

The Analysis and Forecasting COVID-19 Cases in the United States Using Bayesian Structural Time Series Models

Liming Xie

¹ North Dakota State University, Department of Statistics, Fargo, ND 58108, USA

***Corresponding Author:** Liming Xie, North Dakota State University, Department of Statistics, Fargo, ND 58108, USA, **Email:** limingxie2018@gmail.com

Citation: The Analysis and Forecasting COVID-19 Cases in the United States Using Bayesian Structural Time Series Models. Am J of Viro and Dis. 2020; 2(3): 01-08.

Submitted: 13 August 2020; **Approved:** 17 August 2020; **Published:** 19 August 2020

Abstract:

In this paper the Bayesian structural time series model (BSTS) is used to analyze and predict total confirmed cases who infected COVID-19 in the United States from February 28, 2020 through April 6, 2020 using the collect data from CDC (Center of Disease Control) in the United States. It includes variables of days, total confirmed cases, confirmed cases daily, death cases daily, and fatality rates. The author exploits the flexibility of Local Linear Trend, Seasonality, Contemporaneous covariates of dynamic coefficients in the Bayesian structural time series models. In addition, Causal Impact function in R programming is applied to analyze the model and read report of model. The results of the model show that the total confirmed cases who infected COVID-19 will be still most likely to increase straightly, the total numbers infected COVID-19 would be broken through 600,000 in the United States in near future (in the subsequent months). And then arrive at the peak around mid-May 2020. The cumulative prediction values are $\pm 1e+05$ of cumulative trend. Also, the model suggests that the probability of variable Recovered cases daily is 0.07.

Key words BSTS COVID-19 Causal Impact Trend Seasonality Forecast

Introduction

At present, COVID-19 is a brutal enemy who has killed tens of thousands of people. It posed a serious threat to human being. Moreover, people do not have effective ways to annihilate it. In the pandemic period, most of countries can currently take these actions just using passive protect measures like social distance, taking masks, and home isolation to reduce the infected risk. Many researchers analyzed china's epidemic period that China banned travel to and from Wuhan city for more than 50 days the cases who infected COVID-19 have been dropped down [1]. Some researcher even urged that the home isolation should keep until 2022 [2]. However, all economic activities have been shown down now. Thus, people expect that bending the curve of which keeps rising confirmed cases infected COVID-19 and death cases can come up early. Thus, the author tries to apply BSTS model to analysis U.S. COVID-19

case data from CDC in the United States by using Local linear trend, Seasonality, Contemporaneous covariates of dynamic coefficients. On the other hand, the author takes techniques such as prior distribution and prior elicitation, and posterior inference to give some interpretations of the model. The research gains more effective results.

In addition, many institutions and scholars conducted a lot research for COVID-19. Some researchers used an objective approach to predict the continuation of the COVID-19. They produced ten-day ahead point forecasts and prediction intervals [3]. Some people used 1334 cases from China that were line-list data to receive age-stratified estimation. The results displayed that the case fatality ratio was likely to be strongly influenced by the availability of health-care facilities [4]. Also, some scholars chose exponential smoothing family to test forecasting accuracy. The family suggested that

Cite this article: The Analysis and Forecasting COVID-19 Cases in the United States Using Bayesian Structural Time Series Models. Am J of Viro and Dis. 2020; 2(3): 01-08.

there were good effects [5]. In the initial phase, the author in this paper applied general time series forecasting methods followed by his previous techniques such as SARIMA model (1,1,2) (Seasonal autoregressive integrated moving average) and GARCH (generalized autoregressive conditional heteroskedasticity) to tested and analyzed the trend and seasonality of COVID 19 data [6,7]. The results showed good accuracy over a couple of forecasting competition. Also, the author tested to least absolute shrinkage and selection operator to estimate parameters of COVID-19 data by introducing into previous his method [8]. And then, used clustering method to allow “the eigenvalues of the correlation or covariance to test the parameters” [9]. But it was not positive effects.

On the other hand, some pursued new angle of research on COVID-19. For example, German and British scholars’ research found that it is probable to have three types of viruses: A, B, and C for COVID-19. U.S. cases are likely to come from them [10]. But epidemiologists often apply time series models such as ARIMA model (Autoregressive Integrated Moving Average), spectral analysis and the periodogram. For instance, an epidemiologist used the Lomb periodogram and the classic periodogram to Philadelphia Whooping cough timer series to achieve project research goal [11].

