Analysis and forecasts for patterns of COVID-19 in Pakistan using Bayesian models (#50949)

First submission

Guidance from your Editor

Please submit by **3 Sep 2020** for the benefit of the authors (and your $200 publishing discount) .

Structure and Criteria
Please read the 'Structure and Criteria' page for general guidance.

Raw data check
Review the raw data.

Image check
Check that figures and images have not been inappropriately manipulated.

Privacy reminder: If uploading an annotated PDF, remove identifiable information to remain anonymous.

Files

Download and review all files from the materials page.

4 Figure file(s)
3 Table file(s)
1 Scheme file(s)
2 Raw data file(s)

Structure and Criteria

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

When ready submit online.

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your guidance page.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to PeerJ standards, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (see PeerJ policy).

EXPERIMENTAL DESIGN

- Original primary research within Scope of the journal.
- Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. Negative/inconclusive results accepted. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- All underlying data have been provided; they are robust, statistically sound, & controlled.
- Speculation is welcome, but should be identified as such.
- Conclusions are well stated, linked to original research question & limited to supporting results.

Standout reviewing tips

The best reviewers use these techniques

Tip	*Example*
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

Analysis and forecasts for patterns of COVID-19 in Pakistan using Bayesian models

Navid Feroze [Corresp., 1], **Kamran Abbas** [2], **Farzana Noor** [3], **Amjid Ali** [4]

[1] Department of Statistics, The University of Azad Jammu and Kashmir, Muzaffarabad, Pakistan, Muzaffarabad, State of Azad Jammu and Kashmir, Pakistan

[2] The University of Azad Jammu and Kashmir, Muzaffarabad, Pakistan, Muzaffarabad, State of Azad Jammu and Kashmir, Pakistan

[3] Department of Mathematics and Statistics, International Islamic University, Islamabad, Pakistan

[4] Islamia College, Peshawar, Khyber Pakhtunkhwa, Pakistan, Peshawar, Pakistan

Corresponding Author: Navid Feroze
Email address: navidferoz@gmail.com

Background. Though Pakistan is among the countries where the COVID-19 entered quite later, but currently it is on full flow in the country. Given the health facilities in the country, there are serious threats that upcoming moths can be very testing for all the stakeholders. To tackle with challenges in the coming months, there is a need to analyze and forecast the trends of COVID-19 in Pakistan.**Methods.** We have analyzed and forecasted the patterns of this pandemic in the country, for next thirty days, using Bayesian structural time series (BSTS) models. The causal impacts of lifting lockdown have also been investigated using intervention analysis under BSTS models. The forecasting accuracy of the proposed models has been compared with frequently used auto regressive integrated moving average (ARIMA) models. **Results.** We observed the improved forecasting accuracy of BSTS models as compared to frequently used ARIMA models. As far as the forecasts are concerned, on August 10, 2020, the country is expected to have 333,693 positive cases with 95% prediction interval [267,470; 393,702]. Similarly, the number of deaths in the country is expected to reach 7,178 [6,003; 8,368] and recoveries may grow to 265,619 [244,902; 290,985]. The lifting of lockdown has caused an absolute increase of 98,768 confirmed cases with 95% interval [85,544; 111,018]. The positive aspect of the forecasts is that the number of active cases is expected to decrease to 68,344 [22,568; 102,717]. This is the time for the concerned authorities to further restrict the active cases so that the recession of the outbreak continues in the next month.

Analysis and Forecasts for Patterns of COVID-19 in Pakistan using Bayesian Models

Navid Feroze[1], Kamran Abbas[1], Farzana Noor[2], Amjad Ali[3]

[1]Department of Statistics, University of Azad Jammu and Kashmir, Muzaffarabad, Pakistan
[2]Department of Mathematical and Statistics, International Islamic University, Islamabad, Pakistan.
[3]Department of Statistics, Islamia College, Peshawar, Khyber Pakhtunkhwa, Pakistan

Corresponding Author
Department of Statistics, University of Azad Jammu and Kashmir, Muzaffarabad, 13100
Pakistan
Email address: navidferoz@gmail.com

Abstract

Background. Though Pakistan is among the countries where the COVID-19 entered quite later, but currently it is on full flow in the country. Given the health facilities in the country, there are serious threats that upcoming moths can be very testing for all the stakeholders. To tackle with challenges in the coming months, there is a need to analyze and forecast the trends of COVID-19 in Pakistan.
Methods. We have analyzed and forecasted the patterns of this pandemic in the country, for next thirty days, using Bayesian structural time series (BSTS) models. The causal impacts of lifting lockdown have also been investigated using intervention analysis under BSTS models. The forecasting accuracy of the proposed models has been compared with frequently used auto regressive integrated moving average (ARIMA) models.
Results. We observed the improved forecasting accuracy of BSTS models as compared to frequently used ARIMA models. As far as the forecasts are concerned, on August 10, 2020, the country is expected to have 333,693 positive cases with 95% prediction interval [267,470; 393,702]. Similarly, the number of deaths in the country is expected to reach 7,178 [6,003; 8,368] and recoveries may grow to 265,619 [244,902; 290,985]. The lifting of lockdown has caused an absolute increase of 98,768 confirmed cases with 95% interval [85,544; 111,018]. The positive aspect of the forecasts is that the number of active cases is expected to decrease to 68,344 [22,568; 102,717]. This is the time for the concerned authorities to further restrict the active cases so that the recession of the outbreak continues in the next month.

