Model Selection to Forecast the Trend of COVID-19 for the Counties Near Houston, Texas

Yoonsung Jung*

Abstract
After an outbreak from Wuhan, China, in late 2019, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is called COVID-19. The COVID-19 pandemic has had comprehensive consequences. This research project will explore 254 Texas counties' COVID-19 data to provide a statistical model to predict the future COVID-19 trend and unemployment, which is a major element in deciding socioeconomic status. Via a prediction model into a dataset for the daily cumulative test, daily cases, daily fatal case, and weekly unemployment data of Harris county and seven adjacent counties from 254 counties in Texas from March to July 2020, this study aims to provide the value and rationale behind the use and non-use of the data-driven prediction model in the State Health Service Department and State Workforce Commission.

Key Words: COVID-19, Time Series, ARIMA, Holt-Winter Additive Model, TBAT

1. Introduction

A COVID-19, called severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), started an outbreak from Wuhan, China, in late 2019 (Lu et al., 2020). By January 2020, the confirmed cases from more than 21 countries were 9976 cases. The first confirmed case of SARS-CoV-2 in the USA was on January 20, 2020. As of September 20, 2020, the USA cases are 6,833,800, 21.9% of global cases. Because of SARS-CoV-2, the economic effects of the worldwide recession are not only for economics but also for education and human health. Most countries implemented the general guideline to prevent further spread, such as reducing social contact, social distancing, wearing masks, and self-isolation.

Most of the government, official agents, and media reports are focused on the cumulative epidemic cases. Texas Department of State Health Services (dahs.texas.gov) is working closely with the Centers for Disease Control and Prevention (CDC) to control the spread of the new coronavirus disease 2019 (COVID-19). Contrarily, the official has almost no attention to describing, quantifying, comparing, and forecasting the cases within and between cities, counties, and states using case datasets. The capability to identify the COVID-19 rate at any given point in time is vital to prevent further spread. Another critical component is the efficient short-term prediction method to accurately forecast the active, fatality, and cumulative cases at any time point. The use of time series forecasting is the best approach to predict future patterns and trends of COVID-19 cases based on current case datasets.

The rest of this paper is organized as follows. Section 2 summarises previous scientific works related to this study. Three different forecasting methods for data analysis: ARIMA, the Holt-Winter Additive Model, and TBAT, are introduced in Section 3. Section 4 presents the summary of the performance of three time series forecasting methods. The conclusion is given in Section 5.

2. Related Work

A time series uses a series of data with time order and of equally spaced points in time. The infectious disease data, such as active cases and deaths, is a time series because it measures

*Prairie View A&M University, 620 E.E. O'Banion St., Prairie View, Texas 77446

daily, weekly, or monthly points in time. Time series analysis analyzes time series data to extract significant information and other characteristics of the data.

The goal of Dominguez et al. (1996) was to detect the behavior of influenza activity indicators and to evaluate the time series models to improve the influenza epidemic detection. Soebiyanto et al. (2010) showed the importance of climatological parameters on predicting seasonal influenza transmission in two warm-climate regions, Hong Kong, and Maricopa County in Arizona, USA. Hanf et al. (2011) applied the time series forecasting to investigate temporal correlations between the monthly Plasmodium falciparum case and El Niño Southern Oscillation (ENSO). Chadsuthi et al. (2012) evaluate the association between seasonal leptospirosis transmission and climate factors such as rainfall and temperature. They determine leptospirosis's seasonal pattern model.

Song et al. (2016) applied the seasonal ARIMA (SARIMA) model to construct a time series model for short-time prediction with the monthly seasonal influenza incidence. Zhang et al. (2019) developed a multivariate seasonal autoregressive integrated moving average (SARIMA) model to predict seasonal influenza epidemics.

Yin et al. (2020) proposed a robust and efficient time series mutation prediction model, Tempel, for the mutation prediction of seasonal influenza A viruses. Roosa et al. (2020) used three phenomenological models previously applied to infectious diseases such as SARS, Ebola, Pandemic influenza, and dengue. They emphasized the significance of real-time short-term forecast of the cumulative confirm cases. Yang et al. (2020) used a symmetrical function for the daily and total number of infections and deaths and the pandemic's corresponding turning points. Hu et al. (2020) developed an artificial intelligence forecasting model, called a modified stacked auto-encoder, which can model the outbreak's transmission dynamics and forecast the confirmed cases.

3. Statistical Methods

Forecasting is an activity to calculate or predict some future event or condition. Forecasting is widely used today in many fields, such as industry, marketing, economy, and finance. Time series forecasting only requires the previous data of a time series to generalize the forecast. Several available time series forecasting approaches are available such as the moving averages method, linear regression with the time factor, exponential smoothing, and more.