In despite of these research on COVID-19, no institutions and scholars use BSTS to analyze and predict COVID-19. Therefore, the author would like to test to use BSTS to analyze and forecast the total Confirmed cases in the United States that collects data from CDC from February 29, 2020 through April 6, 2020. The author thinks that the Causal Impact function in R programming is the feature selection in BSTS model that may perform causal inference by counterfactual predictions. It has good characteristics such as trend, seasonal, regression, holiday. It is one of good default model [12]. It is supposed that the time series of the treated unit is explained with a set of covariates while do not allow themselves to be impacted by the intervention. This is currently popular analysis method of time series model. Typically, it contains pre-period, post-period arguments and / or alpha, etc. for this paper, a time series model is automatically built and estimated.

Cite this article: The Analysis and Forecasting COVID-19 Cases in the United States Using Bayesian Structural Time Series Models. Am J of Viro and Dis. 2020; 2(3): 01-08.

The argument models provide over the model computations.

For most models of BSTS we know that “niter” is number of Markov Chain Monte Carlo (MCMC) sample to draw. The model can generate time series model for short- and long-term forecasting [13]. If there is higher number, it is more accurate inferences. “standardized. data” allows all columns of the data with moments estimated for the pre-intervention period prior to fitting the model to be standardized. It means that empirical Bayes accessing setting the priors so that the results will be linear transformations of the data. “prior.level.sd.” denotes in terms of data standard deviations that does have the Gaussian random walk of the local level. It can choose and let dataset with low residual volatility when regressing out known predictors such as total confirmed or fatality in this data. “nseasons” is period of the seasonal components. In general, it contains a seasonal component, set this to entire number larger than 1. For example, when there are daily observations, e.g. 1 for a day from a component of data, it interfaces currently only supports up to one seasonal component. In addition, it can let observations specify a number of seasonal components. Therefore, BSTS. model defaults to 1 that means no used seasonal component. “season. duration” is a kind of duration of each season. For example, if adding a day-of-week component to data with daily granularity, and offers the arguments model.args=list(nseasons=7. Season, duration=1), etc. “dynamic. Regression” includes the coefficients of time varying regression. It may combine local trend or local level that effects one of over specification.

2. Bayesian structural time series models.

Typically models of structural time series data are state space models. Most of them can be regarded as a set of equations. Structural models are much easier to generalize, such as covariates and we are not hard to process missing values with structural models [14]. For example, the local level model is

$$Y_t = W_t^T u_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \quad (1)$$

$$u_{t+1} = T_t u_t + S_t \omega_t, \quad \omega_t \sim N(0, \sigma_\omega^2) \quad (2)$$

The above equations assume under independent of all other unknowns. For 1st equation it is the observation function that connects with the observed data Y_t to a latent d-dimensional state u_t .

Second one is the state function that involves in the process of vector u_t through time. Thus, second one called “state” equation sometimes. Also, Y_t is as to a scalar observation and W_t is usually d-dimension output vector by transition matrix $d \times d$. S_t is often a matrix by $d \times q$. ε_t is regarded as a scalar observation error that could be noise variance σ_ε^2 and ω_t is a q-dimensional system error with a $q \times q$ state diffusion matrix σ_ω^2 . Here, note that $q \leq d$ exists.

Since structural timer series model has some advantages of flexible and modular and is statistical method for feature selection, it is more and more valued by time series researchers. For their flexible, it is probably because the model includes all of ARIMA models (Autoregressive Integrated Moving Average). Modular does combine with W_t , T_t and ω_t etc. from a library of component sub-models to capture significant structures in the data. Some of component models have been applied to capturing, the trend, seasonality, or else. For example,

(1) Local linear trend. The 1st components model. It is a form of the following equations:

$$Y_t = u_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \quad (3)$$

$$u_{t+1} = u_t + \beta_t + \omega_t, \quad \omega_t \sim N(0, \sigma_\omega^2) \quad (4)$$

$$\beta_{t+1} = \beta_t + \varphi_t, \quad \varphi_t \sim N(0, \sigma_\varphi^2) \quad (5)$$

This is a selection of modeling trends that can fit for local variation immediately. Especially for short term prediction requirement, it seems to appear in the flexible levels, but it is not appropriate to long term predictions. Here, u_t is trend, β_t is the seasonal component. Moreover, the model is faster to enters the state space framework [15].

(2) Seasonality.

