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

Samer A Kharroubi [a, b], PhD, CStat

[a] Department of Nutrition and Food Sciences, American University of Beirut, Lebanon
[b] Health Economics and Decision Science, School of Health and Related Research, The University of Sheffield, UK

## 1. Introduction

**Background:** As the Coronavirus (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.

**Objectives**: We aim to model and predict the number of COVID-19 infections in Lebanon along time using Bayesian methods.

**Methods**: Two different models were developed using Bayesian Markov chain Monte Carlo simulation (MCMC) methods. The models included Poisson autoregressive as a function of a short-term dependence only and as function of both a short-term dependence and a long-term dependence. The two models are compared in terms of root mean squared error (RMSE) and deviance information criterion (DIC).

**Results**: The Poisson autoregressive model that allows capturing both short-term and long-term components performs best under all criterions used, with RMSE (7.68) and DIC (444.7). The Poisson model also provided better predictions of observed values.

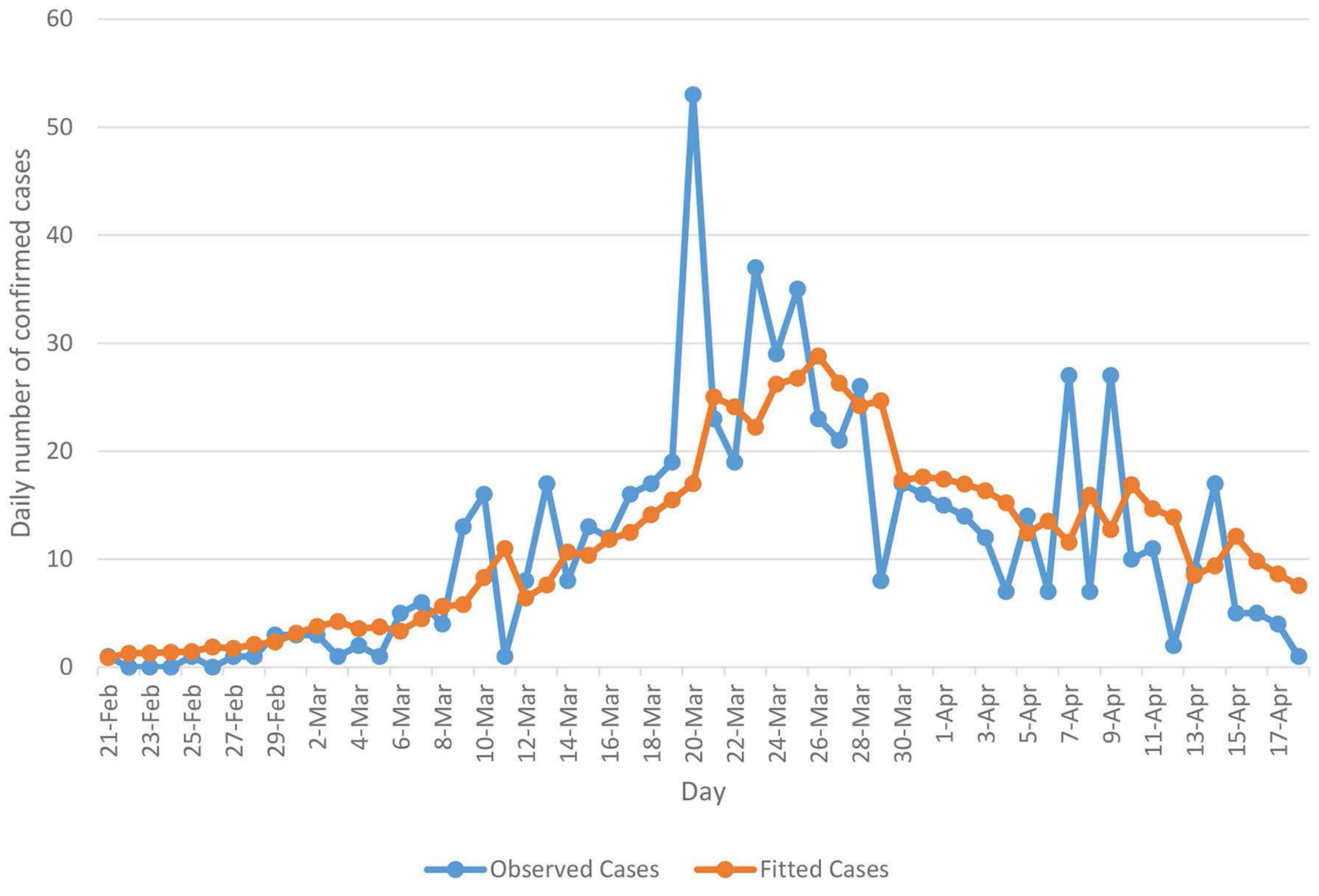
**Conclusion**: The use of such a model can greatly improve the estimation of number of new infections, and can indicate whether disease has an upward/downward trend, and where about every country is on that trend, so that containment measures can be applied and/or relaxed.

