JDSSV

Journal of Data Science, Statistics, and Visualisation

November 2022, Volume II, Issue VII. *doi: 10.52933/jdssv.v2i7.50*

Multiple Changepoint Analysis of COVID-19 Infection Progression and Related Deaths in the Small Island State of Malta

Gianluca Ursino
University of Malta

David Suda
University of Malta

Monique Borg Inguanez
University of Malta

Abstract

In December 2019, in the city of Wuhan (China), Severe Acute Respiratory Syndrome Coronavirus - 2 (SARS-CoV−2), a virus that causes what is known as Coronavirus Disease 2019 (better known as COVID-19), emerged. In a few months, the virus spread around the world becoming a global pandemic that has impacted the world. In Malta (a nation consisting of an archipelago of islands of approximately 500, 000 people), which is the case study of this analysis, the first case was identified on 7/3/2020. In this paper, an ensemble approach is adoped to fitting a piecewise linear trend model to the log-scale of cumulative cases and deaths due to COVID-19 in Malta by implementing the SN-NOT changepoint model. This model combines the self-normalisation (SN) technique, which is used to test whether there is a single change-point in the linear trend of a time series, with the Narrowest Over Threshold algorithm (NOT) to achieve multiple change-point in the linear trend. Through analysis of news reports and other sources of information, estimated change-points are then compared to potential factors such as health restrictions, mass events, government policy and population behaviour that may have affected these changes, in order to determine the effect of these factors on the spread of the disease.

Keywords: changepoint analysis, COVID-19, time series analysis, R.

1. Introduction

Severe Acute Respiratory Syndrome Coronavirus−2 (SARS-CoV−2) is a virus that causes what is known as Coronavirus Disease 2019 (COVID-19), which is a global pandemic that has severly impacted the world since 2020, and of which the world is still feeling the consequences. It first emerged in December 2019 when unusual pneumonia-like cases were identified in the city of Wuhan, China. On 30/12/2019, the Wuhan local health authority issued an epidemiological alert, and on 1/1/2020 the Huanan seafood market was closed since it was associated with the early cases, as noted in Huang et al. (2020). However, other investigations such as Nishura et al. (2020) suggest that the transmissions that had occurred within the market could have been due to secondary transmissions, that is, human-to-human transmissions. In the first few months of 2020, the virus spread to other parts of China and beyond, reaching even Europe and eventually Malta, where the virus was first identified on 7/3/2020. The World Health Organisation (WHO) officially declared this outbreak as a pandemic on 11/3/2020. This study will focus on the impacts of events and government interventions in Malta as a case study. Malta is a particular case study, as it is a small island nation of approximately 500,000 people consisting of an archipelago of islands with only two mainly inhabited: the island of Malta being the main one and Gozo being the smaller sister island. This brings with it some advantages and disadvantages in pandemic circumstances. The first main advantage is that, due to its size, government interventions are more easy to implement nationwide, and their impact more easily understood. The second main advantage would be that Malta has no land borders, and since all international arrivals arrive through the airport or through the main sea port, border control is easier. On the flipside, the main disadvantage is the population density of Malta - at 1628 inhabitants per square kilometre [1], Malta is the sovereign state with the 4$^{\text{th}}$ highest population density in the world (after Monaco, Singapore and Bahrain) - and COVID-19 infection is known to thrive more in urbanised environments.

People who are infected with COVID-19 may exhibit symptoms such as fever, dry cough, fatigue, loss of taste or smell, aches and pains, sore throat and diarrhea. Most people who have contracted the virus have experienced mild to moderate respiratory illness and have recovered without the need of special treatments such as intensive care units (ICUs). However, older people and those who have pre-existing medical conditions such as cardiovascular disease, diabetes, chronic respiratory disease, and cancer, may develop serious illness which may require the use of ICUs, and may even be fatal. Since the start of the pandemic, there has been a worldwide response to reduce the spread of the disease. In most countries, local and international travel had to be restricted or banned, lockdowns or shutdowns were imposed, and social distancing, the use of face masks and hand hygiene were encouraged. In Malta, a number of measures were taken to reduce the spread of COVID-19 such as the closure of schools, the airport and non-essential shops, restriction of group size in public gatherings, mandatory self-isolation for those who contracted the virus or were in close contact to infected persons, and mandatory wearing of masks in public places. As a result of these measures, the number of new cases and the number of active cases were, for a period of time, reduced significantly or were curbed. Previous studies on daily COVID-19 figures in Wuhan and

[1] https://www.nso.gov.mt [Accessed on 18$^{\text{th}}$ June 2021]

in Europe (see e.g., Pan et al. (2020), Flaxman et al. (2020)) showed that interventions such as social distancing, shutdowns and self-isolation among others have reduced the transmission of the virus. This has led to reduction of certain health measures in Europe including travelling. Li et al. (2020) found that the ease of restrictions such as school reopening, lifting ban on public events and public gatherings of more than ten people have contributed to the increase of the reproduction number, which caused a second wave of COVID-19 infections in certain countries. From a Maltese perspective, there is a study by Cuschieri et al. (2021) which attributes mass events as the trigger of a significant second wave following a well-managed first wave.

Since the start of the pandemic, a number of academics around the world have applied statistical techniques on COVID-19 data. Early research papers were more focused on the modelling of the epidemic in mainland China. Zhan et al. (2020) used constrained nonlinear optimisation procedure to estimate the dynamics of the epidemic modelled as a modified Susceptible-Exposed-Infected-Remove (SEIR) model. Zhao et al. (2020) estimated the basic reproduction number using the exponential growth model method. In Tsang et al. (2020), a study of whether the changes in the case definition for COVID-19 in mainland China affected its inferences on the transmission dynamics by using exponential growth models, is presented. As the pandemic progressed, there have been more studies focusing on regions other than China. Sahafizadeh and Sartoli (2020) fitted the Susceptible-Infected-Removed (SIR) epidemic model to describe the epidemic curve and to estimate the reproduction number of COVID-19 in Iran. Sebastiani et al. (2020) analysed the evolution of the COVID-19 cumulative incidence of different provinces in Italy by using a Bayesian approach which takes into account the presence of asymptomatic cases and the effect of the measures implemented by the Italian Government. Yu et al. (2020) analysed the dynamics of COVID-19 by age and gender groups - their study involved fitting a semi-parametric generalized additive model (GAM) to obtain fitted daily case counts in South Korea. Chu (2020) modelled the incidence of COVID-19 in Italy and Spain by using the Susceptible-Infectious-Recovered (SIR) model and the log-linear regression model, and also estimated the basic reproduction number, growth rate, and doubling time. Wang et al. (2020) fitted a logistic regression model to the Brazilian, Russian, Indian, Peruvian, Indonesian and global COVID-19 data and used a machine learning technique to predict the trend of the epidemic. Ogundokun et al. (2020) fitted an ordinary least squares regression model to measure the impact of travelling history and contacts on the spread of COVID-19 in Nigeria. To the knowledge of the authors, there is not considerable focus on the statistical modelling of COVID-19 in Malta. A paper by Cuschieri and Grech (2021) is the only literature found applying a statistical approach to Maltese epidemiological data related to COVID-19, in which regression modelling is used to determine the characteristics and associations of those who have at least two chronic diseases, to enable adequate priority and policy planning to brace the pandemic in Malta. The aim of this paper will be that of analysing changes in what is called the pandemic's exponential growth rate in the local context, which is a measure of the speed of the spread, using an ensemble approach to multiple changepoint analysis. Furthermore, we also aim to identify potential factors that have affected these changes.

