Big Data and Information Analytics

Research article

Modeling the COVID-19 epidemic in Bolivia

Antonio N. Bojanic[1],* and Alejandro Jordán[2]

1 Economics Department, Tulane University, 6823 St. Charles Ave., New Orleans, LA 70118, USA
2 Sam M. Walton College of Business, University of Arkansas, Fayetteville, AR 72701, USA

* Correspondence: Email: abojanic@tulane.edu.

Abstract: The first cases of COVID-19 appeared in Bolivia on March 10. Since then and until September 24, 132,618 people have been infected, 7765 have died, and 92,101 have recovered; the virus has spread throughout the country, but the departments of Santa Cruz, La Paz, and Cochabamba account for the overwhelming percentage of infected, recovered and dead. We analyze the spread of the virus utilizing a SIR model. We find that a maximum fraction of infected individuals occurs on June 17, when around 28% of the population is infected; when the epidemic begins to subside, 54% of the population is in the recovered category, indicating that more than half of the population will have been infected by COVID-19 during some period of the epidemic. There is an uptick in the number of susceptible individuals after July 1, highlighting that the relaxing of mitigation measures might have happened too soon.

Keywords: epidemics/epidemiology; COVID-19; Bolivia

1. Introduction

In Bolivia, the first reported cases of COVID-19 appeared on March 10 in the cities Santa Cruz and Oruro. Since then and until September 24, 132,618 people have been infected throughout the country. Though reliable data is always an issue with epidemics of this magnitude[1], certain aspects are becoming clearer. At the time of writing, 7765 people have died and 92,101 have recovered; the virus has spread throughout the country, but the departments of Santa Cruz, La Paz, and Cochabamba, account for around 68% of infected cases; 80% of recovered; and 77% of the dead[2]. There is no reliable information on whether the virus affects certain groups more than others—i.e., men and

---

[1] This problem is compounded in developing countries like Bolivia
[2] Bolivia is divided into 9 Departments: La Paz, Cochabamba, Santa Cruz, Oruro, Chuquisaca, Pando, Tarija, Beni, and Potosí

women; old and young; urban and rural; healthy and sick—but anecdotal evidence seems to suggest that, as in other countries, the virus is particularly prevalent in older enclaves and in people with underlying health conditions. The government of Bolivia began applying mitigation measures, including a strict quarantine order, on March 22, earlier than most other countries in Latin America; it closed its borders to international travel on March 21 and severely restricted domestic travel as well. On June 1 it began relaxing some of these measures, and at the time of writing was considering implementing them once again after an uptick in new infections and deaths.

Although not explicitly analyzed here, it is important to note that the healthcare system in the country has been completely overwhelmed by the epidemic, both in terms of providing adequate and opportune testing that might help in containing the virus; and in providing basic healthcare services to people infected by the virus. Unfortunately, as is often the case in countries like Bolivia, people who are in greatest need of the system—the uninsured, the poor, and more generally, anyone living outside the principal urban centers—are suffering the most and are paying the consequences of the mismanagement, underfunding, and corruption that have afflicted the public health system in the country for generations.

An additional complicating factor in the spread of the virus is the political situation. The present government—headed by Jeanine Añez and in power since November 2019—is a transitional administration tasked with organizing democratic elections in 2020. Given its transitional nature and the story of Bolivian politics, the current government is inherently fragile and limited in its ability to impose and enforce proper mitigation measures.

We analyze the projected spread of COVID-19 at the national and departmental levels utilizing the Susceptible-Infective-Recovered (SIR) model [3]. Though several assumptions are made, particularly in terms of transmission and recovery rates and the number of contacts between infected and susceptible individuals, we find that our predicted model fits the data well. Further, the model allows us to make predictions regarding the impact of public interventions—including the establishment of quarantines and the introduction of effective antiviral treatments—on the trajectory of the virus.

