

Article

An Epidemiological Analysis for Assessing and Evaluating COVID-19 Based on Data Analytics in Latin American Countries

Víctor Leiva ^{1,*} , Esdras Alcudia ² , Julia Montano ²  and Cecilia Castro ³ ¹ School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, Valparaíso 2362807, Chile² Faculty of Statistics and Informatics, Universidad Veracruzana, Xalapa 91140, Mexico; esdras2304.ea@gmail.com (E.A.); julmontano@uv.mx (J.M.)³ Centre of Mathematics, University of Minho, 4710-057 Braga, Portugal; cecilia@math.uminho.pt

* Correspondence: victor.leiva@pucv.cl or victorleivasanchez@gmail.com

Simple Summary: In this research, we investigate the COVID-19 spread in Latin American countries using time-series and epidemic models. We highlight the diverse outbreak patterns and the crucial role of the reproduction number in modeling pandemic scenarios. Our findings underscore the need for ongoing epidemic surveillance and accurate data handling.

Abstract: This research provides a detailed analysis of the COVID-19 spread across 14 Latin American countries. Using time-series analysis and epidemic models, we identify diverse outbreak patterns, which seem not to be influenced by geographical location or country size, suggesting the influence of other determining factors. Our study uncovers significant discrepancies between the number recorded COVID-19 cases and the real epidemiological situation, emphasizing the crucial need for accurate data handling and continuous surveillance in managing epidemics. The absence of a clear correlation between the country size and the confirmed cases, as well as with the fatalities, further underscores the multifaceted influences on COVID-19 impact beyond population size. Despite the decreased real-time reproduction number indicating quarantine effectiveness in most countries, we note a resurgence in infection rates upon resumption of daily activities. These insights spotlight the challenge of balancing public health measures with economic and social activities. Our core findings provide novel insights, applicable to guiding epidemic control strategies and informing decision-making processes in combatting the pandemic.

Keywords: data science; epidemic models; reproduction number; SARS-CoV-2; time-series models



Citation: Leiva, V.; Alcudia, E.; Montano, J.; Castro, C. An Epidemiological Analysis for Assessing and Evaluating COVID-19 Based on Data Analytics in Latin American Countries. *Biology* **2023**, *12*, 887. <https://doi.org/10.3390/biology12060887>

Academic Editor: Ligen Yu

Received: 25 May 2023

Revised: 14 June 2023

Accepted: 16 June 2023

Published: 20 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus responsible for the COVID-19 pandemic, was first identified in the city of Wuhan, China, in December 2019 [1,2]. Almost three and a half years after its detection, it can be said that this epidemic has plunged humanity into a state of confusion. The governments worldwide monitor the spread of the disease and take measures to contain and control the outbreaks without prior information. Therefore, it becomes crucial to assess and evaluate the best course of action to face this pandemic based on data analytics.

Every country has implemented measures to prevent the spread of COVID-19, with population isolation through quarantines gaining significant momentum. This study aims to assess the effectiveness of these measures in controlling the spread of the virus. In world history, this is the first time that a pandemic has compelled us into a complete state of global quarantine [3]. The coronavirus has infected over 750 million people and has caused the death of almost 7 million people as of the time of the present study [4]. A reliable and accurate dataset of the disease is crucial for scientists to conduct research and make informed decisions regarding policy development. Unfortunately, errors can occur in the data collection process, especially during a pandemic.

The coronavirus pandemic represents one of the most serious public health crises faced by the world, with Latin America being particularly hard-hit. The specific challenge lies in adapting existing protocols from previous epidemics to the unique characteristics of this virus. However, effectively controlling and managing outbreaks is vital due to the complexity involved. COVID-19 has had diverse impacts on both the globe and specifically in Latin America, extending beyond public health to affect various aspects, such as the economy of each country, as indicated by the gross domestic product, which represents the total output of goods and services within a nation [5]. The pandemic has left significant marks on the global level, but its economic and financial impact has been particularly profound in Latin America due to its unique historical backdrop.

In this study, we present a refined feedback process aimed at guiding governments in adopting effective health strategies to combat COVID-19. This process relies on robust data visualization sourced from multiple databases. Most countries have been tracking daily confirmed cases since the beginning, allowing for early reporting of disease incidence. The main challenge lies in ensuring an appropriate response while balancing political, health, and economic measures. Despite limited prior experience with COVID-19, certain crucial steps must be considered based on our accumulated knowledge and ongoing learning as the pandemic unfolds. Ongoing research offers the potential for a deeper understanding of virus's behavior, transmission methods, and potential preventive measures, enabling the creation of predictive scenarios for similar situations in the future. This improved understanding can facilitate more informed decision-making, helping to mitigate the social impacts of the pandemic. After all, the global community was largely caught off guard by this pandemic, and measures used to curb its spread may have been implemented later than ideal. Previous research on similar events has demonstrated its value in decision-making for managing disease spread, contributing to more effective responses.

Time-series models, as explored and applied in [6–8], are particularly relevant in forecasting epidemic diseases, as evidenced in [9,10]. Studies into different aspects of epidemics is being conducted by numerous scientists, each striving to provide helpful insights that may benefit mankind. These studies employ a variety of methodologies, with some relying on epidemic models [11,12], while others predominantly use statistical and mathematical methods, as in [13], where the authors applied mathematical models to analyze and predict the timeline and phases of an epidemic, specifically focusing on COVID-19 in Italy. In addition, control theory [14] has been applied to epidemic models to derive optimal strategies for easing restrictive measures, as showcased in [15].

In addition to time-series and epidemic models, it is worthwhile to mention an alternative approach based on the geometric Brownian motion (GBM). In the context of disease spread modeling, the exponential growth of epidemic cases—as observed, for example, in the initial stages of spreading based on epidemic models—can be modeled with the GBM. Certain quarantine measures for such exponential-growth stochastic processes are then modeled as the so-called resetting, a partial reduction approach in the process magnitude at specifically chosen times (distributed, for instance, according to a Poisson model) [16–18]. Although this approach is not explored in the current study, it offers a valuable theoretical alternative that could provide additional insights into disease spread dynamics.

Nonpharmacological measures to combat COVID-19 have also been modeled in [19,20]. The same was performed in [21] using branching processes. Other research has focused on growth curves to analyze mortality and second waves of the pandemic [22,23]. Similarly, in [24], the generalized Richards and growth models were used to analyze the COVID-19 infected cases in China. In [25], a cluster analysis of COVID-19 mortality according to sociodemographic factors was carried out at municipal level in Mexico.

Machine learning models have been also employed, which are known for their flexibility and adaptability. Machine learning models have been successfully applied to complex phenomena, including cardiovascular diseases [26], and to make reliable predictions for emerging COVID-19 variants. For instance, as in [27], where the authors introduced a novel interpretable deep learning architecture to predict SARS-Cov-2 disease severity.

In the context of epidemics and pandemics, researchers have used machine learning models such as logistic regression, neural networks, and support vector machines to understand and predict the progression of COVID-19. A feature of these studies is their use of data from diverse regions around the world, highlighting the global reach of these epidemics [28–30]. Consequently, the goal of this study is to provide tools to bolster disease control efforts. As we approach the third year since the emergence of the pandemic, our primary objective is to enhance epidemiological surveillance across Latin America. We endeavor to comprehend the current status of the disease, evaluate the strategies deployed by different governments in response to this health crisis, and investigate the disease propagation in various Latin American nations that have implemented government-enforced quarantines. Furthermore, we aim to delineate the key parameters for managing the COVID-19 outbreak within these countries.

The threat level of COVID-19 is heavily dependent on the virus's interaction with sensory receptors, meaning the myriad of coronavirus variants is determined by the interactions of the viral proteins with human receptors [31,32]. To date, COVID-19 has exhibited a variety of strains with diverse levels of contagion. Given the uncertainties surrounding the varying contagion levels of these strains, accurate identification of the virus lifecycle is vital for making informed health decisions, such as managing quarantines, designing effective vaccines [33], and setting appropriate national policies.

Considering the vulnerabilities of Latin American healthcare systems exposed by COVID-19, as demonstrated by the overburdening of healthcare systems and lack of access to treatments for a significant proportion of the population, the importance of thoroughly analyzing the virus's propagation behavior and evaluating the implemented policies cannot be overstated. This critical evaluation is a prerequisite for devising the best strategies for disease containment. A notable approach [34] emphasized the need for efficient distribution of healthcare centers within each country based on parameters such as accessibility, demand, and equity. This optimization framework importantly should encompass the strategic allocation of vaccination centers to ensure the most effective response to the disease. Building upon these insights, our contributions include:

- (i) Investigation of COVID-19 behavior in Latin America based on confirmed cases and deaths reported up until 31 December 2021.
- (ii) Mapping of the incidence rate by country to assess COVID-19 in Latin America.
- (iii) Forecasting of COVID-19 cases in Latin American countries until January 2022.
- (iv) Comparison of the trend changes in COVID-19 by country, observing and describing the number of infection waves each country experienced.
- (v) Formulation of the basic (instantaneous or effective) reproduction number (R_0) with values across different countries and the analysis of the effects of quarantine measures on transmission rates [35].
- (vi) Proposal of an epidemic model to predict future disease spread, which can serve as a tool for developing predictive scenarios.

The rest of this article is organized as follows. Section 2 outlines the work strategy employed to accomplish the aforementioned objectives, including descriptions of each statistical method utilized. In Section 3, we detail the case study on our epidemiological analysis for assessing and evaluating COVID-19 in Latin America countries, showing the results obtained from implementing these methods. Finally, in Section 4, we offer a discussion and conclusions based on the findings, as well as suggestions for future research.

2. Methodology

In this section, we initially explain the process of estimating the instantaneous (or basic) reproduction number, followed by a detailed application of the susceptible, exposed, infectious, and recovered (SEIR) model [36]. Subsequently, we present the applied techniques based on statistical analysis of stochastic processes over time [37]. This includes a focused approach to time-series and trend estimation, with particular attention given to identifying shifts in these trends.