2. The Changepoint Model for Growth Rate

Changepoint analysis dates back to the 1950s, when Page (1954) applied the CUSUM (cumulative sum) statistic to detect changes in the parameter of interest. From there onwards, changepoint analysis has evolved to detect multiple changepoints in various model settings. There are two main branches of changepoint analysis: online and offline. The former performs the analysis in real time as more data becomes available, and ideally detects a changepoint as soon as possible after it occurs, while the latter considers the whole data at one go. The focus will be on the latter, as the aim of this paper will be that of analysing the changepoints retrospectively. In most literature, the time series is assumed to be potentially non-stationary but can be partitioned into stationary segments. Parameters of interest include the mean, the variance, or both of them simultaneously, and these types of changepoint models can be found in e.g., Picard et al. (2005), Chen and Gupta (2012), Killick et al. (2012), Haynes et al. (2017). In literature, one can also find non-parametric approaches (see e.g., Matteson and James (2014), Haynes et al. (2016)), Bayesian approaches (see e.g., Baker et al. (2016), Thies and Molnár (2018)), and multivariate time series approaches (see e.g., Jandhyala et al. (2013), Ma and Yau (2016)).

The cumulative daily cases and deaths are assumed to have an exponential growth rate that varies throughout the pandemic. The changepoint model which shall be implemented is a parametric one proposed by Jiang et al. (2020) which is presented as a changepoint detection problem where the underlying model is a piecewise linear trend model applied to the log-scale of daily cumulative cases and daily cumulative deaths, and the main interest is to find the points in the time series where the exponential growth rate changes. The piecewise linear model can interpret the state of the pandemic in different phases which correspond to different segments in the piecewise linear trend model. Specifically, the phase transitions of the exponential growth rate would be measured as the changes in the slope parameter in the piecewise linear trend model. The piecewise linear trend model has the advantage of being simple, intuitive, interpretable and useful to determine the effectiveness of interventions. Furthermore, the method by Jiang et al. (2020) applies a self-normalisation (SN) statistic for the estimation of a single changepoint which can be extended to multiple changepoints by applying it to randomly drawn subsamples. This method is a combination of the SN statistics proposed by Shao (2010) and the Narrowest-Over-Threshold (NOT) algorithm proposed by Baranowski et al. (2019), and is termed as the SN-NOT algorithm. Jiang et al. (2020) focused on up to 30 representative countries while comparing the evolution of the pandemic in its initial stages (up till 27/05/2020). This paper does not involve a comparative study, but this model shall be applied only to Malta and take a longer time span (up till 31/01/2021), thus allowing us to monitor in detail several stages of the first eleven months of the pandemic on the island, prior to the introduction of more gradual effects such as vaccination. Furthermore, a detailed analysis of relevant events and government interventions that may have contributed to these changepoints shall simultaneously be provided.

Since the log-scale of the cumulative daily cases and deaths of COVID-19 are being analysed to assess the exponential growth rate of the pandemic, the time series should exhibit a non-decreasing trend, hence the assumption of piecewise stationarity is not

adequate. Suppose Y_t represents the log-scale of cumulative daily cases. Then the exponential growth is given by $\beta^* \approx \log(Y_{t+1}) - \log(Y_t) \equiv \log(Y_{t+1}/Y_t)$ (see e.g., Locatelli et al. (2021)). Now suppose that β^* is constant for some time partition $\{\tau+1,\ldots,\tau'\}$, where $\tau' = \tau + n^*$. Since $\log(Y_{\tau'}) - \log(Y_{\tau+1})$ can be viewed as a telescoping sum of terms $\log(Y_{\tau+k+1}) - \log(Y_{\tau+k})$ then it is reasonable to assume that $\log(Y_{\tau'}) - \log(Y_{\tau+1}) \approx (n^*-1)\,\beta^*$ and thus the increase is linear. Hence, to cater for the setting where the exponential growth is assumed to be piecewise constant over specific partitions, the following piecewise linear trend model of time series for cumulative number of cases $\{Y_t\}_{t=1}^n$ is formulated:

$$Y_t = \begin{cases} \beta_0^{(1)} + \beta_1^{(1)}\left(\frac{t}{n}\right) + u_t, & \text{if } t \in \{1,\ldots,\tau_1\} \\ \beta_0^{(2)} + \beta_1^{(2)}\left(\frac{t}{n}\right) + u_t, & \text{if } t \in \{\tau_1+1,\ldots,\tau_2\} \\ \vdots & \\ \beta_0^{(m+1)} + \beta_1^{(m+1)}\left(\frac{t}{n}\right) + u_t, & \text{if } t \in \{\tau_m+1,\ldots,n\} \end{cases} \tag{1}$$

where $\boldsymbol{\tau} = (\tau_1,...,\tau_m)'$ are the m changepoint locations such that $\tau_1 < \cdots < \tau_m$ with the convention that $\tau_0 = 1$ and $\tau_{m+1} = n$, the last time point. The vector $\boldsymbol{\beta}^{(i)} = \left(\beta_0^{(i)}, \beta_1^{(i)}\right)'$ contains the intercept and slope of $\mathbb{E}[Y_t]$ in the i^{th} segment for all segments $i = 1,...,m+1$ and $\{u_t\}_{t=1}^n$ is a weakly dependent stationary error process. It is required that two consecutive intercept and slope parameters must be different, that is, $\beta_j^{(i)} \neq \beta_j^{(i+1)}$ for all $i = 1,...,m$ and $j = 0, 1$.

The piecewise linear model in (1) can also be used to model the cumulative number of deaths. In the period studied, deaths can be assumed to follow a relatively stable proportion of total cases within approximately three weeks of a spike in cases. Indeed, a pre-vaccination study by Zhou et al. (2020) states that the median time from illness onset till death is 18.5 days. Furthermore, the case fatality rate in Malta also stood consistently between 1% and 2% during this period. Nonetheless, in practice one would not be realistic to expect that changepoints for the cumulative deaths process will always occur at the same lag after changepoints for cumulative cases, as this would require one to assume homogenous mixing between susceptible and infected. In reality, during the period of study, it is known that this has not been the case throughout, as there have been periods where certain segments of the population were the cause of outbreaks (e.g., in elderly homes or in entertainment districts) and were thus more infected than others.

The SN-NOT algorithm used for detecting changepoints requires the use of asymptotic results of the SN test (Theorem 2.1 and Theorem 2.2) which can be found in Jiang et al. (2020). In order to ensure that the invariance principle holds so that the mentioned asymptotic results are satisfied, one needs to assume that the error process $\{u_t\}_{t=1}^n$ in the piecewise linear trend model in (1) is strictly stationary such that $\mathbb{E}[u_t] = 0$, $\mathbb{E}[u_t^4] < \infty$, and the long-run variance satisfies

$$\Gamma^2 = \lim_{n\to\infty} \operatorname{Var}\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n} u_t\right] \in (0,\infty).$$

In practice, it is difficult to check for strict stationarity, and one can only check for weak dependence via the inspection of the autocorrelation function (ACF) and partial

autocorrelation function (PACF) plots of the residuals, by determining whether the autocorrelation and partial autocorrelation coefficients approach zero fast enough. Hence, the reason why the weak dependence assumption is made on the model.