Keywords: BSTS models, lockdown, Ljung Box test, ARIMA models

Introduction

During December, 2019, the historical pandemic started in Wuhan, China (Paules et al., 2020). This pandemic was named novel corona virus or COVID-19 by the World Health Organization (WHO, 2020). Though this virus has lower mortality rates as compared to s severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS), this virus has higher transmission rates (Tomar and Gupta, 2020). Due to higher transmissibility, the virus has covered almost the whole world (WHO, 2020). In Pakistan, the first positive case was observed on February 26, 2020. The country imposed the complete lockdown on March 23, 2020. However, the struggling economy of the country forced the government to lift the lockdown on May 9, 2020. During the lockdown, the trajectory of the positive cases was low and slow in the country (Yousaf, et al., 2020), however the outbreak has got the pace since lifting of the lockdown. And there are further expectations that number of cases and deaths will increase more rapidly in future (Yousaf, et al., 2020).

The repeated short term forecasts for the patterns of the pandemic are fundamental (Li et al., 2020; Wang and Zhang, 2020 and Zhou et al., 2020). These short term forecasts help the policy makers to make the informed decisions in accordance with the evolving scenarios (Ippolito et al., 2020 and McCloskey et al., 2020). There are number of contributions regarding the short terms forecasts for the various parameters of the pandemic in different countries. Some important studies have been conducted in China (Li et al., 2020; Perc et al., 2020 and Fanelli and Piazza, 2020), France (Fanelli and Piazza, 2020), Germany (Perc et al., 2020), India (Tomar and Gupta, 2020 and Gupta and Pal, 2020), Iran (Perc et al., 2020, Zhan et al., 2020 and Moftakhar et al., 2020), Italy (Fanelli and Piazza, 2020; Zhan et al., 2020), Nigeria (Majeed et al., 2020), South Korea (Zhan et al., 2020) and United States (Perc et al., 2020).

There has been an earlier attempt to forecast the number of cases, deaths and recoveries in Pakistan (Yousaf, et al., 2020). However, these forecasts were based on quite smaller datasets which is serious issue for producing reliable predictions (Moftakhar et al., 2020). And timeline of these forecasts have passed now. Further these forecasts under estimated the number of confirmed cases and deaths, may be due to changing post lockdown trends in the country. In addition, the said contribution used the autoregressive integrated moving average (ARIMA) models for the forecasts. The ARIMA models have been frequently used to forecast the patterns of this pandemic (Gupta and Pal, 2020; Majeed et al., 2020; Benvenuto et al., 2020 and Kumar et al., 2020). However, these models have some limitations. Firstly, the forecasts from these models are dependent on the previous behavior of the data along with preceding forecast errors, which means the forecasting error accumulates over time. In addition, the forecasting accuracy of these models is affected in presence of covariates (Brockwell and Davis, 2002). To avoid such issues, the Bayesian structural time series (BSTS) models can be used (Scott and Varian, 2013 and Brodersen et al. 2015). The BSTS models (i) allow the inclusion of prior information (ii) allow

the model parameters to evolve over time (iii) can handle large number of covariates using spike and slab prior (iv) are least dependent on certain hypothesized specifications (v) can investigate different components of the time series separately and (vi) do not require the linear regression component of the model (McQuire et a.l, 2019). These models have the capability to provide reliable forecasts for future outbreak of different diseases (Scott and Varian, 2014). These models have already been used for forecasting the health harms by drinking alcohol and alcohol licensing policies (de Vocht et al., 2017; McQuire et al., 2019). The proposed models provided 14% improvement (as compared to ARIMA models) in forecasting the consumer sentiments (Scott and Varian, 2014).

The strict social distancing in Wuhan facilitated China to restrict the spread of the pandemic in other provinces (Li et al., 2020). However, in Pakistan the lockdown was lifted at quite earlier stage of the pandemic. Therefore, the analysis of the impacts of lifting the lockdown in the country is very important.

We have conducted a study to analyze and forecast the various parameters of COVID-19 in Pakistan, for next thirty days. We have proposed more flexible BSTS models to obtain the said forecasts. The causal impacts of lifting the lockdown in the country have also been investigated using intervention analysis under BSTS models. We have observed the improved forecast accuracy for BSTS models, as compared to repeatedly used ARIMA models. The results from the study suggest that the cumulative number of confirmed cases and deaths is expected to increase exponentially in next thirty days. However, the recoveries are expected to increase faster than confirmed cases, due to which the active number of cases are expected to decrease during the next month. The study also revealed that lifting the lockdown at the earlier stage has seriously increased the trajectory of the outbreak in the country.

Materials & Methods

The data have been obtained from the published reports of the National Institute of Health (NIH), Islamabad, Pakistan. The data contain the details regarding cumulative (and daily) number of cases, deaths, recoveries and tests in the country (NIH, 2020). NIH updates the data on daily basis since February 26, 2020. As we have used the published data by NIH, no ethical approval is required for the study.

We have used R software to conduct the analysis regarding the study. The computations for ARIMA models have been obtained using auto.arima function available in forecast package, while bsts package has been used to obtain the forecasts for the BSTS models (Scott, 2020). The forecasting accuracy of BSTS models and ARIMA models have been compared based on different measures of forecast accuracy, such as, root mean square error (RMSE), mean absolute error (MAE), root mean square percentage error (RMSPE) and root median square percentage

error (RMdSPE). These are quite efficient measure of forecast accuracy of a model (Hyndman and Koehler, 2006 and Bowerman et al., 2004). For a parsimonious BSTS model, the residuals from the model should be white noise. We have tested this assumption using Ljung Box test at different lags. As we observed the improved forecasting capacity of BSTS models, the detailed forecasts have only been reported under BSTS models for brevity. The forecasts under BSTS models are based on prior information and current data (likelihood function). The bsts package uses spike and slab prior for the analysis. The prior information is combined with the likelihood function to produce the posterior distribution. The posterior distribution is estimated using Gibbs sampler (George and Mcculloch, 1997). In addition the Kalman filter and Bayesian model averaging were used to produce the final forecasts. The causal impacts of lifting the lockdown have been investigated using intervention analysis under BSTS models. These causal impacts have been computed using CausalImpact package in R (Brodersen and Hauser, 2020).

