

# A Poisson autoregressive model to understand COVID-19 contagion dynamics

Arianna Agosto and Paolo Giudici\*  
University of Pavia, Italy  
\*giudici@unipv.it

## Abstract

We present a statistical model which 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. Model results are presented from the observed series of China, Iran, Italy and South Korea.

## 1 Methods

We aim to build a data-driven model which can provide support to policy makers engaged in contrasting the spread of the COVID-19. To this aim, we propose a statistical model that can help to understand when the peak of contagion is reached, so that preventive measures (such as mobility restrictions) can be applied and/or relaxed.

To be built the model requires, for each country (or region), the daily count of new infections. In the study of epidemics, it is usually assumed that infection counts follow an exponential growth, driven by the reproduction number  $R$  (see, e.g., [1]). The latter can be estimated by the ratio between the new cases arising in consecutive days: a short-term dependence. This procedure, however, may not be adequate: incubation time is quite variable among individuals and data occurrence and measurement is not uniform across different countries (and, sometimes, along time): these aspects induce a long-term dependence.

From the previous considerations, it follows that it would be ideal to model newly infected counts as a function of both a short-term and a long-term component. A model of this kind has been recently proposed by [2], in the context of financial contagion. We propose to adapt this model to the COVID-19 contagion.

Formally, resorting to the log-linear version of Poisson autoregression, introduced by [3], we assume that the statistical distribution of new cases at time (day)  $t$ , conditional on the information up to  $t - 1$ , is Poisson, with a log-linear autoregressive intensity, as follows:

$$Y_t | \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t)$$

$$\log(\lambda_t) = \omega + \alpha \log(1 + y_{t-1}) + \beta \log(\lambda_{t-1}),$$

where  $y \in \mathbb{N}$ ,  $\omega \in \mathbb{R}$ ,  $\alpha \in \mathbb{R}$ ,  $\beta \in \mathbb{R}$ . Note that the inclusion of  $\log(1 + y_{t-1})$ , rather than  $\log(y_{t-1})$ , allows to deal with zero values.

In the model,  $\omega$  is the intercept term, whereas  $\alpha$  and  $\beta$  express the dependence of the expected number of new infections,  $\lambda_t$ , on the past counts of new infections. Specifically, the  $\alpha$  component represents the short-term dependence on the previous time point. The  $\beta$  component represents a trend component, that is, the long-term dependence on all past values of the observed process. The inclusion of the  $\beta$  component is analogous to moving from an ARCH ([4]) to a GARCH ([5]) model in Gaussian processes, and allows to capture long memory effects. The advantage of a log-linear intensity specification, rather than the linear one known as integer-valued GARCH (see, e.g., [6]), is that it allows for negative dependence. From an inferential viewpoint, [3] show that the model can be estimated by a maximum likelihood method.

## 2 Results

We have applied the model to the available data, which cover the period from January 20 to March 15, 2020. The data sources are the daily World Health Organisation reports (see [7]), from which we have extracted the "Total confirmed new cases". Figure 1 presents the observed evolution of the daily new cases of infection: for China (starting from January, 20th), Iran, South Korea and Italy (starting from February, 20th).

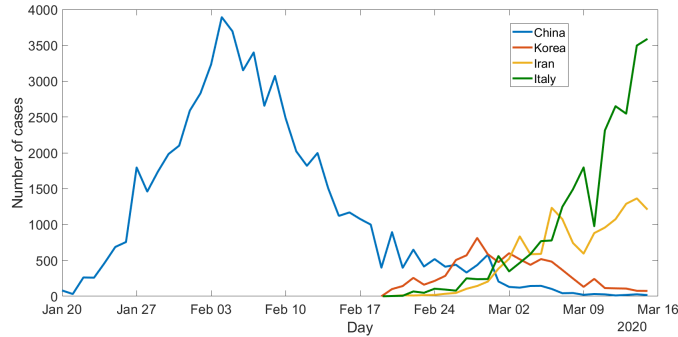


Figure 1: Observed infection counts

Figure 1 shows that COVID-19 contagion in China has completed a full cycle, with an upward trend, a peak, and a downward trend. South Korea seems to be leading to a similar situation, with a much smaller intensity. Italy seems to be

in a still increasing trend. The contagion dynamics in Iran is more difficult to interpret.

The application of our model can better qualify these conclusions. The estimated model parameters for China, using all data, are shown in Table 1.

Parameter	Estimate	Std error (p-value)
$\omega$	0.402	0.338 (0.120)
$\alpha$	0.815	0.073 (0.000)
$\beta$	0.131	0.062 (0.016)

Table 1: Model estimates for China, with standard errors and p-values.