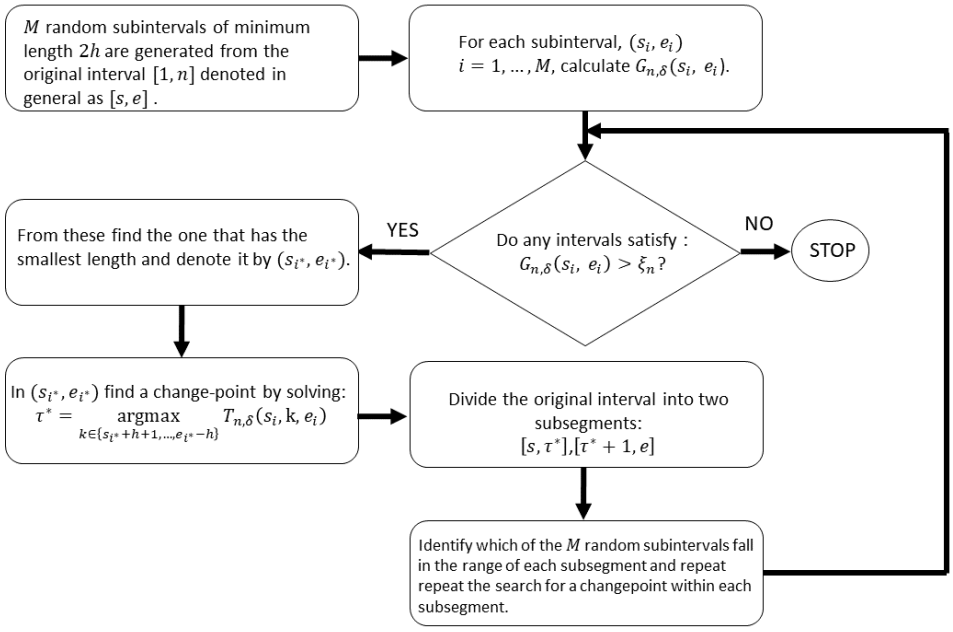

Figure 1: Schematic of Algorithm 1.

The main objective is to estimate the unknown number of changepoints m and their locations $\boldsymbol{\tau}$, and to relate them to the interventions that have occurred from 7$^{\text{th}}$ March 2020 until 31$^{\text{st}}$ January 2021 during the pandemic in Malta. The method proposed by Jiang et al. (2020) shall be used. These authors have applied this method to model the trajectory of the log-scale of COVID-19 cumulative confirmed cases and deaths in 30 major countries. Figure 1 gives a schematic of the SN-NOT algorithm, of which a brief overview follows. The core idea is to draw a number of random intervals of time points (s_i, e_i) where $1 \leq s_i < e_i \leq n$, and apply a single changepoint detection method on each of the subsamples $\{Y_t\}_{t=s_i}^{e_i}$. Those subsamples in which the single changepoint detection method exceeds a user-specified threshold are retained, and the one in which $e_i - s_i + 1$ is the smallest is selected, that is, the subsample that is drawn on the narrowest interval. This ensures that the narrowest subsample in which the single changepoint detection method exceeds the threshold has a high probability that it contains at most one changepoint. Afterwards, the time series is split into two at the estimated changepoint and the process repeats on both. One of the main advantages of this algorithm is that it can be applied to a wide range of changepoint models, including the model in (1). The SN-NOT algorithm shall be discussed in more mathematical detail in the next section.

2.1. SN-NOT Procedure

Denote $F_n^M = \{(s_i, e_i) : i = 1, ..., M\}$ as the set of M random intervals of time points

such that each pair of integers (s_i, e_i) are drawn uniformly from $\{1, ..., n\}$ and satisfy $1 \leq s_i < e_i \leq n$ and $e_i - s_i + 1 \geq 2h$, where $h = \lfloor \varepsilon n \rfloor$ is the minimum segment length and $0 < \varepsilon < 1/2$ is a trimming parameter. The trimming parameter ε is used because if a changepoint τ is very near to the start and end points of a segment it would be difficult to detect it. Such a trimming parameter has been used in changepoint models applied in Hawkins (1987), Andrews (1993) and Bai and Perron (1998). Consider one subsample $\{Y_t\}_{t=s}^{e}$ where $(s, e) \in F_n^M$. $\hat{\boldsymbol{\beta}}_{(s,e)}$ denotes the OLS estimator of the intercept and slope parameters on $\{Y_t\}_{t=s}^{e}$, which can be written in terms of the following summation:

$$\hat{\boldsymbol{\beta}}_{(s,e)} = \left[\sum_{t=s}^{e} \mathbf{F}\left(\frac{t}{n}\right) \mathbf{F}\left(\frac{t}{n}\right)'\right]^{-1} \left[\sum_{t=s}^{e} \mathbf{F}\left(\frac{t}{n}\right) Y_t\right],$$

where $\mathbf{F}(s) = (1, s)'$. The contrast statistic and self-normaliser matrix respectively are calculated for subsample $\{Y_t\}_{t=s}^{e}$ given a potential changepoint k, where $s + h - 1 \leq k \leq e - h$:

$$\begin{aligned} \mathbf{D}_n(s, k, e) &= \tfrac{(k-s+1)(e-k)}{(e-s+1)^{\frac{3}{2}}} \left(\hat{\boldsymbol{\beta}}_{(s,k)} - \hat{\boldsymbol{\beta}}_{(k+1,e)}\right), \\ \mathbb{V}_{n,\delta}(s, k, e) &= \mathbb{L}_{n,\delta}(s, k, e) + \mathbb{R}_{n,\delta}(s, k, e), \end{aligned}$$

where $0 < \delta < \epsilon/2$ is a local trimming parameter and

$$\begin{aligned} \mathbb{L}_{n,\delta}(s, k, e) &= \textstyle\sum_{j=s+1+\lfloor n\delta \rfloor}^{k-2-\lfloor n\delta \rfloor} \frac{(j-s+1)^2(k-j)^2}{(k-s+1)^2(e-s+1)^2} \left(\hat{\boldsymbol{\beta}}_{(s,j)} - \hat{\boldsymbol{\beta}}_{(j+1,k)}\right)^{\otimes 2}, \\ \mathbb{R}_{n,\delta}(s, k, e) &= \textstyle\sum_{j=k+3+\lfloor n\delta \rfloor}^{e-1-\lfloor n\delta \rfloor} \frac{(j-1-k)^2(e-j+1)^2}{(e-s+1)^2(e-k)^2} \left(\hat{\boldsymbol{\beta}}_{(j,e)} - \hat{\boldsymbol{\beta}}_{(k+1,j-1)}\right)^{\otimes 2}, \end{aligned}$$

where for any vector $\mathbf{x}$, $\mathbf{x}^{\otimes 2} = \mathbf{x}\mathbf{x}'$. The SN test statistic $G_{n,\delta}(s, e)$ for subsample $\{Y_t\}_{t=s}^{e}$ is then calculated as

$$G_{n,\delta}(s, e) = \max_{k \in \{s+h-1, \ldots, e-h\}} T_{n,\delta}(s, k, e), \tag{2}$$

where $T_{n,\delta}(s, k, e) = \mathbf{D}_n(s, k, e)' \mathbb{V}_{n,\delta}(s, k, e)^{-1} \mathbf{D}_n(s, k, e)$. Among those subsamples where the SN test statistic in (2) exceeds a given threshold ζ_n, the one which has the narrowest interval, denoted as (s_{i^*}, e_{i^*}), is chosen. The changepoint location is then estimated as

$$\hat{\tau} = \underset{k \in \{s_{i^*}+h-1, \ldots, e_{i^*}-h\}}{\arg\max} T_{n,\delta}(s_{i^*}, k, e_{i^*}).$$

Once a changepoint $\hat{\tau}$ is identified, SN-NOT then divides the time series into two subsamples, $\{Y_t\}_{t=1}^{\hat{\tau}}$ and $\{Y_t\}_{t=\hat{\tau}+1}^{n}$, and the same procedure is applied to both of them. In each recursive step, previously drawn intervals could be reused, provided that they fall within the subsample considered. The process is implemented recursively until no more changepoints can be detected. The main reason that the intervals of time points are drawn randomly is to avoid making a choice of a particular fixed design.