2.1. Estimation of the Instantaneous Reproduction Number

In [10], a methodology was proposed assuming that the infectious profile of a patient only depends on the time that has passed since the patient acquired the illness, rather than the time that has elapsed since the epidemic started. Given the instant t , we can represent the distribution of the number of infected people as

$$I(t) | (I(0), \dots, I(t-1), R(t), w(t)) \sim \text{Poisson}(R(t) \sum_{s=1}^t I(t-s)w(s)),$$

where $w(t)$ is the probability distribution of the generation time of the outbreak, which can be considered as the probability distribution of the interval between successive cases of the illness [38]; and $R(t)$ is the instantaneous reproduction number represented as

$$R(t) = \frac{E(I(t))}{\sum_{s=1}^t I(t-s)w(s)},$$

which is estimated by replacing the incidence expected by its observed value given by

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

The estimation of the reproduction number can be highly variable when the time interval is small at the time of interpretation. To address this issue, in [39], a process was proposed for parameter estimation using a Bayesian approach, assuming a specific probability distribution for the reproduction number. In this approach, a prior distribution is chosen to be a gamma probability model with a mean and standard deviation of 5, and a posterior distribution is obtained as

$$R(t) | (I(t-\tau+1), \dots, I(t), w(t)) \sim \text{Gamma}\left(1 + \sum_{s=t-\tau+1}^t I(s), \frac{1}{5} + \sum_{s=t-\tau+1}^t \sum_{r=1}^s I(s-r)w(r)\right).$$

We use a window width of $\tau = 7$ days. The instantaneous reproduction number is estimated assuming a log-normal distribution with a mean of 4.7 days and a standard deviation of 2.9 days for the interval between successive cases. The implementation in the R software is provided by the function `estimate_r` from the `EpiEstim` package [10].

2.2. SEIR Model

One of the most critical challenges posed by the pandemic is the need to project the spread of the disease in the future and provide tools for better outbreak management in subsequent situations. In addition to the SEIR model, numerous other modeling methodologies have been utilized for COVID-19. For instance, susceptible, infectious, and recovered models, classical tools in epidemiology, have been refined and adapted for COVID-19 in several studies [40]. Machine learning models, capable of learning from complex, high-dimensional data, have been employed for forecasting the virus's spread, using techniques such as regression, decision trees, and random forests [41]. Further, deep learning models, a subset of machine learning structures, have leveraged artificial neural networks to simulate the virus's spread with a high degree of accuracy [42]. Each model has its unique strengths and limitations, and its applicability can depend on the specific objectives and constraints of a study. Given these considerations, we chose to utilize the SEIR model for our study and fit it as

$$\begin{aligned} S'(t) &= -\beta S(t)I(t)/N \\ E'(t) &= \beta S(t)I(t)/N - \sigma E(t) \\ I'(t) &= \sigma E(t) - (\gamma + \mu)I(t) \\ R'(t) &= \gamma I(t), \end{aligned}$$

where $S'(t) = dS(t)/dt$ and similarly for $E'(t), I'(t), R'(t)$; β denotes the transmission rate of the disease; σ represents the rate at which individuals transit from being infected to being infectious; μ signifies the rate of mortality due to the disease; γ represents the recovery rate of individuals; and N represents the total population size. The SEIR model is adjusted using the transmission functions, denoted as

$$\beta_t = \begin{cases} \beta, & \text{if } t \leq t_i; \\ \beta \left(1 - \frac{\delta_d(t-t_i)^2}{1+\delta_d(t-t_i)^2} \right), & \text{if } t_i < t < t_e; \\ \beta_{\min} + (\beta - \beta_{\min}) \left(\frac{\delta_i(t-t_e)^2}{1+\delta_i(t-t_e)^2} \right), & \text{if } t > t_e; \end{cases}$$

where β is the initial transmission rate; t is the time from the first detected case; t_i is the initial day of confinement minus a value equal to 2; t_e is the end of confinement; δ_d is the rate of decrease of transmission; 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 .

If we consider the variation rate of the accumulated infected people as the rate at which individuals complete their exposed period, as suggested in [43], we can state that

$$C'(T) = \sigma E(t),$$

where $C(T)$ is an auxiliary variable that keeps track of the cumulative number of infectious individuals, and $C'(T)$ keeps track of the curve of new cases (incidence). Therefore, the SEIR parameters are estimated by least squares [44] from the fit of the accumulated infected number and the number of accumulated confirmed cases until 31 December 2021, which is given as

$$(\hat{\beta}, \hat{\mu}, \hat{\sigma}, \hat{\gamma}, \hat{\delta}_d, \hat{\delta}_i) = \operatorname{argmin} \sum_{i=1}^T (C(t) - y_{t_i})^2,$$

where y_{t_i} is the time series that represents the observed number of accumulated confirmed cases at time t_i and $T = n$ days have elapsed since the first confirmed case [45].

2.3. Time-Series Models and Forecasting

ARIMA, standing for autoregressive integrated moving average, is employed in time-series analysis to comprehend and anticipate future trends [46–48]. In [49], it was argued that the ARIMA models must be suitable for dealing with complex and dynamic problems. These models use past observations for future predictions and incorporate unit root tests to check the stationarity of the series. The parameters of the ARIMA model are ascertained via the maximum likelihood (ML) estimation method, analogous to least squares, leading to efficient estimators. The ARIMA model is symbolized as $ARIMA(p, d, q)$, where $AR(p)$ signifies the autoregressive part of order p , $I(d)$ is the degree d of differencing for stationarity, and $MA(q)$ is the moving average part of order q . Then, respectively, p , d , and q represent the number of lagged observations, the level of differencing, and the lagged forecast errors. The $ARIMA(p, d, q)$ model is defined as

$$\varphi(B)(1 - B^d)Y_t = \varrho + \vartheta(B)\varepsilon_t,$$

where $\varphi(B)$ and $\vartheta(B)$ denote polynomials of orders p, q ; Y_t is a random variable with an observed value denoted by y_t ; and ε_t is the model random error. These polynomials should not have roots inside the unit circle, that is $|B| < 1$, to ensure the model causality and invertibility. Note that ϱ is a constant, and a differencing polynomial of order d is included in the forecast when $\varrho \neq 0$ [6]. It is pertinent to note the backshift operator B , a notation used for lagged sequences. For a time series y_t , the lagged series is written as $By_t = y_{t-1}$, and in a broader sense, $B^k y_t = y_{t-k}$.

The ARIMA model assumes stationarity of the time series, implying a consistent mean and variance over time. The parameter d in the ARIMA model refers to differencing value, which is used to achieve stationarity if the series is initially non-stationary.

Differentiation converts the series into differences between consecutive observations ($y_t - y_{t-1}$), and it can be applied more than once if the series remains non-stationary, as indicated by the parameter d in the ARIMA(p, d, q) model.

It is common to present the ARIMA(p, d, q) model in an alternate form as

$$Y_t = c + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} + \varepsilon_t,$$

where Y_t is a random variable at time t and its observed values at $t-1, \dots, t-p$ are y_{t-1}, \dots, y_{t-p} , respectively; ϕ_i are the autoregressive parameters of the model; θ_i are the moving average parameters; and ε_t is the model random error with observed values (residuals) at $t-1, \dots, t-p$ being e_{t-1}, \dots, e_{t-p} , respectively. The parameter p , the order of the autoregressive part, represents the number of lags of Y to be used as predictors. The parameter d is the order of integration, representing the number of times the data have had past values subtracted (also known as differencing), to make the time series stationary. Then, q is the order of the moving average part, representing the number of lagged forecast errors that should go into the ARIMA model.

The autocorrelation function (ACF) and partial autocorrelation function (PACF) are pivotal tools in time-series analysis, particularly for setting the parameters (p, q) of an ARIMA model. The ACF quantifies the correlation between time-series observations at different time points relative to the time lag between them, while the PACF determines the correlation between these observations when considering any correlations due to the values at shorter lags.

After identifying the model order (p, d, q), the parameters of the ARIMA model can be estimated, often via ML estimation. This estimation selects parameters that optimize the likelihood of the observed data given the model.

Suppose y_1, \dots, y_n is a time series of n observations. The likelihood function for an ARIMA model is defined as

$$L(\Theta; y_1, \dots, y_n) = f(y_1, \dots, y_n | \Theta),$$

where Θ denotes the parameter vector to be estimated, and f is the joint probability density function of the observed data for the parameter Θ . Often, the log-likelihood function is used due to its mathematical ease, which is in our case given by

$$l(\Theta; y_1, \dots, y_n) = \log(L(\Theta; y_1, \dots, y_n)).$$

The likelihood function, L namely, quantifies how well a particular statistical model explains the observed data. In other words, it is a measure of how likely the observed data are, given the specific parameters of the model.

For an ARIMA model, the likelihood function assumes the errors at each time point follow a normal (or Gaussian) distribution. The likelihood function for the model parameters is calculated by evaluating the normal probability density function at each data point and multiplying these densities for the errors because they are a white noise (independent). Mathematically, for a sample of size n and errors (independent) $\varepsilon_1, \dots, \varepsilon_n$, the likelihood function is stated as

$$L(\phi, \theta, \zeta^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\zeta}} \exp\left(-\frac{\varepsilon_i^2}{2\zeta^2}\right).$$

The ML estimates are the parameter values that maximize this log-likelihood function. The parameters to be estimated in an ARMA(p, q) model include autoregressive coefficients ϕ_j , for $j \in \{1, \dots, p\}$, moving average coefficients θ_j , for $j \in \{1, \dots, q\}$, and the error term variance ζ^2 .

Different approaches to the initial values of the estimation process, which are usually unknown, can result in different likelihood versions and parameter estimates. Due to the high dimensionality of the parameter space and potential multiple local maxima, numerical optimization methods are generally used to find the ML estimates.

For ARIMA models, automatic model selection can optimize predictive accuracy. The model order (p, d, q) can be selected minimizing the values of the Akaike (AIC) or Bayesian (BIC) information criteria expressed as

$$\text{AIC} = -2 \log(L) + 2(p + q + k + 1), \quad \text{BIC} = \text{AIC} + (\log(n) - 2)(p + q + k + 1),$$

where $k = 1$, if $\mu \neq 0$, and $k = 0$, if $\mu = 0$.

Through an automatic ARIMA modeling process, beneficially implemented in R [50], models were selected based on precision, using criteria such as mean absolute percentage error (MAPE), mean absolute deviation (MAD, also known as mean absolute error), and mean squared deviation (MSD, also known as mean squared error) to distinguish the best forecasts. These are common metrics used in statistics and machine learning to measure the accuracy of predictions or forecasts, especially in the context of regression and time-series analysis. This automatic process, outlined in Algorithm 1, enables an efficient and systematic approach to data modeling. The baseline model, representing future COVID-19 case forecasts, is presented as in [51].

