~∿~ **Signa Vitae**

ORIGINAL RESEARCH

A statistical analysis for the epidemiological surveillance of COVID-19 in Chile

Nixon Jerez-Lillo[1], Bernardo Lagos Álvarez[1], Joel Muñoz Gutiérrez[1],
Jorge I. Figueroa-Zúñiga[1], Víctor Leiva[2],*

[1]Department of Statistics, Universidad de Concepción, 3349001 Concepción, Bio Bio, Chile
[2]School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, 2362807 Valparaíso, Chile

*Correspondence
victor.leiva@pucv.cl;
victorleivasanchez@gmail.com
(Víctor Leiva)

Abstract

The emergence of COVID-19 so far and in the immediate future has brought significant uncertainties that negatively impact institutions and individuals in developing and planning their activities worldwide. The uncertainty of the effectiveness of vaccines has forced the authorities to adopt different protocols, the most relevant of which is the isolation of people through quarantine to avoid contagion, drastically affecting our way of life. For this reason, it is crucial to evaluate the effectiveness of quarantines. In this paper, we analyze the spread of the disease in Chile according to the quarantines decreed by the sanitary authority. An inferential study is used to estimate the trend changes in COVID-19 cases and their basic and instantaneous reproduction numbers, which allows us to evaluate the decreed measures and establish vaccination policies. According to the data obtained until 03 March 2021 of confirmed COVID-19 cases disaggregated at a regional level in Chile, we observe a heterogeneous spread in most Chilean regions. When incorporating the dynamic quarantines decreed, effectiveness is detected in most regions, except in a few of them. Our results indicate that we are unable to identify the measures in the step-by-step protocols partly responsible for non-compliance with quarantines. However, our specific findings that can be extrapolated to daily practice and enlighten the ways of other countries are as follows. On the one hand, a measure that has been effective in curbing the spread of the disease is the strict early quarantine as detected in some Chilean regions. Therefore, indexes are needed to measure the mobility of citizens. On the other hand, as time passes without stopping infections, quarantines lose effectiveness even if the estimated instantaneous reproduction number is negligible and stable. In addition, other factors can cause this number to not be within the expected ranges, which must be further studied. Also, we have estimated the basic reproduction number whose value confirms the suitability of the pandemic declaration.

Keywords

Basic and instantaneous reproduction numbers; Data science; PCR; R software; SARS-CoV-2; Time-series

1. Introduction

The COVID-19 pandemic, associated with the respiratory syndrome coronavirus 2 (SARS-CoV-2), has undoubtedly been one of the most extreme adverse events of the 21st century. The course of the disease, discovered on 17 November 2019, in Wuhan, China, threatens the lives of the entire world's population and has forced us into a new way of life that will leave its mark on society forever [1–5].

To confirm suspected cases of COVID-19, the polymerase chain reaction (PCR) test is often applied, based on the detection of a certain amount of genetic fragments other than the virus in an individual [6]. Other alternative diagnostic tools for detecting COVID-19 cases have been recently proposed [7].

Most countries have reported the confirmed cases (positive PCR tests) daily since the first day, which makes it possible to account for the disease incidence. However, this is underestimated due to the difficulties that arise in a health problem of this magnitude, ranging from the patient's refusal to attend a health system that may eventually collapse to the limitations of this system itself that does not allow us to provide care to all its patients [8].

The relevance of this health event, and its severe consequences, requires the government authorities to take an active role in controlling and preventing the spread of the disease. Protocols established by the World Health Organization are applied, such as containment measures, isolation of infected patients through mandatory quarantines, and mitigation measures. These protocols aim to prevent a possible collapse of the health system and take suppression measures to interrupt the chain of contagion, but not necessarily stop the

transmission of the disease [9]. Thus, we expect that the number of positive PCR tests diagnosed each day will evolve by applying these measures to obtain good public response levels. One way to evaluate the effectiveness of these measures is with the instantaneous (or effective) reproduction number, which considers the speed with which a disease spreads in a population. This epidemiological parameter quantifies the average of contagions that an infected person causes, assuming that disease transmissibility is constant in a window of time [10, 11]. Thus, an effective quarantine restricts the free spread of the disease and maintains this parameter at values less than one to achieve a controlled spread [9].

Another fundamental concept of the epidemic models is the basic reproduction number (R_0), a quantity that is useful in establishing vaccination policies [12]. Note that R_0 is defined as the expected number of secondary (new) infected cases, produced by a primary infected case, during the entire period of infectiousness (epidemic), when this infected case is introduced into a large population of susceptible cases [13]. Thus, estimates of R_0 facilitate proper vaccination policies. An epidemic can start with a positive probability if R_0 is greater than one. However, an epidemic will quickly die out if R_0 is at most one. Hence, R_0 often serves as a threshold parameter or benchmark.

The step-by-step plan implemented by the government of Chile has four phases that apply mobility restrictions at a district level throughout the country. These phases change according to epidemiological indicators, healthcare networks, and traceability. Hence, in Chile, the phase changes vary from Phase 1 (total quarantine) to Phase 4 (which represents an initial re-opening phase). This last phase has not been applied in any district in Chile since these measures were initiated at the end of July 2020. Full details about these phases are available in the link at: `https://www.gob.cl/pasoapaso` (accessed on 11 March 2021).

More than a year after the arrival of COVID-19, this paper aims to contribute to the analysis of epidemiological surveillance in Chile, understand its realities, and evaluate the Chilean Ministry of Health measures during this sanitary emergency. The objective of this study is to analyze the spread of the disease in Chile according to the quarantines decreed by the Chilean sanitary authority. Consequently, we estimate the trend changes in COVID-19 cases and the instantaneous and basic and reproduction numbers using Chilean official data. These numbers permit us to evaluate the decreed measures and establish vaccination policies.

2. Materials and methods

Using the number of confirmed cases reported up to 03 March 2021, statistical methodologies are applied, which are often used in epidemiology. First, a moving average model is employed to estimate the non-stationarity in the number of confirmed cases and identify the moments when there was a significant change in the time-series trend. The visualization of these moments is shown utilizing colored panels, which indicate those intervals where the time-series trend was stable. Second, the instantaneous reproduction number is estimated through a Bayesian approach represented by its posterior me-

dian, superimposing this number a tile whose filling intensity is proportional to the percentage of the quarantined population (dividing the sum of the inhabitants of the quarantined district by the region's total population) to evaluate its effectiveness. Also, we estimate the basic reproduction number through the likelihood-based methods using a branching process with Poisson distribution [2].

2.1 Materials

We use the official data corresponding to the number of confirmed COVID-19 cases disaggregated at the regional level in Chile. These data were extracted from https://github.com/MinCiencia/Datos-COVID19 (accessed on 22 April 2021) and are handled by the Chilean Ministry of Science, Technology, Knowledge, and Innovation. Several organizations have warned that the data related to COVID-19 must rigorously follow standards and protocols. In the Chilean case, it is possible to deduce that some levels of relaxation occurred in meeting such standards, provoking abrupt peaks of contagions displayed in the region; see Figs. 1 and 2. This can be due to a discontinuous supply of data by the pandemic tracking system. Indeed, the Chilean Ministry of Health reports the cases registered in the system until 9 PM of the previous day. However, these reports may contain the register of more days, depending on the laboratory that carries out the PCR tests, the conditions of work, and whether the record belongs to a weekend. These and other reasons can explain the appearance of peaks at the number of confirmed cases in each Chilean region, as shown in Figs. 1 and 2. The trend changes of the time-series are henceforth called cut-off points. The graphs, data summaries, and implementation of the methods used were developed in the R software 4.2.0 (R Core Team and the R Foundation for Statistical Computing, Vienna, Austria), which can be secured freely from https://www.R-project.org/ (accessed on 22 April 2021).