Algorithm 1 SN-NOT

Input: Data $\{Y_t\}_{t=1}^n$, threshold ζ_n, trimming size $d = \lfloor \delta n \rfloor$ and $h = \lfloor \varepsilon n \rfloor$, random intervals F_n^M.
Output: Estimated number of changepoints $\hat{m}$ and their estimated locations $\hat{\boldsymbol{\tau}}$.
Initialisation: SN-NOT$(1, n, \zeta_n)$
Procedure: SN-NOT(s, e, ζ_n), $\hat{m} = 0$, $\hat{\boldsymbol{\tau}} = \emptyset$
if $e - s + 1 < 2h$ **then**
  Stop;
else
  $\mathcal{M}_{(s,e)} := \left\{ i : [s_i, e_i] \in F_n^M, [s_i, e_i] \subset [s, e], e_i - s_i + 1 \geq 2h \right\}$
  if $\mathcal{M}_{(s,e)} = \emptyset$ **then**
    Stop;
  else
    $\mathcal{O}_{(s,e)} := \left\{ i \in \mathcal{M}_{(s,e)} : G_{n,\delta}(s_i, e_i) > \zeta_n \right\}$
    if $\mathcal{O}_{(s,e)} = \emptyset$ **then**
      Stop;
    else
      $i^* = \arg\min_{i \in \mathcal{O}_{(s,e)}} |e_i - s_i + 1|$
      $\tau^* = \arg\max_{k \in \{s_{i^*}+h-1, \ldots, e_{i^*}-h\}} T_{n,\delta}(s_{i^*}, k, e_{i^*})$
      $\hat{\boldsymbol{\tau}} = \hat{\boldsymbol{\tau}} \cup \tau^*$
      $\hat{m} = \hat{m} + 1$
      SN-NOT(s, τ^*, ζ_n)
      SN-NOT$(\tau^* + 1, e, \zeta_n)$
    end
  end
end

Even though a deterministic scheme can be used to draw intervals such as in Rufibach and Walther (2010), according to Fryzlewicz (2014) and Baranowski et al. (2019) the difference in performance between the random and deterministic designs is likely to be minimal when considering a very large M. Jiang et al. (2020) have proposed selecting the threshold ζ_n as follows.

1. Generate B sequences of i.i.d $N(0,1)$ random variables $\{\epsilon_t^b\}_{t=1}^n, b = 1, ..., B$;
2. For every b^{th} sample calculate $\zeta_n^b = \max_{i=1,\ldots,M} G_{n,\delta}(s_i, e_i)$;
3. The threshold ζ_n is set as the $(1-\alpha)$ sample quantile of $\{\zeta_n^b\}_{b=1}^B$.

Repeated numerical runs reveal that for fixed ε, the computational cost for the threshold search is $O(BMn)$. The value of δ has no significant impact on the computational time. Since the SN test statistic is asymptotically pivotal, this threshold is expected to well approximate the $(1-\alpha)$ quantile of the finite sample distribution of the maximum SN test statistic on the M random intervals under the null hypothesis. The pseudo-code of SN-NOT, also found in Jiang et al. (2020), is given in Algorithm 1. In this algorithm, the set $\mathcal{M}_{(s,e)}$ contains the intervals considered in the current step and

the set $\mathcal{O}_{(s,e)}$ contains the intervals in which the SN statistic exceed the threshold in the current step. According to Baranowski et al. (2019), the benefits of the NOT algorithm, other than its applicability on a variety of changepoint models, include the simplicity of its implementation, the fact that it yields optimal rates of convergence for the estimators of changepoint locations, and that it has a linear computational cost of $O(Mn)$. The NOT algorithm is a variant of a popular multiple changepoint detection method known as the binary segmentation proposed by Vostrikova (1981). In this method, the single changepoint detection is applied to the whole time series, and if a changepoint is detected, the time series is split into two subsamples and the single changepoint detection is applied to both of them. This procedure goes on until there are no changepoints left to be detected. However, Fryzlewicz (2014) showed that the binary segmentation is only consistent when the spacing between any two consecutive changepoints is of order greater than $n^{3/4}$. A variant of the binary segmentation is the wild binary segmentation proposed by Fryzlewicz (2014) which also applies single changepoint detection in random subsamples of the time series, but it does not focus on the narrowest interval and hence it is not suitable for models which are not piecewise stationary.

The description of the SN-NOT Algorithm to detect multiple changepoints in the linear trend model in (1) has been provided, and this can be extended to detect multiple changepoints in a polynomial structure. However, Jiang et al. (2020) applied the SN-NOT for the piecewise quadratic trend model as the COVID-19 log-scale cumulative daily cases and deaths in eight major countries to compare their changepoints to the ones in the piecewise linear trend model. They observed that the estimated changepoints in the piecewise quadratic trend model are not well associated with the health measures, as the quadratic function may have absorbed the intervention effects, thus the piecewise linear trend model is preferred. Thus, only the piecewise linear trend model will be considered. In the Section 3, this method is applied to the epidemiological spread and infection deaths in Malta, simultaneously using COVID-19 related news reports to determine the cause.

Since it is observed that the algorithm results may vary according to the random intervals generated, an ensemble approach shall be taken when applying the changepoint model, rather than the single model approach implemented by Jiang et al. (2020). The ensemble approach consists of taking n_{sim} initial seeds which yield a set of changepoints, and the different sets of changepoints are aligned in time. Only aligned changepoints which occur in at least 50% of the runs are retained in the final sample of changepoints. For the retained aligned changepoints, a 90% confidence interval is obtained by extracting from the sample the 0.05- and 0.95-quantiles of valid runs (we shall have n_{sim} valid runs for aligned changepoints with 100% occurrence, or less when this is not the case), unless all runs have yielded one single date. In Section 3, Tables 3 and 4 display these date intervals. Moreover, the changepoint yielded by a single run of the SN-NOT algorithm determines the end of a segment, and since each run may yield different segments, we give the collective result of the ensemble by presenting median start and end date of segments, obtained from the sample of n_{sim} runs (or less). In Section 3, refer to Figure 2, and Tables 1 and 2 for where median changepoints are used. Finally, it must be noted that the computational cost of the ensemble approach is of $O(n_{sim}Mn)$, which means that the computational cost increases linearly with each

run.

3. Results

In this section, the multiple changepoint detection approach described in Section 2 is applied to the log-scale of both the daily cumulative cases and the cumulative deaths of COVID-19 in Malta. Since the SN-NOT algorithm's results may be influenced by the random intervals generated, the algorithm is run $n_{sim} = 150$ times using randomly selected seeds, and an ensemble result is given. A 90% confidence interval for each changepoint is provided and, furthermore, residual diagnostics are based on the median changepoint. For both time series, the estimated slope is compared from one segment to the next, and speculation is made on which health measures, significant events and other factors may have caused these changepoints. The time series for the log scale of cumulative daily cases of COVID-19 in Malta ranges from 7/3/2020 till 31/1/2021 for a total of $n = 331$ days. On the other hand, the time series considered for cumulative deaths ranges from 30/8/2020 till 31/1/2021 for a total of $n = 155$ days. The reason for starting the cumulative deaths time series at a later date was due to the fact that, prior to 30/8/2020 there was not a significant number of deaths from COVID-19 (only nine up till that point); this created computational issues resulting from matrix singularities. The reason for considering both time series up to 31/1/2021 is due to the fact that beyond this date, the gradual effect of vaccination may make the changepoint model less suitable for the analysis. Both time series were obtained from the European Centre for Disease Prevention and Control website [2] up until 13/12/2020 when the data used to be updated daily. After that, daily data was collected from the Maltese Ministry for Health's infographics on the *Saħħa* Facebook page. Malta consistently had a very robust testing regime, and particularly in the first wave, Malta had a very high testing capacity. According to an article published in the Times of Malta on April 14th, 2020, Malta had the third highest testing capacity in the world (38.74 per 1000 people) after Iceland and Luxembourg. [3] Furthermore, Health Minister Christopher Fearne confirmed with Maltese media that the Maltese health authorities follow WHO guidelines in the reporting of deaths, stating that these would be deceased individuals who were COVID-19 positive at the time of death.[4]