Algorithm 1 Automatic ARIMA modeling procedure

Step 1: Select d , for $0 \leq d \leq 2$, using the Kwiatkowski–Phillips–Schmidt–Shin unit root test.

Step 2: Obtain p , d , and q by minimizing the AIC after differencing the data d times.

Step 3: Apply a stepwise search through model space to generate a simple model.

In this work, we apply ARIMA models to forecast time series of COVID-19 cases in a representative sample of Latin American countries, that is, Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Paraguay, Peru, and Uruguay, in South America; Dominican Republic in the Caribbean; Costa Rica and Belize in Central America; and Mexico in North America. This selection of countries was driven by the availability of COVID-19 data.

The time series is broken down into its components (data, trend, seasonality, and remainder) for visualization purposes and was used to predict COVID-19 cases for January 2022. In the process, the Dickey–Fuller and Ljung–Box tests [52] were employed to evaluate stationarity and examine autocorrelation, respectively [53]. Following these tests, we generated the ACF and PACF to initiate the order determination for the ARIMA models.

Another prevalent method for forecasting time series employed in this study is the Holt–Winters method [54]. Renowned for its efficacy with time-series data exhibiting trends and seasonal patterns, it serves as a valuable tool in our analysis. We used this approach as a comparative measure, scrutinizing its predictive abilities against those of the ARIMA models. The Holt–Winters method is characterized by three parameters: level (α), trend (ζ), and seasonality (λ). The model consists of two variations: additive and multiplicative. On the one hand, in the additive model, the forecast equation at $t + h$ is represented as

$$\hat{y}_{t+h} = l_t + hb_t + s_{t-m} + h_m^+,$$

where l_t is the level, b_t is the trend, and s_t is the seasonal component. Note that h is the forecast horizon; m represents the seasonal length; and h_m^+ is the smallest integer greater than or equal to h/m .

The level, trend, and seasonal equations at t are formulated as

$$\begin{aligned} l_t &= \alpha(y_t - s_{t-m}) + (1 - \alpha)(l_{t-1} + b_{t-1}) \\ b_t &= \beta(l_t - l_{t-1}) + (1 - \zeta)b_{t-1} \\ s_t &= \lambda(y_t - l_t) + (1 - \lambda)s_{t-m}. \end{aligned}$$

The multiplicative model, on the other hand, uses the forecast equation at $t + h$ established as

$$\hat{y}_{t+h} = (l_t + hb_t)s_{t-m} + h_m^+$$

The level, trend, and seasonal equations at t are slightly different and presented as

$$\begin{aligned} l_t &= \alpha \left(\frac{y_t}{s_{t-m}} \right) + (1 - \alpha)(l_{t-1} + b_{t-1}) \\ b_t &= \zeta(l_t - l_{t-1}) + (1 - \zeta)b_{t-1} \\ s_t &= \lambda \left(\frac{y_t}{l_t} \right) + (1 - \lambda)s_{t-m}. \end{aligned}$$

In all these equations, α , ζ , and λ are constants that must be optimized.

2.4. Trend Estimation

To estimate the daily trend of COVID-19 cases in Latin America, we utilize a time-series model [9] expressed as

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

where y_t represents the confirmed COVID-19 cases at time t , and the estimated trend, \hat{m} say, is determined by a moving average [9,10]. The constant d governs the range and degree of smoothing [9]. Decreasing d results in smoother moving averages that capture trends effectively but may induce false alarms. Conversely, increasing d minimizes false alarms but may compromise the identification of ongoing trends [9,51]. Given the nature of our data, we set $d = 7$, which corresponds to a 15-day window for trend estimation.

2.5. Detection of Trend Shifts

In epidemiological studies, the utilization of statistical models is crucial for understanding the direct impacts of diseases on various social dimensions, including global markets [5]. A valuable tool in these endeavors is the breakpoint methodology, used for estimating shifts in trends. In the context of the COVID-19 pandemic, we adopted this approach to discern trend changes across Latin American countries reporting confirmed cases.

Given the unique disease progression in each country, modulated by factors such as initial exposure date and implemented containment measures, we were able to identify shifts in trends regarding confirmed cases and fatalities.

Following [10], the detection of trend periods is simplified by identifying the breakpoints that divide the series. For a time series, y_t namely, with a linear trend, we hypothesize multiple segments, each potentially displaying a distinct trend. This is represented using a piecewise linear regression model stated as

$$Y_t = \begin{cases} \alpha_1 + \zeta_1 t + \epsilon_t, & t \in \{1, \dots, p\}; \\ \alpha_2 + \zeta_2 t + \epsilon_t, & t \in \{p + 1, \dots, T\}. \end{cases}$$

In this formulation, $\alpha_1, \alpha_2, \zeta_1, \zeta_2$ are the regression model coefficients, and $\epsilon(t)$ represents a random perturbation with zero mean and constant variance.

To establish p as a change point, we test the hypotheses presented as

$$\begin{aligned} \mathcal{H}_0: & \alpha_1 = \alpha_2, \quad \zeta_1 = \zeta_2 \\ \mathcal{H}_1: & \alpha_1 \neq \alpha_2, \quad \zeta_1 \neq \zeta_2. \end{aligned}$$

The corresponding F-statistic enables us to contrast the residuals of the piecewise model against an unsegmented model given by

$$Y_t = \alpha + \zeta t + \epsilon_t.$$

As pointed out in [55,56], we might not know the exact number of breakpoints in a real-world time series. Therefore, a test is proposed to verify if the series has only m breakpoints, leading to a piecewise linear model formulated as

$$Y_t = \alpha_j + \zeta_j t + \epsilon_t, \quad t = t_{j-1} + 1, \dots, t_{ji}, \quad j \in \{1, \dots, m+1\},$$

where $t_0 = 0$ and $t_{m+1} = n$ [56].

Denote the m breakpoints identified in the test as i_1, \dots, i_m . Then, we define the sum of squared residuals (SRS) for the segments of the model as

$$\text{SRS}(i_1, \dots, i_m) = \sum_{j=1}^{m+1} \text{SRS}(i_{j-1}, i_j),$$

where $\text{SRS}(i_{j-1}, i_j)$ represents the SRS for a given segment, between i_{j-1} th and i_j th breakpoints. This SRS is calculated by fitting the segmented model to that specific interval and summing the squares of the differences between the observed values and the values predicted by the model. As a result, we have that

$$(\hat{i}_1, \dots, \hat{i}_m) = \text{argmin SRS}(i_1, \dots, i_m).$$

This methodology is applied as a complementary element of the trend estimation, enhancing our understanding of the time-series trend shifts. We used the function `breakpoints` from the `strucchange` package of the R statistical software for this purpose [50,57]. The identified breakpoints are visually represented in the time-series plot using blue and red panels, where each panel denotes a period with the trend remaining consistent.

3. Case Study

In this section, we present the main results obtained based on the methodology outlined in Section 2 and the analyses conducted throughout that section. We conduct an exploratory phase of our analysis, which involved the creation of choropleth maps to represent the distribution of COVID-19 cases across the Latin American countries upon study. To estimate non-stationarity, we used a moving average model, enabling us to identify significant shifts in the time-series trends. Subsequently, we calculated the real-time reproduction number using a Bayesian method. This number was superimposed on the quarantine timelines for each country in our study to assess the behavior of R_0 during and post confinement. To conclude our analysis, we deploy a SEIR model to determine the epidemiological curves based on realistic estimates of exposed and infected individuals.

3.1. Data, Methodology, and Software

The data used for this analysis comprise cases and fatalities confirmed by COVID-19 on a national level in Latin America, drawn from public data repositories at John Hopkins University and Our World in Data. The open-source repositories can be secured from <https://github.com/CSSEGISandData/COVID-19> (accessed on 20 May 2023), <https://github.com/owid/covid-19-data> (accessed on 15 April 2022).

We have also consulted the institutional websites of each country to gather officially recognized dates for the implementation of quarantines. The dataset encompasses the period from 23 February 2020, when the first COVID-19 case was confirmed among the thirteen countries included in this study, to 31 December 2021.

Our methodology is summarized in Figure 1.

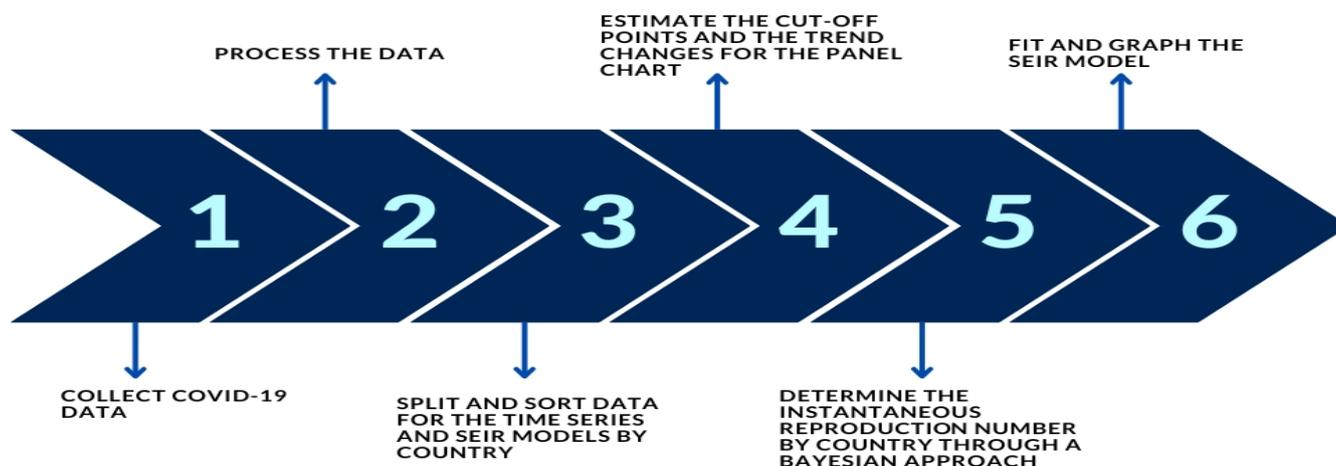


Figure 1. Proposed methodology. Source: The authors.

