

POTENTIAL IMPACT OF A DEMONSTRATION ON COVID-19 CONTAGION: AN APPLICATION OF A METHOD

Conceição Leal^{1,2*} Leonel Morgado^{1,2} Teresa A. Oliveira^{1,2}

¹ Universidade Aberta, Portugal

² CEAUL, FCUL, Portugal

³ INESC TEC, Portugal

^{1,2*} conceicao.leal2010@gmail.com

Abstract: There is evidence that some outdoor events may have contributed to the spread of COVID-19. We updated an empirical methodology based on regression modeling and hypothesis testing to analyze the potential impact of a demonstration that took place in Lisbon, within the scope of the 'Black Lives Matter' context, on the contagion pattern in the region where this event occurred. We find that in the post-impact period there was no acceleration in the number of cases in the region, unlike in a prior event in the region. The proportion of counties where there was a potential impact of the event is not statistically significant. This result demonstrates that not all outdoor events contributed to the spread of COVID-19 and exemplifies how to apply the selected empirical methodology.

keywords: Time series segmentation; Modeling COVID-19 data; Geographic heterogeneous impacts.

MSC2020: 49-XX; 34-XX; 92-XX.

1 Introduction

The COVID-19 coronavirus disease is characterized by a high level of human-to-human transmission hazard. Some outdoor events played an important role in the spread of COVID-19 [1]. A retrospective analysis of the change in the incidence rate of COVID-19 during the two weeks after outdoor mass gatherings in low-incidence US counties revealed a mean increase of 1.5 times [4]. To analyze the potential impact of large-scale public events, in an earlier work [3], we implemented an empirical analysis based on regression modeling and hypothesis testing for two periods: before the potential impacts of the event are reported (pre-impact) and after potential impacts have been reported (post-impact). We use this approach to analyze the potential impact in COVID-19 contagion of the demonstration of thousands of people in Lisbon, joining the "Black Lives Matter" movement, on June 6, 2020. We first observed that there was no acceleration in contagion in the region where the event took place (Lisbon Metropolitan Area – AML) after it was held (Fig 1b) similarly to what happened in the rest of the country. This is in contrast to what was seen in the May 1st event studied in the previous study [3] (Fig. 1a). Then, we analyzed whether this pattern of no acceleration of contagion was geographically homogeneous across the AML counties.

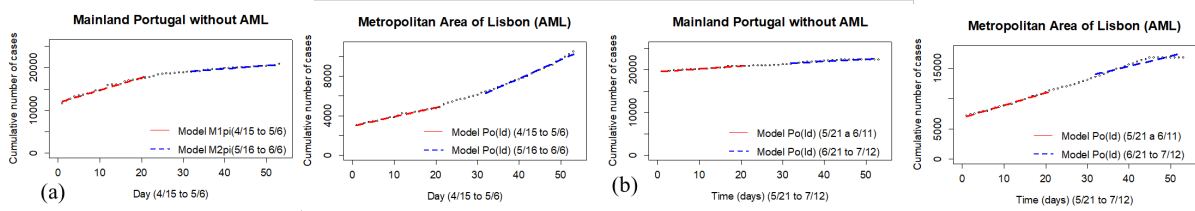


Figure 1: Number of cumulative cases: Portugal & Lisbon region (AML, Portuguese-language acronym): (a) May 1st event; (b) "Black Lives Matter" event

2 Results and discussion

2.1 Data

Data on the cumulative number of COVID-19 cases in Portuguese counties come from the daily situation reports issued from May 21 to July 12, 2020, by the National Health Directorate. Data were collected daily by the authors, directly from official reports and from the online repository of the "Data Science for Social Good Portugal" community, which extracted these data daily and made them publicly available (<https://github.com/dssg-pt/covid19pt-data>). Following the method presented in an earlier work [3], the analysis was implemented using R software tools and applied to data matching two timeframes: an S1 period where possible event contagion had not yet been reported (pre-impact) and an S3 period where any such contagion would expectably have been reported (post-impact). In the intermediate period S2, any event-related contagion reporting would be developing and we ignore it because, to detect if an impact occurred, it is not relevant to detect its ongoing development, only its existence after the impact is complete.