Like local linear trend, seasonality is also state component model. The functions are as follows:

$$Y_t = u_t + \theta_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \quad (6)$$

$$\theta_{t+1} = -\sum_{s=0}^{S-2} \theta_{t-s} + \gamma_{\theta,t} \quad (7)$$

Here, S expresses the number of seasons and $\gamma_{\theta,t}$ that indicates joint contribution to the observed response Y_t . Please note that $S-2$ is most recent seasonal effects. Also, there is a scalar for the error term. Moreover, the equation is one of state model that does have less than full rank. For θ_{t+1} there is the mean that has zero of the total seasonal effect with summed over S seasons. For instance, when $S=3$ so that we can capture 3 seasons per time unit, the mean of the spring coefficient might be $-1 \times (\text{summer} + \text{fall} + \text{winter})$

then, for the seasonal model the transition matrix M_t will be $(S-1) \times (S-1)$ matrix with -1's with the top row, which the sub-diagonal and 0's else.

(3) Contemporaneous covariates of dynamic/static coefficients

The latter is to control time series when it gets no treatment. It is used to solve accurate counterfactual predictions and effects other unobserved causes. In this paper, it is mainly used of former, contemporaneous covariates with dynamic coefficients. It is a regression component to explain time varying ties for example, suppose covariates $a=1, 2, \dots, n$, was imported in the dynamic regression component, then we have the following expression:

$$Z_m^M \rho_m = \sum_{n=1}^n z_{n,m} \rho_{m,n} \quad (8)$$

$$\rho_{m,n+1} = \rho_{m,n} + \tau_{\rho,m,n} \quad (9)$$

Here, the above equation $\tau_{\rho,m,n}$ belongs to $\tau_{\rho,m,n} \sim N(0, \sigma_{\rho m}^2)$ the coefficient for the m th control series and $\sigma_{\rho m}$ is the standard deviation of its connected with random walk. If setting $Q_m = a_m$, $P_m = L_m$, on the other hand, we set the corresponding the transition matrix into $V_m = K_{n \times n}$, and $D_m = \text{diag}(\sigma_{\rho m}^2)$, then we can complete the dynamic regression components.

(4) Prior distributions and prior elicitation

Suppose γ expresses all of model parameters, φ is a set of the full state sequences and $\varphi = \varphi_1, \dots, \varphi_n$. They are allowed to a prior distribution $x(\gamma)$ with the model parameters and exist in $x(\varphi_0 | \gamma)$ is regarded as the initial state values. So, we use MCMC method to gain the sample results through $p(\varphi, \gamma | d)$. If there are a matrix τ with zero of elements and suppose $\theta_t = 1$, $\tau_t \neq 0$ and $\theta_t = 0$, otherwise Also, if τ_n expresses the nonzero scalars for the vector τ . If sequences $q = q_1, q_2, \dots, q_n$. Suppose $\sum_{-q} -1$ is the rows. The columns are be nonzero entries for q . Then, we can use the spike and slab in the following function:

$$p(q, \tau, 1/\sigma_\varepsilon^2) = p(q) p(\sigma_\varepsilon^2 | q) p(\tau_m | q, \sigma_\varepsilon^2) [16] \quad (10)$$

In addition, we can also write a formula for slab of the prior if using Gamma distribution:

$$\tau_q | \sigma_\varepsilon^2 \sim N(b_q \sigma_\varepsilon^2 (\sum_q^{-1})^{-1}) \quad [16] \quad (11)$$

(5) Posterior inference

For the full likelihood function, if the law of total probability is $p(A) = \sum_n P(A | B_n) P(B_n)$, then we have the function:

$$p(\delta, \tau, \sum_q \pi) = p(\delta, \tau, \sum_q \pi) p(\tau | \pi) p(\sum_q | \tau) p(\pi) \quad (12)$$

Here, δ expresses the set of values to multiple time series that is components of time series with trend, seasonality and others). Also, $\sum q$ is a correlated error. It is involved in a normal prior to have the observations standardize, and then generate a posterior distribution. Thus, we could use MCMC to compute the posterior probability distribution [16]

(6) Forecasting

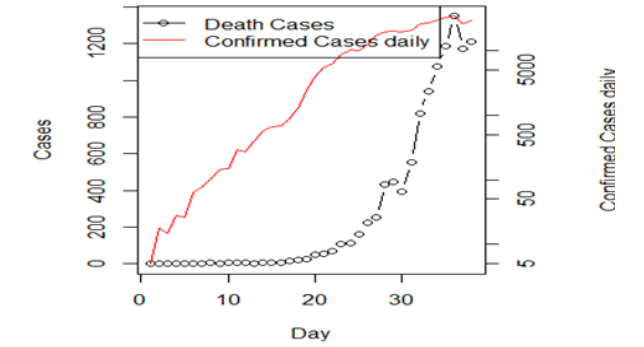
Since the forecasts are associated with the mode with the posterior predictive distribution, Bayesian analysis and forecasts have a formula that is like posterior inference. If we define δ as the set of values, then we can get the formula of posterior predictive distribution is as follows:

$$p(\delta | \omega) = \int p(\delta | \beta) p(\beta | \omega) d\beta. \quad [16] \quad (13)$$