A summary of findings is this: on its current path and at the national level, a maximum fraction of infected individuals occurs around June 17, when approximately 28% of the population (3.2 million) is infected by the virus; when the epidemic begins to subside, about 54 % of the population (6.3 million) is in the recovered group, indicating that more than half of the population will have been infected by COVID-19 during some period of the epidemic. There is an uptick in the number of susceptible individuals after July 1, highlighting that the relaxing of mitigation measures might have happened too soon.

The rest of the paper is organized as follows. Section 2 presents a short review of the literature; section 3 analyzes the data and the methodology; section 4 reports the principal results, and section 5 concludes.

2. Literature review

Predictive mathematical models for epidemics are fundamental to understand the trajectory of an epidemic and to plan effective control strategies to contain it. Developed by Kermack and

---

[3] In the SIR model, 'Recovered' includes those individuals that recover naturally or through medication from the disease and those that die (i.e., are removed) because of the disease. We did not choose a more detailed model due to lack of reliable data.

McKendrick [1], the SIR model for human-to-human transmission of a disease is a commonly used methodology that describes the flow of individuals through three mutually exclusive stages of infection: Susceptible, infected, and recovered.

There have been many extensions to the SIR model and its utilization to describe the COVID-19 epidemic. Lin, et al. [2] for instance, utilize an SEIR (susceptible, exposed, infectious, removed) model considering risk perception and a cumulative number of cases; Anastassopoulou, et al. [3] analyze a discrete-time SIR model that includes dead individuals; Casella [4] develops a control-oriented SIR model that stresses the effects of delays and compares the outcomes of different containment policies; Giordano, et al. [5] compare simulation results with real data from Italy utilizing a novel SIDARTHE (susceptible, infected, diagnosed, ailing, recognized, threatened, healed, and extinct) model; Wu, et al. [6] use transmission dynamics to estimate the clinical severity of COVID-19; and Tam, et al. [7], utilizing a SIR model but adding the number of quarantined people, analyze the projected spread of COVID-19 in Louisiana without including the effects of the stay-at-home order. Some authors have also considered stochastic transmission models, including Hellewell, et al. [8] and Kucharski, et al. [9].

For the specific case of Bolivia, a handful of working papers, reports, and policy briefs have been developed on the theme of COVID-19. Utilizing an SEIR[4] model and with national data only, Birbuet and López [10] present an analysis of the spread of the virus before and after the establishment of a quarantine. A report from the MRC Centre for Global Infectious Disease Analysis[5] focuses on back-calculating an 'inferred number of COVID-19 infections' to estimate the number of people that have been infected and to make short-term projections for future healthcare needs; two additional documents—a report from the World Bank [6] and a policy brief from United Nations[7]—focus on the impact of COVID 19 on the economic, social, and political performance of countries in Latin America and the Caribbean, but neither analyzes the evolution of the virus in each country.

As the preceding review makes clear, there is a gap concerning a formal analysis of COVID-19 in Bolivia, particularly at the departmental level. The contribution of this paper rests on filling this gap by presenting a mathematical model that describes the evolution of the epidemic from its inception point in early March of 2020 to its eventual decline, including how the trajectory of the virus might be altered if mitigation policies are put into place to control the spread of the virus.

3. Data and methodology

3.1. Data

The primary source of data for the number of infections, dead, and recovered is the government of Bolivia. It has set up an official website—Bolivia Segura[8]—where all statistical data is placed daily. The website also serves to inform the public on all matters related to the epidemic. A second

---

[4] We are not comfortable utilizing a latency number to predict the number of exposed individuals; we feel that an SIR model is more accurate in describing the spread of COVID-19 in Bolivia. If we had included a latent state in addition to the susceptible, infected, and recovered states, we would have to utilize two additional parameters – i.e., time it takes for an infected person to become infectious; and the average time during which an infected person is able to transmit the disease – which may not add significantly to the basic findings since, as it is the case with all parameters, much is still uncertain about COVID-19

[5] Imperial College, London. Available at: https://mrc-ide.github.io/global-lmic-reports/BOL/

[6] Available at: https://openknowledge.worldbank.org/handle/10986/33555

[7] Available at: https://www.un.org/sites/un2.un.org/files/sg_policy_brief_covid_lac.pdf

[8] https://www.boliviasegura.gob.bo/

webpage built by a group of university students—COVID Bolivia[9]—processes the data collected in *Bolivia Segura* and cross-references it with other private sources of data to control for the reliability of government-produced statistics. The data utilized here comes from *Bolivia Segura* but it is continually cross-referenced with data from *COVID Bolivia*[10].

