelhalter DJ. *Markov Chain Monte Carlo in practice.* London: Chapman and Hall (1996). doi: 10.1201/b14835
7. Spiegelhalter DJ, Thomas A, Best NG, Lunn D. *WinBUGS User manual: Version 1.4.* Cambridge: MRC Biostatistics Unit (2003).
8. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Statistical Sci.* (1992) 7:457–72. doi: 10.1214/ss/1177011136
9. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. *J Royal Statistical Soc Ser B.* (2002) 64:583–639. doi: 10.1111/1467-9868.00353
10. Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. Bayesian methods in health technology assessment: a review. *Health Technology Assessment.* (2000) 4:4380. doi: 10.3310/hta4380

**Conflict of Interest:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2020 Kharroubi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*
APPENDIX

WinBUGS Code—Bayesian Analysis of Poisson Autoregressive Model (Model 2)
model{
  # AR(1) model
  log(lambda [1]) <- w + alpha*log(1+y0)
  + beta*log(lambda0)
  y[1] ~ dpois(lambda [1])
  for(t in 2:N) {
    log(lambda[t]) <- w + alpha*log(1+y[t-1])
    + beta*log(lambda[t-1])
    y[t] ~ dpois(lambda[t])
  }
  # Prior distribution
  w ~ dnorm(0.0, 1.0E-6)
  alpha ~ dnorm(0.0, 1.0E-6)
  beta ~ dnorm(0.0, 1.0E-6)
  y0 ~ dunif(0.001, 1000)
  lambda0 ~ dunif(0.001, 1000)
}