3. Analysis and forecasting of empirical data

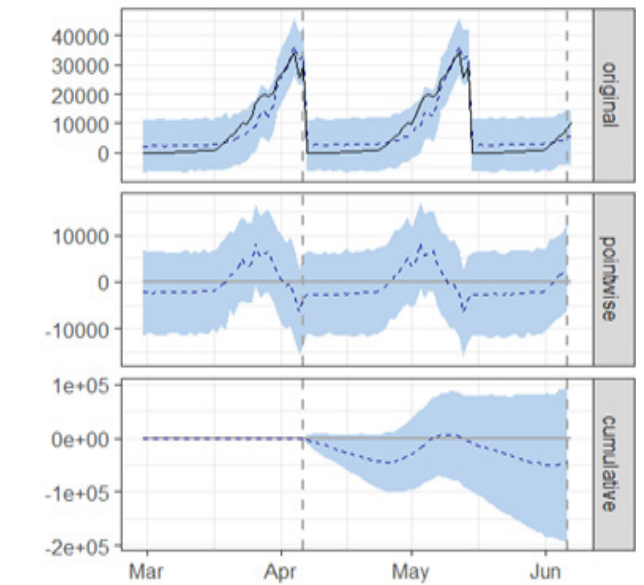
The data of this paper is used from CDC (Center of Disease Control) in the United States. The author collects variables is that (1) The days are time that gets to start to report the first death case, from February 29, 2020 through April 6, 2020. (2) Total Confirmed cases. This is counting same as the days by which patients have been tested the positive Corona virus Nucleic Acid testing no matter whether or not they have symptoms with cough, fever, and other upper respiratory tract infection. (3) Recovered cases daily who have been treated or in hospital or out outpatients, or self-heal without any treatment. (4) Death cases who were computed without vital signs. (5) Fatality rate. It is the numbers that computed by the formula=100 % (Death cases daily/ Confirmed cases daily). To find the relationship with the confirmed case daily and death case daily in the United States, the author made plot of time series for the both. The results showed that there is significant linear regression between the both. Please see the following figure 1.

Figure 1. The plot of time series with conformed cases daily and death cases daily.



For predict struct, plots are three panels: The 1st one shows that the data and a counterfactual prediction for the post-treatment period. The 2nd one displays the difference between observed data and counterfactual prediction trends, which is the pointwise case effect via the model estimations. The last one is plus of the pointwise contributions by the 2nd panel that generates a plot of the cumulative effect of the intervention. They are based mainly on the assumption that most of covariates impacted by the intervention. Also, the model is assumed by which the relationship between covariates and treated time series as established while the pre-period, remains stable to the post-period.

Figure 2. The prediction plot for original data, pointwise, and cumulative.



In the Figure 2 the model contains that the relationship between Confirmed cases daily in the and Deaths cases daily in the United States from February 29, 2020 throughout April 6, 2020. Clearly, it shows that the cases of confirmed data will be dropped as the end of April 2020, and then slightly rebounded in mid-May 2020 but decreased after that. The numbers of cases daily is fluctuated around 10,000 cases.

Causal Impact function is one of feature selections in BSTS model that may perform causal inference by counterfactual predictions. It is supposed that the time series of the treated unit is explained with a set of covariates while do not allow themselves to be impacted by the intervention. It is currently popular analysis method of time series model. Typically, it contains pre-period, post-period, mode.arg and/ or alpha, etc. for this paper, a time series model is automatically built and estimated.

The argument mode. arguments provide over the model.

In Table 1 R output suggests that the predictive values are the range of (73594, 439814) for total confirmed cases that infected COVID-19 virus in the United States. For cumulative prediction the value is $\pm 1e+05$ of cumulative trend.