2.2 Statistical methods

We estimate the trend of the number of confirmed COVID-19 cases per day in Chile with a time-series model as follows. Let $y(t)$ represent the number of confirmed cases at instant t. Then, the trend of this series, $m(t)$ say, can be estimated through the moving average given by

$$\hat{m}(t) = \frac{1}{2q+1} \sum_{i=-q}^{q} y(t+i),$$

where q is a constant that regulates the width of the data interval with which the average is calculated, and therefore, controls the degree of smoothing. Note that, as q decreases, the moving average is more sensitive, so it can capture recent trends very well, although it might produce false alerts. However, when q is large, these false alerts are avoided, but the identification of current trends is more transient [14]. In this study, we use a 15-day window ($q = 7$). The choice of the value of q can be made according to the seasonality of the minor scale detected and considering that the confinement measures for this period of observation were the weekends.
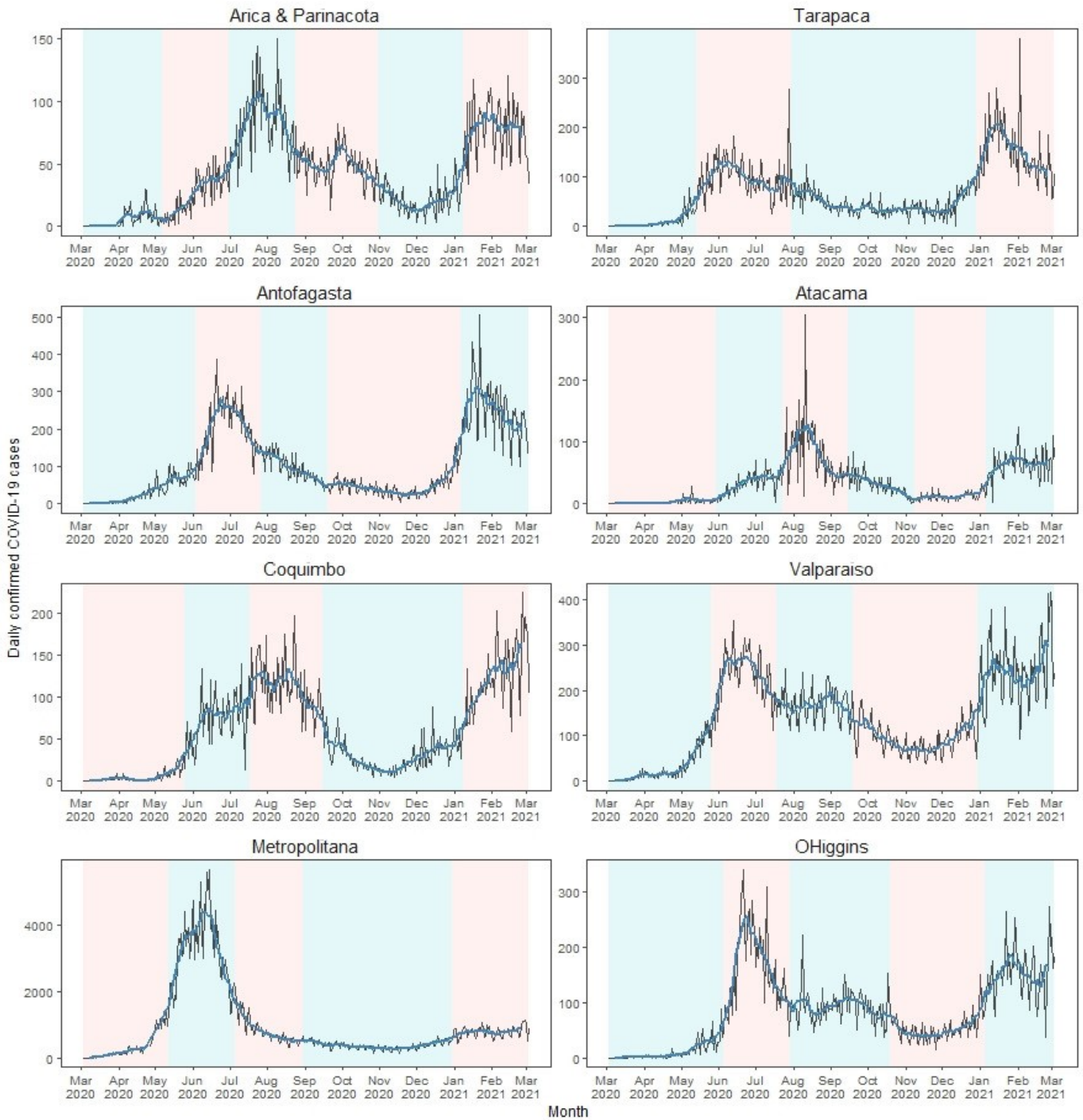
FIGURE 1. Plots of the number of confirmed COVID-19 cases reported daily, disaggregated at the regional level from the indicated region (Arica & Parinacota to O'Higgins). A moving average is superimposed with a blue line, and the segments are constructed from the cut-off points identified in Table 1. The scale of the vertical axis is different for each region.

We identify significant trend changes as follows. First, let the set $\Lambda_{m,T} = \{t_1, ..., t_m\}$ denote an m-partition of the times-series containing these trend changes or cut-off points. The linear regression representing a time-series with $m+1$ segments [15] is defined by