Papastefanopoulos et al. (2020) compared six different time series approaches such as ARIMA (Box and Jenkins, 1970), the Holt-Winter Additive Model (Chatfield, 1978), and TBAT (De Livera et al., 2011), Facebook's Prophet (Taylor and Letham, 2018), N-BEATS (Oreshkin et al., 2019) and DeepAR (Salinas et al., 2019; Alexandrov et al., 2019). Three approaches outperform Facebook's Prophet and the deep learning methods such as N-BEATS and DeepAR. This section introduces three different time series forecasting methods: ARIMA, the Holt-Winter Additive Model, and TBAT.

3.1 Autoregressive Integrated Moving Average (ARIMA)

The autoregressive integrated moving average (ARIMA) model developed for economics application is the generalized type of an autoregressive moving average (ARMA) model.(Yule, 1926; Box and Jenkins, 1970) ARIMA's popularity comes from implementing various exponential smoothing models (Yule 1926, McKenzie 1984) and the well-known Box-Jenkins methodology (Wold, 1939) .

There are two advantages to using ARIMA model. First, ARIMA is well-understood and easily explains the relationship between the independent variables and the dependent variables. The second is that the model section for ARIMA models can be implemented in an automated way to maximize prediction accuracy (Kane et al., 2014). However, ARIMA also has a weak point, which cannot deal with non-linear patterns or relationships.

3.2 Holt-Winter Additive Model (HWAAS)

Holt-Winter Additive Model is an extension of Holt's method to capture seasonality. The Holt-Winters seasonal method comprises the forecast equation and three smoothing equations for the level, trend, and seasonality with corresponding smoothing parameters. The Holt-Winter forecasting is a popular projection method to deal with the trend and seasonal variation. It is a simple yet powerful because empirical studies using automatic versions of the method are not as accurate as of the more complicated Box-Jenkins procedure. Chatifield (1978) shown how to improve the automatic Holt-Winter forecasts by subjective modifications. He suggests general recommendations for the choice of a univariate forecasting procedure and the implementation of the Holt-Winters procedure. Kalekar (2004) discussed two exponential smoothing models, the multiplicative seasonal model, and the additive seasonal model to separate the pattern's trend and seasonality. Chatfield and Yar (1988) provided detailed suggestions for implementing the Holt-Winters model for an automatic and non-automatic approach. Gelper et al. (2010) presented the robust forecasting of the exponential and Holt-Winters smoothing method for univariate time series with outliers.

3.3 TBAT

The TBATS model is a forecasting model based on exponential smoothing. The TBAT model's name is an acronym for Trigonometric (Harvey et al., 1997), Box-Cox transform (Box and Cox, 1964), ARMA errors (Adhikari and Agrawal, 2013), Trend and Seasonal components. De Livera (2010) applied the BATS model to improve the prediction performance compared to the simple Sate Space Model. However, the BATS model does not have good performance when the seasonality is complex and high frequency. De Livera (2011) proposed a TBATS model, a combination of the BATS model and Trigonometric Seasonal. Livera et al. (2011) proposed a modified trigonometric formulation to decompose complex seasonal time series. It used exponential smoothing to identify and extract complex seasonal components.

The TBAT modeling has many key advantages: 1) Box-cox transformation can deal with data with non-linearity and then somewhat makes the variance becomes constant, 2) ARMA model on residuals can solve the autocorrelation problem, 3) No need to worry about initial values, 4) It can get not only point prediction but also interval prediction, 5) The performance is better than simple state space model, 6) It can deal with data with non-integer seasonal period, non-nested periods, and high-frequency data, and 7) It can do multi-seasonality without increasing too many parameters. However, the weakness of TBAT is that 1) The assumption of $\epsilon_t \sim NID(0, \sigma^2)$ may not hold, 2) it can not add explanatory variables, 3) the performance for long-term prediction is not very well, and 4) the computation cost is big if the data size is large.

4. Data Application

The data analyses for ARIMA, Holt-Winter additive model, and TBAT were conducted in R v.4.0.3 (R foundation for Statistical Computing) using the *readxl*, *TTR*, and *forecast* packages.

4.1 Data Description

As a result of the COVID-19 outbreak, state government and many organizations have created publicly available datasets for research and analysis. Texas government website also provides publicly available data. Texas Department of State Health Services (dshs.texas.gov/coronavirus/additionaldata/) provides the daily data for Cases over Time by County, Fatalities over Time by County, Estimated Active Cases over Time by County, and Cumulative Tests over Time by County. The other is Texas Workforce Commission (www.twc.texas.gov), which provides economic data, including the weekly Unemployment Claims by County, zip code, Texas House, Texas Senate, US Congress, and Workforce Development Area.

In this study, The data for cumulated total cases and active cases over Time by County was used for comparing time series models to forecast efficiently at weekly short-term period. The total confirmed case data obtained from Texas Department of State Health Services is from March 4, 2020 to July 31, 2020, which has three missing dates, March 7, 8, and 14. The active case data is from March 7,2020 to July 31, 2020.