We utilize the Tableau software [58] for the development of the exploratory analysis graphs, and the R software [50] was instrumental in producing data summaries and executing the aforementioned methodologies. Data handling was performed using the `dplyr`, `openxlsx`, `reshape2`, `xtable`, and `tidyverse` packages. Forecasting of time series incorporated the `forecast`, `tseries`, `TTR`, `lubridate`, and `zoo` packages, while the `ggplot2` package was utilized for the design of graphs.

3.2. Exploratory Data Analysis

The results presented next highlight the distribution of confirmed COVID-19 cases across Latin America. Figure 2 illustrates that Brazil had the highest number of confirmed cases throughout the first two years of the pandemic. It is notable that neighboring countries generally exhibited lower infection rates compared to Brazil, with the exception of Colombia, which experienced a significant number of cases. Moreover, Belize, with its smaller population size, had the lowest number of cases, as evident when considering the incidence rate.

In Latin America, the deaths have been high [59], particularly in Brazil [9] and Chile [10]. However, there is no consistent relationship between the number of deaths and the number of confirmed cases in each country, except for Brazil and Belize, which have the highest and lowest numbers of confirmed cases and deaths, respectively (Figure 3). For instance, Peru and Chile have similar numbers of confirmed cases (2,296,831 and 1,806,494, respectively), but Peru has reported 203,399 confirmed deaths compared to Chile that reported 38,271. Similarly, Costa Rica and Ecuador both have around 570,556 confirmed cases, but Costa Rica reported 7353 deaths while Ecuador informed 21,043.

To accurately assess the situation in each country, it is important to consider the national incidence rate, which takes into account the population size. It is crucial to recognize that having 30,000 cases in a population of 300,000 is different from having 30,000 cases in a population of 3,000,000. Failing to calculate the incidence rate can lead to incorrect conclusions. Therefore, the incidence rate is calculated by adjusting the number of confirmed cases and deaths based on the population size in each country.

When examining the incidence rate adjusted by the population size, it reveals interesting insights into the impact of COVID-19 across different countries. Surprisingly, countries such as Argentina, Uruguay, Costa Rica, and Colombia show higher incidence rates compared to Brazil, despite not having the highest number of cases. In addition, Mexico, Colombia, and Ecuador have lower incidence rates, indicating a more consistent relationship between confirmed cases and population size in these countries (Figure 4).



Figure 2. Latin America choropleth map of COVID-19 accumulative confirmed cases. The colors show a scale of confirmed cases, being the color with the lowest tonality the country with the fewest cases, until 31 December 2021. Source: The authors produced with 2022 Mapbox OpenStreetMap.



Figure 3. Latin America choropleth map of COVID-19 accumulative deaths. The colors show a scale of confirmed deaths, being the color with the lowest tonality the country with the fewest cases, until 31 December 2021. Source: the authors produced with 2022 Mapbox OpenStreetMap.



Figure 4. Latin America dot map of COVID-19 incidence rate by country. The size of every dot shows the incidence rate level, being the dot with the highest size the country with the highest incidence rate, until 31 December 2021. Source: the authors produced with 2022 Mapbox OpenStreetMap.

3.3. Epidemic Model

Figures 5 and 6 illustrate the behavior of the instantaneous reproduction number (R_0) by presenting its posterior median along with the corresponding posterior distribution of the parameter. The blue panels superimposed on the figures represent the periods of quarantine, allowing us to observe the impact of these measures on R_0 during and after the implemented confinement.

In countries such as Argentina, Belize, Bolivia, Brazil, Chile, Dominican Republic, and Ecuador, the implementation of quarantine measures led to a significant decrease in the instantaneous reproduction number. However, in Peru and Uruguay, despite having similar quarantine periods (and longer in some cases), R_0 did not decrease significantly.

After the end of the quarantine period, Uruguay experienced a notable increase in R_0 , while Peru exhibited a varying behavior with frequent fluctuations throughout the studied period. The highest values of R_0 were observed in Mexico and Belize, ranging from 10 to 20 units, indicating that one infected individual could potentially transmit the virus to a range between 10 and 20 other people. In addition, Peru, Paraguay, Costa Rica, Bolivia, and Argentina had lower reproduction values, with a maximum value of 4 units, implying lower transmission rates. For the SEIR model, Colombia was considered due to its relatively small population size compared to other countries such as Mexico and its high level of contagion. An initial confinement period was defined from the 19th day (counting from the first confirmed case) and lasted until the 178th day.

The SEIR model parameters were fitted based on the criteria earlier described using data up to 31 December 2021. Figure 7 shows the prevalence of COVID-19 infected cases in Colombia and its relationship with the daily confirmed cases. The estimates of the SEIR parameters are: $\hat{\beta} = 0.3250$, $\hat{\mu} = 0.0002$, $\hat{\sigma} = 1.3608$, $\hat{\gamma} = 0.2363$, $\hat{\delta}_d = 0.0000$, and $\hat{\delta}_i = 0.0603$. The fitting and estimation of the SEIR model provide valuable insights into the prevalence of the disease, capturing the true number of infections at a given time.

The model estimates a significant difference between the confirmed cases reported and the COVID-19 number of infections. This disparity is evident, with the model estimating up to 60,000 infected individuals compared to the recorded peak of nearly 30,000 cases during the epidemic. These findings highlight the importance of accounting for undetected or unreported cases and emphasize the need for comprehensive testing and surveillance measures to accurately assess the true extent of the epidemic.

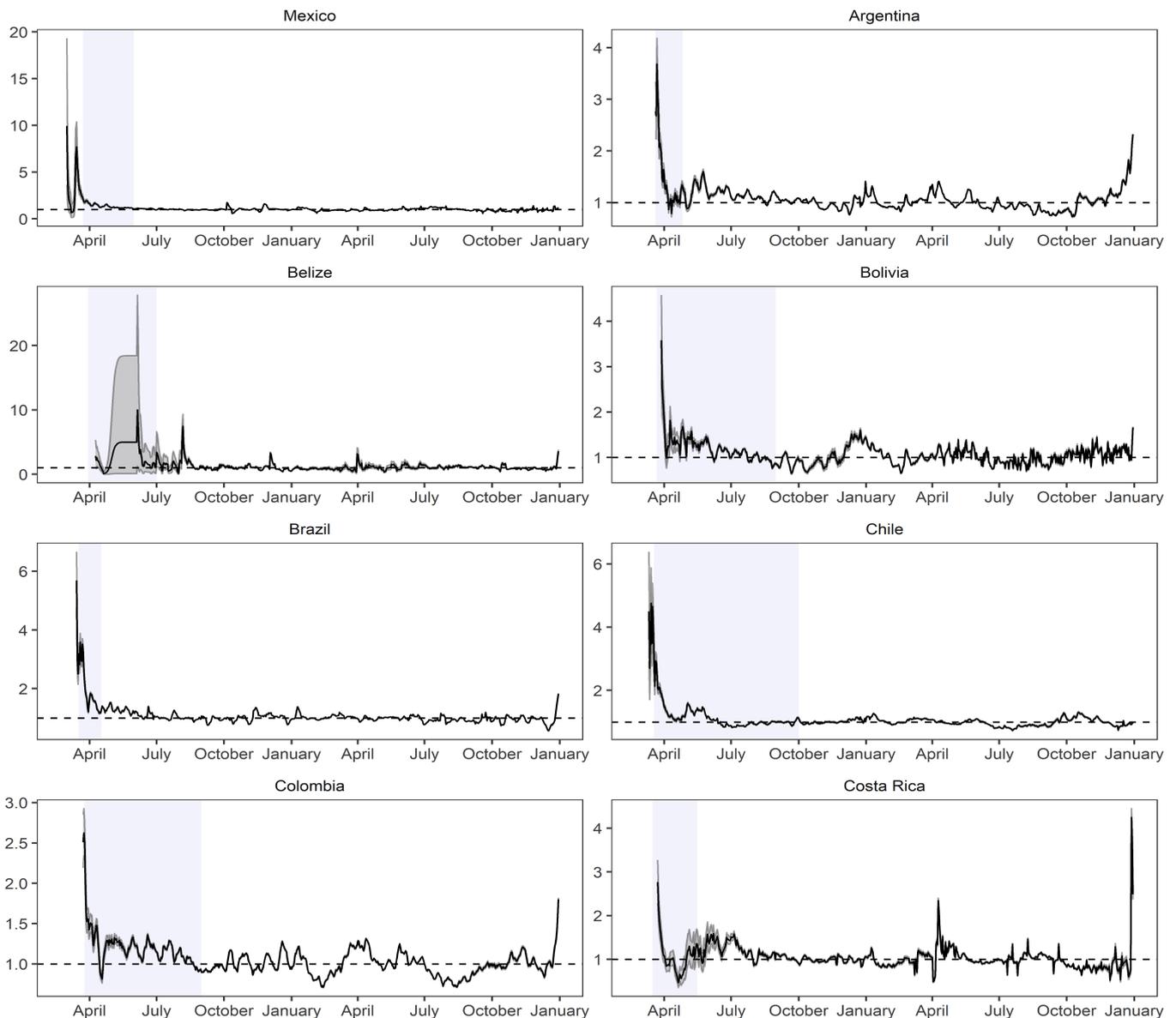


Figure 5. Instantaneous reproduction number estimated implementing its posterior median and disaggregated by country ranging from Mexico to Costa Rica. Source: The authors.

3.4. Main Results

Next, we present the main results obtained based on the methodology outlined in Section 2 and the analyses conducted throughout that section.

For forecasting purposes, Uruguay was chosen as an example. ARIMA models with various orders were fitted, and the AIC was used to select the most appropriate model. The model with the lowest AIC value was chosen. The accuracy of the models was evaluated using the MAPE and the ARIMA (3,1,1) structure.