$$y(t) = \alpha_j + \beta_j t + \varepsilon(t),$$

where α_j and β_j are the intercept and slope of the trend in the segment j, respectively, and $\varepsilon(t)$ is a model error with zero mean and an uncorrelated structure. Note that $t = t_{j-1} + 1, ..., t_j$ and $j = 1, ..., m$ indicate the time and segment index, respectively, where, by convention, we take $t_0 = 0$ and $t_{m+1} = T$ (number total of cases). Observe that the parameters defining the linear trend of the model above for $y(t)$ are constant in each of the m periods. For example, in the first segment ($j = 1$), t varies between zero and t_1, the second varies between t_1 and t_2, and so on until the last segment where t goes between t_m and T. Thus, t cannot take non-negative integer values.
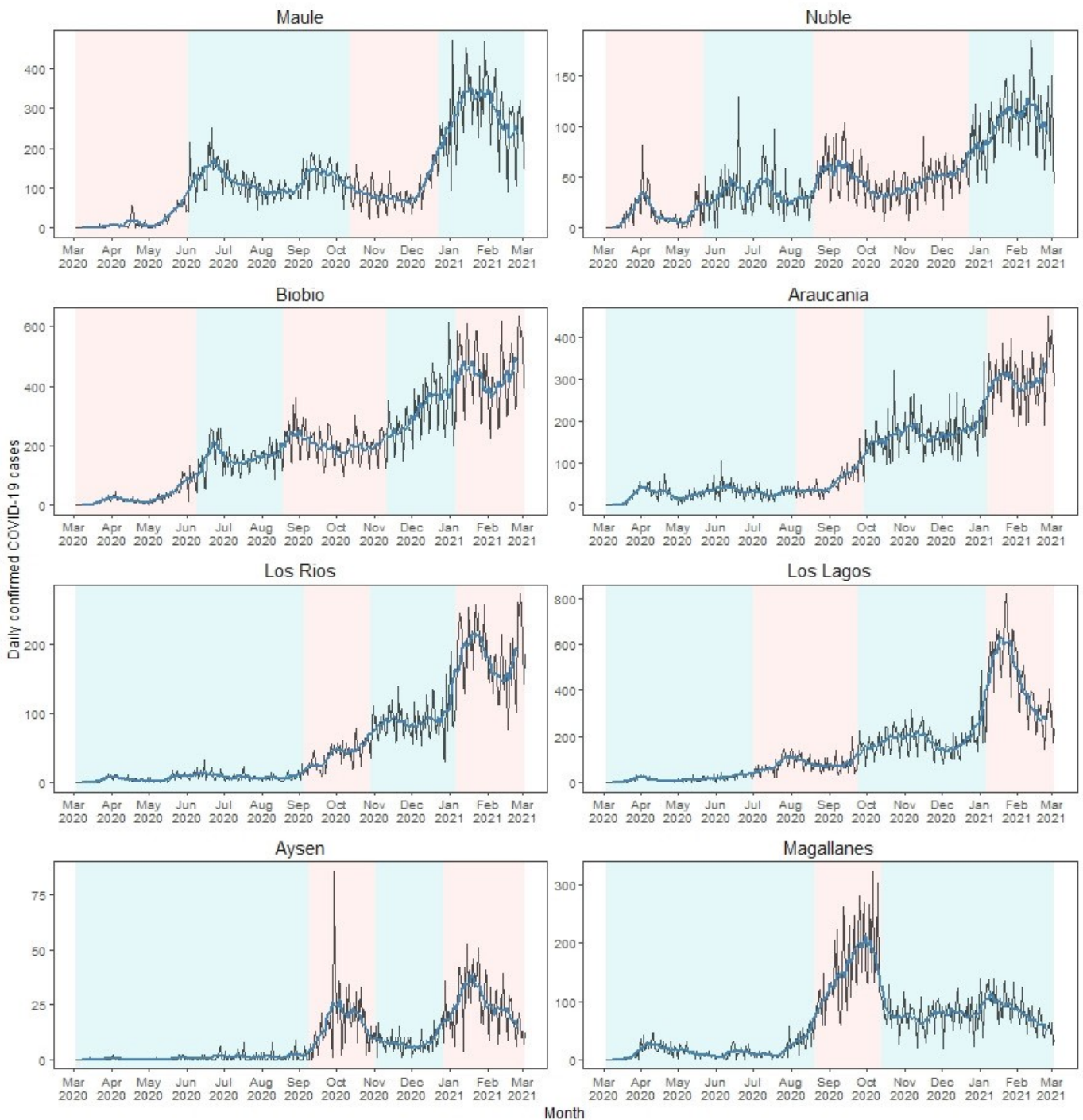
FIGURE 2. Plots of the number of confirmed COVID-19 cases reported daily, disaggregated at the regional level from the indicated region (Maule to Magallanes). A moving average is superimposed with a blue line, and the segments are constructed from the cut-off points identified in Table 1. The scale of the vertical axis is different for each region.

Regarding the number of cut-off points that a time-series has, a statistical test to prove that it has m points is proposed in [16, 17]. Thus, following [15], the m cut-off points are chosen to minimize the sum of residual squares (SRS) of the segmented model, obtained for each of the $m+1$ segments that comprise it and defined as

$$(\hat{i}_1, \ldots, \hat{i}_m) = \operatorname{argmin}_{(i_1, \ldots, i_m)} \sum_{j=1}^{m+1} \text{SRS}\,(i_{j-1} + 1, i_j).$$

Then, we must calculate the SRS in each segment, from initial and final cut-off points, and obtain this summation. The above methodology is implemented in the R software through the function breakpoints{strucchange} of the strucchange package.

We estimate the parameter corresponding to the instantaneous reproduction number as follows. At instant t, $I(t)$ indicates the number of infected individuals which is assumed to follow a Poisson distribution [18]. This distribution depends on (i) the incidences of the previous days, $I(0), \ldots, I(t-1)$

T A B L E 1. Dates for trend changes in the time-series of daily COVID-19 cases for the indicated region.

Region	Change of trend				
	1st	2nd	3rd	4th	5th
Arica & Parinacota	07-May-20	30-Jun-20	24-Aug-20	31-Oct-20	08-Jan-21
Tarapacá	13-May-20	30-Jul-20	29-Dec-20	N/A	N/A
Antofagasta	03-Jun-20	27-Jul-20	19-Sep-20	06/Ene/2021	N/A
Atacama	30-May-20	23-Jul-20	15-Sep-20	08-Nov-20	06-Jan-21
Coquimbo	25-May-20	18-Jul-20	15-Sep-20	08-Jan-21	N/A
Valparaíso	25-May-20	18-Jul-20	19-Sep-20	30-Dec-20	N/A
Metropolitana	12-May-20	05-Jul-20	30-Aug-20	30-Dec-20	N/A
O'Higgins	05-Jun-20	29-Jul-20	19-Oct-20	05-Jan-21	N/A
Maule	02-Jun-20	12-Oct-20	23-Dec-20	N/A	N/A
Ñuble	22-May-20	19/Ago/2020	23-Dec-20	N/A	N/A
Biobío	09-Jun-20	19/Ago/2020	11-Nov-20	06-Jan-21	N/A
Araucanía	05-Aug-20	29-Sep-20	07-Jan-21	N/A	N/A
Los Ríos	05-Sep-20	29-Oct-20	06-Jan-21	N/A	N/A
Los Lagos	01-Jul-20	24-Sep-20	06-Jan-21	N/A	N/A
Aysén	09-Sep-20	02-Nov-20	27-Dec-20	N/A	N/A
Magallanes	20-Aug-20	13-Oct-20	N/A	N/A	N/A

say; (ii) the statistical distribution of the time of disease generation, *w(t)* namely; and (iii) the instantaneous reproduction number, $R_\tau(t)$ say. Assume that the population conditions remain constant over τ days, that is, $I(t)$ conditional on $I(0),\ldots,I(t-1),R_\tau(t),\,w(t)$ is Poisson distributed as

$$I(t) \mid I(0),\ldots,I(t-1),R_\tau(t),w(t) \sim \text{Poisson}\left(R_\tau(t)\lambda(t)\right),$$

where

$$\lambda(t) = \sum_{s=1}^{t} I(t-s)w(s).$$

Then, the likelihood function of $R_\tau(t)$ given the occurrences in the time indexes of the interval $[t-\tau+1,t]$ is stated as

$$l\left(I(t-\tau+1),\ldots,I(t) \mid I(0),\ldots,I(t-\tau),R_\tau(t),w(t)\right) =$$
$$\prod_{s=t-\tau-1}^{t} \frac{\exp\left(-R_\tau(s)\lambda(s)\right)\left(R_\tau(s)\lambda(s)\right)^{I(s)}}{I(s)!}.$$