3.2. Methodology

The SIR dynamical system consists of three ordinary difference equations describing the evolution of the population in each state over time[11]:

$$s_{t+1} = s_t - \alpha\gamma s_t i_t + n - ms_t \tag{1}$$

$$i_{t+1} = i_t + \alpha\gamma s_t i_t - ki_t - mi_t$$

$$i_{t+1} = (1 + \alpha\gamma s_t - k - m)i_t \tag{2}$$

$$r_{t+1} = r_t + ki_t - mr_t \tag{3}$$

where $s(t)$, $i(t)$, and $r(t)$ represent fractions of the total population, per time, in the susceptible, infected, and recovered states (or categories), respectively. The parameters are defined as follows:

- α denotes the percent of contacts resulting in an infection; also known as the transmission rate
- γ represents the number of non-infected people contacted by each infected person
- k represents the fraction of the population that recovers from the disease each period; also known as the recovery rate
- and n and m denote the birth and death rates, respectively

If we let $\beta = \alpha\gamma$, Eqs (1)–(3) can be restated as:

$$s_{t+1} = s_t - \beta s_t i_t + n - ms_t \tag{4}$$

$$i_{t+1} = (1 + \beta s_t - k - m)i_t \tag{5}$$

$$r_{t+1} = r_t + ki_t - mr_t \tag{6}$$

The SIR models (4)–(6) are a bilinear system with three difference equations. The system is positive: All the state variables take non-negative values for $t \geq 0$ if initialized at time 0 with non-negative values. Additionally, it assumes a constant population, hence $n = m$. The system is compartmental, and therefore the sum of the states (total population) is constant. Because the variables denote population fractions, we assume

$$s_{t+1} + i_{t+1} + r_{t+1} = s_t + i_t + r_t = 1 \tag{7}$$

Note that we are above the epidemic threshold whenever $i_{t+1} > i_t$; from Eq (5), this inequality is equivalent to $\beta s_t > k - m$, which implies that whenever this inequality holds we will be above the

---

[9] https://www.covid-bolivia.com/
[10] At the onset of the epidemic, there were small discrepancies between these sources of data, but as time elapsed these discrepancies have almost disappeared.
[11] The SIR model presented is based on the description by Tassier [11]

epidemic threshold and the number of infected individuals in the population will increase. Alternatively, if $\frac{\beta s_t}{k-m} > 1$, the level of infected individuals increases; if $\frac{\beta s_t}{k-m} < 1$, the level of infected individuals decreases.

Since at the beginning of the epidemic $s_t \approx 1$, $\frac{\beta s_t}{k-m} \approx \frac{\beta}{k-m}$. The latter fraction is known as the Reproduction Number (R_0) and it denotes the average number of people who will contract a contagious disease from one person with that disease. It specifically applies to a population of people who were previously free of infection and have not been vaccinated, hence this number is particularly relevant at the beginning of an epidemic.

Finally, note that, from Eq (5), if $1 + \beta s_t - k - m < 1$, the epidemic never materializes. Rearranging, the epidemic will disappear if $s_t < \frac{k+m}{\beta}$. Thus, herd immunity is achieved if $1 - \frac{k+m}{\beta}$ of the population are recovered and immune.