4.2 Comparison Process for COVID-19 data

This study aims to compare time series models and find a time series model with high accuracy to forecast at a short-term period as weekly. Papastefanopoulos et al. (2020) investigated the accuracy of six time series models for coronavirus outbreak detection in ten different countries. As a result of the study, they demonstrated that ARIMA, TBAT, and Holt-Winters models outperform the deep learning methods such as N-BEATS and GlutonTS, and Prophet developed by Facebook. This study considers only three models, ARIMA, TBAT, and Holt-Winters, demonstrated by Papastefanopoulos et al. (2020). The steps for the model comparison of COVID-19 time series data are as follows:

Step 1. Data Import: The first thing to analyse time series data is to read it into R. The *read_excel()* function, which is in the *readxl* R package, was used for importing Excel data from a local drive.

Step 2. Create time series object: The function *ts()* is used to create time-series objects with the imported data.

1. create first time series objects with time series data by April 20, 2020.
2. After finished the following Step 3 to 6, update time series objects as of the size of forecast (seven per one week).
3. Keep repeat the step 2 to 6 for 15 coming weeks.

Step3. Time series plot: The *plot.ts()* in R is designed specifically for ts objects was used for making a plot of the time series data.

Step 4. Fit a Predictive Model: Three time series models such as Holt-Winters, ARIMA, and TBATS, are tested for COVID-19 data:

- The *HoltWinters()* function is for fitting a simple exponential smoothing predictive model.
- The *arima()* function was used to find the appropriate ARIMA model.
- The *tbats()* function is for forecasting time series with complex seasonal patterns using exponential smoothing.

Step 5. Forecasting: Use *forecast()* function for forecasting from time series or time series models (Step 4). The *forecast()* function provides information about the forecasting method, the data used, the point forecasts obtained, prediction intervals, residuals and fitted values.

Step 6. Model Accuracy Measure: With forecasting data at Step 5, measure the performance of the accuracy (RMSE) of each model for each county by *accuracy()* function, which returns range of summary measures of the forecast accuracy.

4.3 Basic Summary of Data

Harris County is the biggest COVID-19 impacted county in Texas. The following results are for Harris county and seven counties nearby Harris county. Two data are used in the analysis for the total confirmed cases per county from March 4, 2020 to July 31, 2020, and for the number of active cases from April 4 to July 31, 2020. The plot of two different time series data can help get a deeper understanding of the spread of COVID-19 in each of eight counties.

Figure 1 and 2 are for the spread of the pandemic for eight counties about the total confirmed cases and the number of active cases. The trend between the confirmed case and active cases has a similar trend because each county's active case for each day was the rest after subtracting both recovered and the death's patients from the confirmed cases. After the outbreak in Texas, March 30 and June 18 are two important time points because the trend has been dramatically changed.

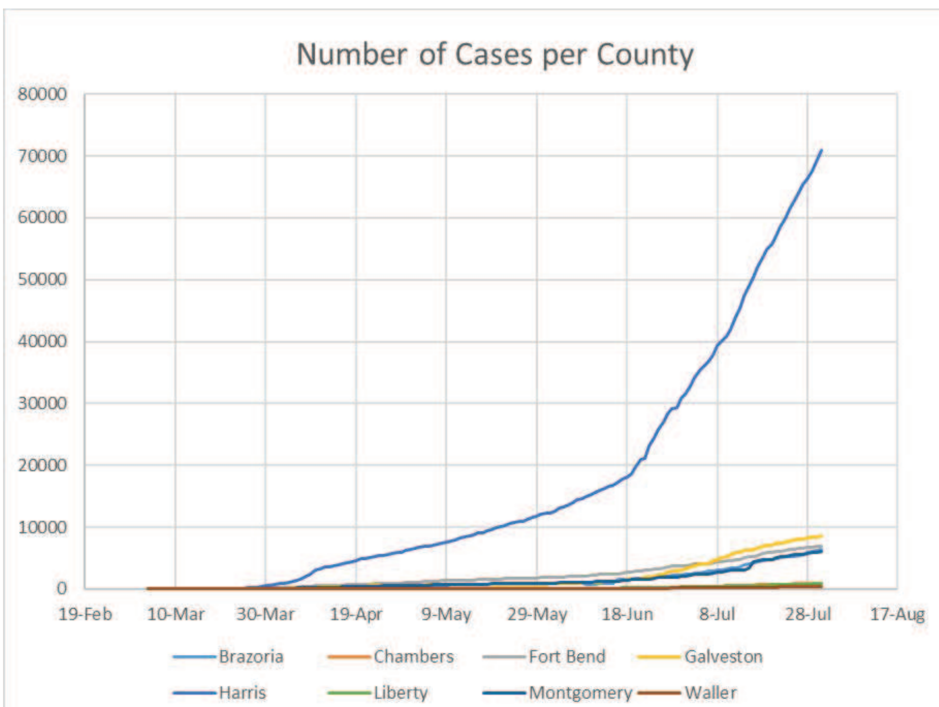

Figure 1: Number of Confirmed Cases per County

Especially, Harris county's confirmed cases show us a clear trend. Four counties, including Brazoria, Fort Bend, Galveston, and Montgomery, have an increasing trend after June 18 as Harris county. Finally, the bottom three lines are for Chamber, Liberty, and Waller County, which counts also continue to increase, but it looks stable and under control.