Note that the estimator of the parameter $R_\tau(t)$, obtained by the maximum likelihood method, turns out to be very nonstable. This instability reduces the precision of the estimator when the time interval of the data is small. A methodology to estimate $R_\tau(t)$ using a Bayesian approach was proposed in [19]. This approach initially specifies a gamma prior distribution (traditionally with mean equal to 5 and standard deviation equal to 5), obtaining a gamma posterior distribution expressed as

$$R_\tau(t) \mid I(t-\tau+1),\ldots,I(t),w(t) \sim \text{Gamma}(A(t),B(t)),$$

with $A(t) = 1 + \sum_{s=t-\tau+1}^{t} I(s)$ and

$$B(t) = \left(\frac{1}{5} + \sum_{s=t-\tau+1}^{t} \lambda(s)\right)^{-1}.$$

In this work, as mentioned, we use a window width of one week ($\tau = 7$ days, as suggested in [20]), assuming a log-normal distribution of mean 4.7 days and standard deviation 2.9 days for the interval of successive cases [21]. The methodology is implemented in the R software through the function estimate_r{EpiEstim} of the EpiEstim package. The basic reproduction number is estimated with a Poisson branching process that utilizes a serial interval distribution based on the log-normal model above mentioned and the number of new cases reported. The methodology is applied using the first two weeks (03 March to 17 Mach 2020) through of the earlyR package implemented in the R software [22].

3. Results

Fig. 3 shows the autocorrelations of the number of confirmed cases, after removing the linear trend, for each region. From this figure, note that there is a seasonal component of $q = 7$ days in most of the time series. This value is consequent with what was indicated suggested in [20].

The application of the test to identify the cut-off points in the trend of cases in each Chilean region clearly shows heterogeneity between the regions. More specifically, this heterogeneity is related to the date when the change occurs and its order; see Table 1. In addition, although there are pairs of regions such as Coquimbo-Metropolitana, Los Ríos-Aysén, and Valparaíso-O'Higgins, whose trend change dates are very similar, this does not necessarily imply that their time-series
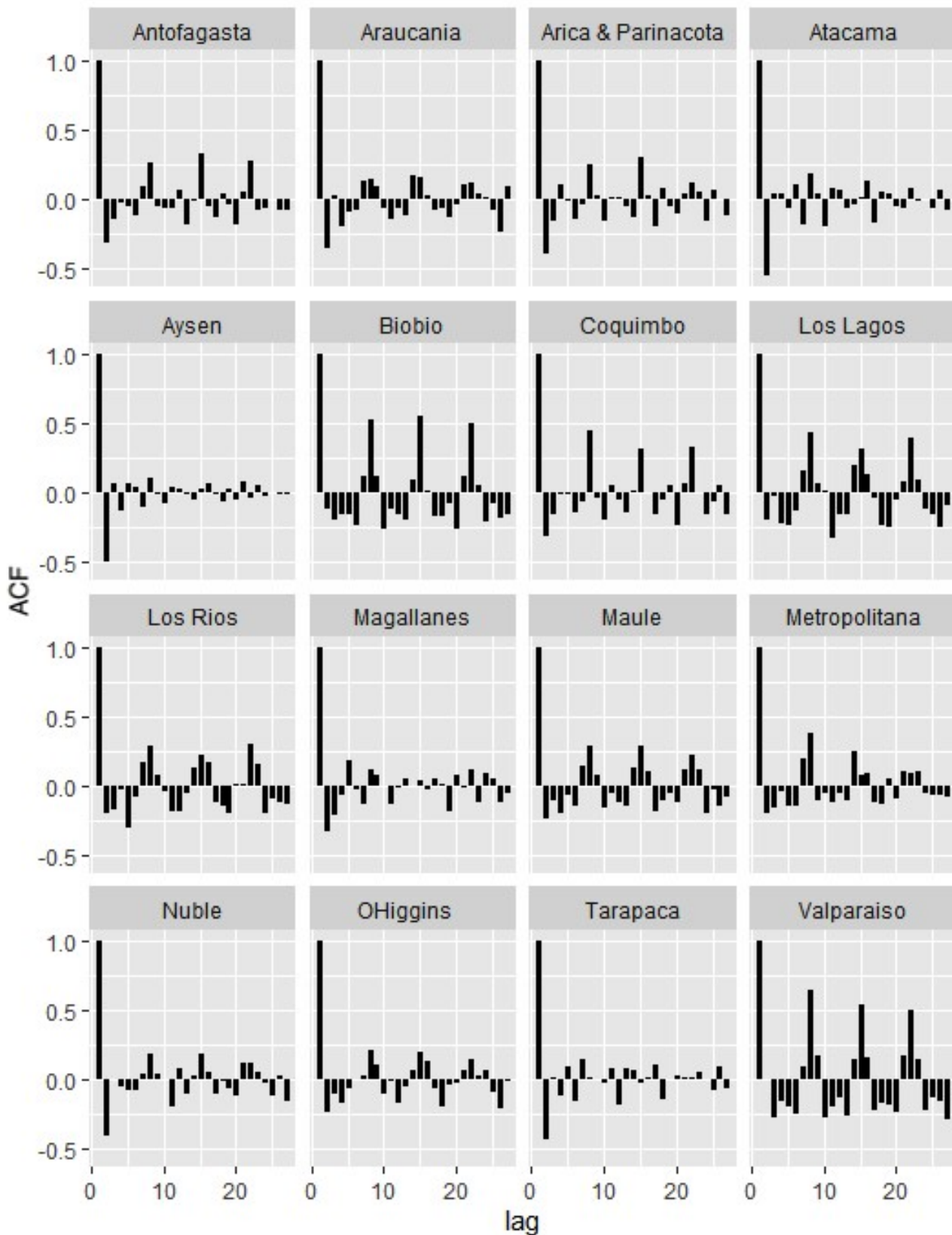
FIGURE 3. Auto-correlation functions (ACFs) of confirmed COVID-19 cases, after removing the linear trend, for the indicated region.

are similar, such is the case of the first two pairs of regions mentioned; see Table 1 and Fig. 1.