Results & Discussions

On July 11, 2020, Pakistan had 246,351 number of total confirmed cases, 93,217 active cases, 5123 deaths and 153,134 recoveries regarding COVID-19. The ratio of cumulative confirmed cases per 100 tests (RCT) stood at 16.01. Similarly, the ratio of cumulative deaths per 100 confirmed cases (RDC) was 2.08, and the ratio of cumulative number of recoveries per 100 confirmed cases (RRC) was 62.12. These figures have already challenged the healthcare system of Pakistan having total number of 215,436 doctors, 108,137 nurses, 19,218 community health workers (CHWs) and 41,689 labs across the country (WHO, 2020). The government figures also suggest that there are 6664 crucial care beds and about 2500 Ventilators for the COVID-19 patients in the country (NIH, 2020). These figures simply indicate that Pakistan has very limited healthcare infrastructure to cope up this pandemic. Hence, analysis of current and future trends regarding different parameters of this pandemic is important to make informed decisions in the country. Using the dataset obtained from NIH, we compared the forecasting accuracy of BSTS models with frequently used ARIMA models. The comparison has been made on the basis of RMSE, MAE, RMSPE and RMdSPE. The results regarding this comparison have been placed in Table 1. The diagnostic checking for the proposed models, using Ljung Box test, has been reported in Table 2. The forecasts for different parameters of the outbreak have been presented in Table 3 and in Figure 2. The separated analysis of trend, seasonality and regression components has been presented in Figure 3. On the other hand, the causal impact of lifting lockdown in the country has been discussed in Figure 4.

Figure 1 reports the current trends regarding different parameters of the COVID-19 in Pakistan, where the zeroth day denotes February 25, 2020. In particular, this figure includes the trends of RCT, RDC, RRC and RAC. Though RAC has decreasing trend and RRC has increasing trend, the issue for the country is that RDC is constant over time, which means that rate of deaths is directly proportional to number of cases. In addition, the RCT has increased over time, which is

quit alarming. It means the relative pace of spread has increased over time. The model comparison, reported in Table 1, simply suggests that forecast accuracy of BSTS models is better than that of ARIMA models, with few exceptions. Further, the results regarding diagnostic checking of the proposed models, presented in Table 2, elucidate that the residuals from the proposed models are white noise. Hence, the models can efficiently be used to obtain forecasts for different parameters of the pandemic. Table 3 and Figure 2 represent the forecasts (for next thirty days) for cumulative number of confirmed cases, cumulative number of deaths, cumulative number of recoveries and cumulative number of active cases, using BSTS models. It should be noted that this figure has logarithmic scale for y-axis, hence straight lines depict the exponential trend. From this figure, it can be assessed that fitted values (using BSTS models) are very close to the observed data (mostly overlapping the observed series), which indicates the model efficiency. Further, the cumulative number of cases, cumulative number of deaths and cumulative number of recoveries are expected to increase exponentially during the next thirty days. However, the growth in number of recoveries is expected to be faster than that of confirmed case, which is a good sign for the country. To be more specific, on August 10, 2020, the expected number of positive cases in Pakistan will be 333,693 with 95% prediction interval [267,470; 393,702]. Similarly, the number of deaths in the country is expected to reach 7,178 [6,003; 8,368] and recoveries may grow to 265,619 [244,902; 290,985]. However, it was very encouraging to observe that the number of active cases is expected to decrease to 68,344 [22,568; 102,717], which is lower as compared to current number. This relief is mainly due to rapid growth in the recoveries in the country. The demand for the healthcare facilities is expected to be slightly lower in the next month. The health administration should utilize this opportunity to improve the quality of health services regarding the COVID-19.

The additional advantage of using the BSTS models is that these models allow us to investigate the patterns of trend, seasonality and regressions separately. We have investigated the contribution of these components for cumulative number of positive cases, as the patterns for the other parameters were alike. As the cumulative number of positive cases depends on the cumulative number of tests, we considered the cumulative number of tests conducted in the country as covariate. The contributions of the said components have been presented in Figure 3. This figure elucidates that the cumulative number of positive cases is increasing exponentially in the country. The contribution of seasonality is slightly increasing over time. In addition, the contribution of regression component is also has an upward trend, this is mainly due to increase in rate of confirmed cases with respect to number of tests conducted.

The causal impact of lifting the lockdown in the country has also been discussed using intervention analysis under BSTS models. The results have been given in Figure 4. This figure simply indicates the significant increase in pace of the outbreak (in black lines) as compared to expected, had the lockdown not lifted (in grey shade). The probability of causal impact was 0.9989, which was quite high indicating the significance of the causal impacts of lifting the

lockdown in the country. On July 11, 2020, the total number of confirmed cases was 246,351; however under lockdown the number of cases would have been 147,583 with 95% interval [135,333; 160,807]. So, there is an absolute increase of 98,768 confirmed cases with 95% interval [85,544; 111,018]. Hence, the country has conceded 56% increase (in the confirmed cases) than the expected, if the lockdown would have continued. In addition, during the lockdown, the average rate of positive cases per 100 tests was 8.68%, while for the post lockdown period this rate stands at 13.21%. Similarly, the rate of deaths per 100 positive cases, for the lockdown period, was 1.46% which increased to 2.06% in the post lockdown period. On the whole, the lifting of lockdown has significantly increased the load on the Pakistani healthcare system. It is also a truth that the country cannot afford a prolonged lockdown. However, identifying the hotspots and enforcing lockdown in those areas may help. In addition, the serious effort to induce the people to abide by the SOPs in the country is fundamental.