Figure 2 shows the plots of the log-scale of the cumulative cases and log-scale of the cumulative deaths respectively, including the median changepoints obtained from multiple runs of the SN-NOT algorithm explained in Section 2. For the log-scale of cumulative cases, the trimming parameters chosen for the algorithm were taken to be $(\varepsilon, \delta) = (0.045, 0.005)$, and for the log-scale of cumulative deaths, these were taken to be $(\varepsilon, \delta) = (0.095, 0.01)$. Since $h = \lfloor \varepsilon n \rfloor$, the value of ε needs to be chosen in such a way that the desired minimum segment length is satisfied. The choice for h to be equal to 14 days has been based on the maximum reported incubation period of COVID-19. On the other hand, δ must satisfy $\lfloor \delta n \rfloor \geq 1$, but it was also observed that smaller values

[2] https://www.ecdc.europa.eu [Accessed on 25th February 2021]

[3] https://timesofmalta.com/articles/view/malta-covid-19-testing-rate-is-among-top-three-in-the-world-data.785597 [Accessed on 25th April 2022]

[4] https://lovinmalta.com/news/patients-who-tested-covid-19-positive-at-time-of-death-included-in-maltas-total-count/ [Accessed on 25th April 2022]

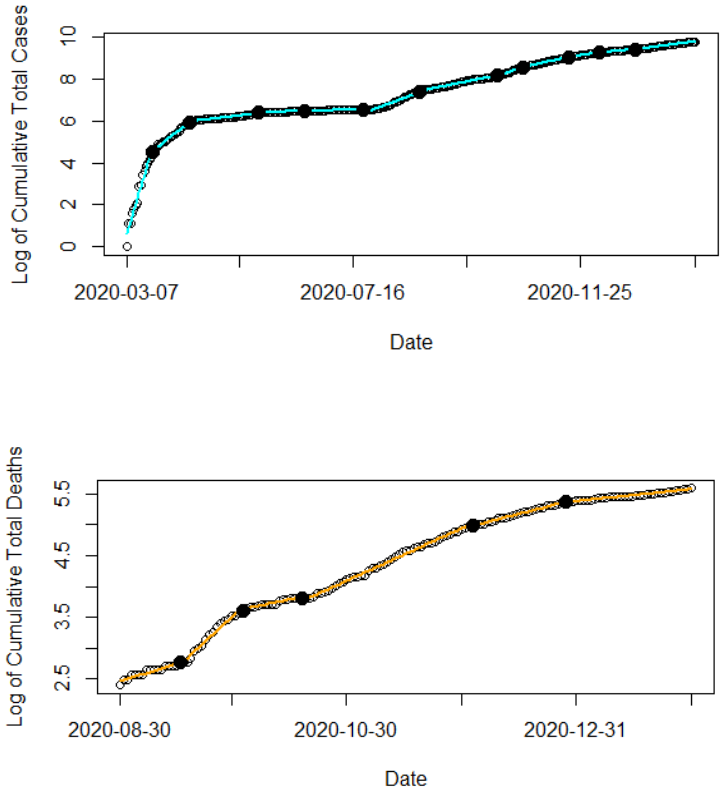

Figure 2: Log-scale of Cumulative Daily Cases of COVID-19 in Malta from 7/3/2020 to 31/1/2021 (top) and Log-scale of Cumulative Daily Deaths of COVID-19 in Malta from 30/8/2020 to 31/1/2021 (bottom) with the median changepoints represented by large black dots.

of δ (satisfying this constraint) yielded less variability in the number of changepoints between different runs of the algorithm using different seeds. Furthermore, there is no added benefit to using smaller values of δ which yield the same value for $\lfloor \delta n \rfloor$, so a value of δ such that $\lfloor \delta n \rfloor = 1$ is chosen arbitrarily. Due to the computational intensiveness of the algorithm, a thorough sensitivity analysis on ε and δ was not conducted, but it can be an avenue for possible future research regarding this algorithm. The threshold value for the SN-NOT algorithm was found by generating $B = 1000$ samples of i.i.d. standard normally distributed values and $M = 5000$ intervals from 1 to 331 (for the log-scale of cumulative cases), and 1 to 155 (for log-scale of cumulative deaths), calculating $\zeta_n^b = \max_{i=1,\ldots,M} G_{n,\delta}(s_i, e_i)$ for every $b = 1, ..., B$, and finding the 95% sample quantile of $\{\zeta_n^b\}_{b=1}^B$, as described in Section 2. In both cases, the OLS estimator was used to find the intercept and slope of each segment, where the same M is maintained. Given the aforementioned settings for both algorithms, and using an 11th Gen Intel (R) COR(TM) i9-11900K 3.5GHz processor with 64GB of RAM, the threshold search for the cumulative cases time series is expected to take (on average) 35.4 hours while the threshold search for the cumulative deaths time series is expected to take 16.6 hours. On the other hand, a single run of the SN-NOT algorithm for the cumulative cases

Table 1: Median start and end date, length in days, intercept and slope estimates, and percentage slope change of the log-scale of the daily cumulative cases of COVID-19 in Malta for each segment.

i	Start of Segment (Median)	End of Segment (Median)	Days	$\hat{\beta}_0^{(i)}$	$\hat{\beta}_1^{(i)}/n$	% Change
1	7/3/2020	22/3/2020	16	0.3198	0.2805	NA
2	23/3/2020	12/4/2020	21	3.5760	0.0638	-77%
3	13/4/2020	22/5/2020	40	5.6252	0.0094	-85%
4	23/5/2020	18/6/2020	27	6.1528	0.0032	-66%
5	19/6/2020	22/7/2020	34	6.4414	0.0005	-83%
6	23/7/2020	24/8/2020	33	2.0180	0.0314	5652%
7	25/8/2020	8/10/2020	45	4.6008	0.0165	-47%
8	9/10/2020	23/10/2020	15	2.0025	0.0284	72%
9	24/10/2020	18/11/2020	26	4.3393	0.0183	-36%
10	19/11/2020	6/12/2020	18	5.9049	0.0122	-33%
11	7/12/2020	27/12/2020	21	7.2625	0.0073	-40%
12	28/12/2020	31/1/2021	35	6.1194	0.0111	53%

time series is expected to take 298 seconds while a single run of the SN-NOT algorithm for the cumulative deaths time series is expected to take 140 seconds.

Table 2: Median start and end date, length in days, intercept and slope estimates, and percentage slope change of the log-scale of the daily cumulative deaths of COVID-19 in Malta for each segment.

i	Start of Segment (Median)	End of Segment (Median)	Days	$\hat{\beta}_0^{(i)}$	$\hat{\beta}_1^{(i)}/n$	% Change
1	30/08/2020	15/09/2020	17	2.4467	0.0185	NA
2	16/09/2020	02/10/2020	17	1.7628	0.0562	204%
3	03/10/2020	18/10/2020	16	3.2348	0.0118	-79%
4	19/10/2020	03/12/2020	46	2.3992	0.0273	132%
5	04/12/2020	28/12/2020	25	3.4915	0.0155	-43%
6	29/12/2020	31/01/2021	34	4.6096	0.0062	-60%

Tables 1 and 2 are a summary of the segments in between median changepoints for the log-scale of cumulative cases time series and the log-scale of cumulative deaths time series respectively. They detail the median start/end date of each segment, together with the length (in days), the estimated intercept and slope parameters (up to four decimal places), and the percentage change in the slope from the previous segment. On the other hand, Tables 3 and 4 give a summary of the potential causes of changepoints for the log-scale of cumulative cases time series and the log-scale of cumulative deaths time series respectively. As potential causes, we identify interventions and events that have happened up to 14 days prior to the changepoint interval for the log-scale of cumulative cases, and interventions and events that have happened up to five weeks

prior to the changepoint interval for the log-scale of cumulative deaths. The former is based on the maximum incubation time of COVID-19, and the latter takes into account the incubation time and the aforementioned median time from onset of illness till death. A marker also indicates whether the change led to an increase or a decrease in the slope. The causes, as much as possible, are also provided with a date. These dates may refer to the date of the intervention, or the date when a particular potential cause was reported. Information on this was obtained from the Times of Malta website [5]. Extra detail to the information given in Tables 3 and 4 in the subsequent discussion shall be added where relevant.