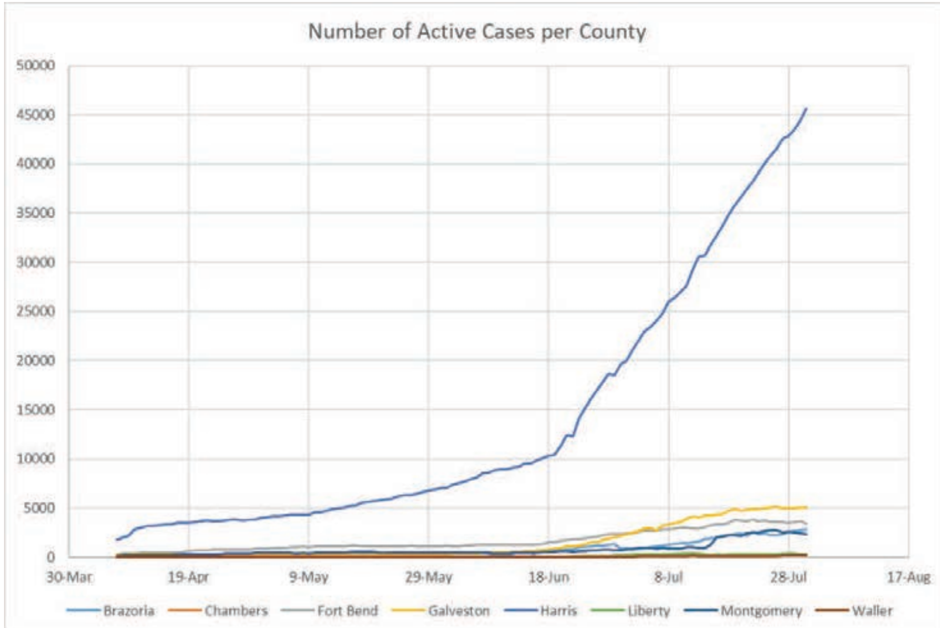

Figure 2: Number of Active Cases per County

With the confirmed cases, it isn't easy to compare the trend among eight counties because the number of confirmed cases in Harris county is so big different from seven other counties. This study considers the percentage of confirmed and active cases for each county's total population. The percentage change per county population could help us get more clear comparisons of the spread of COVID-19 between counties.

Figure 3 and 4 show the percentage of confirmed and active cases for the total population per county. The percentage of confirmed and active cases in Figure 2 has a dilatory increase trend by June 18. After June 18, the trend of confirmed and active cases is as follows:

First, the percentage of confirmed cases in eight counties have increased quickly. The change for confirmed cases at Harris county is 1.1% from about 0.4% to 1.5%. However, Galveston and Chambers county have a high impact of COVID-19 more than Harris county. Galveston county has the biggest change from 0.5% to 2.5% for the percentage of confirmed cases by the end of July. The percentage of Chambers county has changed 1.7% from 0.3% to almost 2.0% by the end of July. Chambers county has passed the percentage of Harris county in July. Brazoria county also has passed Harris county in the middle of July. Four counties, including Fort Bend, Liberty, Montgomery, and Waller county, also increased the speed but did not overpass the percentage of Harris county.

Second, the change of active cases at Galveston county is from 0.2% up to 1.5%. But Harris county was up to 1.0% at the end of July. Brazoria county increased up to 0.75%. The other counties also have an increasing trend, but after July 13, the active percentage stays between 0.3% and 0.5%.

4.4 Performance Comparison

Performance results of the three time series model for eight counties are summarized in Table 1 to 4. This study aims to find the best fit time series model for the root mean squared error (RMSE).

Table 1 and 2 show the result for RMSE of three models' performance for each county.