The results from the study are in close agreement with the earlier studies conducted to forecast the patterns of COVID-19 in different countries such as (Tomar and Gupta, 2020; Li et al., 2020; Perc et al., 2020; Fanelli and Piazza, 2020; Majeed et al., 2020 and Moftakhar et al., 2020). Our results are also comparable with the earlier study carried out in Pakistan (Yousaf, et al., 2020). However, this study did not cover the impact of lifting the lockdown in the country. In addition, we have obtained these forecasts using a more flexible model.

Conclusions

The study has been conducted to analyze and forecast the different parameters of COVID-19 in Pakistan. The findings of the study suggest that the lifting of lockdown has increased the speed of the pandemic in the country. The number of positive cases is directly proportional to the number of tests carried out in the country. The total number of confirmed cases and deaths is expected to increase exponentially. However, the trajectory of recoveries is expected to be higher than that for confirmed cases. Therefore, the expected number of active cases is likely to decrease in the next month. We suggest that this is the right time where concerned authorities may put more efforts to further restrict the number of active cases in the country. If it happened, there are expectations that country will see the contraction in the outbreak of this pandemic. It should also be noted that Pakistan cannot afford the complete lockdown due to its frail economy; hence we cannot suggest the complete lockdown in the country again. However, properly identifying hotspots and enforcing lockdown partially may help. The government also needs to follow a strict tracking, tracing and quarantine strategy. And more importantly, all possible efforts and resources should be mobilized to improve the current healthcare infrastructure in the country to meet the needs of the coming month.

Our study has some limitations too. Firstly, the data may be underreported, as the sufficient random testing has not been considered in the country, and not all the patients report themselves

to health officials with a fear of staying fourteen days in quarantine. Secondly, the forecasts are based on the assumption that the current trends will follow in the next month, the forecasts can be misleading if this assumption is violated. Thirdly, no risk factors have been investigated as the data regarding the demography and social networks of the patients were not available.

References

Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S. and Picozzi, S. 2020. Application of the ARIMA model on the COVID-2019 epidemic dataset, Data in Brief, 29, 105340. https://doi.org/10.1016/j.dib.2020.105340.

Bowerman, B. L., O'Connell, R. T. and Koehler, A. B. 2004. Forecasting, time series and regression: An applied approach. Belmont, CA7 Thomson Brooks/Cole. p.18.

Brockwell, P. J. and Davis R. A. (2002). Introduction to Time Series and Forecasting, 39, New York; NY: Springer.

Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., & Scott, S. L. (2015). Inferring causal impact using Bayesian structural time-series models. Annals of Applied Statistics, 9(1), 247–274. https:// doi.org/10.1214/14-AOAS788.

Brodersen, K. H. and Hauser, A. (2020). Inferring Causal Effects using Bayesian Structural Time-Series Models, https://cran.r-project.org/web/packages/CausalImpact.

de Vocht, F., Tilling, K., Pliakas, T., Angus, C., Egan, M., Brennan, A. et al. (2017). The intervention effect of local alcohol licensing policies on hospital admission and crime: A natural experiment using a novel Bayesian synthetic time–series method. J Epidemiol Community Health, 71, 912-8.

Fanelli, D. and Piazza, F. 2020. Analysis and forecast of COVID-19 spreading in China, Italy and France. Chaos, Solitons & Fractals, 134, 109761.

Gupta, R.. and Pal, S. K. 2020. Trend Analysis and Forecasting of COVID-19 outbreak in India. MedRxiv. https://doi.org/10.1101/2020.03.26.20044511.

Hyndman, R. J. and Koehler, A. B. 2006. Another look at measures of forecast accuracy. International Journal of Forecasting, 22(4), 679-688.

Ippolito, G., Hui, D. S., Ntoumi, F., Maeurer, M. and Zumla, A. 2020. Toning down the 2019-nCoV media hype – and restoring hope. Lancet Respir Med. 8:230-1. doi: 10.1016/S2213-2600(20)30070-9.

Kumar, P., Kalita, H., Patairiya, S., Sharma, Y. D., Nanda, C., Rani, M., Rahmai, J. and Bhagavathula A. S. 2020. Forecasting the dynamics of COVID-19 Pandemic in Top 15 countries in April 2020 through ARIMA Model with Machine Learning Approach. MedRxiv. https://doi.org/10.1101/2020.03.30.20046227.

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y. et al. 2020. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. N Engl J Med., 382:1199–207. doi: 10.1056/NEJMoa20 01316.