Table 1. The R outputs of posterior inference using causal impact function

Posterior Inference (Causal Impact)		
Average		Cumulative
Actual	6480	395258
Prediction(s.d.)	7210 (1206)	439814 (73594)
95% CI	[4883, 9624]	[297834, 587088]
Absolute effect (s.d.)	-730 (1206)	-44556 (73594)
95% CI	[-3145, 1597]	[-191830, 97424]
Relative effect (s.d.)	-10% (17%)	-10% (17%)
95% CI	[-44%, 22%]	[-44%, 22%]
Posterior tail-area probability p: 0.25301		
Posterior prob. of a causal effect: 75%		

> summary (impact, „report“)
Analysis report {Causal Impact}

During the post-intervention period, the response variable had an average value of approx. 6.48K. In the absence of an intervention, we would have expected an average response of 7.21K. The 95% interval of this counterfactual prediction is [4.88K, 9.62K]. Subtracting this prediction from the observed response yields an estimate of the causal effect the intervention had on the response variable. This effect is -0.73K with a 95% interval of [-3.14K, 1.60K]. For a discussion of the significance of this effect, see below. Summing up the individual data points during the post-intervention period (which can only sometimes be meaningfully interpreted), the response variable had an overall value of 395.26K. Had the intervention not taken place, we would have expected a sum of 439.81K. The 95% interval of this prediction is [297.83K, 587.09K].

The above results are given in terms of absolute numbers. In relative terms, the response variable showed a decrease of -10%. The 95% interval of this percentage is [-44%, +22%].

This means that, although it may look as though the intervention has exerted a negative effect on the response variable when considering the intervention period as a whole,

this effect is not statistically significant, and so cannot be meaningfully interpreted. The apparent effect could be the result of random fluctuations that are unrelated to the intervention. This is often the case when the intervention period is very long and includes much of the time when the effect has already worn off. It can also be the case when the intervention period is too short to distinguish the signal from the noise. Finally, failing to find a significant effect can happen when there are not enough control variables or when these variables do not correlate well with the response variable during the learning period. The probability of obtaining this effect by chance is $p = 0.274$. This means the effect may be spurious.

Table 2. The summary on predictive average values and the ranges.

	Actual	Pred	Pred.lower	Pred. upper
Average	6479.639	7210.071	4808.34	9627.339
Cumulative	395258.000	439814.318	293308.71	587267.675
	Pred.sd	AbsEffect	AbsEffect.lower	
Average	1216.347	-730.4314	-3144.747	
Cumulative	74197.197	-44556.3184	-191829.583	
	AbsEffect.upper	AbsEffect.sd	RelEffect	
Average	1671.3	1216.347	-0.1013071	
Cumulative	101949.3	74197.197	-0.1013071	
	RelEffect.lower	RelEffect.upper	RelEffect.sd	
Average	-0.4365699	0.2318008	0.1673295	
Cumulative	-0.4365699	0.2318008	0.1673295	
	Alpha	P		
Average	0.05	0.2740964		
Cumulative	0.05	0.2740964		

In Table 2 we can see that pre.lower and Pred. upper are 293308.71 and 587267.67. For time series data of response variables and other covariates, data is often given a zoo object, a vector, a matrix or others. The response variable could be the 1st column and other covariates distributes to other columns; a zoo object will be regarded as its time that indicates the x-airs in plot technique. Option “response” is used to describe the observed response as supplied to Causal Impact function. Option “cum.reponse” provides cumulative response with the modeling time. Option “point.pred” is posterior mean of counterfactual predictions. “Point.pred.lower” and “point.pred.upper” are a set of lower limit or upper limit (1-alpha) posterior intervals. “Pomt.effect” is point-wise posterior causal effect. It may include lower and upper of

the poster intervals “Cum.effect.lower” or “Cum.effect.Upper” are similar as the above. The results show that the range of total cumulative confirmed cases infected COVID-19 virus in the United States should be around 4808 to 9627 daily and most likely to be (293309, 587268). That is, total numbers infected COVID-19 virus would be nearly 600,000 in the United States in subsequent months.

Figure 3. Smoothed values for components in the parameters of death and total confirmed.