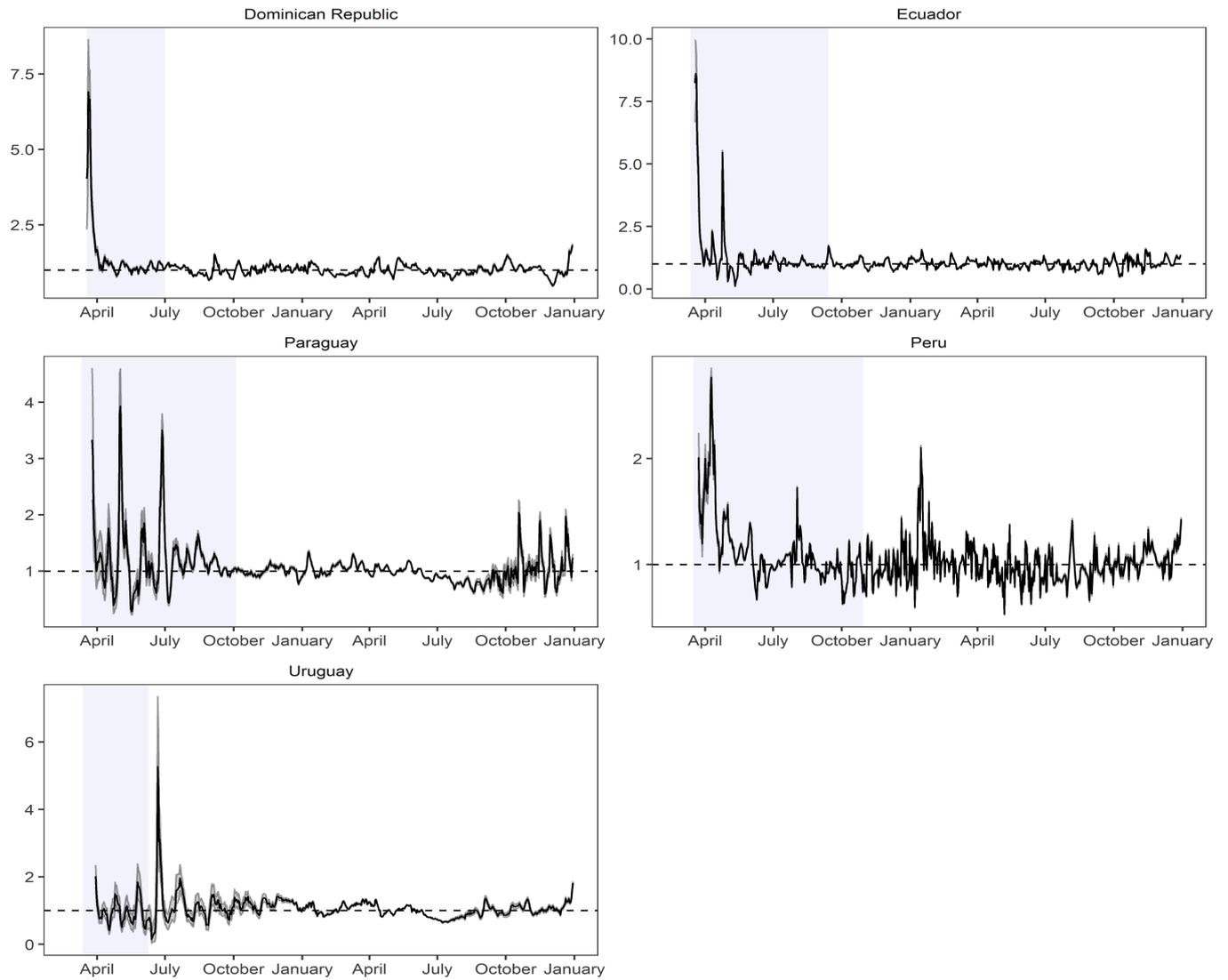


Figure 6. Instantaneous reproduction number estimated implementing its posterior median and disaggregated by country ranging from Dominican Republic to Uruguay. Source: The authors.

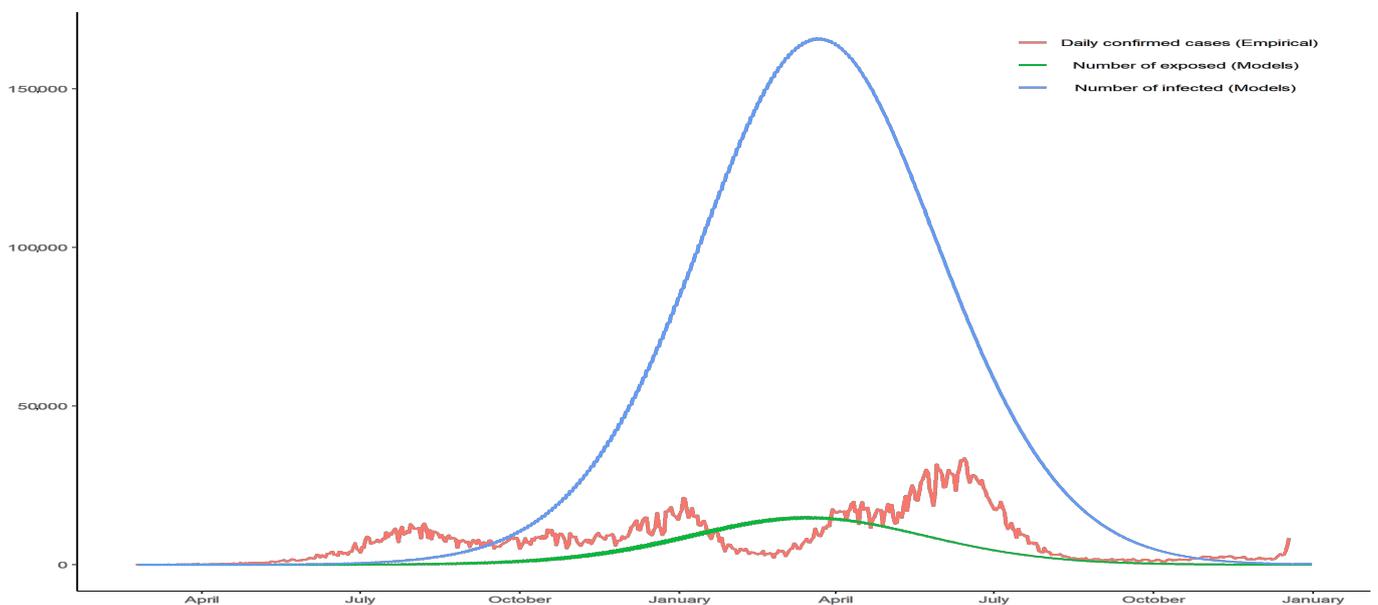


Figure 7. Prevalence of infected cases in Colombia and its relationship with the daily confirmed cases. Source: the authors.

To determine the best forecast model, an automatic ARIMA model was also fitted and compared to the previously selected model. Three models were evaluated for forecasting: Holt-Winters, automatic ARIMA, and ARIMA (3,1,1). After conducting evaluations using the same parameters, Holt-Winters was identified as the best-fitted model and provided the most accurate forecast. Consequently, the forecast was generated using the Holt-Winters model up until January 2022.

By comparing the forecast made with the model to the behavior of the disease in the first month of 2022, we can observe a remarkably accurate fit to the empirical data. The forecasted values closely align with the observed trends and patterns, validating the effectiveness of the chosen model in capturing the dynamics of the disease. Note that in countries where the first wave occurred between June and July, there was a noticeable pattern of sharp increases and decreases in COVID-19 cases throughout 2020 and 2021. This can be observed in the distinct trends displayed by each country in Figures 8 and 9, with Dominican Republic exhibiting one of the most significant trend changes. Ecuador and Belize experienced relatively stable case numbers without a prominent surge or decline. Furthermore, we identify plateau periods in all countries, indicating the impact of implemented quarantines by the respective health authorities. These stability periods can be compared to the estimated instantaneous reproduction number superimposed on the quarantine periods [60], revealing the effectiveness of quarantine measures in controlling the spread of the disease. It is worth highlighting the situation in Uruguay, where the onset of COVID-19 waves occurred later compared to other countries. Despite experiencing a significant wave when the disease initially entered the country, Uruguay generally maintained a low infection in subsequent periods.

The present analysis of each Latin American country clearly demonstrates heterogeneity, as evident from the identified cutoff dates presented in Table 1. Notably, there is similarity in the trend changes among Mexico, Brazil, Chile, and Dominican Republic, as observed in Figures 8 and 9. Table 1 reveals that the first wave of COVID-19 cases occurred in Mexico, Bolivia, Brazil, Chile, Colombia, and Peru in the early days of June. Similarly, Colombia, Costa Rica, and Ecuador experienced their first trend change in late June and early July. Argentina, Belize, and Paraguay encountered their first wave in late August and early September. Uruguay had the latest onset of the COVID-19 wave, with cases emerging in early December. The impact of implemented quarantines on the population is reflected in the dynamics of the instantaneous reproduction number [56].

Table 1. COVID-19 confirmed cases cut-off points by country. Source: the authors.

Country	Cut-Off Point				
	1st	2nd	3rd	4th	5th
Argentina	3 August 2020	13 November 2020	5 April 2021	16 September 2021	N/A
Belize	29 September 2020	09 November 2020	15 August 2021	N/A	N/A
Bolivia	8 June 2020	17 September 2020	27 December 2020	7 April 2021	17 June 2021
Brazil	2 June 2020	21 November 2020	2 March 2021	23 September 2021	N/A
Chile	2 June 2020	6 December 2020	17 March 2021	28 June 2021	N/A
Colombia	8 July 2020	10 April 2021	23 June 2021	N/A	N/A
Costa Rica	21 June 2020	7 January 2021	18 April 2021	20 September 2021	N/A
Dominican Republic	10 June 2020	18 November 2020	27 February 2021	12 July 2021	N/A
Ecuador	24 June 2020	12 January 2021	15 May 2021	24 August 2021	N/A
Mexico	2 June 2020	18 November 2020	27 February 2021	11 June 2021	20 September 2021
Paraguay	11 August 2020	23 November 2020	7 March 2021	09 July 2021	N/A
Peru	3 June 2020	25 September 2020	19 January 2021	5 June 2021	N/A
Uruguay	6 December 2020	17 March 2021	26 June 2021	N/A	N/A