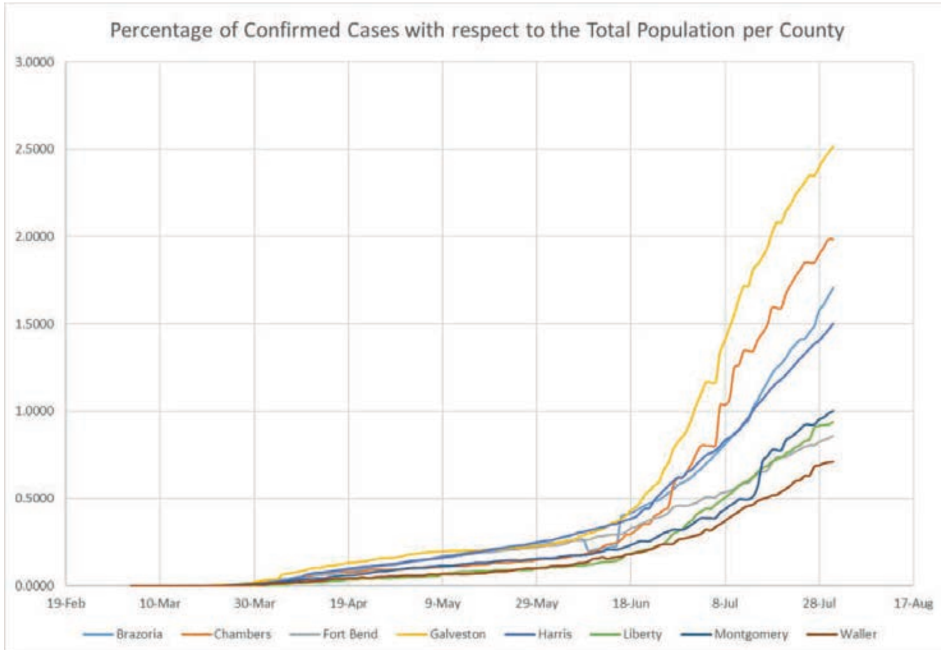

Figure 3: Percentage of Confirmed Cases with respect to the total population per County

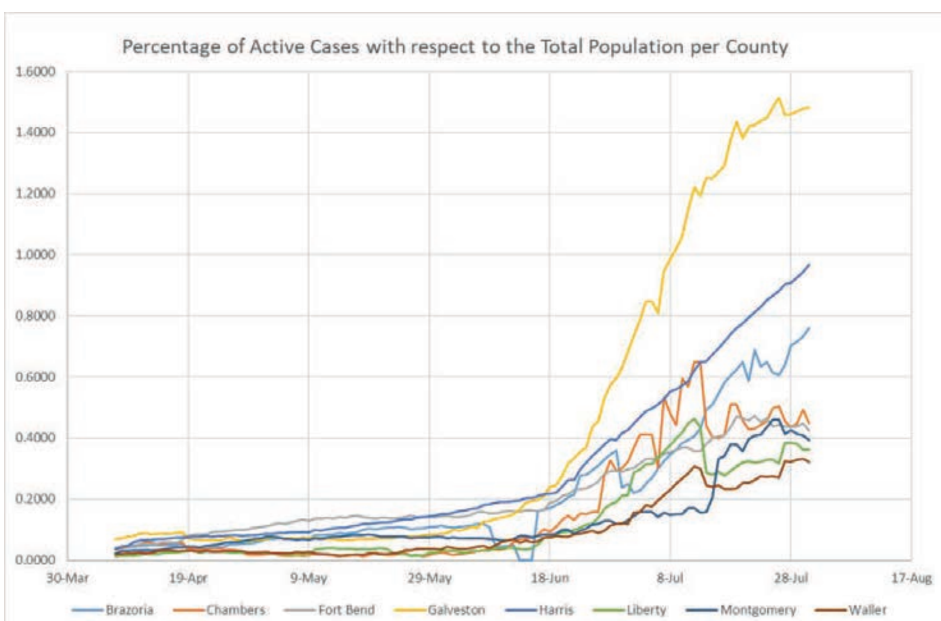

Figure 4: Percentage of Active Cases with respect to the total population per County

Firstly ARIMA and secondly TBAT seems to be the best performing methods overall in terms of RMSE in Fort Bend, Galveston, Harris, Mongomery, and Waller county. TBAT achieved the best results and ARIMA next in Chambers and Liberty county. However, Holt-Winters forecasting procedure does not work for eight counties because it is probably due to the lack of high volumes of data to cope with the trend and seasonal variations.

	Brazoria County			Fort Bend County		
Week	ARIMA	H-W	TBAT	ARIMA	H-W	TBAT
1	**4.88729**	4.99758	5.18260	**20.34316**	22.78541	22.77436
2	**5.94691**	6.10316	6.21489	**20.30437**	22.62022	22.63180
3	**6.79819**	6.82130	6.93505	**19.50527**	21.77083	21.77047
4	**7.24655**	7.33863	7.41597	**19.72941**	21.60928	21.65050
5	**7.38378**	7.47272	7.55777	**21.38225**	21.43955	21.51781
6	**7.38378**	7.47272	7.55777	**20.18949**	21.30689	21.39641
7	**7.75893**	7.85222	7.91423	**21.00295**	21.01899	21.11603
8	**7.56636**	7.64928	7.71188	**21.96794**	22.18984	22.24804
9	27.03813	26.22991	**26.06720**	**22.99200**	23.23987	23.26644
10	67.31305	**66.55012**	67.12384	**28.34264**	28.61159	28.47063
11	**64.25789**	64.81922	64.30164	32.17870	32.46565	**29.45489**
12	**62.89434**	63.41571	62.92309	**30.88627**	34.78246	33.93648
13	**61.60139**	62.08522	61.62094	**33.78799**	36.69270	36.03916
14	**62.97660**	63.43948	62.98053	**40.63314**	43.30837	42.92604
15	**62.44267**	62.88541	62.45351	**43.82014**	47.28596	44.20112
	Chamber County			Galveston County		
Week	ARIMA	H-W	TBAT	ARIMA	H-W	TBAT
1	**1.29786**	1.32255	1.30770	**12.94178**	13.16153	13.52428
2	**1.29291**	1.31868	1.30824	**12.29334**	12.47966	12.88856
3	**1.24119**	1.26320	1.25335	**11.83909**	12.02929	12.48802
4	1.43614	1.21657	**1.21034**	**11.83909**	12.02929	12.48802
5	1.39415	1.17994	**1.17483**	**10.81160**	10.95345	11.51844
6	1.39596	1.18275	**1.17798**	**10.33729**	10.49069	11.05990
7	1.41837	1.18110	**1.17691**	**10.00863**	10.14511	10.72684
8	1.29531	1.18282	**1.17714**	**10.21823**	10.30912	10.84587
9	**1.43277**	1.44736	1.43887	**10.90304**	11.33333	11.77790
10	1.67237	1.77369	**1.58921**	**13.23154**	13.91705	14.23893
11	**4.40397**	4.49864	4.46164	**21.57482**	21.95039	22.07135
12	**4.76405**	5.29612	5.24852	25.90820	26.12643	**22.15313**
13	10.29714	10.82303	**9.76809**	**50.24525**	55.39974	55.05728
14	10.85522	11.42568	**10.21379**	58.71156	60.41990	**56.20390**
15	10.86983	11.47907	**10.77223**	60.62775	61.69388	**57.33909**