4. Empirical results

The parameters for the baseline SIR model are set at $\alpha = 0.03$; $\gamma = 10$; $k = 0.10$; and $n = m = 0.025$[12]. The resulting basic reproduction number is $R_0 = 4$, a clear indication that an epidemic is set to take off. Figure 1 shows the model evolution at the national level for the period March 10 to September 24, 2020.
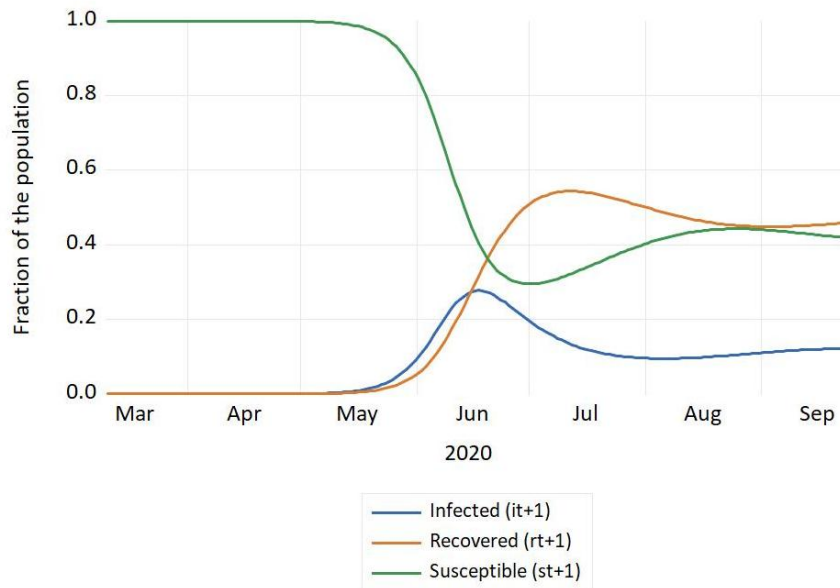

Figure 1. Bolivia: susceptible, infected, and recovered populations.

---

[12] The values for α, γ, and k are estimates based on available information. The transmission rate (α) remains highly uncertain and the value utilized here represents a lower bound in two comparable studies: one by Munayco, et al. [12] on Peru and the other one by Birbuet and López [10] on Bolivia. The recovery rate (k) also represents a lower bound value based on the number of daily and accumulated number of recovered individuals with respect to daily and accumulated infected individuals in Bolivia. The birth rate (n) was obtained from the World Bank (https://data.worldbank.org/indicator/SP.DYN.CBRT.IN?locations=BO).

As illustrated in Figure 1, with the baseline parameters the epidemic peaks on June 17, when approximately 28% of the population (3.2 million) becomes infected. Subsequent to the peak, the share of infected individuals decreases until around August 4, when 9% of the population are infected, but it gradually increases after that, breaking the 12% mark on September 16. On July 12, 54% of the population (6.3 million) are part of the recovered population, implying that more than half of the people in the country become infected with COVID-19 at some point during the epidemic. Beginning on July 13, the share of recovered individuals starts decreasing gradually but stabilizes at around 45% after September 10. The share of susceptible individuals also experiences an uneven evolution: after reaching a low of 29% on July 1, it begins to increase on July 2 reaching a level of 44% in August 28 when it once again retakes its downward trend. These fluctuations in the fractions of recovered and susceptible individuals highlight that the loosening of mitigation measures—which began on June 1—might have caused some recovered individuals to become susceptible again, denting the progress made in terms of preventing susceptible individuals from coming into contact with people carrying the disease[13]. Either way, the duration of the epidemic is likely to be extended by fluctuations in the number of susceptible individuals[14]. On present trends, about 58% of the population $(1 - \frac{0.10 + 0.03}{0.30} = 0.58)$ will have to be immunized to achieve herd immunity.