## 2. Highlights

- The Bayesian statistical model can detect the contagion curve dynamics, so can shed some light on the understanding of its possible future path.
- The model can also provide support to policy makers engaged in contrasting the spread of the COVID-19, and their economical consequences
- The model is flexible in characterizing inputs to regression models and more comprehensive in characterizing the uncertainty in the model outputs.
- 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 until similar studies are conducted in the region.

## 3. Data

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) [1] and worldometer website [2]. The overall temporal distribution of daily counts of COVID-19 cases (blue line) is presented in Figure 1 below.



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.

## 4. Modelling

**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 [3], 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:

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

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

where $\alpha$ is 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 at time $(t-1)$. The $\gamma$ component in model 2 corresponds to a trend component and represents the long-term dependence of $\lambda_t$ on all past counts.

**Model Estimation**

Both models were implemented from a Bayesian perspective using Gibbs sampling MCMC simulation methods using WinBUGS software [4]. For every model, an initial 10,000 iterations were run as a "burn-in" to reach convergence, and was then followed by 50,000 iterations for parameter estimation purposes. Finally, vague prior distributions were defined for all regression parameters:

$$\alpha, \beta, \gamma \sim N(0, 10^6)$$

**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.

## 5. Results

**Model Estimation and Validation**

Table 1 shows the estimated autoregressive coefficients for both models together with their associated 95% CI. All coefficients had the expected positive sign and their CIs excluded 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.

**Table 1.** Model coefficients and model performance

| Parameter | Model 1 | Model 2 |
|---|---|---|
| $\boldsymbol{\alpha}$ | 0.743 (0.476, 1.003) | 0.169 (0.038, 0.301) |
| $\boldsymbol{\beta}$ | 0.704 (0.611, 0.798) | 0.608 (0.514, 0.693) |
| $\boldsymbol{\gamma}$ | NA | 0.332 (0.241, 0.429) |
| **RMSE** | 8.56 | 7.68 |
| **DIC** | 517.5 | 444.7 |