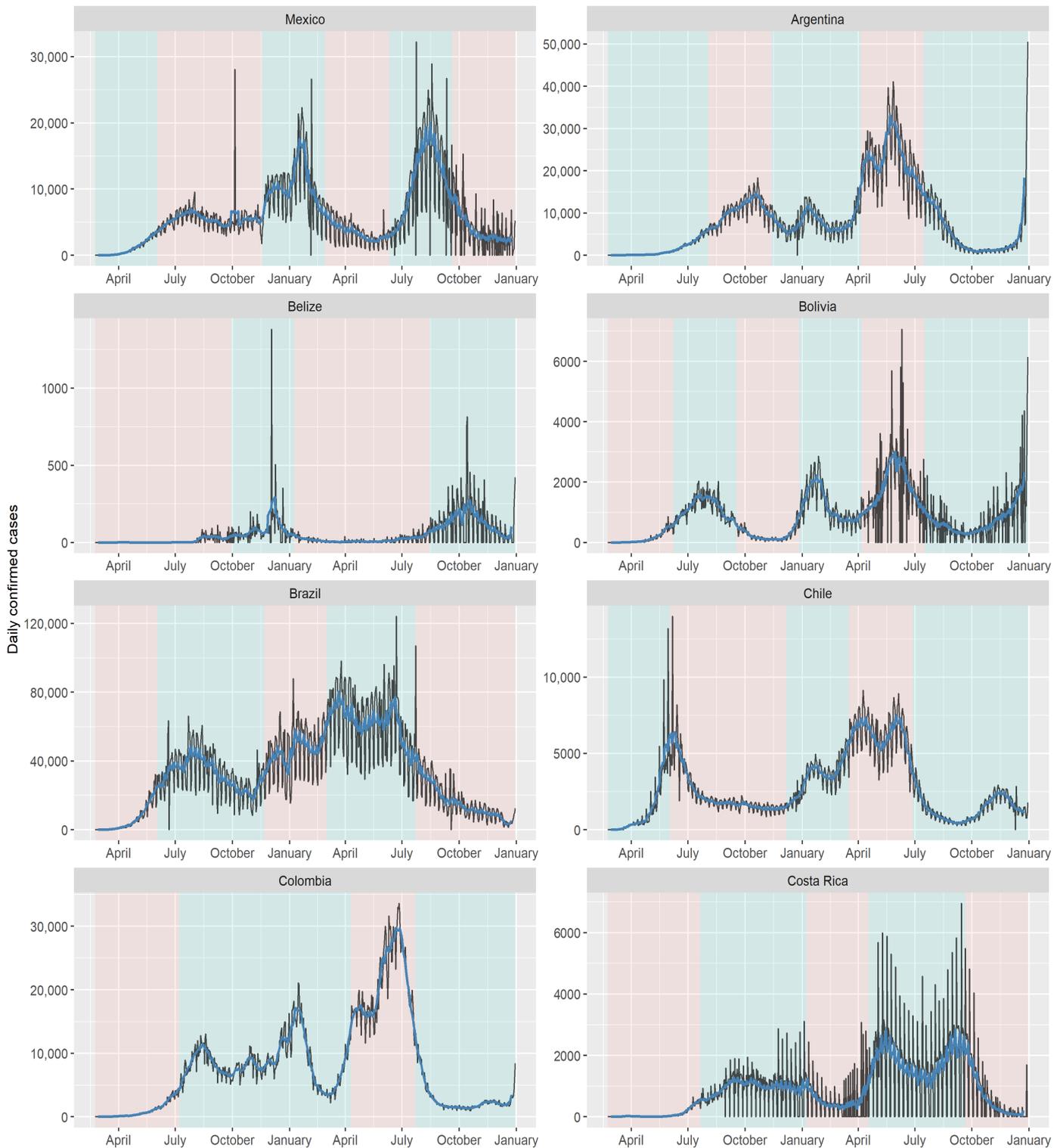


Figure 8. COVID-19 cases from Mexico to Costa Rica with the moving average superimposed and the cut-off points from Table 1. The vertical axis of each graph has an own scale according the country. Source: the authors.

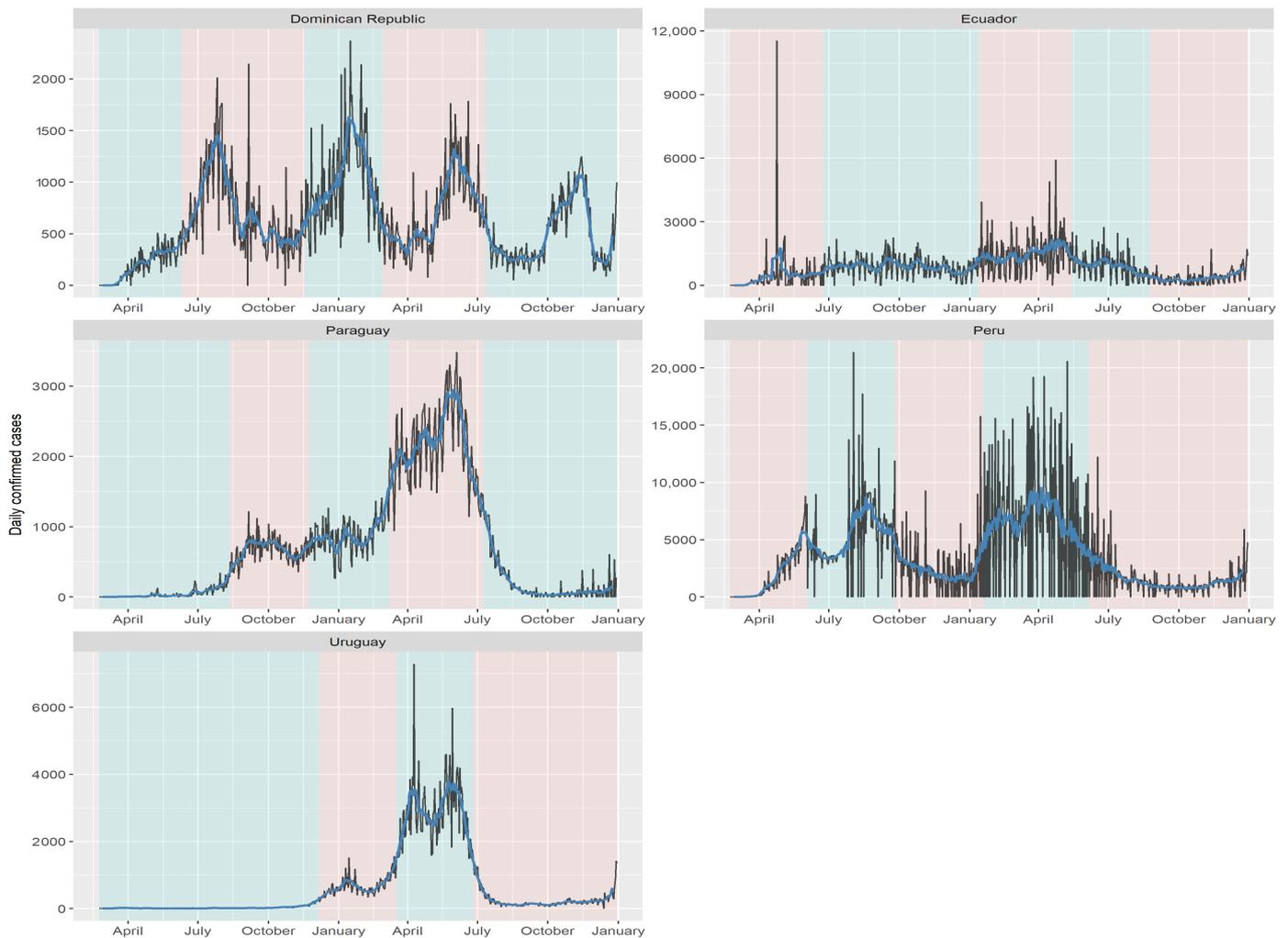


Figure 9. COVID-19 cases from Dominican Republic to Uruguay with the moving average superimposed and the cut-off points from Table 1. The vertical axis of each graph has an own scale according the country. Source: the authors.

4. Discussion and Conclusions

The central objective of this article was to scrutinize the spread of COVID-19 across Latin America and assess the effectiveness of implemented quarantines. Through comprehensive time-series analysis and the evaluation of trend alterations via moving averages, we uncovered distinct patterns in COVID-19 cases across neighboring countries. A compelling example of this emerged in the comparison of Mexico and Belize, which exhibited significant differences in their outbreak spread behaviors. These disparities could be attributed to variations in health infrastructure and economic strength, with Mexico typically having more robust systems in place. Similarly, our examination revealed parallel trend patterns between Colombia and Peru, as well as with Peru and Bolivia. Such patterns could be influenced by unique government strategies and policies enacted in these countries, highlighting the significance of diverse factors such as the timing and stringency of measures when assessing COVID-19 dynamics.

A unique contribution of our research lies in the methodological approach we adopted, utilizing time-series analysis and moving averages. This provided a refined understanding of the pandemic's trajectory and enabled us to dissect differences in outbreak spread behaviors among neighboring countries. It also allowed us to study the impact of various factors, including population size and public health policies, on these behaviors.

In assessing the data, it was evident that larger nations, such as Brazil and Mexico, experienced a high burden of infections. Yet, smaller countries, such as Colombia, outpaced Mexico in confirmed cases. This discrepancy underscores the fact that factors beyond population size can influence the spread and severity of the disease. Furthermore, the absence of a directly proportional relationship between the number of cases and deaths in some nations warrants further exploration. This observation underscores the necessity for in-depth research and more comprehensive studies like our current work, to understand the complex dynamics of COVID-19 spread and the multitude of influencing factors.

Our research underscores the effectiveness of the forecast models used, as demonstrated in the comparison presented in Figures 10 and 11. These models have proven to be valuable tools in understanding the disease’s behavior and forecasting its future trajectory. The novelty of our contribution lies in applying these models specifically within the Latin American context, providing a deeper understanding of the region’s situation.