Figure 2 illustrates the impact of loosening and tightening quarantine restrictions on the share of infected individuals at the national level.
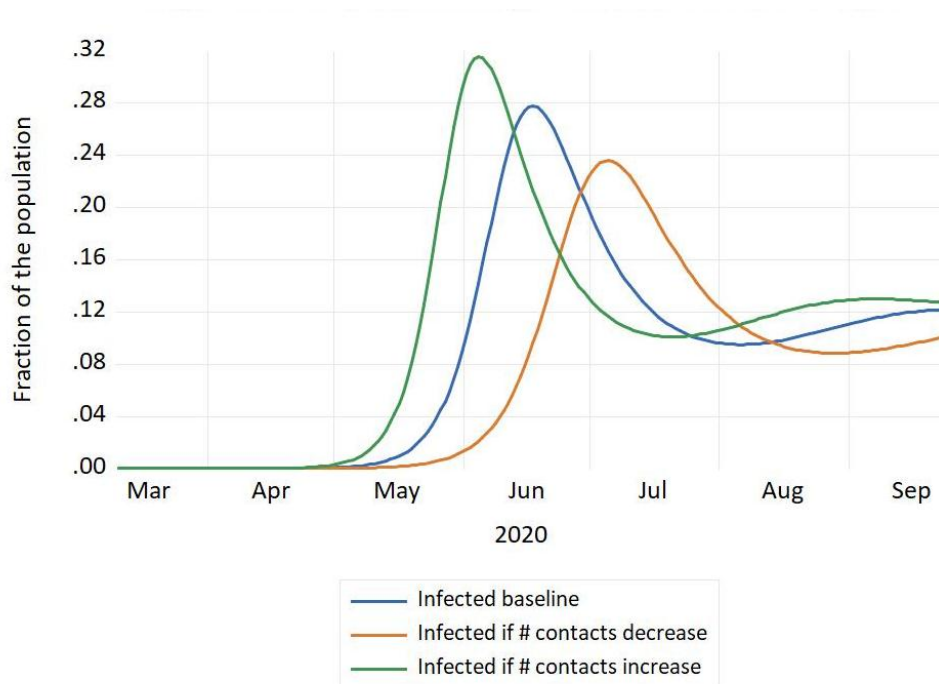

Figure 2. Bolivia: infected with higher and lower number of contacts.

---

[13] It is also possible that fluctuations in the number of recovered individuals is the result of inexact testing; that is, infected individuals that eventually tested negative for the virus might have not recovered.

[14] An alternative explanation for the fluctuation in susceptible individuals is that, with new births, the number of people in state S will always rise once the reproduction number falls below 1. Then, as S increases, the reproduction number rises and the number of infections increase as well. The result being a new peak of infections and a continuation of the disease dynamics in a cyclical pattern. Absent a permanent cure (i.e., the introduction of an effective vaccine), the process generates a steady state (endemic) level of infections over time.

If quarantine restrictions were tightened and the number of contacts between infected and susceptible individuals decreases by a single person (from a baseline value $\gamma = 10$ to $\gamma = 9$), a lower fraction of the population becomes infected (from a baseline peak of 28% on June 17 to a lower high peak of 24% on July 5); likewise, if quarantine restrictions are loosened and the number of contacts increases by 1 (from $\gamma = 10$ to $\gamma = 11$), a larger percentage of the population becomes infected (32% on June 4), highlighting that quarantine restrictions can help control the spread of the virus and reduce the level of infections, though at the cost of extending the duration of the epidemic. A similar exercise can be done with the recovery rate. If an effective treatment is found and k increases—say, from $k = 0.10$ to $k = 0.15$—so that individuals recover more quickly, the peak level of infections decreases to approximately 14% of the population (1.58 million) and the number of total infections decreases[15]. The behavioral pattern of the epidemic becomes more apparent if changes to γ and k are more significant than those reported here.