Table 1 shows that all estimated autoregressive coefficients are significant, confirming the presence of both a short-term dependence and a long-term trend. From an interpretational viewpoint, the estimate of  $\alpha$  shows that, if the expectation of new cases for yesterday was close to 0, 100 new cases observed yesterday generate about 40 new expected cases today. According to the value estimated for  $\beta$ , an expectation of 100 new cases for yesterday generates instead about 2 new expected cases today, if no cases were observed yesterday. Additionally, we remark that the goodness of fit of the model is quite high, as the Root Mean Squared Error (RMSE) is equal to 315.36, against an overall mean of 1087.48.

With the aim of better interpreting the still short time series of the other countries, we repeatedly fit the model to the Chinese data, using increasing amounts of data, in a retrospective way. More precisely, we first fit the model on the first 15 counts from China (a minimal requirement for statistical consistency of the results), then on the first 16, and so on. For each fit we plot the estimated  $\alpha$  and  $\beta$  parameters in Figure 2.

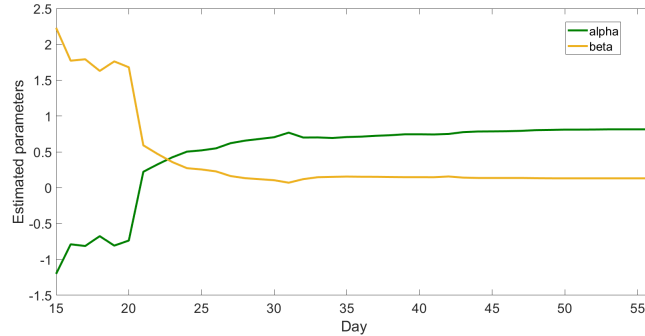


Figure 2: Evolution of the  $\alpha$  and  $\beta$  parameters for Chinese daily infection counts.

Figure 2 shows that, until February 11th (the 23rd day reported)  $\beta$  is greater than  $\alpha$ , indicating the presence of a still increasing trend (the  $\beta$  component) that absorbs the short-term component. After that time, downward trend data is accumulated,  $\beta$  starts decreasing and  $\alpha$  increasing. The results approximate

the values in Table 1 around February, 20th: after this date the estimated parameters become stable, as the difference between subsequent estimates becomes lower than 0.01.

What obtained from the Chinese data suggests to similarly estimate, in a recursive way, the  $\alpha$  and  $\beta$  parameters for all other countries, from day 15 onwards. This means, at the day of writing, up to 25 days. Our results show that, for Iran, the  $\omega$  parameter prevails, with an estimated value equal to 1.02, indicating a strong erratic behaviour of the reported infection counts. The estimated  $\beta$  parameter for South Korea is not significant, indicating absence of a trend effect on the daily counts, consistently with what observed in Figure 1. For Italy, instead,  $\alpha$  is positive (0.14 on day 25) but still lower than  $\beta$  (0.78 on day 25), suggesting that the peak of the trend is not reached yet, with a time delay from China (which peaked on Day 23).

To conclude, we believe that our model can constitute a useful statistical tool for decision makers: in each country, once a minimal series of data is collected (we suggest 15 days) the values of  $\omega$ ,  $\alpha$  and  $\beta$  can be monitored along time, to reveal at what stage the contagion cycle is: in an upward trend (as Italy); in a downward trend (as China); in a no trend cycle (as South Korea); or in a situation which needs more data to be understood (as Iran). The full reproducibility of our model can easily extend its application to more countries and time periods as data becomes available.

## References

- [1] Biggerstaff, M., Cauchemez, S., Reed, C., Gambhir M. and Finelli L. (2014) Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature. *BMC Infect Dis.* 2014 Sep 4, 14:480.
- [2] Agosto, A., G. Cavaliere, D. Kristensen, A. Rahbek (2016), Modeling Corporate Defaults: Poisson Autoregressions with Exogenous Covariates (PARX), *Journal of Empirical Finance* 38 (B), 640-663.
- [3] Fokianos, K. and Tjøstheim, D. (2011). Log-linear Poisson autoregression, *Journal of Multivariate Analysis*, 102, 563-578.
- [4] Engle, R.F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U. K. Inflation. *Econometrica*, 50, 987-1008.
- [5] Engle, R.F and Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric Reviews*, 5:1, 1-50.
- [6] Ferland, R., Latour, A., Oraichi, D. (2006). Integer-valued GARCH processes. *J. Time Ser. Anal.* 27, 923-942.
- [7] World Health Organisation (2020). Novel Coronavirus (2019-nCoV) situation reports, 1-49.