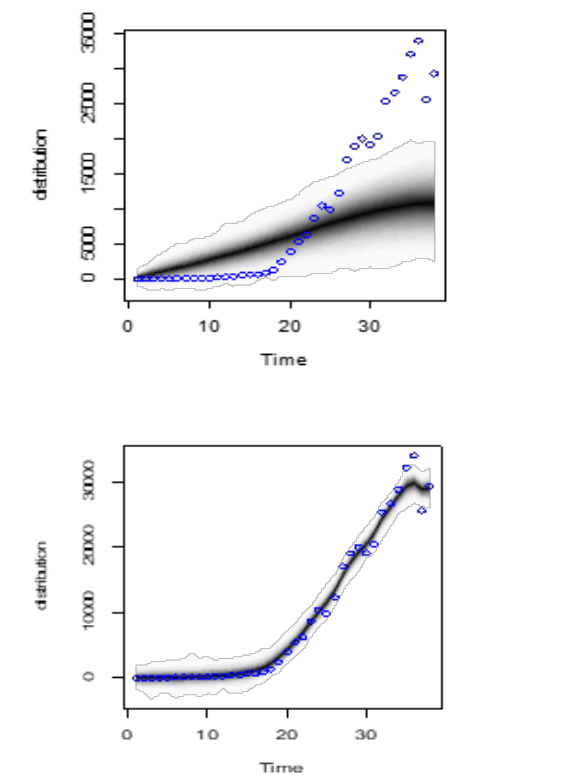


Figure 3 shows the data and estimated smoothed values. the actual data are shown as solid lines, the points denote smoothed values with ± 2 standard error bounds are shown as a blue color; tick marks indicate days of observation.

Figure 4. Structural models: the trend and seasonal parameters of the total confirmed variable.

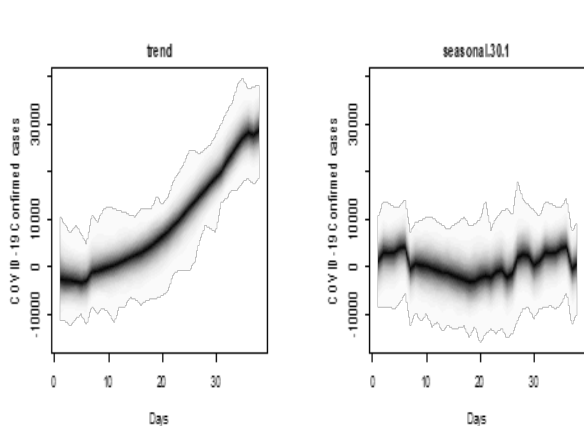
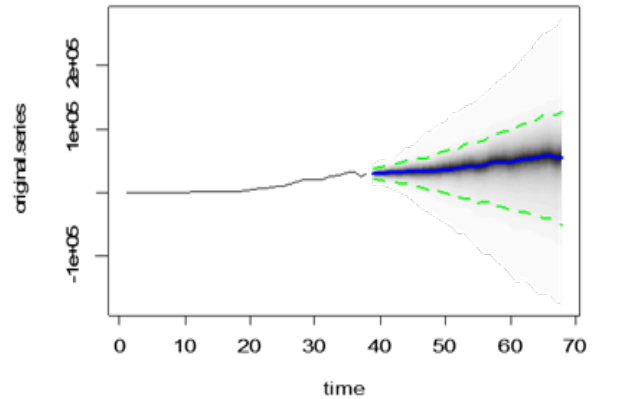


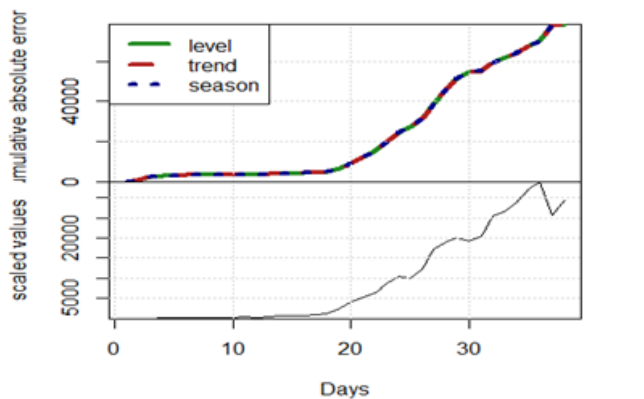
Figure 4 shows that estimated trend component and seasonal component of the total confirmed variable with the variable days. Gray areas represent some root MSE bounds.

Figure 5. Visualization of BSTS predicting original series.



The Figure 5 is a visualization of predictive values for original series. It suggests that, after 38th day, which the last day collect dataset, the total confirmed cases infected COVID-19 would be fluctuated with the range of $\pm 1e+05$. That is, it conforms that total confirmed cases could be increased in subsequent months. Therefore, it meets with the forecast of Figure 2 (The prediction plot for original data, point-wise, and cumulative), which is $\pm 1e+05$ of cumulative trend.

Figure 6. Plots for scale value and cumulative absolute errors of level, trend and season with days that started in February 29.