Figure 3 illustrates the evolution of COVID-19 in Santa Cruz and La Paz, two of the departments most affected by it. The parameters are equal to the baseline parameters utilized at the national level with the only difference being the number of contacts γ; the value of γ is a function of the level of unrest in each department.
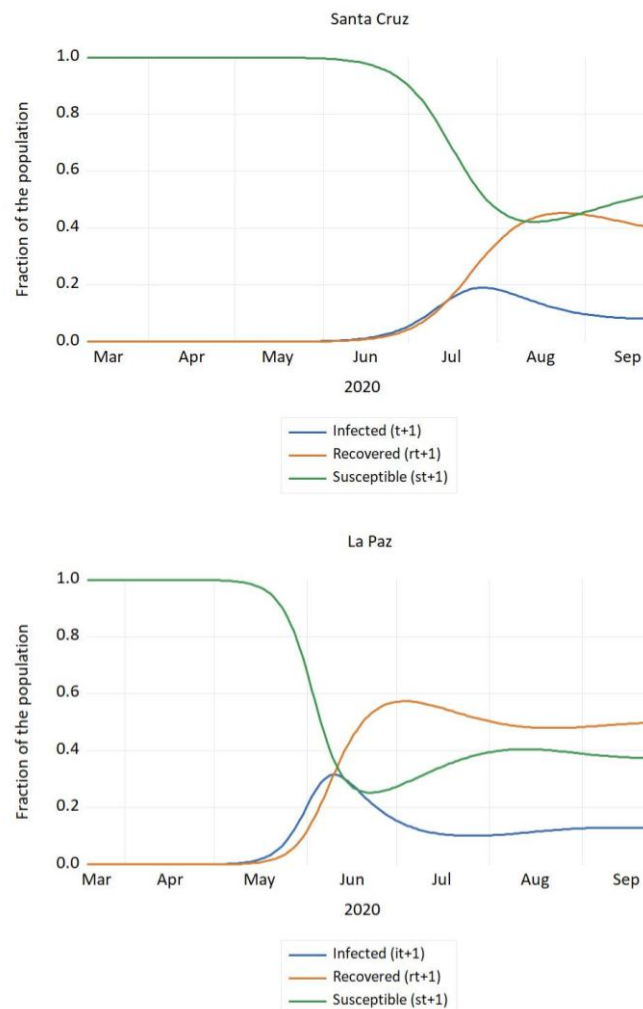

Figure 3. Evolution of COVID-19 in Santa Cruz and La Paz.

---

[15] The graphs showing how the trajectory of the virus changes with varying values of k are available upon request.

Public demonstrations against mitigation measures have been particularly prevalent in La Paz[16]. Additionally, demands for national elections, which the current government has postponed until October 18 due to concerns for greater propagation of the virus, have also been more visible and widespread in La Paz. As both types of public manifestations require large gatherings of people, we assume that the number of contacts with infected individuals is likely higher in this department than in relatively more peaceful Santa Cruz. If $\gamma = 11$ in La Paz and $\gamma = 8$ in Santa Cruz, we observe that in La Paz the virus is likely to infect a larger share of the population but will last a shorter period of time. Conversely, in relatively more peaceful Santa Cruz, a smaller fraction of the population is infected, with the epidemic lasting longer.