Table 1: Model Performance of RMSE for ARIMA, Holt-Winters, and TBAT at Brazoria, Fort Bend, Chamber, Galveston County

	Harris County			Montgomery County		
Week	ARIMA	H-W	TBAT	ARIMA	H-W	TBATS
1	**104.25854**	105.18695	106.88164	**8.45859**	8.65024	8.55556
2	**100.08714**	100.78077	103.05876	**8.44314**	8.61499	8.48572
3	**94.62712**	95.18869	97.43360	**8.61759**	8.94277	8.77148
4	**90.19340**	90.77685	93.46970	**8.80332**	9.15385	8.93985
5	**88.81097**	89.18057	91.44696	**8.88573**	9.26740	9.02493
6	**86.50977**	88.50532	90.64113	**8.30352**	9.50366	9.28519
7	**87.76281**	87.99397	89.85659	**8.43860**	9.57026	9.29718
8	**90.48430**	90.79090	92.38886	8.92368	10.10495	**8.89307**
9	**91.21025**	91.23006	92.61355	**10.01163**	11.26325	10.67573
10	**116.90181**	118.01510	118.74942	12.21440	13.10568	**10.65887**
11	**171.85351**	179.61576	179.44131	**12.37381**	16.32121	13.65958
12	**212.33030**	222.58717	221.79088	**15.41106**	21.81653	18.60867
13	**235.57062**	237.44058	236.42294	**20.25520**	26.52571	20.59514
14	253.28187	253.86216	**252.72288**	**65.75111**	73.92191	70.63936
15	**252.69117**	253.94603	252.86238	**64.51900**	78.72152	64.53092
	Liberty County			Waller County		
Week	ARIMA	H-W	TBAT	ARIMA	H-W	TBAT
1	1.06873	1.20730	**1.03684**	**0.79344**	0.80946	0.81677
2	**1.15512**	1.26901	1.24785	**0.92191**	0.98440	0.98627
3	**1.12216**	1.22537	1.20553	**0.93162**	0.97616	0.97868
4	1.89689	1.92573	**1.89521**	**0.97061**	1.01974	1.02045
5	1.85496	1.88151	**1.85363**	**0.95380**	1.01913	1.02088
6	1.85876	1.88338	**1.85412**	**1.00957**	1.05511	1.05516
7	1.99769	2.02289	**1.98729**	**1.11343**	1.12654	1.12598
8	1.97269	1.99650	**1.96773**	**1.16955**	1.18227	1.18079
9	2.02620	2.04815	**2.02281**	1.59724	1.60138	**1.59254**
10	**2.54800**	2.60182	2.54834	**1.61967**	1.63517	1.62631
11	**3.21918**	3.30566	3.21930	**1.76662**	1.77757	1.76755
12	3.68966	3.72069	**3.68522**	2.07041	2.06819	**2.05554**
13	4.07067	4.10295	**4.06054**	**2.27768**	2.29569	2.28144
14	4.24349	4.27537	**4.22807**	**2.35409**	2.37173	2.35741
15	4.26041	4.29082	**4.25493**	**2.51450**	2.53239	2.51735

Table 2: Model Performance of RMSE for ARIMA, Holt-Winters, and TBAT at Harris, Montgomery, Liberty, Waller County