279 Majeed, K. A., Adeleke, M. Popoola, L. 2020. Online Forecasting of Covid-19 Cases in Nigeria
280 Using Limited Data, Data in Brief, doi: https://doi.org/10.1016/j.dib.2020.105683.
281 McCloskey, B., Zumla, A., Ippolito, G., Blumberg, L., Arbon, P., Cicero, A. et al. 2020. Mass
282 gathering events and reducing further global spread of COVID19: a political and public
283 health dilemma. Lancet. 395, 1096-9. doi: 10.1016/S0140-6736(20)30681-4.
284 McQuire, C., Tilling, K., Hickman, M. and de Vocht, F. (2019). Forecasting the 2021 local
285 burden of population alcoholrelated harms using Bayesian structural time–series, Addiction,
286 114, 994-1003. DOI:10.1111/add.14568.
287 Moftakhar, L., Seif, M. and Safe, M. S. 2020. Exponentially Increasing Trend of Infected
288 Patients with COVID-19 in Iran: A Comparison of Neural Network and ARIMA
289 Forecasting Models, Iran J Public Health, 49(1), 92-100.
290 National Institute of Health (NIH) Islamabad Pakistan. https://www.nih.org.pk/. Accessed July
291 11, 2020.
292 Paules, C. I., Marston, H. D. and Fauci, A. S. 2020. Coronavirus infections-more than just the
293 common Cold. JAMA, 323(8):707–8.doi:10.1001/ jama.2020.0757.
294 Perc, M., Gorisek, M. N., Slavinec, M. and Stozer, A. 2020. Forecasting COVID-19. Front.
295 Phys. 8:127. doi:10.3389/fphy.2020.00127.
296 Scott, S. L. (2020). Bayesian Structural Time Series, https://www.rdocumentation.org/packages
297 /bsts/versions/0.9.5.
298 Scott S. L. and Varian H. (2014). Bayesian variable selection for nowcasting economic time
299 series. In: Goldfarb A., Greenstein S. M., Tucker C. E., editors. Economic Analysis of the
300 Digital Economy. Cambridge, MA: National Bureau of Economic Research, Inc; 119-35.
301 Scott, S. L. and Varian, H. (2013). Bayesian variable selection for nowcasting economic time
302 series. NBER Working Paper Series. Cambridge. https://doi.org/10.3386/w19567.
303 Tomar, A. and Gupta, N. 2020. Prediction for the spread of COVID-19 in India and effectiveness
304 of preventive measures, Science of the Total Environment, 728, 138762.
305 Wang, F. and Zhang, C. 2020. What to do next to control the 2019-nCoV epidemic?. The Lancet
306 395:10222, 391-393.
307 World Health Organization (WHO). 2020. Novel coronavirus (2019-nCoV); Available from
308 URL: https://www.who.int/ emergencies/diseases/novel-coronavirus-2019/situation reports.
309 Yousaf, M., Zahir, S., Riaz, M., Hussain, S. M. and Shah, K. (2020). Statistical analysis of
310 forecasting COVID-19 for upcoming month in Pakistan, Chaos, Solitons and Fractals, 138,
311 109926.
312 Zhan, C., Chi, K. T., Lai, Z., Hao, T. Su, J. (2020). Prediction of COVID-19 Spreading Profiles
313 in South Korea, Italy and Iran by Data-Driven Coding. medRxiv.
314 Zhou, T., Liu, Q., Yang, Z., Liao, J., Yang, K., Bai, W. et al. (2020). Preliminary prediction of
315 the basic reproduction number of the Wuhan novel coronavirus 2019-nCoV. J Evid Based
316 Med. 13:3-7. doi: 10.1111/jeb m.12376.
317

Table 1(on next page)

Measures of forecast accuracy for BSTS models and ARIMA models

Item	Results under BSTS Models				Results under ARIMA Models			
	RMSE	MAE	RMSPE	RMdSPE	RMSE	MAE	RMSPE	RMdSPE
Total Cases	453.36	302.43	0.0937	0.0958	468.57	311.41	0.1057	0.0102
Deaths	10.86	9.88	0.0630	0.0116	16.77	11.27	0.0614	0.0102
Recoveries	1186.16	469.49	0.1093	0.0268	1222.12	498.21	0.1242	0.0271
Active Cases	1196.98	627.50	0.0934	0.0188	1252.21	625.50	0.1085	0.0205

1

Table 2(on next page)

Results of Ljung Box test (Q-statistic) for diagnostic checking of the proposed models

1

BSTS Model for	Lag-5		Lag-10		Lag-30	
	Q-Statistic	P-Value	Q-Statistic	P-Value	Q-Statistic	P-Value
Total Cases	2.297	0.807	6.478	0.774	28.205	0.560
Deaths	10.088	0.073	12.414	0.258	42.046	0.071
Recoveries	3.108	0.683	3.750	0.958	11.894	0.999
Active Cases	7.658	0.176	11.639	0.310	32.847	0.411

2

Table 3(on next page)

Summary of forecasts and expected required resources after one month (on August 10, 2020)

1

Item	Point Forecast	95% Prediction Interval
No. of expected cases	333,693	(267,470 -- 393,702)
No. of expected deaths	7,178	(6,003 -- 8,368)
No. of expected recoveries	265,619	(244,902 -- 290,985)
No. of active cases	68,344	(22,568 -- 102,717)
No. of required crucial care beds	10,251	(3,385 -- 15,407)
No. of required ventilators	3,417	(1,128 -- 5.136)

2

Figure 1

Current Trends of the outbreak

The data points represent the ratios for different parameters of the pandemic. The green-red line indicates the Ratio of cumulative number of cases to cumulative number of tests (RCT). Pink-blue line indicates the **ratio of cumulative number of deaths to the cumulative number of cases (**RDC). **The dark.green-dark.magenta line shows the ratio of cumulative number of recoveries to the cumulative number of cases (** RRC).**The orange-green line represents the ratio of cumulative number of active cases to the cumulative number of confirmed cases (** RAC).