Figure 4 illustrates the evolution of the recovered state utilizing actual data (i.e., the share of recovered individuals in the population) and the predicted values of the SIR model (i.e., the share of recovered individuals as reflected in Eq (6)) for the departments of Santa Cruz and La Paz. The parameters are those utilized for describing the evolution of the virus in Figure 3.
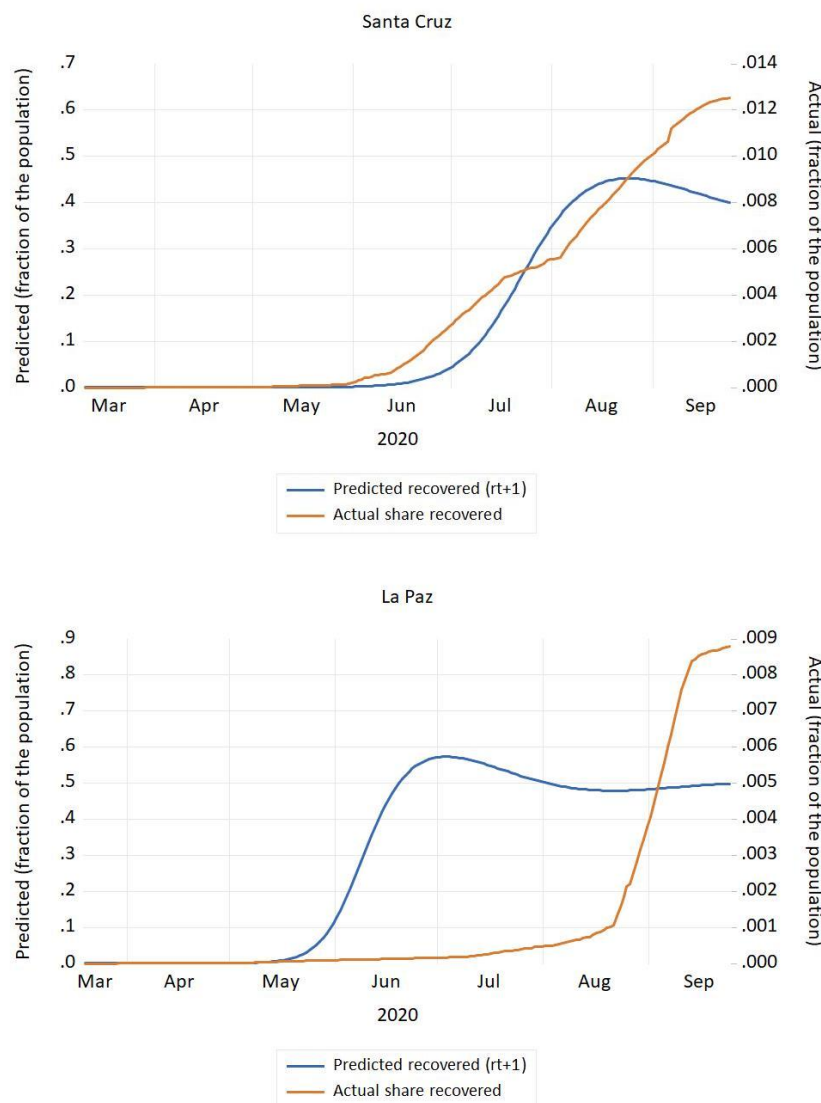
Figure 4. Actual and predicted values for share of recovered individuals.

---

[16] Social unrest has also occurred in other departments, but La Paz, as the administrative capital of the country, is where public demonstrations have been fiercest.

As is evident, the pattern of behavior for the predicted value of the recovered share of the population mimics the evolution of the actual share of recovered individuals in both departments. The correlation coefficient between both variables in Santa Cruz is +0.95 and +0.39 in La Paz, and though there seems to be lag between the variables—especially in La Paz, where the predicted increase in recovered cases lags by almost 3 months the actual increase in cases—the model does seem to describe the actual evolution of the share of recovered individuals relatively well[17].

Finally, it has been noted that the virus has spread throughout the country. Summary statistics for all departments are shown in Table 1.

Table 1. Summary statistics by department.

Department	Infections started on:	Infections predicted to peak on:	Share of Infections (%)	Share of Recovered (%)	Share of Dead (%)
Santa Cruz	March 10	July 27	31.95	41.59	50.02
La Paz	March 19	June 9	25.77	26.88	12.80
Cochabamba	March 13	June 1	10.17	11.58	14.11
Oruro	March 10	June 2	4.28	2.94	3.80
Potosi	March 26	June 22	6.04	2.45	3.08
Pando	March 28	July 16	1.78	0.54	2.11
Tarija	March 30	June 23	9.57	7.46	3.76
Chuquisaca	March 27	June 20	5.22	4.56	6.17
Beni	April 20	August 12	5.36	2.50	4.55

Notes: 1. Date for peak on infections based on baseline parameters and levels of unrest. 2. Share of infections as a percentage of total infections. 3. Share of recovered as a percentage of total recovered. 4. Share of dead as a percentage of total dead. 5. Infections, recovered, and dead figures as of September 24, 2020.

As Table 1 illustrates, though the brunt of the epidemic is being felt in Santa Cruz, La Paz and Cochabamba, it has become a national epidemic. There has been a constant increase in the number of dead in all departments, and this increase has become particularly visible in Chuquisaca and Beni. Additionally, the share of infections is increasing everywhere, being especially evident in the departments of Potosí and Tarija. As the preceding analysis makes clear, the epidemic in Bolivia is nowhere near being contained, and mitigation measures—particularly the mandatory quarantine in place until June 1—were lifted too soon.