2.2 Times series segmentation

To define the S1, S2, and S3 periods from publicly available data, previous work [3] estimated the reporting delay using the country-level data (i.e., nationally) on new daily cases per symptom emergence date, via graphical measuring of the only data source: bar charts in public daily reports. This was done from April 17 to May 9, 2020, and later confirmed by a public statement of the National Health Directorate (ibid.).

To estimate the number of days, X , until the number of reported cases stabilised on day t , we used a Shewhart Control Chart targeting the zero daily variation of the reported daily number. We took the value 4.1 for the reporting delay [3]. Combining the expectation that less than 2.5% of infected people will show symptoms within 2.2 days after infection [3] and the average reporting delay of 4.1 days, an extremely small number of cases of infection on June 6, 2020 will have been reported until June 12. So, the first segment (S1) refers to the period from May 21 to June 11. By Lauer et al., [2], 97.5% of cases will show symptoms within 11.5 days after infection. So, the intermediate period, in which any contagions that occurred during the event will be reported, runs from June 12 to June 20. This period corresponds to the second segment (S2). S3 corresponds to the period from June 21 to July 12, 2020.

2.3 Models

For the S1 and S3 periods under analysis, we defined Poisson regression models, with Identity (model 1) and Log link functions, adjusted to $S1 \cup S3$ data, for the Lisbon region, for the rest of the country (Fig. 1), and for individual counties in the AML (Lisbon) region (Fig. 2). In the

models, we considered the interaction between time and a dummy variable that identifies the periods S1 and S3, to test the β homogeneity coefficient between periods S1 and S3.

$$\begin{aligned}
 Po(Id) : \quad & g(E(NCases)) = \beta_0 + \beta_1 t + \beta_2 \text{Dummy} + \gamma(\text{Dummy} \times t) \\
 & g(E(NCases)) = E(NCases) \quad (\text{Identity link function}) \\
 & NCases \sim Po(g(E(NCases)))
 \end{aligned} \tag{1}$$

t is the time, in days, between May 21 and July 12, 2020, $g()$ designates the link function and $E()$ is the expected value.

2.4 Testing homogeneity of regression coefficients - some results

In the methodology, we systematically used one-way ANCOVA to test the homogeneity of coefficients in the S1 and S3 periods for the $Po(Id)$ and $Po(Log)$ models. When the homogeneity hypothesis was rejected, we employed graphical visualisation to determine the nature of the differences and used the Likelihood Ratio test to statistically assess whether the rate of change in the number of COVID-19 cases decreased from S1 to S3.

Counties where the hypothesis of homogeneity of variation rates was not rejected for $Po(Id)$ (i.e., the hypothesis of maintaining the growth pattern after the event was not rejected), and counties where this hypothesis was rejected but graphical visualisation and the Likelihood Ratio test did not indicate an acceleration of case growth from S1 to S3, even suggesting a reduction, were included in the group of counties for which the growth pattern of cases did not accelerate from S1 to S3.

For other counties, we tested the homogeneity of the $Po(Log)$ model coefficients. Subsequently:

- Counties where the homogeneity of $Po(Log)$ model coefficients was also rejected were placed in the group of counties which exhibited case growth acceleration from the S1 to the S3 period.
- For counties where there was an acceleration in the case growth rate from S1 to S3 (the hypothesis of homogeneity of $Po(Id)$ coefficients was rejected), but where the hypothesis of coefficient homogeneity for $Po(Log)$ was not rejected, it can be inferred that such exponential growth had already been occurring in S1, and that the growth rate could already be accelerating exponentially. Thus, it is expected that the exponential model will continue to be the best fit for S3.

Using the AIC criterion, we can check which model better fits each period. If $Po(Log)$ fits better to S3, then there was no acceleration; if $Po(Log)$ fits better to S1 and $Po(Id)$ fits S3 better, we extend the S1 period by one month earlier to evaluate which model fits each period best. If both are linear ($Po(Id)$), then it is assumed that the growth pattern has changed. If $Po(Id)$ has the best fit for S1 and $Po(Log)$ has the best fit for S3, then we assume that the growth pattern has not changed from one period to the other.