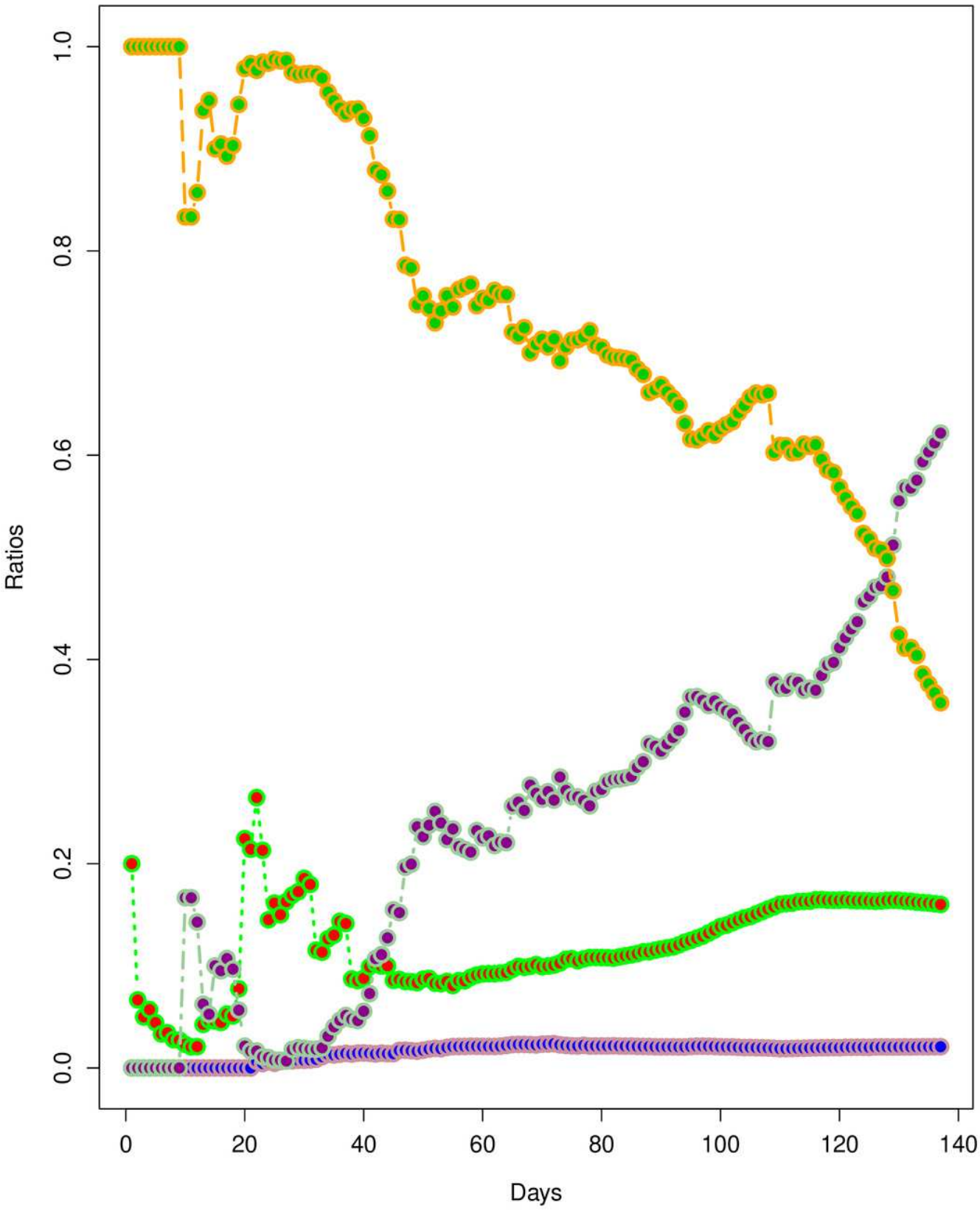

Ratios
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
120
140
Days

Figure 2

Forecasts for different parameters of the pandemic using BSTS models

Each line represents the observed versus fitted and forecast values. The cyan line represent the forecasts for cumulative number of cases. The magenta line shows the fitted values for cumulative number of cases (which have overlapped the observed numbers of cases given in cyan color). The green line indicates the forecasts for the cumulative number of deaths. The red line represents the fitted values for the cumulative number of deaths (which have overlapped the observed numbers of deaths presented in green color). The blue line shows the forecasts for the cumulative number of recoveries. The yellow lines depicts the fitted values for the recoveries (which have overlapped the observed numbers of recoveries presented in blue color). The dark.green line shows the cumulative number of active cases. The yellow-green line represents the fitted values for the active cases (which have overlapped the observed numbers of active cases presented in dark.green color)