Potential causes cannot always be linked to specific government interventions or significant events, with specific dates. For example, Malta has always had a very robust testing regime with one of the highest per capita rates in the world - currently 18th worldwide [6]. This, coupled with rigorous contact tracing and the closure of ports and airports (except for repatriation flights), may have potentially been effective in keeping the cases low and reducing the infection rate when the numbers were low. This appears to have been the cause for the reduction in the slopes for changepoints 3 and 4. The reduction occurred despite the fact that segment 3 saw the resumption of non-essential travel between Malta and Gozo and the reopening of some non-essential shops (4/5/2020), and segment 4 saw the the reopening of restaurants, hotels, hairdressers, barbers, beauticians, outdoor sports facilities, bars, gyms, child care centres and clubs (between 22/5/2020 and 5/6/2020). In addition, in between these dates, the limit of crowds for outdoor social gatherings was increased and funeral ceremonies with restricted numbers were allowed. The same cannot be said for when the numbers were high during the pandemic, which saw contact tracing being much less effective due to the exponential increase in potentially infected people.

Furthermore, changepoints can sometimes also be attributed to the population's collective behaviour - in particular, increases in the exponential growth rate can possibly be a consequence of collective violation of COVID-19 health recommendations at the time due to pandemic fatigue. One such instance in particular can be attributed to the population's behaviour related to the Christmas period in segment 11 of Table 1 - since no additional restrictions were made by the authorities (though health recommendations were given), this led to retail establishments in Malta being consistently crowded, and families and friends attending gatherings during this period. Thus, changepoint 11 (on 27/12/2020) in Table 3 led to considerable increase in the exponential growth. However, this did not coincide with a change in the exponential growth of deaths in the following weeks. Indeed, one can see a negative change in exponential growth of deaths on 29/12/2020 due to the lack of apparent new restrictive measures, however this also coincides with the commencement of the vaccine rollout on 27/12/2020 to the highest elderly bracket and health care workers, so efforts by health authorities to keep the case count low in the most vulnerable segment of the population at this time may have been affected.

Finally, one also needs to check if the residuals of the fitted piecewise linear trend model satisfy assumption of weak dependence. Figure 3 show the ACFs and PACFs for the residuals of the piecewise linear models for both the log-scale of cumulative

[5] https://www.timesofmalta.com [Accessed on 18th June 2021]
[6] https://www.worldometers.info/coronavirus/[Accessed on 18th June 2021]

Table 3: List of changepoints with 90% confidence interval for date of occurrence and potential causes (dated) for the log-scale of the cumulative daily cases of COVID-19 in Malta. The arrow next to the changepoint number indicates whether the changepoint led to an increase or a decrease in the slope. Changepoint marked with † occurred in 76.7% of the runs. Changepoints with less than 50% occurrence are not considered.

Changepoint (90% C. I.)	**Potential Causes**
1 ↓ (21/3/2020 - 23/3/2020)	· banning of various religious, cultural and crowded outdoor events (10/3/2020-14/3/2020) · suspension of international connections/closure of ports (10/3/2020-21/3/2020) · banning of visiting hours at elderly homes (12/3/2020) · imposition of fine for people breaking quarantine rules (16/3/2020)
2 ↓ (10/4/2020 - 18/4/2020)	· public gathering restrictions (28/3/2020) · banning of non-essential travel to Gozo (3/4/2020) · robust testing regime (one of the highest per capita worldwide)
3† ↓ (20/5/2020 - 24/5/2020)	· no new restrictive measures within 14 days prior to changepoint · robust testing regime with low case count/continued closure of ports
4 ↓ (16/6/2020 - 19/6/2020)	· no new restrictive measures within 14 days prior to changepoint · robust testing regime with low case count/continued closure of ports
5 ↑ (20/7/2020 - 25/7/2020)	· all existing Covid-19 preventative measures lifted (including reopening of ports and resuming of international connections) (1/7/2020) · superspreader events (pool party, religious festival) (17/7/2020-26/7/2020)
6 ↓ (19/8/2020 - 25/8/2020)	· restrictions on crowd size in indoor/outdoor events · restrictions on hospital/elderly home visits · banning of dance floor use and boat parties · enforcement of mask use in buses, shops, enclosed public areas · seated wedding receptions only allowed · number limit on crowds in public areas (all above measures introduced 7/8/2020-17/8/2020) · introduction of amber list for international arrivals (21/8/2020)
7 ↑ (8/10/2020)	· re-opening of the majority of schools in the last week of September
8 ↓ (23/10/2020)	· wearing of masks made mandatory in all public areas (16/10/2020)
9 ↓ (18/11/2020)	· no new restrictive measures within 14 days prior to changepoint
10 ↓ (6/12/2020)	· no new restrictive measures within 14 days prior to changepoint
11 ↑ (27/12/2020)	· no particular event or government policy within 14 days can be attributed to changepoint, but lack of additional restrictions during Christmas season may have been the cause

daily cases and the log-scale of cumulative daily deaths of COVID-19. One can see that in both cases, the autocorrelation coefficients and partial autocorrelation coefficients of the residuals for both models decay quickly to zero. Furthemore, the best model for the residuals of the log-scale of cumulative daily cases model was found to

Table 4: List of changepoints with their dates and potential causes (dated) for the log-scale of the cumulative daily deaths of COVID-19 in Malta. The arrow next to the changepoint number indicates whether the changepoint led to an increase or a decrease in the slope.

Changepoint (90% C. I.)	**Potential Causes**
1 ↑ (13/9/2020 - 18/9/2020)	· elderly homes clusters (identified 18/8/2020)
2 ↓ (28/9/2020 - 4/10/2020)	· updated safety guidelines for elderly homes (17/9/2020)
3 ↑ (18/10/2020 - 21/10/2020)	· no particular event/government policy from previous days can be attributed to changepoint · increase potentially due to lack of restrictions on family gatherings due evidence of family clusters (reported on Times of Malta, 16/9/2020)
4 ↓ (03/12/2020)	· wearing of masks made mandatory in all public areas (16/10/2020) · introduction of rapid testing (28/10/2020) · all bars closed (29/10/2020)
5 ↓ (29/12/2020)	· no new restrictive measures within 5 weeks prior to changepoint