Based on Table 1, note that by the end of February 2021, the regions from Arica & Parinacota to Maule have experienced

the first wave of COVID-19 cases in mid-June 2020 and the second wave beginning in late December 2020, after an evident decrease in cases between the two events. A similar situation occurs in Aysén, except that the first wave occurred in October
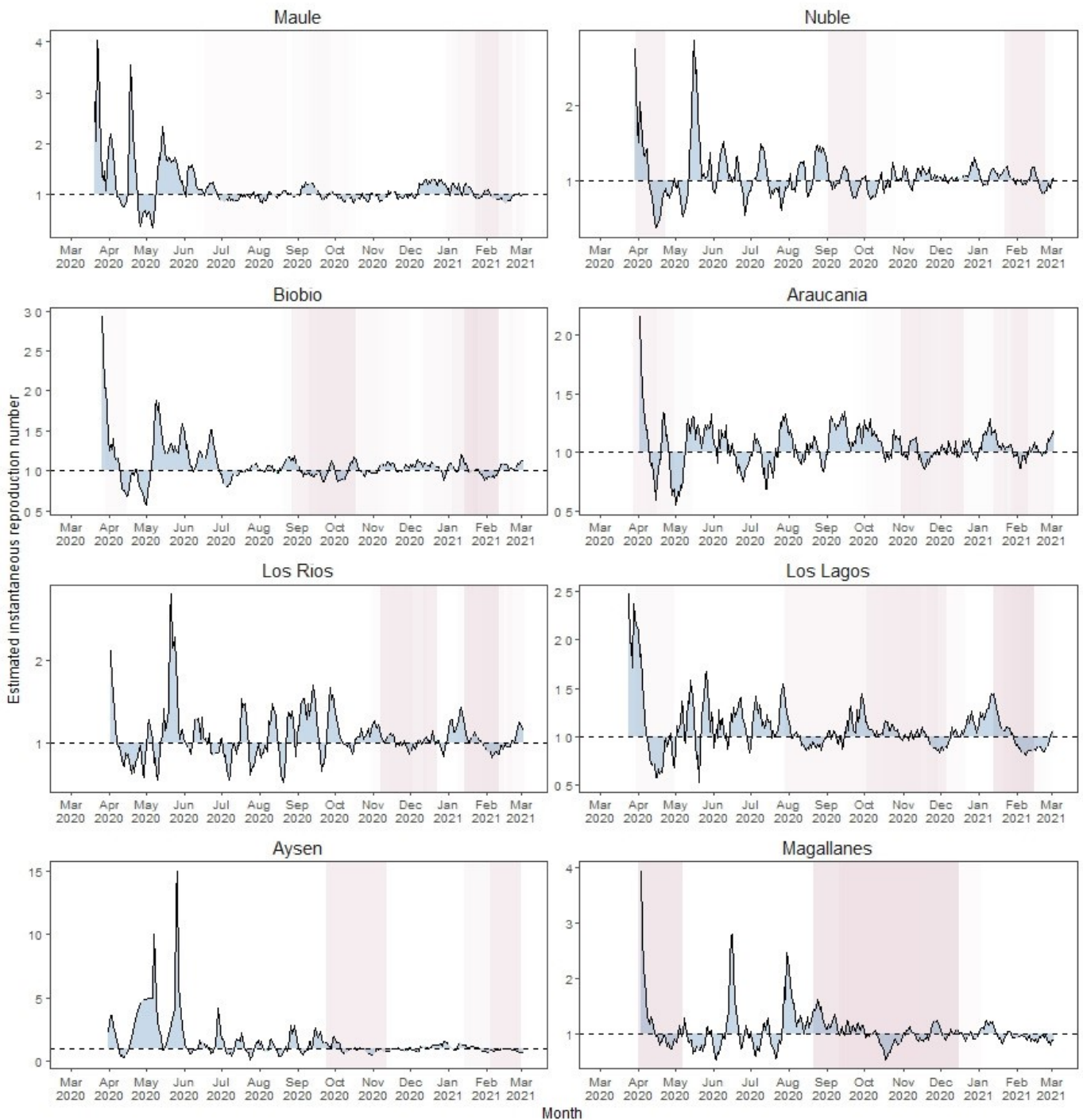
FIGURE 4. Instantaneous reproduction number estimated through its posterior median, disaggregated at the regional level from Maule to Magallanes. The periods superimposed by shades in red represent the percentage of the population in quarantine according to its intensity. The scale of the vertical axis is different for each region.

2020; see Figs. 1 and 2. In addition, observe that in Ñuble repeated ups and downs of COVID-19 cases, that at the end of 2020 begin an upward trend, seem to have been controlled in February 2021. Los Lagos, which experiences a gradual rise in cases until January 2021, registers an important pick of cases that, as in Ñuble, seem to be controlled quickly. The Magallanes region stands out for an abrupt decrease in COVID-19 cases in mid-October 2020, when the trend seemed to be alarmingly increasing but was successfully controlled; see Fig. 2.

Note that the above results show a common situation in the spread of disease presented to date [23]. Furthermore, an increase in COVID-19 cases was interrupted by the decree of some measures, but in the absence of pharmaceutical interventions, it inevitably increases again, causing successive waves of contagion. However, some regions completely disobey this rule, such as Biobío, Araucanía, and Los Ríos, whose COVID-19 cases trend has not decreased since mid-2020. Although they are not permanently increasing, there is evidence that the measures decreed in these regions have been neither sufficient

nor much less effective; see Fig. 2. As the instantaneous reproduction number is not decreasing in quarantine, the disease transmission does not stop, and the number of COVID-19 cases does not decrease either.

Our suspicions of good or bad control of the spread can be confirmed through the median of the instantaneous reproduction number, which should decrease after applying a quarantine. This explains the decrease in COVID-19 cases for the second half of 2020 in the regions from Arica & Parinacota to Maule, and the abrupt decline in Los Lagos, Aysén, and Magallanes, which translates into a successful mitigation measure; see Figs. 1 and 2. Nevertheless, note that, in the regions from Biobío to Los Ríos, the estimated instantaneous reproduction number decreases, but COVID-19 cases have been kept in check. However, there is no downward trend in the number of confirmed cases. Considering the use of intensive care unit beds (https://www.icovidchile.cl/capacidad-hospitalaria, accessed on 09 July 2021), we postulate that these regions have remained with a high hospital load as a result of such a situation; see Fig. 4. An interesting piece of information is the dates when the quarantines were imposed. However, these dates are complex because the quarantines are decreed by districts and not by region. Figs. 4 and 5 show the percentage of the population in each region in quarantine (Phase 1) over time. When the percentage is high, this is represented in the figure as a rectangle, but if the percentage is low, then the rectangle is transparent. In our GitHub repository: https://github.com/NJEREZLILLO/proyecto_titulo (accessed on 16 June 2021), the interested reader can find the file cuarentena.xlsx with some of these dates.

A heterogeneous spread is observed for the number of confirmed cases and the instantaneous reproduction number throughout the regions of Chile, which is explained by the dynamic quarantines decreed by the sanitary authority and other factors not considered in this work. Regarding the effectiveness of these measures, note that the spread has been successfully controlled in most regions, which permits us to observe at least two waves of contagion. This situation was simulated in [23] and modeled with a susceptible-exposed-infectious-recovered (SEIR) structure (see Appendix and [12] for details of the SEIR model). In that study, after adding the adaptive confinement for the transmission of the disease (in which quarantines are decreed and lifted when cases exceed or fall below a threshold of cases), successive waves of contagion were registered whose amplitudes decreased over time. Therefore, we conclude that this is an entirely expected situation. The chain of contagion is successfully interrupted without necessarily eliminating the disease from the population due to the absence of pharmacological intervention (such as vaccines) or an effective and efficient traceability system. However, the Biobío, Araucanía, and Los Ríos regions have not responded favorably to this type of measure, which is why we conjecture they have not had a downward trend in the number of COVID-19 cases since mid-2020. Thus, these measures are insufficient and have only partial effects for infection control, which is confirmed by the instantaneous reproduction number values that are greater than one. Notorious negative effectiveness is detected in the Biobío, Araucanía, and Los Ríos regions. Note that the

instantaneous reproduction number in the Araucania and Los Lagos regions is very similar, but the scale is different (Los Lagos has a bigger scale).