In Figure 6, we can see clearly that, using “AddlocalLinearTrend” and seasonal functions could allow us analyze level, trend and season change with variable “days” after making logarithm of data. The results show that predictive values regressed with original data.

It explains enough that the advantages of flexibility of local linear trend, level and season of BSTS state component models for the short-term predictions.

Figure 7. The plot of average coefficients from six parameters.

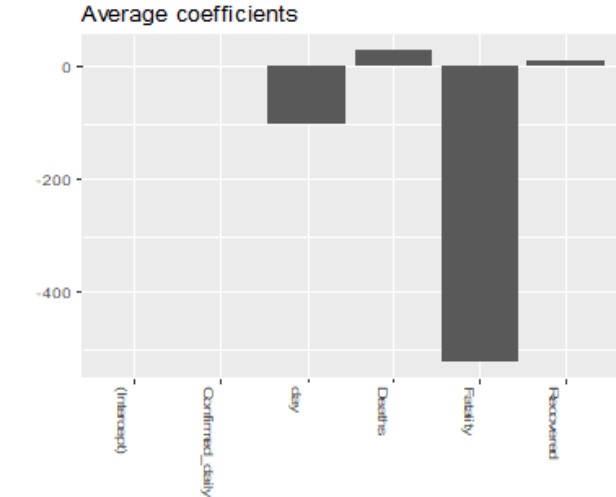


Figure 7 generates a histogram that displays average coefficients from Fatality, Day, Deaths, Recovered, and Confirmed daily. It shows that average coefficient of Fatality seems to be more significant than other variables in the model. Hence, the average coefficient of Fatality rate should be worthy in our analysis on these parameters. The author would like to point that average coefficient from Conformed Daily looks like to not have any status in the model.

Figure 8. Histogram of probabilities for data variables

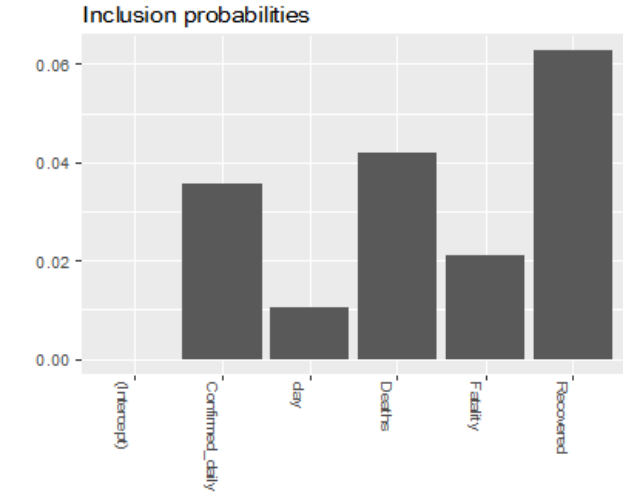


Figure 8 provides us inclusion probabilities comparing each variable in the analysis system. The results shows that the probability from variable Recovered is nearly 0.07. It looks like to play more important role than other variables in the data.

Second one is more than 0.04 of variable Death; around 0.038 of Confirmed Daily, lastly, other two variables, Fatality and Day, are equal to or lower than 0.02. Thus, it suggests that in the empirical estimation and analysis on the model we should eye on them.

4.Discussion

For most models of BSTS we know that “niter” is number of MCMC sample to draw. If there is higher number, it is more accurate inferences. “standardized data” allows all columns of the data with moments estimated for the pre-intervention period prior to fitting the model to be standardized. It means that empirical Bayes accessing setting the priors so that the results will be linear transformations of the data. “priort.level.sd.” denotes in terms of data standard deviations that does have the Gaussian random walk of the local level. It can choose and let dataset with low residual volatility when regressing out known predictors such as total confirmed or fatality in this data. “nseasons” is period of the seasonal components. In general, it contains a seasonal component, set this to entire number larger than 1. Thus, BSTS model defaults to 1 that means no used seasonal component. “season. duration” is a kind of duration of each season. “dynamic. Regression” includes the coefficients of time varying regression. It may combine local trend or local level that effects one of overspecification. Also, for posterior distribution, it can summarize the proportion if observing the data [17].

In this paper Causal Impact function of BSTS model is applied to analyze and forecast the model of the total confirmed cases in the United States from February29 to April 6.The author thinks that the total conformed cases that infected COVID-19 virus in the United States will be most likely to increase straightly, total numbers infected COVID-19 virus would be nearly 600,000 in the United States in near future (in the subsequent months). And then will appear in the peak around the md-May 2020. The empirical results suggest that the flexibilities of local linear trend, seasonality, and contemporaneous covariates of dynamic/ static coefficients have good effects on short term predictions.