	ARIMA Model							
	County							
Week	Brazoria	Chamber	Fort Bend	Galveston	Harris	Liberty	Montgomery	Waller
1	4.8872	1.2978	20.3431	12.9417	104.2585	1.0687	8.4585	**0.7934**
2	5.9469	1.2929	20.3043	12.2933	100.0871	1.1551	8.4431	**0.9219**
3	6.7981	1.2411	19.5052	11.8390	94.6271	1.1221	8.6175	**0.9316**
4	7.2465	1.4361	19.7294	11.8390	90.1934	1.8968	8.8033	**0.9706**
5	7.3837	1.3941	21.3822	10.8116	88.8109	1.8549	8.8857	**0.9538**
6	7.3837	1.3959	20.1894	10.3372	86.5097	1.8587	8.3035	**1.0095**
7	7.7589	1.4183	21.0029	10.0086	87.7628	1.9976	8.4386	**1.1134**
8	7.5663	1.2953	21.9679	10.2182	90.4843	1.9726	8.9236	**1.1695**
9	27.0381	**1.4327**	22.9920	10.9030	91.2102	2.0262	10.0116	1.5972
10	67.3130	1.6723	28.3426	13.2315	116.9018	2.5480	12.2144	**1.6196**
11	64.2578	4.4039	32.1787	21.5748	171.8535	3.2191	12.3738	**1.7666**
12	62.8943	4.7640	30.8862	25.9082	212.3303	3.6896	15.4110	**2.0704**
13	61.6013	10.2971	33.7879	50.2452	235.5706	4.0706	20.2552	**2.2776**
14	62.9766	10.8552	40.6331	58.7115	253.2818	4.2434	65.7511	**2.3540**
15	62.4426	10.8698	43.8201	60.6277	252.6911	4.2604	64.5190	**2.5145**
	Holt-Winters Model							
	County							
Week	Brazoria	Chamber	Fort Bend	Galveston	Harris	Liberty	Montgomery	Waller
1	4.9975	1.3225	22.7854	13.1615	105.1869	1.2073	8.6502	**0.8094**
2	6.1031	1.3186	22.6202	12.4796	100.7807	1.2690	8.6149	**0.9844**
3	6.8213	1.2632	21.7708	12.0292	95.1886	1.2253	8.9427	**0.9761**
4	7.3386	1.2165	21.6092	12.0292	90.7768	1.9257	9.1538	**1.0197**
5	7.4727	1.1799	21.4395	10.9534	89.1805	1.8815	9.2674	**1.0191**
6	7.4727	1.1827	21.3068	10.4906	88.5053	1.8833	9.5036	**1.0551**
7	7.8522	1.1811	21.0189	10.1451	87.9939	2.0228	9.5702	**1.1265**
8	7.6492	1.1828	22.1898	10.3091	90.7909	1.9965	10.1049	**1.1822**
9	26.2299	**1.4473**	23.2398	11.3333	91.2300	2.0481	11.26325	1.6013
10	66.5501	1.7736	28.6115	13.9170	118.0151	2.6018	13.1056	**1.6351**
11	64.8192	4.4986	32.4656	21.9503	179.6157	3.3056	16.3212	**1.7775**
12	63.4157	5.2961	34.7824	26.1264	222.5871	3.7206	21.8165	**2.0681**
13	62.0852	10.8230	36.6927	55.3997	237.4405	4.1029	26.5257	**2.2956**
14	63.4394	11.4256	43.3083	60.4199	253.8621	4.2753	73.9219	**2.3717**
15	62.8854	11.4790	47.2859	61.6938	253.9460	4.2908	78.7215	**2.5323**
	TBAT Model							
	County							
Week	Brazoria	Chamber	Fort Bend	Galveston	Harris	Liberty	Montgomery	Waller
1	5.1826	1.3077	22.7743	13.5242	106.8816	1.0368	8.5555	**0.8167**
2	6.2148	1.3082	22.6318	12.8885	103.0587	1.2478	8.4857	**0.9862**
3	6.9350	1.2533	21.7704	12.4880	97.4336	1.2055	8.7714	**0.9786**
4	7.4159	1.2103	21.6505	12.4880	93.4697	1.8952	8.9398	**1.0204**
5	7.5577	1.1748	21.5178	11.5184	91.4469	1.8536	9.0249	**1.0208**
6	7.5577	1.1779	21.3964	11.0599	90.6411	1.8541	9.2851	**1.0551**
7	7.9142	1.1769	21.1160	10.7268	89.8565	1.9872	9.2971	**1.1259**
8	7.7118	**1.1771**	22.2480	10.8458	92.3888	1.9677	8.8930	1.1807
9	26.0672	**1.4388**	23.2664	11.7779	92.6135	2.0228	10.6757	1.5925
10	67.1238	**1.5892**	28.4706	14.2389	118.7494	2.5483	10.6588	1.6263
11	64.3016	4.4616	29.4548	22.0713	179.4413	3.2193	13.6595	**1.7675**
12	62.9230	5.2485	33.9364	22.1531	221.7908	3.6852	18.6086	**2.0555**
13	61.6209	9.7680	36.0391	55.0572	236.4229	4.0605	20.5951	**2.2814**
14	62.9805	10.2137	42.9260	56.2039	252.7228	4.2280	70.6393	**2.3574**
15	62.4535	10.7722	44.2011	57.3390	252.8623	4.2549	64.5309	**2.5173**

Table 3: Model Performance of RMSE for eight counties at ARIMA, Holt-Winters and TBAT model

Table 1 to 3 compare counties for each model with a weekly update by ARIMA, Holt-Winters, and TBAT methods, in terms of RMSE. Overall, three time series methods achieved superior performance in Waller county. Secondly, three models partially performed in Chamber county.