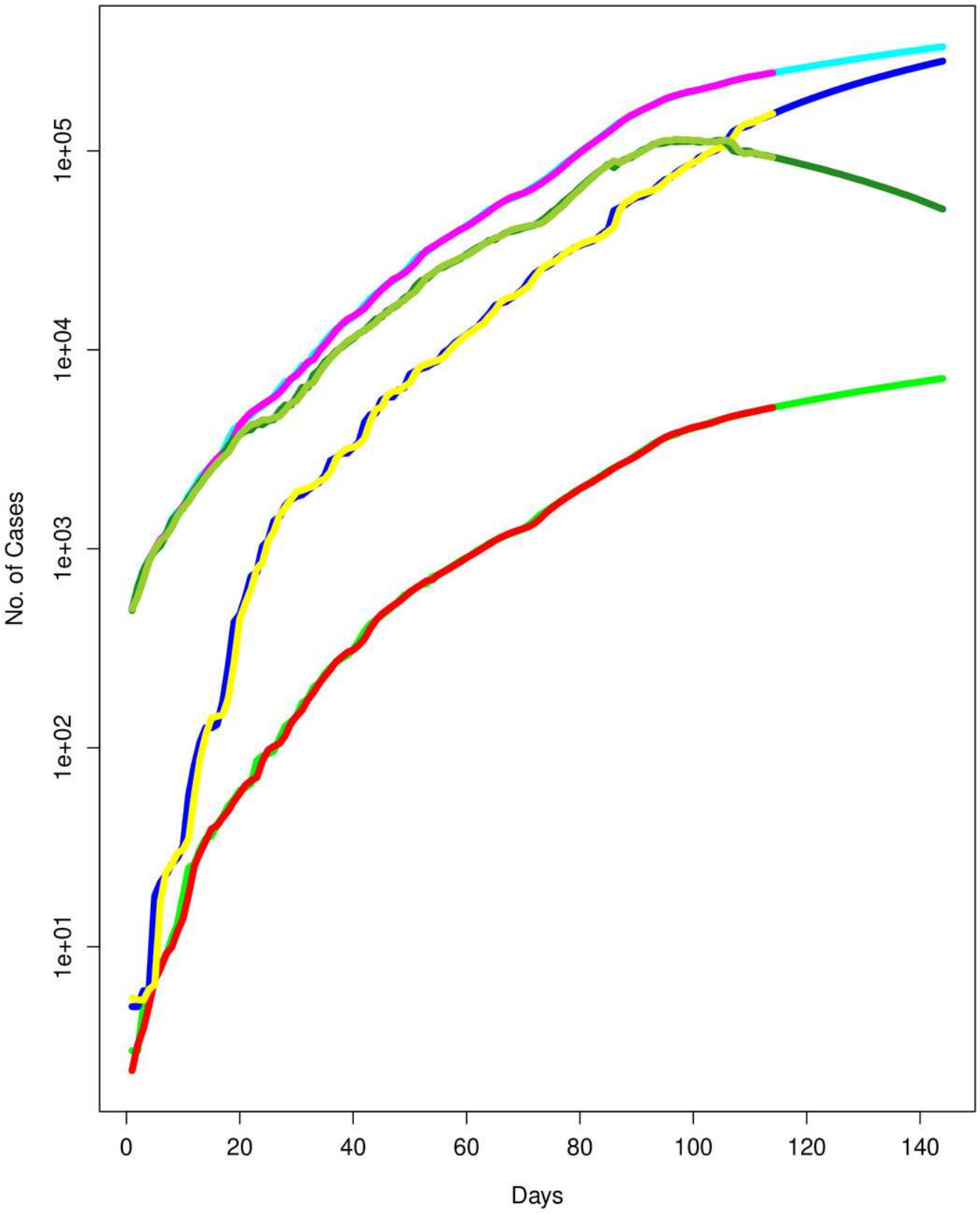

No. of Cases
1e+01
1e+02
1e+03
1e+04
1e+05
0
20
40
60
80
100
120
140
Days

Figure 3

Analysis of trend, seasonality and trend components

The upper panel shows the contribution of trend component. The central panel shows the contribution of seasonality and the bottom panel shows the contribution of regression component of the series.