References

1. Huajyu Tiam. Yomhong Liu, Yidan Li. Chieh-His Wu, Bin Chen, Moritz U. G. Kraemer, Bingying Li, Jun Cai, Bo Xu, Qiqi Yang, Ben Wang, Peng Yang, Yujun Cui, Yimeng Song, Pai Zheng, Quanyi Wang, Ottar N. Bjornstad, Ruifu Yang, Bryan T. Grenfell, Oliver F. Pybus, and Christopher Dye (2020). An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. Science. 10.1126/science abb6105.

2. Stephen M. Kissler. Christine Tedijanto, Edward Goldstein, Yonatan H. Grad, and Mare Lipsiteh (2020). Projecting the transmission dynamics of SARS-CoV-2 through the post-pandemic period. Science 10.1126/science. abb.5793.

3. Fotios Petropoulos and Spyros Makridakis (2020). Forecasting the novel coronavirus COVID-19. PLOS One. 15(3): e0231236. <https://doi.org/10.1371/journal.pone.0231236>.

4. Robert Verity, Lucy C Okell, Ilaria Dorigatti, Peter Winskill, Charles Whittaker, Nat-suko Imai, Gina Cuomo-Dannenburg, Hayley Thompson, Patrick G T Walker, Han Fu, Amy Dighe, Jamie T Griffin, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Anne Cori, Zulma Cucunubá, Rich FitzJohn, Katy Gaythorpe, Will Green, Arran Hamlet, Wes Hinsley, Daniel Laydon, Gemma Nedjati-Gilani, Steven Riley, Sabine van Elsland, Erik Volz, Haowei Wang, Yuanrong Wang, Xiaoyue Xi, Christl A Donnelly, Azra C Ghani, and Neil M Ferguson (2020). Estimates of the severity of coronavirus diseases 2019: a model-based analysis. Lancet Infectious Disease March 30. [https://doi.org/10.1016/S1473-3099\(20\)30243-7](https://doi.org/10.1016/S1473-3099(20)30243-7).

5. Taylor JW (2003). Exponential smoothing with a damped multiplicative trend. International Journal of Forecasting: 19(4): 715-725.

6. Liming Xie (2017). Time series analysis and prediction on cancer incidence rates. Journal of Medical Discovery: 2(3); jmd17030. DOI:10.24262/jmd.2.3.17030.

7. Liming Xie (2018). Analyzing and forecasting HIV data using hybrid time series models. Asian Journal of Probability and Statistics: 2(3); 1-12; Article no. AJPAS. 46566.

8. Liming Xie (2019). Estimation of parameters on Texas reservoirs using least absolute shrinkage and selection operate. Springer; 3 (2):93-104. <https://doi.org/10.1007/s42108-019-00018-8>.

9. Liming Xie (2019). Clustering analysis of the survey for mobility reasons in the US 1999-2017. Asian Journal of Probability and Statistics; 3(1): 1-12, Article no. AJPAS.47064.

10. Peter Forster, Lucy Forster, Colin Renfrew, and Michael Forster (2020). Phylogenetic network analysis of SARS-CoV-2 genomes. Proceedings of the National Academy of Sciences of the United States of America: Latest Articles 1-3. www.pnas.org/cgi/doi/10.1073/pnas.2004999117.

11. Ottar N. Bjornstad (2018). Epidemics models and data using R: Springer Nature. Switzerland AG, Gewerbestrasse 11, 6330 Cham, Switzerland.

12. Steven L. Scott and Hal R. Varian (2013). Predicting the present with Bayesian structural time series. Google; Predicting the present, November 22.

13. Sunghae Jun (2019). Bayesian structural time series and regression modeling for sustainable technology management. Sustainability, MDPI: 11, 4945; doi:10.3390/su11184945.

14. Rob J. Hyndman (2014). Forecasting: Principles & Practice. University of Western Australia. September 23-25.

15. Robert H. Shumway, and David S. Stoffer (2017). Time series analysis and its applications with R examples (4th Edition). Springer: Switzerland AG, Gewerbestrasse 11, 6330 Cham, Switzerland. p 315.

16. Steven L. Scott and Hal Varian (2014). Predicting the Present with Bayesian Structural Time Series. International Journal of Mathematical Modeling and Numerical Optimization. 5 pp 3-23.

17. Avril Coghlan (2017). A little book of R for Bayesian statistics (Release 0.1). Nov.07 2017.