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**

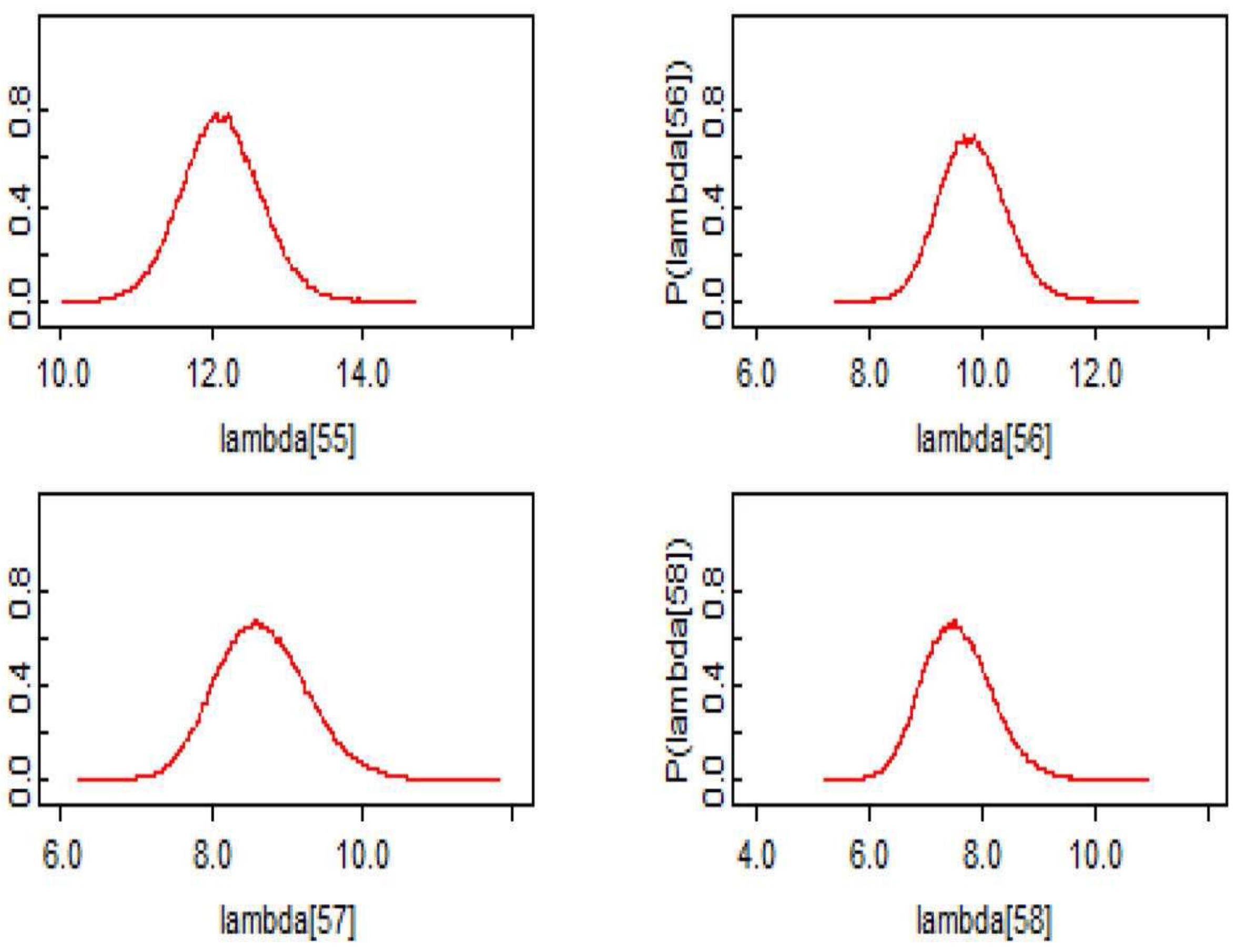
The orange line in Figure 1 represents the resulting fitted mean occurrences obtained from Model 2. As shown, the Poisson autoregressive model as a function of both a short-term dependence and a long term dependence predicts the data quite well.

Note from Table 1 that $\gamma$ is lower than $\beta$, so a downward trend data is accumulated. Further, we split the data into two time periods (Feb 23 - Mar 23 and from March 24 onwards), and we separately fit model 2 for each data set. Analysis from 1st data set revealed that $\gamma$ was larger than $\beta$, confirming the presence of an upward trend, After this date, $\gamma$ became lower, so a downward trend data is accumulated.

**Uncertainty in Model Predictions**

Figure 2 shows the probability distributions for a sampler of daily count predictions ($\lambda_{55}$–$\lambda_{58}$) from model 2. From these distributions, the mean, median, standard deviation, and corresponding 95% credible intervals along with MC error can be computed.

**Figure 2.** Example Probability Distributions for a sampler of daily counts predictions of COVID-19 cases



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.

## 6. Discussion

- We have developed a Bayesian statistical model that can be employed to understand the contagion dynamics of the COVID-19.
- The model is a Poisson autoregression of the daily new observed cases, and can reveal whether contagion has a trend, and where is each country on that trend.
- This may be of great usefulness for public decision makers to better plan health policy interventions , and their economical consequences
- The model is also applicable to other countries and more time periods as data becomes available. Ongoing research for the US, UK, China and Italy has preliminary results that are very promising.

## 7. References

1. Ministry of Public Health. Novel Coronavirus 2019. (2020). Available online at: https://www.moph.gov.lb/en/Pages/2/24870/novel-coronavirus-2019.
2. Worldometer. Coronavirus Cases. (2020). Available online at: https://doi.org/10.1101/2020.01.23.20018549V2.
3. Agosto A, Giudici P. A poisson autoregressive model to understand COVID-19 contagion dynamics. Risks. (2020) 8:77.
4. Spiegelhatler DJ, Thomas A, Best NG, Lunn D. WinBUGS User manual: Version 1.4. Cambridge: MRC Biostatistics Unit (2003).