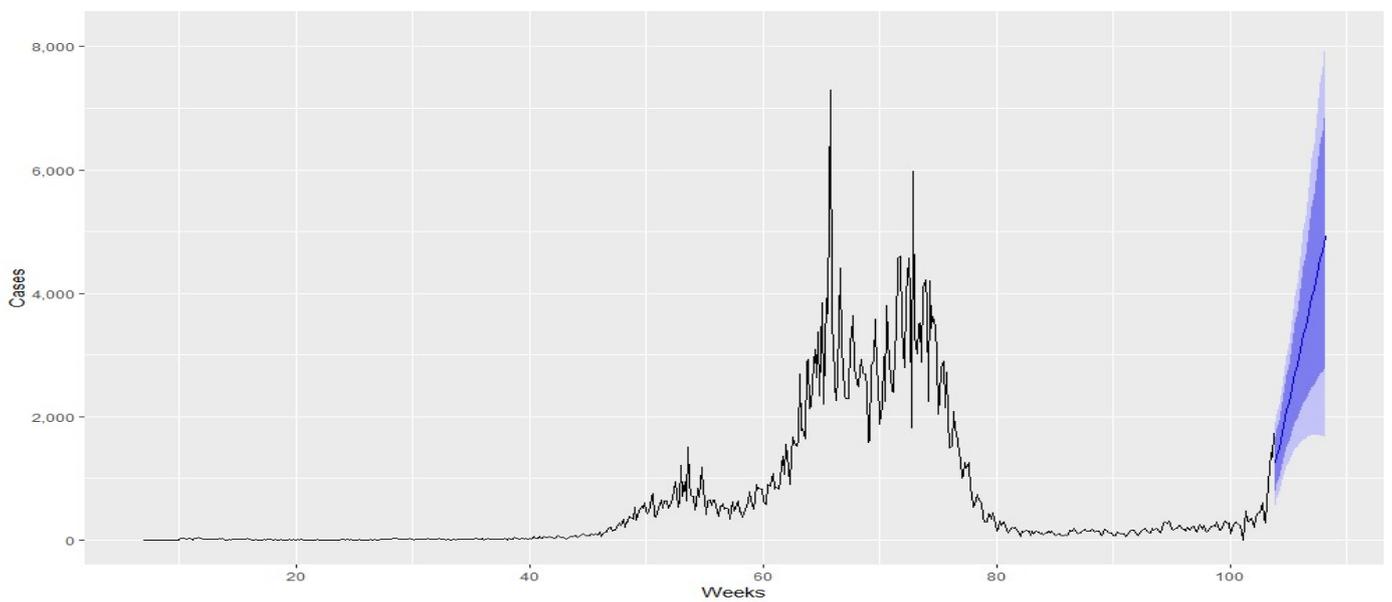


Figure 10. Forecast through the Holt-Winters model in Uruguay to January 2022. Source: the authors.

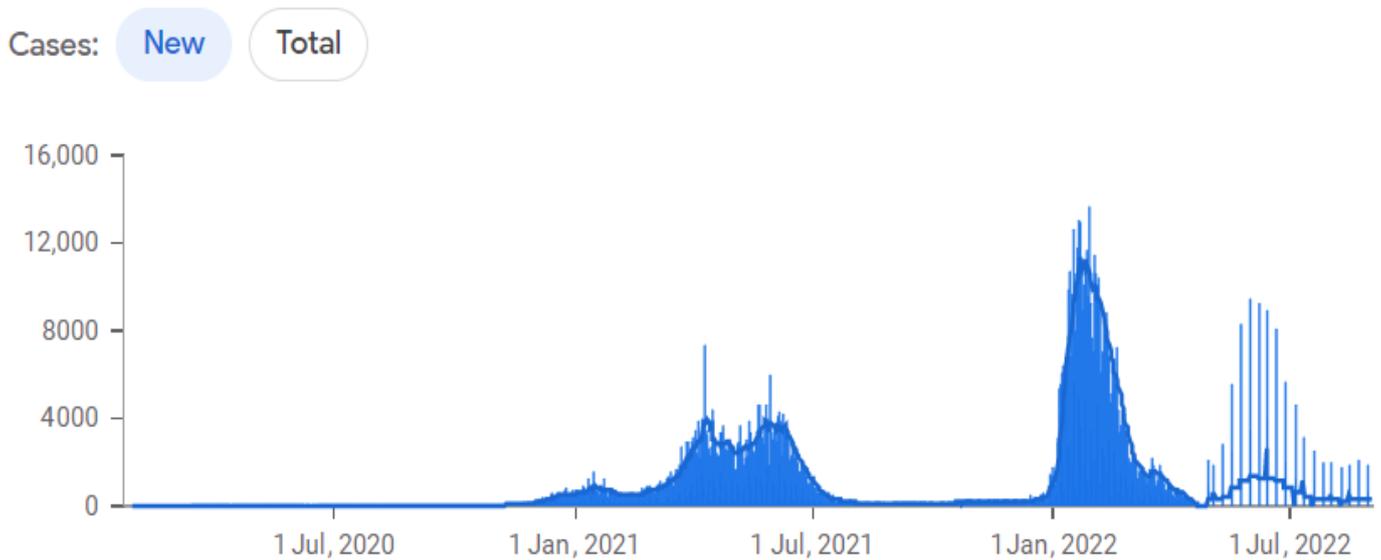


Figure 11. COVID-19 cases in Uruguay until 20 July 2022. Source: John Hopkins University.

The study draws parallels with the observation given in [39] regarding the pivotal role of the reproduction number in epidemic control strategies. Despite a decrease in this number following quarantine measures, we found a consistent increase in confirmed COVID-19 cases, indicating the disease's spread was not effectively halted. This was particularly notable in Uruguay, where a more extended quarantine period could potentially have suppressed both the infection surge and reproduction number. These findings suggest concrete recommendations for policy-makers and public health officials, underlining the importance of a careful and gradual resumption of everyday activities after the quarantine.

This article considered the COVID-19 incidence by accounting for new cases in Latin America and considering the incidence rate and population size per country. The primary focus was to analyze and compare the spread of the disease between countries. However, we recommend a comprehensive analysis that incorporates the incidence rate for more accurately evaluating COVID-19 severity in future research.

By modifying the SEIR model to incorporate time-dependent transmission rates, we provided a closer approximation of the epidemiological situation. In Colombia, for instance, the estimated infection curve was approximately double the number of registered cases, revealing a critical discrepancy that needs to be addressed for improved epidemiological surveillance.

Throughout this work, the different realities experienced during the pandemic in Latin America were demonstrated. This work reaffirms the critical role of statistical and mathematical methodologies in understanding and addressing outbreaks. Yet, as new data and insights become available, continuous refinement and updating of these methodologies is required.

Future research should consider alternative methodological approaches not explored in this article, such as principal component analysis for country classification [12,59,61], as well as a theoretical approach based on the geometric Brownian motion [16–18]. Such studies can foster the development of more comprehensive, effective, and region-specific response strategies to future outbreaks.

Author Contributions: Conceptualization, V.L., E.A., J.M., and C.C.; data curation, E.A.; formal analysis, V.L., E.A., J.M., and C.C.; investigation, V.L., E.A., J.M., and C.C.; methodology, V.L., E.A., J.M., and C.C.; validation, V.L. and C.C.; writing—original draft, E.A., J.M., and C.C.; writing—review and editing, V.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by FONDECYT grant number 1200525 (V.L.) from the National Agency for Research and Development (ANID) of the Chilean government under the Ministry of Science, Technology, Knowledge, and Innovation.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would also like to thank three reviewers for their constructive comments which led to improvement in the presentation of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dong, E.; Du, H.; Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **2020**, *20*, 533–534. [CrossRef]
2. Chakraborty, T.; Ghosh, I. Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis. *Chaos Solitons Fractals* **2020**, *135*, 109850. [CrossRef]
3. Fraser, C.; Cummings, D.A.; Klinkenberg, D.; Burke, D.S.; Ferguson, N.M. Influenza transmission in households during the 1918 pandemic. *Am. J. Epidemiol.* **2011**, *174*, 505–514. [CrossRef] [PubMed]
4. WH Organization. Coronavirus Disease (COVID-19) Pandemic. 2020. Available online: www.who.int/emergencies/diseases/novel-coronavirus-2019 (accessed on 19 May 2023).