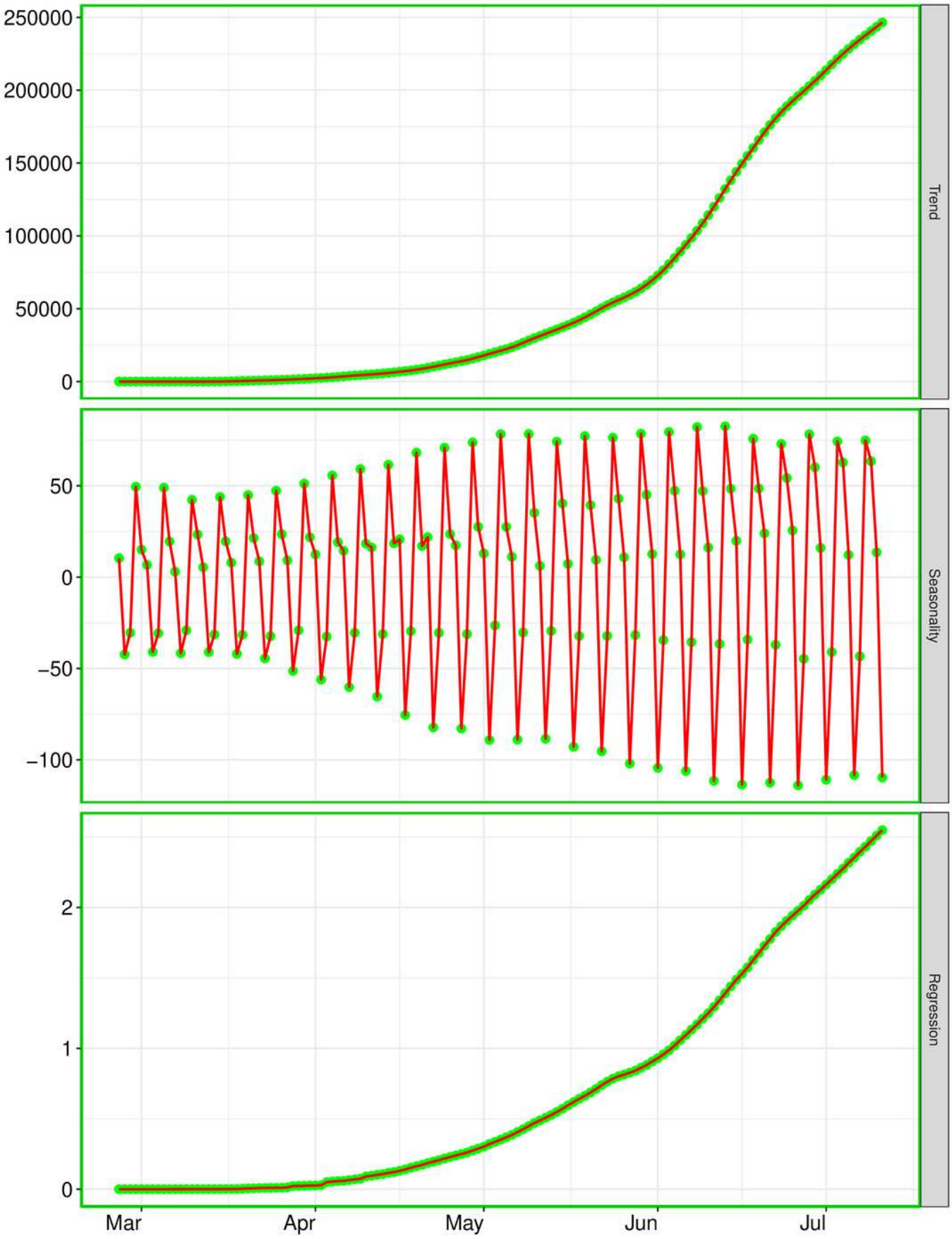
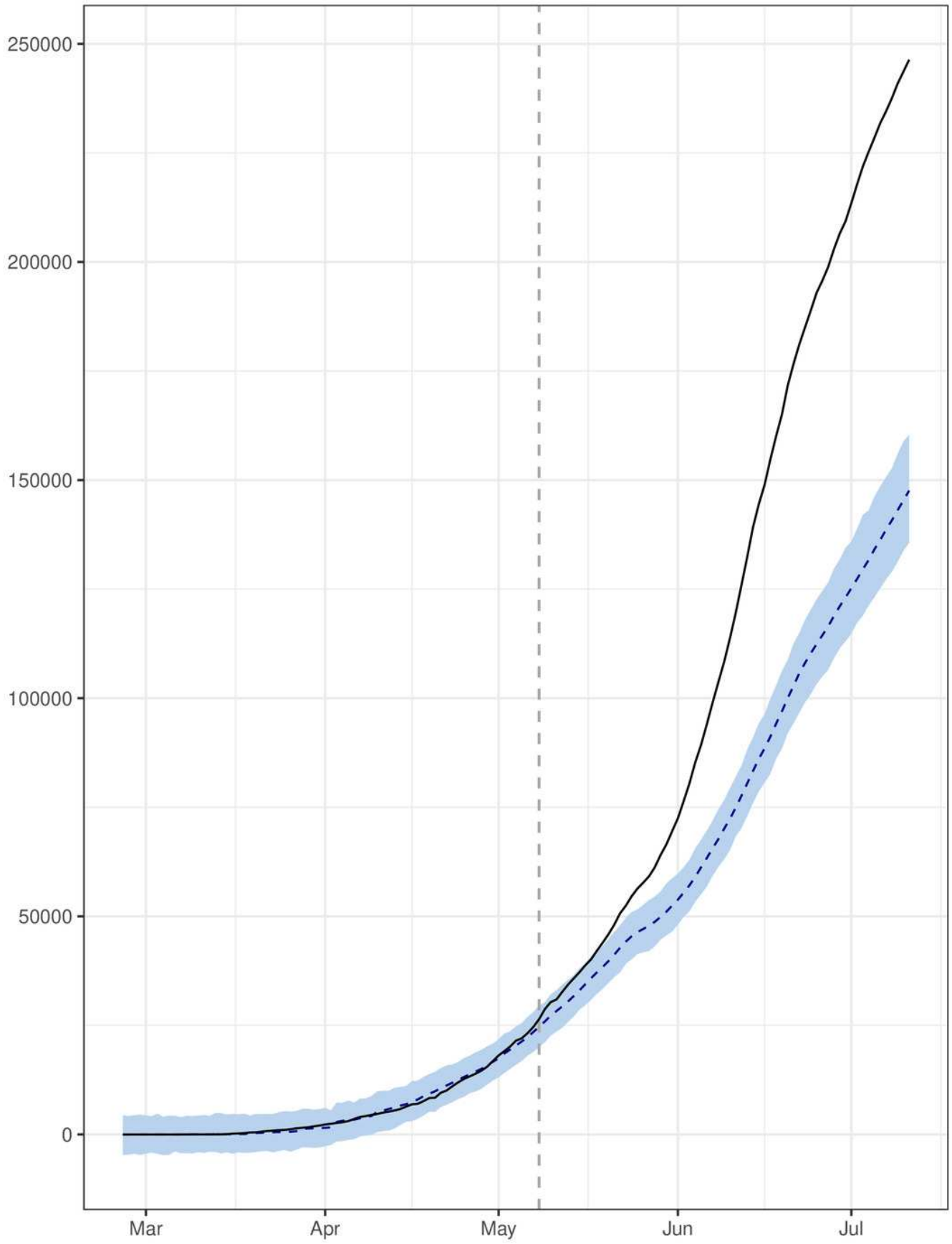

250000
200000
150000
100000
50000
0
Trend
50
0
−50
−100
Seasonality
2
1
0
Regression
Mar
Apr
May
Jun
Jul

Figure 4

Analysis of causal impacts of lifting lockdown in the country

The data points indicate the observed and expected number (hand the lockdown not lifted) of cumulative cases. The black line shows the observed number of cases. The grey line with grey shade indicates the expected number of cases, had the lockdown not lifted.


250000
200000
150000
100000
50000
0
Mar
Apr
May
Jun
Jul