As mentioned, we estimate the expected number of new cases infected by a primary case early during the COVID-19 epidemic when all the population is susceptible to fall ill, R_0 namely, with the maximum likelihood method using a branching process with Poisson distribution. This process employs a serial interval distribution [21], in our case corresponding to a log-normal model of mean 4.7 days and standard deviation 2.9 days for the interval of successive cases, and the number of new cases reports. We utilize the data corresponding to COVID-19 cases of the first two weeks with dates from 03 March to 17 Mach 2020. When this methodology was applied, we obtained an estimate of R_0 equal to $\hat{R}_0 = 12.39$ as observed in Fig. 6. This estimate confirms the adequacy of the pandemic declaration according that it must start if R_0 is greater than one. Furthermore, such an estimate facilitates the establishment of adequate vaccination policies by the authorities.

4. Discussion, conclusions, limitations, and future research

Different theories could explain why the mentioned Chilean regions do not respond after the application of a quarantine. The first one might point out to the fact that the quarantines that have been decreed are not long enough to achieve the expected effect. A question arises: What could the authorities motivate to lift the quarantines decreed there if they do not respond positively?

It is evident that, in the first months, people may obey the confinement and the quarantines to be extended as long as necessary to guarantee their effectiveness. However, in more advanced stages, it is not only health that is at risk, but other factors are added that make it difficult for citizens to comply. Note that it is impressive how other regions quarantined in January 2020, such as Tarapacá and Antofagasta, did reduce the number of COVID-19 cases; see Fig. 5. Therefore, it is necessary to explore more deeply what is happening in these regions.

An immediate suggestion is to study the meaning that a quarantine has on the population. For instance, a mobility index constructed through anonymous mobile tracking has become very useful nowadays. An example of this is applied in [24], where the transmission of the disease is studied concerning mobility indexes provided by Apple and Google. This would help to understand how the population responds to the measures, which can direct authorities on improving them. Therefore, a way to verify that there is obedience is to analyze the mobility data by Apple and Google used in [24]. Note that mobility decreases abruptly in the first months of the pandemic, but it is challenging to keep due to other factors for a long time.

In the present study, the incidence of COVID-19 is measured with the number of new cases without considering the dimension of the population in each area [25]. This might be a factor affecting the spread of the virus over time and a limitation of our investigation. Indeed, the incidence of a disease can be calculated as the number of new cases over the size of the
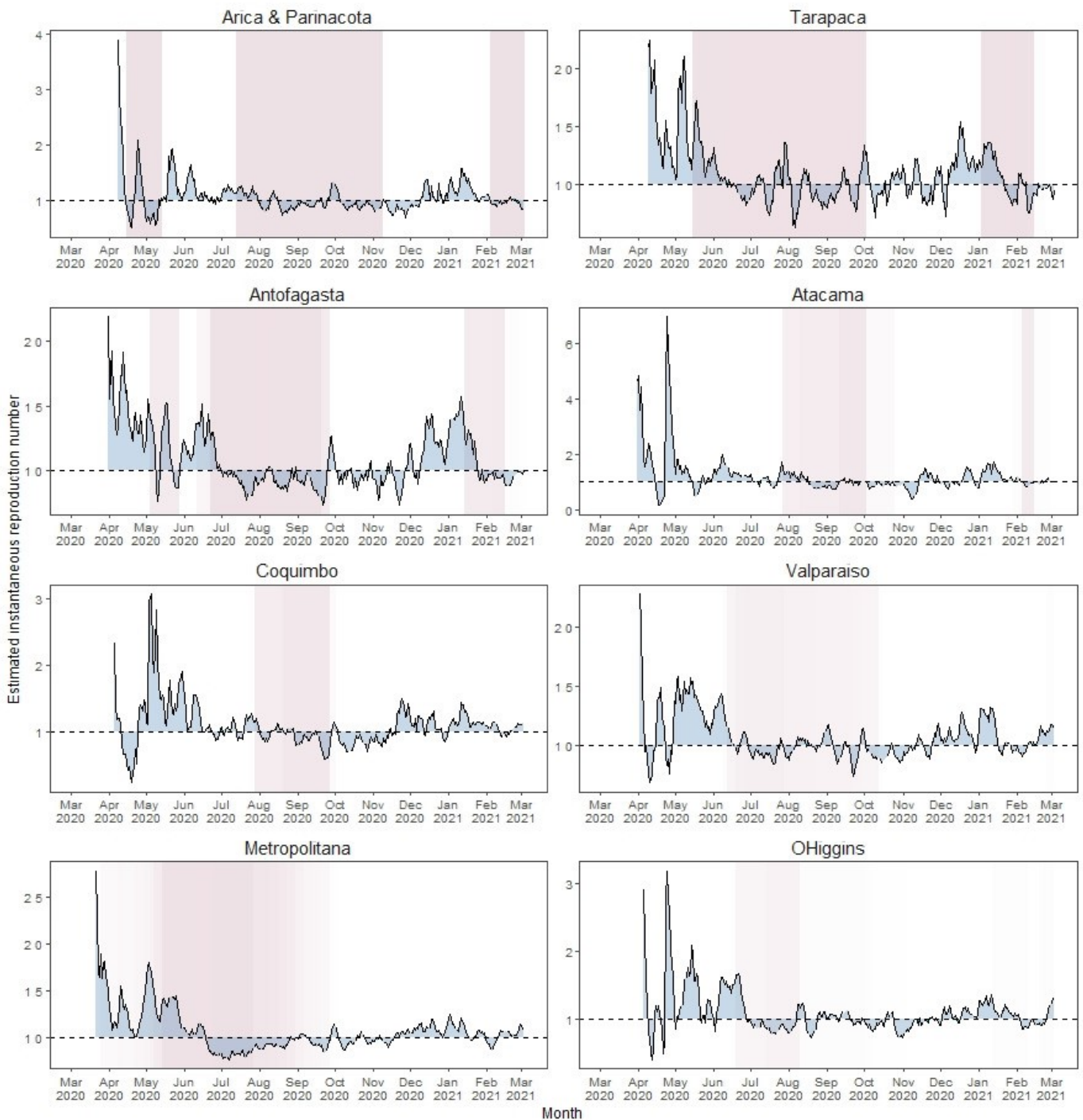
F I G U R E 5. Instantaneous reproduction number estimated through its posterior median, disaggregated at the regional level from Arica & Parinacota to the O'Higgins. The periods superimposed by shades in red represent the percentage of the population in quarantine according to its intensity. The scale of the vertical axis is different for each region.

population. However, in our study, the objective was to detect cases in each region and not compare by gravity. Therefore, we think this is not necessary to compare considering the dimension of the population in each area. Another limitation of the used data is that the number of new daily cases might be underestimated with territorial differences in this phenomenon. The number of newly diagnosed cases depends on several factors, for example, the number of available swabs. One could estimate the model parameters with more robust data such as the number of people in hospitals and deaths. Indeed, under-

estimation is a transversal problem for each country [7], and some authors approach it [26]. Nevertheless, underestimation is often ignored in these models.