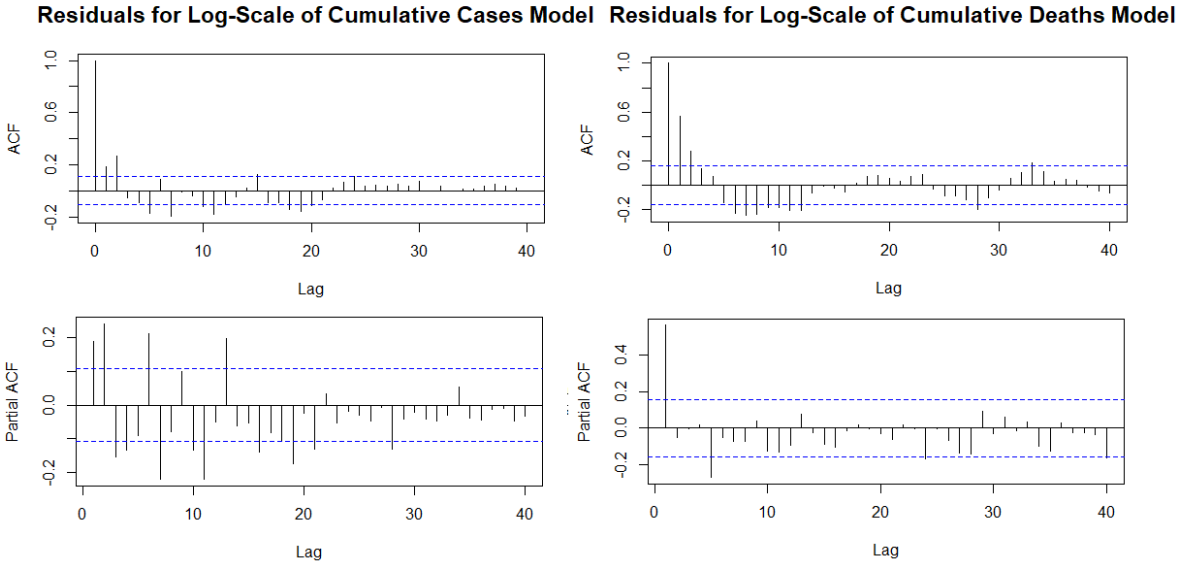

Figure 3: ACFs and PACFs of Residuals of Piecewise Linear Models for Log-scale of Cumulative Daily Cases (left) and Log-scale of Cumulative Daily Deaths (right) of COVID-19.

be an $ARMA\,(4,1)$ process with autoregressive polynomial $1 - 0.2716z - 0.5432z^2 + 0.4037z^3 + 0.1111z^4$ and moving average polynomial $1 - 0.124z$, for which the roots of

both polynomials are found to lie well outside the unit circle. Also, the best model for the residuals of the log-scale of cumulative daily deaths model was found to be an $ARMA\,(2,2)$ process with autoregressive polynomial $1+0.4394z-0.1651z^2$ and moving average polynomial $1+1.14z+0.4689z^2$, for which the roots of both polynomials in this case are also found to lie well outside the unit circle. This means that there is evidence of weak dependence of the residuals in both models, and consequently one can deduce that both changepoint models are reliable.

4. Discussion

In this paper, the SN-NOT changepoint model has been applied to the log-scale of the cumulative cases and deaths of COVID-19 in Malta. The changepoint model has been reasonably successful at detecting changepoints in the log-scale of the cumulative cases up to two weeks after certain events or interventions related to COVID-19 have occurred, even though exogenous information was not used in the model itself. Similarly, it can also be observed how certain events and interventions may have contributed to the increase or decrease of the rate of the log-scale of the cumulative deaths in subsequent weeks. These results may be used to learn which measures, events and population behaviour may lead to a change in the epidemiological dynamic of cases and deaths. The advent of vaccines has certainly provided us with different tools of dealing with the pandemic, however this study may still provide insight to the transmission of future respiratory infectious diseases. The results indicate that, during periods of low case counts, rigorous testing and effective border control have kept the pandemic in check. In fact, careful relaxation of health measures, such as those between May and June 2020, have not led to spikes in numbers. On the contrary, crowded events, the unrestricted reopening of Malta International Airport to certain corridor countries, and collective pandemic fatigue/behavioural changes within the population have potentially triggered spikes of new cases, and in some cases also deaths. The findings corroborate the conclusions in the studies by Jiang et al. (2020) and Cuschieri et al. (2021). In the former, it is postulated that restriction of movement, and a robust and well-prepared health system, were crucial in curtailing the spread of the pandemic. In the latter, on the other hand, we have mass events being mentioned as the trigger to significant spikes. These all form part of the conclusions we have reached in the changepoint analysis implemented in this paper. Other aspects that appear to have had an impact on the exponential growth rate are the re-opening of schools (which was followed by an increase in the exponential growth rate) and the wearing of masks in public places (which was followed by a decrease in the exponential growth rate). Focusing specifically on deaths, the highest contributions to spikes in deaths appeared to arise from elderly home and family clusters, while safety guidelines on elderly homes appear to have resulted in bringing these back down.

As mentioned earlier, the date range considered was 7/3/2020 till 31/1/2021, so phenomena such as the advent of variants of concern and their impact, the effectiveness of the vaccine rollout on the general population (with Malta having one of the highest vaccination rates worldwide [7]), and the implementation of some subsequent restrictive

[7] https://ourworldindata.org/[Accessed on 18th June 2021]

measures to curtail spikes in cases have not been studied here. Although it would be interesting to extend the methodology to assess the presence of changepoints in the pandemic in Malta also throughout the first half of 2021, with the advent of vaccinations - which will have had a more gradual and long-term effect on the course of the pandemic - it may prove to be more challenging to attribute certain changepoints to specific events.

Computational Details

The results in this paper were obtained using R 3.5.1. R itself and all packages used are available from the Comprehensive R Archive Network (CRAN) at https://CRAN.R-project.org/. The data used can be found on: https://github.com/gurs0001/Maltese-COVID-19-Dataset.

Acknowledgements

The authors would like to extend their gratitude to the authors of Jiang et al. (2020) for sharing with us the code, which we have adapted in this study.

References

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4):821–856, DOI: 10.2307/2951764.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47–78, DOI: 10.2307/2998540.

Baker, S. T., Leslie, A. M., Gallistel, C. R., and Hood, B. M. (2016). Bayesian change-point analysis reveals developmental change in a classic theory of mind task. *Cognitive Psychology*, 91:124–149, DOI: 10.1016/j.cogpsych.2016.08.001.

Baranowski, R., Chen, Y., and Fryzlewicz, P. (2019). Narrowest-over-threshold detection of multiple changepoints and changepoint-like features. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 81(3):649–672, DOI: 10.1111/rssb.12322.

Chen, J. and Gupta, A. K. (2012). *Parametric Statistical Change Point Analysis: With Applications to Genetics, Medecine, and Finance*. Birkhäuser, Boston, 2nd edition, DOI: 10.1007/978-0-8176-4801-5.

Chu, J. (2020). A statistical analysis of the novel coronavirus (COVID-19) in Italy and Spain. *Plos One*, 16(3), DOI: 10.1371/journal.pone.0249037.g001.

Cuschieri, S., Balzan, M., Gauci, C., Agius, S., and Grech, V. (2021). Mass events trigger Malta's second peak after initial successful pandemic suppresion. *Journal of Community Health*, 46:618–625, DOI: 10.1007/s10900-020-00925-6.

Cuschieri, S. and Grech, S. (2021). At-risk population for COVID-19: Multimorbidity characteristics of a European small island state. *Public Health*, 192:33–36, DOI: 10.1016/j.puhe.2020.12.012.

Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., Whittaker, C., Zhu, H., Berah, T., Eaton, J. W., Monod, M., Ghani, A. C., Donnelly, C. A., Riley, S., Michaela A. C. Vollmer, N. M. F., Okell, L. C., and Bhatt, S. (2020). Estimating the effects of non-pharmaceutrical interventions on COVID-19 in Europe. *Nature*, 584:257–261, DOI: 10.1038/s41586-020-2405-7.

Fryzlewicz, P. (2014). Wild binary segmentation for multiple changepoint detection. *Annals of Statistics*, 42(6):2243–2281, DOI: 10.1214/14-AOS1245.

Hawkins, D. L. (1987). A test for a changepoint in a parametric model based on a maximal wald-type statistic. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, 49(3):368–376.

Haynes, K., Eckley, I. A., and Fearnhead, P. (2016). A computationally efficient nonparametric approach for changepoint detection. *Statistics and Computing*, 27(5):1293–1305, DOI: 10.1007/s11222-016-9687-5.