5.   Conclusions and policy implications

The first two cases of COVID-19 appeared in Bolivia on March 10 of 2020. Since then, the virus has spread throughout the country and has overwhelmed an already weak public health system. The SIR model presented here accurately describes the trajectory of the virus, both at the national and departmental level, and also demonstrates that this trajectory could readily be altered by mitigation measures, particularly mandatory quarantines, that aim to reduce the number of contacts between infected and susceptible individuals. The model also shows that in the absence of effective mitigation measures the epidemic could become more widespread and last longer.

---

[17] Similar comparisons were made between predicted and actual values for the susceptible and infected states and the model accurately describes the actual evolution of the data in all departments. These graphical comparisons are available upon request.

The principal conclusions are these: Nationally and at present trends, the virus peaks in mid-June when around 28% of the population becomes infected; as the epidemic begins to subside in July, more than half of the population belongs to the 'recovered' group, implying that more than 50% of Bolivians will have become infected with COVID-19 at some point during the epidemic. Beginning in July, there has been a fluctuation in the share of susceptible and recovered individuals, highlighting that the epidemic is nowhere near being contained. The pattern of behavior at the departmental level is similar than at the national level. The date when each department peaks in terms of infection shares varies, as does the timing when the epidemic begins to subside, but in all cases the evidence and the model show that the epidemic is far from over.

A key policy recommendation is that a mandatory quarantine throughout the country should once again be established to contain the latest uptick in infections. Though difficult because of the inherent weakness of the public health sector and the fragility of the current government, trying to isolate those infected with the disease and preventing contact with susceptible individuals seems to be the only current alternative to containing the epidemic. An additional recommendation concerns the state of government entities in charge of managing public health crises. The epidemic has laid bare the limitations of the public sector to deal with crises like the one unfolding now. Of critical importance when the Bolivian government finally addresses the fragility of these institutions is their capacity to collect—and make public—periodic, reliable data that allows the design of proper policy responses, and, from the perspective of this work, the development of more thorough epidemiological models that describe the spread of a disease.

Acknowledgments

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. Kermack WO and McKendrick AG, (1927) A contribution to the mathematical theory of epidemics. *Proc R Soc Lond* 115: 700–721.
2. Lin Q, Zhao S, Gao D, et al. (2020) A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. *Int J Inf Dis* 93: 211–216.
3. Anastassopoulou C, Russo L, Tsakris A, et al. (2020) Data-based analysis, modelling and forecasting of the COVID-19 outbreak. *PloS One* 15: e0230405.
4. Casella F, (2020) Can the COVID-19 epidemic be controlled on the basis of daily test reports?, Preprint, arXiv:2003.06967.
5. Giordano G, Blanchini F, Bruno R, et al. (2020) Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nat Med* 26: 855–860.

6. Wu JT, Leung K, Bushman M, et al. (2020) Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat Med* 26: 506–510.

7. Tam KM, Walker N and Moreno J, (2020) Projected development of COVID-19 in Louisiana. *arXiv* 2004.02859.

8. Hellewell J, Abbott S, Gimma A, et al. (2020) Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Global Health*, 8: e488–e496.

9. Kucharski AJ, Russell TW, Diamond C, et al. (2020) Early dynamics of transmission and control of COVID-19: a mathematical modelling study, *Lancet Infect Dis* 20: 553–558.

10. Birbuet JC and López V R, (2020) Dinámica de expansión del COVID-19 en Bolivia durante las 6 primeras semanas. Working Paper. Available from: https://github.com/rafalopezv/modelo_seir_bolivia.

11. Tassier T, (2013) *The Economics of Epidemiology*, Springer-Verlag Berlin Heidelberg.

12. Munayco CV, Tariq A, Rothenberg R, et al. (2020) Early transmission dynamics of COVID-19 in a southern hemisphere setting: Lima-Peru: February 29th–March 30th, 2020. *Infect Dis Model*, 5: 338–345.