The estimation of the instantaneous reproduction number does not incorporate the source of variation produced by the interventions. Instead, we use the methods that are implemented in the literature. Incorporating interventions in this type of model is part of future work since it is beyond the objective of the present investigation.
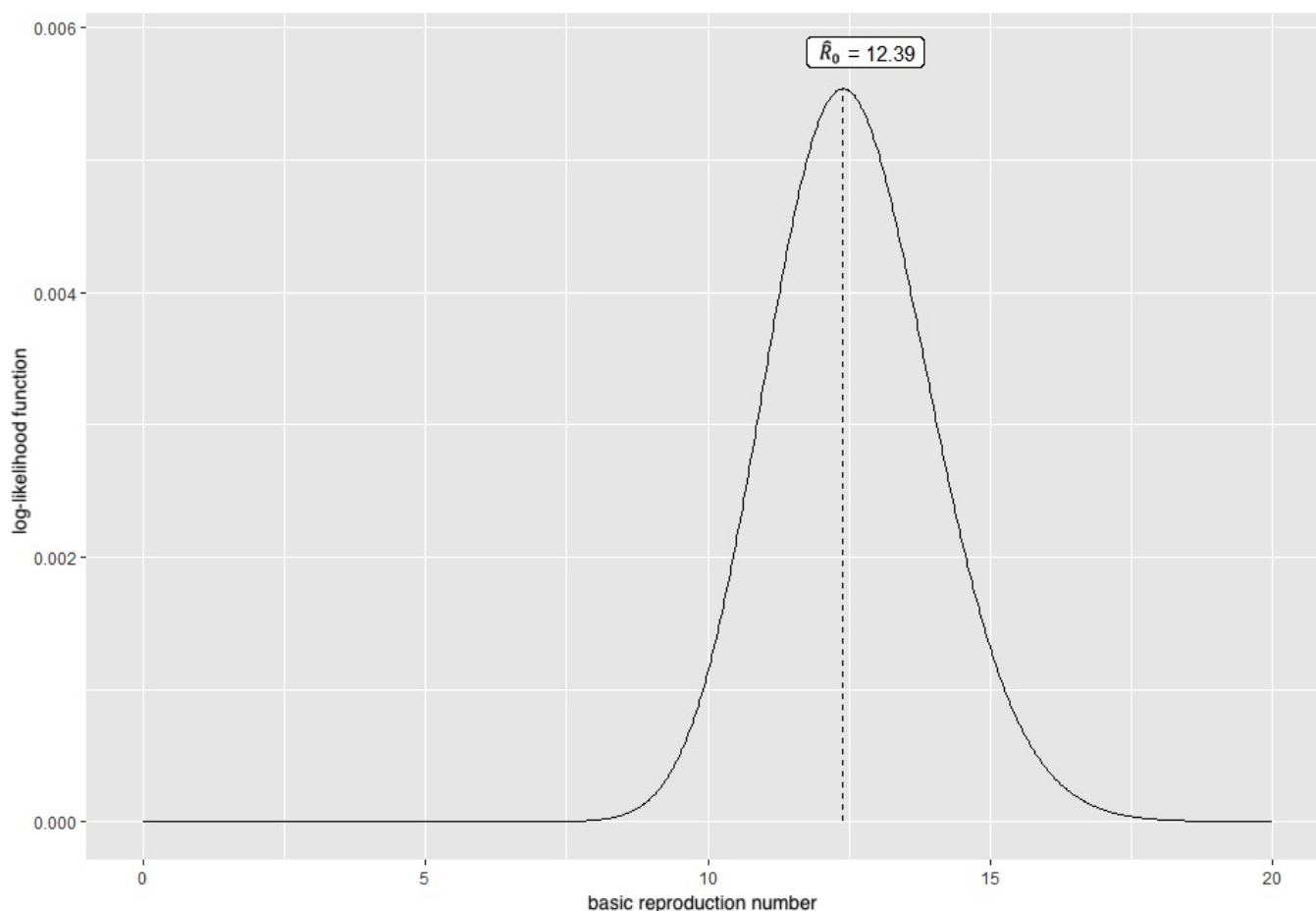
F I G U R E 6. Basic reproduction number estimated through the maximum likelihood method using a branching process with Poisson distribution for Chilean COVID-19 cases from 03 March to 17 Mach 2020.

In summary, we are unable to identify the measures considered in the context of protocols in the step-by-step plan, which is partly responsible for non-compliance with quarantines. However, three specific findings of this investigation that may be extrapolated to everyday practice, and enlighten the way of other countries, are the following:

(i) A measure that has proved to be effective to stop the spread of the disease is the early quarantine strict, such as detected in some Chilean regions. In this way, indexes that measure the mobility of citizens are necessary.

(ii) As time passes without stopping infections, quarantines lose effectiveness even if the estimated instantaneous reproduction number is small and stable. In addition, other factors can cause the instantaneous reproduction number to not be within the expected ranges which must be studied in a future investigation.

(iii) The estimated basic reproduction number confirmed the suitability of the pandemic declaration.

Overall, we highlight the need to modify the current step-by-step plan so that the efforts of the authorities and the population are not in vain. For this, it is necessary to monitor the population that may be more likely to disobey the measures, improve the traceability of outbreaks, and promote self-care of the population while the vaccination plan does not allow us to obtain herd immunity. Furthermore, considering that Chile is one of the countries with the largest immunized population, we immediately call for improving the step-by-step plan or redefining the health system in the face of the pandemic. Moreover, the authorities must improve the traceability of outbreaks and provide clear and persuasive signals to promote the collective self-care of the population and preserve their well-being. Finally, our study visualizes the lack of a solid public health policy that can stop the spread of the disease and keep the population safe.

AUTHOR CONTRIBUTIONS

NJL and BL designed the study, collected the data, and created the models and codes. NJL, BLA, JFZ, and VL performed the data analysis. NJL, BLA, JMG, JFZ, and VL drafted and wrote the final version the paper. NJL, BLA, JMG, JFZ, and VL performed a clinical development and validated the analysis and results. NJL, BLA, JMG, JFZ, and VL revised the paper critically for important intellectual contents. All authors approved the final version to be published.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

CONFLICT OF INTEREST

The authors declare no conflict of interest. Víctor Leiva is serving as one of the Editorial Board members of this journal. We declare that Víctor Leiva had no involvement in the peer review of this article and has no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to OK.

DATA AVAILABILITY

The data used to support the findings of this study are available from the corresponding author upon request.