Haynes, K., Eckley, I. A., and Fearnhead, P. (2017). Computationally efficient changepoint detection for a range of penalties. *Journal of Computational and Graphical Statistics*, 26(1):134–143, DOI: 10.1080/10618600.2015.1116445.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., Xiao, Y., Gao, H., Guo, L., Xie, J., Wang, G., Jiang, R., Gao, Z., Jin, Q., Wang, J., and Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*, 395(10223):497–506, DOI: 10.1016/S0140-6736(20)30183-5.

Jandhyala, V., Fotopoulos, S., MacNeill, I., and Liu, P. (2013). Inference for single and multiple changepoints in time series. *Journal of Time Series Analysis*, 34(4):423–446, DOI: 10.1111/jtsa.12035.

Jiang, F., Zhao, Z., and Shao, X. (2020). Time series analysis of COVID-19 infection curve: a changepoint perspective. *Journal of Econometrics*, In Press, DOI: 10.1016/j.jeconom.2020.07.039.

Killick, R., Fearnhead, P., and Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107:1590–1598, DOI: 10.1080/02621459.2012.737745.

Li, Y., Campbell, H., Kulkarni, D., Harpur, A., Nundy, M., Wang, X., and Nair, H. (2020). The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varing reproduction number (r) of sars-cov-2: a modelling study across 131 countries. *Lancet Infectious Diseases*, 21(2):193–202, DOI: 10.1016/S1473-3099(20)30785-4.

Locatelli, I., Trachsel, B., and Rousson, V. (2021). Estimating the basic reproduction number for COVID-19 in Western Europe. *PLoS ONE*, 16(3):e0248731, DOI: 10.s1371/journal.pone.0248731.

Ma, T. F. and Yau, C. Y. (2016). A pairwise likelihood-based approach for changepoint detection in multivariate time series models. *Biometrika*, 103(2):409–421, DOI: 10.1093/biomet/asw002.

Matteson, D. S. and James, N. A. (2014). A nonparametric approach for multiple changepoint analysis of multivariate data. *Journal of the American Statistical Association*, 109:334–345, DOI: 10.1080/01621459.2013.849605.

Nishura, H., Linton, N. M., and Akhmetzhanov, A. R. (2020). Initial cluster of novel coronavirus (2019-ncov) infections in Wuhan, China is consistent with substantial human-to-human transmission. *Journal of Clinical Medicine*, 9(2):488, DOI: 10.3390/jcm9020488.

Ogundokun, R. O., Lukman, A. F., Kibria, G. B. M., Awotunde, J. B., and Aladeitan, B. B. (2020). Predictive modelling of COVID-19 confirmed cases in Nigeria. *Infectious Disease Modelling*, 5:543–548, DOI: 10.1016/j.idm.2020.08.003.

Page, E. (1954). Continuous inspection schemes. *Biometrika*, 41:100–115, DOI: 10.1093/biomet/41.1-2.100.

Pan, A., Liu, L., Wang, C., Guo, H., Hao, X., Wang, Q., Huang, J., He, N., Yu, H., Lin, X., Wei, S., and Wu, T. (2020). Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China. *JAMA*, 323(19):1915–1923, DOI: 10.1001/jama.2020.6130.

Picard, F., Robin, S., Lavielle, M., Vaisse, C., and Daudin, J.-J. (2005). A statistical approach for array CGH data analysis. *BMC Bioinformatics*, 6(27):100–115, DOI: 10.1186/1471-2015-6-27.

Rufibach, K. and Walther, G. (2010). The block criterion for multiscale inference about a density, with applications to other multiscale problems. *Journal of Computational and Graphical Statistics*, 19(1):175–190, DOI: 10.1198/jcgs.2009.07071.

Sahafizadeh, E. and Sartoli, S. (2020). Epidemic curve and reproduction number of COVID-19 in Iran. *Journal of Travel Medicine*, 27(5), DOI: 10.1101/2020.03.20.20038422.

Sebastiani, G., Massa, M., and Riboli, E. (2020). COVID-19 epidemic in Italy: Evolution, projections and impact of government measures. *European Journal of Epidemiology*, 35(4):341–345, DOI: 10.7326/m20-0504.

Shao, X. (2010). A self-normalized approach to confidence interval construction in time series. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 72(3):343–366, DOI: 10.1111/j.1467-9868.2009.00737.x.

Thies, S. and Molnár, P. (2018). Bayesian changepoint analysis of Bitcoin returns. *Finance Research Letters*, 27:223–227, DOI: 10.1016/j.frl.2018.03.

Tsang, T. K., Wu, P., Lin, Y., Lau, E. H. Y., Leung, G. M., and Cowling, B. J. (2020). Effect of changing case definitions for covid-10 on the epidemic curve and transmission parameters in mainland China: a modelling study. *The Lancet Public Health*, 5(5):e289–e296, DOI: 10.1016/S2468-2667(20)30089-X.

Vostrikova, L. Y. (1981). Detecting disorder in multidimensional random processes. *Soviet Mathematics Doklady*, 24:55–59.

Wang, P., Zheng, X., Li, J., and Zhu, B. (2020). Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics. *Chaos, Solitons and Fractals*, 139:110058, DOI: 10.1016/j.chaos.2020.110058.

Yu, X., Duan, J., Jiang, Y., and Zhang, H. (2020). Distinctive trajectories of the COVID-19 epidemic by age and gender: a retrospective modeling of the epidemic in South Korea. *International Journal of Infectious Diseases*, 98:200–205, DOI: 10.1016/j.ijid.2020.06.101.

Zhan, C., Tse, C. K., Fu, Y., Lai, Z., and Zhang, H. (2020). Modeling and prediction of the 2019 coronavirus disease spreading in China incorporating human migration data. *Plos One*, 15(10), DOI: 10.1371/journal.pone.0241171.

Zhao, S., Lin, Q., Ran, J., Musa, S. S., Yang, G., Wang, W., Lou, Y., Gao, D., Yang, L., He, D., and Wang, M. H. (2020). Preliminary estimation of the basic reproduction number of novel coronavirus (2019-ncov) in China from 2019 to 2020: a data-driven analysis on the early phase of the outbreak. *International Journal of Infectious Diseases*, 92:214–217, DOI: 10.1016/j.ijid.2020.01.050.

Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., Gu, X., Guan, L., Wei, Y., Li, H., Wu, X., Xu, J., Tu, S., Zhang, Y., Chen, H., and Cao, B. (2020). Clinical course and risk factors for mortality of adult inpatients with covid-19 in wuhan, china: a retrospective cohort study. *Lancet*, 395(10229):1054–1062, DOI: 10.1016/S0140-6736(20)30566-3.

Affiliation:

Gianluca Ursino
Department of Statistics and Operations Research
Faculty of Science
University of Malta
Room 507 Mathematics and Physics Building
Msida MSD2080, Malta
E-mail: gianluca.ursino.16@um.edu.mt

David Suda
Department of Statistics and Operations Research
Faculty of Science
University of Malta
Room 511 Mathematics and Physics Building
Msida MSD2080, Malta
E-mail: david.suda@um.edu.mt
URL: https://www.um.edu.mt/profile/davidsuda

Monique Borg Inguanez
Department of Statistics and Operations Research
Faculty of Science
University of Malta
Room 514 Mathematics and Physics Building
Msida MSD2080, Malta
E-mail: monique.inguanez@um.edu.mt
URL: https://www.um.edu.mt/profile/moniqueinguanez

Journal of Data Science, Statistics, and Visualisation https://jdssv.org/
published by the International Association for Statistical Computing
http://iasc-isi.org/

ISSN 2773-0689
November 2022, Volume II, Issue VII *Submitted:* 2021-06-29
doi:10.52933/jdssv.v2i7.50 *Accepted:* 2022-09-29