5. De la Fuente-Mella, H.; Rubilar, R.; Chahuán-Jiménez, K.; Leiva, V. Modeling COVID-19 cases statistically and evaluating their effect on the economy of countries. *Mathematics* **2021**, *9*, 1558. [[CrossRef](#)]
6. Brockwell, P.J.; Davis, R.A. *Introduction to Time Series and Forecasting*; Springer: New York, NY, USA, 2002.
7. Box, G.; Jenkins, G. *Time Series Analysis Forecasting and Control*; Wiley: Hoboken, NJ, USA, 2015.
8. Hamilton, J.D. *Time Series Analysis*; Princeton University Press: Princeton, NJ, USA, 2020.
9. Ospina, R.; Leite, A.; Ferraz, C.; Magalhaes, A.; Leiva, V. Data-driven tools for assessing and combating COVID-19 out-breaks based on analytics and statistical methods in Brazil. *Signa Vitae* **2022**, *18*, 18–32.
10. Jerez-Lillo, N.; Alvarez, B.L.; Gutierrez, J.M.; Figueroa-Zuniga, J.; Leiva, V. A statistical analysis for the epidemiological surveillance of COVID-19 in Chile. *Signa Vitae* **2022**, *18*, 19–30.
11. Fierro, R.; Leiva, V.; Balakrishnan, N. Statistical inference on a stochastic epidemic model. *Commun. Stat. Simul. Comput.* **2015**, *44*, 2297–2314. [[CrossRef](#)]
12. Intissar, A. A mathematical study of a generalized SEIR model of COVID-19. *SciMed. J.* **2020**, *2*, 30–67. [[CrossRef](#)]
13. Boselli, P.M.; Soriano, J.M. COVID-19 in Italy: Is the mortality analysis a way to estimate how the epidemic lasts? *Biology* **2023**, *12*, 584. [[CrossRef](#)]
14. Lenhart, S.; Workman, J.T. *Optimal Control Applied to Biological Models*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2007.
15. Gondim, J.A.M.; Machado, L. Optimal quarantine strategies for the COVID-19 pandemic in a population with a discrete age structure. *Chaos Solitons Fractals* **2020**, *140*, 110166. [[CrossRef](#)]
16. Stojkoski, V.; Sandev, T.; Kocarev, L.; Pal, A. Geometric Brownian motion under stochastic resetting: A stationary yet nonergodic process. *Phys. Rev. E* **2021**, *104*, 014121. [[CrossRef](#)]
17. Vinod, D.; Cherstvy, A.G.; Wang, W.; Metzler, R.; Sokolov, I.M. Nonergodicity of reset geometric Brownian motion. *Phys. Rev. E* **2022**, *105*, L012106. [[CrossRef](#)] [[PubMed](#)]
18. Vinod, D.; Cherstvy, A.G.; Metzler, R.; Sokolov, I.M. Time-averaging and nonergodicity of reset geometric Brownian motion with drift. *Phys. Rev. E* **2022**, *106*, 034137. [[CrossRef](#)]
19. Eikenberry, S.E.; Mancuso, M.; Iboi, E.; Phan, T.; Eikenberry, K.; Kuang, Y.; Kostelich, E.; Gumel, A.B. To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. *Infect. Dis. Model.* **2020**, *5*, 293–308. [[CrossRef](#)] [[PubMed](#)]
20. Gondim, J.A.M. Preventing epidemics by wearing masks: An application to COVID-19. *Chaos Solitons Fractals* **2021**, *143*, 110599. [[CrossRef](#)]
21. Stutt, R.O.J.H.; Retkute, R.; Bradley, M.; Gilligan, C.A.; Colvin, J. A modelling framework to assess the likely effectiveness of facemasks in combination with ‘lock-down’ in managing the COVID-19 pandemic. *Proc. R. Soc. A* **2020**, *476*, 20200376. [[CrossRef](#)] [[PubMed](#)]
22. Vasconcelos, G.L.; Brum, A.A.; Almeida, F.A.G.; Macêdo, A.M.S.; Duarte-Filho, G.C.; Ospina, R. Standard and anomalous waves of COVID-19: A multiple-wave growth model for epidemics. *Braz. J. Phys.* **2021**, *51*, 1867–1883. [[CrossRef](#)]
23. Vasconcelos, G.L.; Macedo, A.M.S.; Duarte-Filho, G.C.; Brum, A.A.; Ospina, R.; Almeida, F.A.G. Power law behaviour in the saturation regime of fatality curves of the COVID-19 pandemic. *Sci. Rep.* **2021**, *11*, 4619. [[CrossRef](#)]
24. Wu, K.; Darcet, D.; Wang, Q.; Sornette, D. Generalized logistic growth modeling of the COVID-19 outbreak: Comparing the dynamics in provinces in China and in the rest of the world. *Nonlinear Dyn.* **2020**, *101*, 1561–1581. [[CrossRef](#)]
25. Pérez-Ortega, J.; Almanza-Ortega, N.N.; Torres-Poveda, K.; Martínez-González, G.; Zavala-Díaz, J.C.; Pazos-Rangel, R. Application of data science for cluster analysis of COVID-19 mortality according to sociodemographic factors at municipal level in Mexico. *Mathematics* **2022**, *10*, 2167. [[CrossRef](#)]
26. Cavalcante, T.; Ospina, R.; Leiva, V.; Cabezas, X.; Martin-Barreiro, C. Weibull regression and machine learning survival models: Methodology, comparison, and application to biomedical data related to cardiac surgery. *Biology* **2023**, *12*, 442. [[CrossRef](#)]
27. Sokhansanj, B.A.; Zhao, Z.; Rosen, G.L. Interpretable and predictive deep neural network modeling of the SARS-Cov-2 spike protein sequence to predict COVID-19 disease severity. *Biology* **2022**, *11*, 1786. [[CrossRef](#)] [[PubMed](#)]
28. Sardar, I.; Akbar, M.A.; Leiva, V.; Alsanad, A.; Mishra, P. Machine learning and automatic ARIMA/Prophet models-based forecasting of COVID-19: Methodology, evaluation, and case study in SAARC countries. *Stoch. Environ. Res. Risk Assess.* **2022**, *37*, 345–359. [[CrossRef](#)] [[PubMed](#)]
29. Cacha, I.H.; Díaz, J.S.; Castrillo, M.; García, A.L. Forecasting COVID-19 spreading through an ensemble of classical and machine learning models: Spain’s case study. *Sci. Rep.* **2023**, *13*, 6750. [[CrossRef](#)] [[PubMed](#)]
30. Marzouk, M.; Elshaboury, N.; Abdel-Latif, A.; Azab, S. Deep learning model for forecasting COVID-19 outbreak in Egypt. *Process. Saf. Environ. Prot.* **2021**, *153*, 363–375. [[CrossRef](#)] [[PubMed](#)]
31. Alkady, W.; ElBahnasy, K.; Leiva, V.; Gad, W. Classifying COVID-19 based on amino acids encoding with machine learning algorithms. *Chemom. Intell. Lab. Syst.* **2022**, *224*, 104535. [[CrossRef](#)]
32. Ullah, A.; Malik, K.M.; Saudagar, A.K.; Khan, M.B.; Hasanat, M.H.; AlTameem, A.; AlKhathami, M.; Sajjad, M. COVID-19 genome sequence analysis for new variant prediction and generation. *Mathematics* **2022**, *10*, 4267. [[CrossRef](#)]
33. Nguyen, P.H.; Tsai, J.F.; Lin, M.H.; Hu, Y.C. A hybrid model with spherical fuzzy-AHP, PLS-SEM and ANN to predict vaccination intention against COVID-19. *Mathematics* **2021**, *9*, 3075. [[CrossRef](#)]
34. Delgado, E.J.; Cabezas, X.; Martin-Barreiro, C.; Leiva, V.; Rojas, F. An equity-based optimization model to solve the location problem for healthcare centers applied to hospital beds and COVID-19 vaccination. *Mathematics* **2022**, *10*, 1825. [[CrossRef](#)]

35. Ito, K.; Piantham, C.; Nishiura, H. Relative instantaneous reproduction number of Omicron SARS-CoV-2 variant with respect to the Delta variant in Denmark. *J. Med. Virol.* **2022**, *94*, 2265–2268. [[CrossRef](#)]
36. Yang, Z.; Zeng, Z.; Wang, K.; Wong, S.S.; Liang, W.; Zanin, M.; Liu, P.; Cao, X.; Gao, Z.; Mai, Z.; et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *J. Thorac. Dis.* **2020**, *12*, 165. [[CrossRef](#)]
37. Lindsey, J.K. *Statistical Analysis of Stochastic Processes in Time*; Cambridge University Press: Cambridge, UK, 2004.
38. Senel, K.; Özdinc, M.; Ozturkcan, S.; Akgul, A. Instantaneous R for COVID-19 in Turkey: Estimation by Bayesian statistical inference. *Turk. Klin. J. Med. Sci.* **2020**, *40*, 127–131.
39. Cori, A.; Ferguson, N.M.; Fraser, C.; Cauchemez, S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am. J. Epidemiol.* **2013**, *178*, 1505–1512. [[CrossRef](#)]
40. Kong, L.; Duan, M.; Shi, J.; Hong, J.; Chang, Z.; Zhang, Z. Compartmental structures used in modeling COVID-19: A scoping review. *Infect. Dis. Poverty* **2022**, *11*, 72. [[CrossRef](#)]
41. Chumachenko, D.; Menailov, I.; Bazilevych, K.; Chumachenko, T.; Yakovlev, S. Investigation of statistical machine learning models for COVID-19 epidemic process simulation: Random forest, K-nearest neighbors, gradient boosting. *Computation* **2022**, *10*, 86. [[CrossRef](#)]
42. Al-Rashedi, A.; Al-Hagery, M.A. Deep learning algorithms for forecasting COVID-19 cases in Saudi Arabia. *Appl. Sci.* **2023**, *13*, 1816. [[CrossRef](#)]
43. Chowell, G. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. *Infect. Dis. Model.* **2017**, *2*, 379–398. [[CrossRef](#)] [[PubMed](#)]
44. Verbesselt, J.; Hyndman, R.; Newnham, G.; Culvenor, D. Detecting trend and seasonal changes in satellite image time series. *Remote Sens. Environ.* **2010**, *114*, 106–115. [[CrossRef](#)]
45. He, S.; Peng, Y.; Sun, K. SEIR modeling of the COVID-19 and its dynamics. *Nonlinear Dyn.* **2020**, *101*, 1667–1680. [[CrossRef](#)]
46. Hyndman, R.J.; Khandakar, Y. Automatic time series forecasting: The forecast package for R. *J. Stat. Softw.* **2008**, *27*, 1–22. [[CrossRef](#)]
47. Cryer, J.D.; Chan, K.S. *Time Series Analysis with Applications in R*; Springer: New York, NY, USA, 2008.
48. Kirchgässner, G.; Wolters, J.; Hassler, U. *Introduction to Modern Time Series Analysis*; Springer: New York, NY, USA, 2012.
49. Papastefanopoulos, V.; Linardatos, P.; Kotsiantis, S. COVID-19: A comparison of time series methods to forecast percentage of active cases per population. *Appl. Sci.* **2020**, *10*, 3880. [[CrossRef](#)]
50. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.
51. Shumway, R.H.; Stoffer, D.S.; Stoffer, D.S. *Time Series Analysis and Its Applications*; Springer: New York, NY, USA, 2000.
52. Burns, P. Robustness of the Ljung-Box Test and Its Rank Equivalent. SSRN 443560. 2002. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=443560 (accessed on 16 June 2023).
53. Dickey, D.A.; Fuller, W.A. Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Stat. Assoc.* **1979**, *74*, 427–431.
54. Winters, P.R. Forecasting sales by exponentially weighted moving averages. *Manag. Sci.* **1960**, *6*, 324–342. [[CrossRef](#)]
55. Bai, J.; Perron, P. Estimating and testing linear models with multiple structural changes. *Econometrica* **1988**, *66*, 47–78. [[CrossRef](#)]
56. Bai, J.; Perron, P. Computation and analysis of multiple structural change models. *J. Appl. Econom.* **2003**, *18*, 1–22. [[CrossRef](#)]
57. Krispin, R. *Hands-On Time Series Analysis with R: Perform Time Series Analysis and Forecasting Using R*; Packt Publishing, Limited: Birmingham, UK, 2019.
58. Tableau (Version 2023.1) [Computer Software]. Tableau Software. Available online: www.tableau.com (accessed on 16 June 2023).
59. Martin-Barreiro, C.; Ramirez-Figueroa, J.A.; Cabezas, X.; Leiva, V.; Galindo-Villardón, M.P. 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. [[CrossRef](#)]
60. Nishiura, H.; Chowell, G. The effective reproduction number as a prelude to statistical estimation of time-dependent epidemic trends. In *Mathematical and Statistical Estimation Approaches in Epidemiology*; Springer: Dordrecht, Germany, 2009; pp. 103–121.
61. Cortés-Carvajal, P.D.; Cubilla-Montilla, M.; González-Cortés, D.R. Estimation of the instantaneous reproduction number and its confidence interval for modeling the COVID-19 pandemic. *Mathematics* **2022**, *10*, 287. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.