APPENDIX

The SEIR model classifies the population into four classes: susceptible (S), exposed (E, infected but not yet infectious), infected (and infectious, I), and recovered (R). This model is represented by the equations given by

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\frac{\beta_t\,\mathrm{SI}}{\mathrm{N}}, \frac{\mathrm{d}E}{\mathrm{d}t} = \frac{\beta_t\,\mathrm{SI}}{\mathrm{N}} - \sigma\mathrm{E}, \frac{\mathrm{d}I}{\mathrm{d}t} = \sigma\mathrm{E} - (\gamma+\mu)\mathrm{I}, \frac{\mathrm{d}R}{\mathrm{d}t} = \gamma\mathrm{I},$$

where $\mathrm{N} = \mathrm{S} + \mathrm{E} + \mathrm{I} + \mathrm{R}$, σ is the loss of latency rate, γ is the recovery rate, μ is the disease-induced mortality rate, and β_t is the time-dependent transmission rate defined as

$$\beta_t = \begin{cases} \beta, & \text{if } t \leq t_\mathrm{i}; \\ \beta\left(1 - \frac{\delta_\mathrm{d}(t-t_\mathrm{i})^2}{1+\delta_\mathrm{d}(t-t_\mathrm{i})^2}\right), & \text{if } t_\mathrm{i} < t < t_\mathrm{e}; \\ \beta_\mathrm{min} + (\beta - \beta_\mathrm{min})\frac{\delta_\mathrm{i}(t-t_\mathrm{e})^2}{1+\delta_\mathrm{i}(t-t_\mathrm{e})^2}, & \text{if } t \geq t_\mathrm{e}; \end{cases}$$

where β is the initial transmission rate, without any measures to limit contact between people; t is the time (in days) from the first detected case; t_i is the initial day of confinement minus two; t_e is the end of confinement; δ_d is the rate of decrease of transmission because of the control measures; β_min is the value of β_t at time t_e; and δ_i is the rate of increase of transmission after t_e. Once the SEIR model is fitted, the basic reproduction number is estimated [27] as

$$R_0 = \frac{r}{\gamma + \mu},$$

where r is the exponential rate of increase of cases at the beginning of the epidemic, which, in turn, is estimated by fitting an exponential function using only the first data.

REFERENCES

[1] Karunathilake K. Positive and negative impacts of COVID-19, an analysis with special reference to challenges on the supply chain in South Asian countries. Journal of Social and Economic Development. 2020. (in press)

[2] Liu Y, Mao C, Leiva V, Liu S, Silva W. Asymmetric autoregressive models: Statistical aspects and a financial application under COVID-19 pandemic. Journal of Applied Statistics. 2021; 1: 1–25.

[3] Chahuan-Jimenez K, Rubilar R, de la Fuente-Mella H, Leiva V. Breakpoint analysis for the COVID-19 pandemic and its effect on the stock markets. Entropy. 2021; 23: 100.

[4] de la Fuente-Mella H, Rubilar R, Chahuan-Jimenez K, Leiva V. Modeling COVID-19 cases statistically and evaluating their effect on the economy of the countries. Mathematics. 2021; 9: 1558.

[5] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, *et al*. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020; 395: 497–506.

[6] Tang YW, Schmitz JE, Persing DH, Stratton CW. Laboratory diagnosis of COVID-19: Current issues and challenge. Journal of Clinical Microbiology. 2020; 58: 512–520.

[7] Bustos N, Tello M, Droppelmann G, Garcia N, Feijoo F, Leiva V. Machine learning techniques as an efficient alternative diagnostic tool for COVID-19 cases. Signa Vitae. 2022; 18: 23-33.

[8] Gibbons CL, Mangen MJ, Plass D, Havelaar AH, Brooke RJ, Kramarz P, *et al*. Measuring underreporting and under-ascertainment in infectious disease datasets: a comparison of methods. BMC Public Health. 2014; 14: 1–7.

[9] Patiño-Lugo DF, Vélez M, Velásquez Salazar P, Vera-Giraldo C, Vélez V, Marín I, *et al*. Non-pharmaceutical interventions for containment, mitigation and suppression of COVID-19. Colombia Medica. 2020; 51: 42–66.

[10] Bautista LA, Conesa MG, Caravaca GR, Balbas BB. COVID-19 effective reproduction ratio determination: An application, and analysis of issues and influential factors. 2020. (in press)

[11] Rovetta A, Bhagavathula AS, Castaldo L. Modeling the Epidemiological Trend and Behavior of COVID-19 in Italy. Cureus. 2020; 12: e9884.

[12] Fierro R, Leiva V, Balakrishnan N. Statistical Inference on a Stochastic Epidemic Model. Communications in Statistics. 2015; 44: 2297–2314.

[13] Nouvellet P, Cori A, Garske T, Blake IM, Dorigatti I, Hinsley W, *et al*. A simple approach to measure transmissibility and forecast incidence. Epidemics. 2018; 22: 29–35.

[14] Shumway RH, Stoffer DS. Time series analysis and its applications: with R examples. New York: Springer. 2017.

[15] Zeileis A, Kleiber C, Krämer W, Hornik K. Testing and dating of structural changes in practice. Computational Statistics & Data Analysis. 2003; 44: 109–123.

[16] Bai J, Perron P. Estimating and Testing Linear Models with Multiple Structural Changes. Econometrica. 1998; 66: 47–78.

[17] Bai J, Perron P. Computation and analysis of multiple structural change models. Journal of Applied Econometrics. 2003; 18: 1–22.

[18] Fraser C. Estimating individual and household reproduction numbers in an emerging epidemic. PLoS ONE. 2007; 2: e758.

[19] Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. American Journal of Epidemiology. 2013; 178: 1505–1512.

[20] Hu FC. The estimated time-varying reproduction numbers during the ongoing pandemic of the Coronavirus disease 2019 (COVID-19) in 12 selected countries outside China. 2020. Available at: https://www.medrxiv.org/content/10.1101/2020.05.10.20097154v1 (Accessed: 24 May 2021).

[21] Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. International Journal of Infectious Diseases. 2020; 93: 284–286.

[22] Thibaut J, Anne C, Pierre N, Janetta S. EarlyR: Estimation of transmissibility in the early stages of a disease outbreak. 2013. Available

at: https://CRAN.R-project.org/package=earlyR (Accessed:19 June 2021).

[23] de Castro F. Modelling of the second (and subsequent) waves of the coronavirus epidemic. Spain and Germany as case studies. 2020. Available at: https://doi.org/10.21203/rs.3.rs-49249/v1 (Accessed: 19 June 2021).

[24] Nouvellet P, Bhatia S, Cori A, Ainslie KEC, Baguelin M, Bhatt S, *et al*. Reduction in mobility and COVID-19 transmission. Nature Communications. 2021; 12: 1–9.

[25] Martin-Barreiro C, Ramirez-Figueroa JA, Cabezas X, Leiva V, Galindo-Villardón MP. Disjoint and functional principal component analysis for infected cases and deaths due to COVID-19 in South American countries with sensor-related data. Sensors. 2021; 21: 4094.

[26] Hortaçsu A, Liu J, Schwieg T. Estimating the fraction of unreported infections in epidemics with a known epicenter: An application to COVID-19. Journal of Econometrics. 2021; 220: 106–129.

[27] Diekmann O, Heesterbeek JA, Metz JA. On the definition and the computation of the basic reproduction ratio Ro in models for infectious diseases in heterogeneous populations. Journal of Mathematical Biology. 1990; 28: 365–382.