Based on the results obtained, it is found that for all counties in the AML, from period S1 (May 21 to June 11) to period S3 (June 21 to July 12), there is no statistical evidence of an increase in the growth rate of the cumulative number of COVID-19 cases. In all counties, either the null hypothesis of homogeneity of growth rates in the two periods was not rejected, or it was rejected in favour of the alternative hypothesis of a decrease in this rate.

Therefore, based on the implemented methodology, it is concluded that there is no statistical evidence to support that the "Black Lives Matter" demonstration had an effect on the spread of COVID-19 in the AML, contrary to what was observed in the May 1st demonstration ([3])

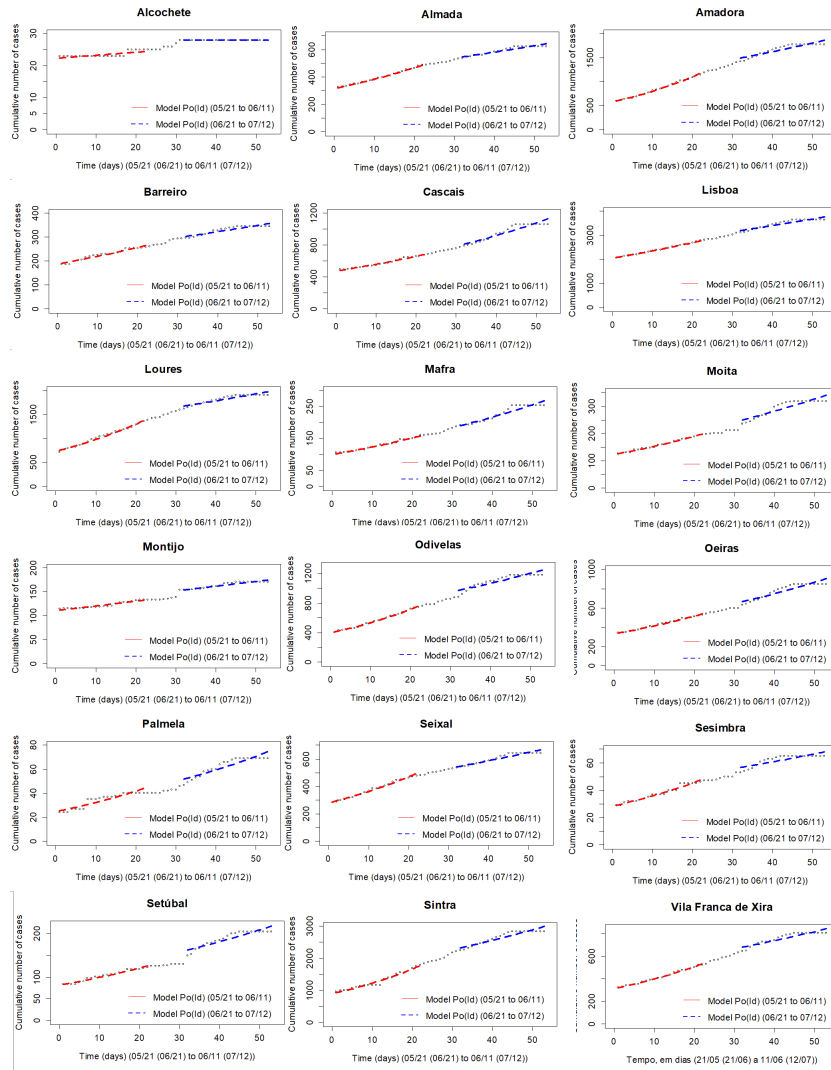


Figure 2: Poisson regression model with *Identity* link function, adjusted to S1 and S3 and observed data.

References

- [1] Kucharski, A. J., Russell, T. W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., ... Flasche, S. (2020). Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The lancet infectious diseases*, 20(5), 553-558.
- [2] Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., ... Lessler, J. (2020). The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of internal medicine*, 172(9), 577-582.
- [3] Leal, C., Morgado, L., Oliveira, T. A. (2023). Mathematical and Statistical Modelling for Assessing COVID-19 Superspreader Contagion: Analysis of Geographical Heterogeneous Impacts from Public Events. *Mathematics*, 11(5), 1156.
- [4] Miron, O., Yu, K. H., Wilf-Miron, R., Davidovitch, N. (2020). COVID-19 infections following outdoor mass gatherings in low incidence areas: retrospective cohort study. *medRxiv*, 2020-10.