5. Conclusions

This study provided a model application to find an efficient method to forecast the trend of COVID-19 with local counties in Texas when there is a major impact location as Houston in Harris county. The proper forecasting method and knowledge of COVID-19 in each location at the county level can potentially reduce the pandemic's impact. State and local government official is enabled to modify their policies for social and health issues ahead of time.

As an assessment approach for the performance of three traditional statistical methods, the root mean square error (RMSE) has been compared. Results as Table 1 and 2 indicate that, although there is no one-size-fits-all method, traditional time series methods such as ARIMA and TBAT overall outperformed to forecast weekly short-term ahead of time. This study's result shows that the Holt-Winters method does not best fit for short-term forecast because of the low volume of data and not clear seasonality. However, if the impact of the pandemic keep grow and any trend and seasonal variation can exist, we could expect the performance of the Holt-Winters method.

Conflict of interest

The author declares that they have no conflict of interest.

REFERENCES

Adhikari, R. and Agrawal, R.K. (2013), "An introductory study on time series modeling and forecasting," arXiv:1302.6613.

Alexandrov, A., Benidis, K., Bohlke-Schneider, M., Flunkert, V., Gasthaus, J., Januschowski, T., Maddix, D.C., Rangapuram, S., Salinas, D., Schulz, J., et al. (2019), "Gluonts: Probabilistic time series models in python," arXiv:1906.05264.

Anastassopoulou, C., Russo, L., Tsakris, A., and Siettos, C. (2020), "Data-based analysis, modelling and forecasting of the COVID-19 outbreak," *PLoS ONE*, 15, e0230405.

Box, G.E. and Cox, D.R. (1964), "An analysis of transformations," *J. R. Stat. Soc. Ser*, 26, 211–243.

Box, G. and Jenkins, G. (2015), "Time Series Analysis Forecasting and Control/ Holden Day, San Francisco, California, 1970," John Wiley & Sons: Hoboken, NJ, USA.

Chatfield, C. (1978), "The Holt–Winters forecasting procedure," *J. R. Stat. Soc. Ser.*, 27, 264–279.

Chatfield, C. and Yar, M. (1988), "Holt-Winters forecasting: some practical issues," *J. R. Stat. Soc. Ser.*, 37, 129–140.

Chadsuthi, S., Modchang, C., Lenbury, Y., Iamsirithaworn, S., and Triampo, W. (2012), "Modeling seasonal leptospirosis transmission and its association with rainfall and temperature in Thailand using time–series and ARIMAX analyses," *Asian Pac. J. Trop. Med.*, 5, 539–546. [CrossRef]

De Livera, Alysha M. (2010), "Automatic forecasting with a modified exponential smoothing state space framework." *Monash Econometrics and Business Statistics Working Papers* 10, no. 10.

De Livera, Alysha M., Rob J. Hyndman, and Ralph D. Snyder. (2011), "Forecasting time series with complex seasonal patterns using exponential smoothing," *Journal of the American Statistical Association* 106, no. 496, 1513-1527.

De Livera, A.M., Hyndman, R.J. and Snyder, R.D. (2011), "Forecasting time series with complex seasonal patterns using exponential smoothing," *J. Am. Stat. Assoc.* **2011**, 106, 1513–1527.

Dominguez, A., Mu noz, P., Martínez, A. and Orcau, A. (1996), "Monitoring mortality as an indicator of influenza in Catalonia," *Spain. J. Epidemiol. Community Health*, 50, 293–298.

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

Fernandes, N. (2020), ”Economic Effects of Coronavirus Outbreak (COVID-19) on the World Economy,” https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3557504

Gelper, S., Fried, R. and Croux, C. (2010), ”Robust forecasting with exponential and Holt–Winters smoothing,” *J. Forecast*, 29, 285–300.

Hanf, M., Adenis, A., Nacher, M. and Carme, B. (2011), ”The role of El Ni no southern oscillation (ENSO) on variations of monthly Plasmodium falciparum malaria cases at the cayenne general hospital, 1996–2009, French Guiana,” *Malar. J.*, 10, 100.

Hu, Z., Ge, Q., Jin, L. and Xiong, M. (2020), ”Artificial intelligence forecasting of covid-19 in china,” arXiv:2002.07112.

Johns Hopkins University CSSE. ”Wuhan coronavirus (2019-nCoV) global cases” (https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html/bda7594740fd40299 423467b48e9ecf6. opens in new tab).

Kalekar, P.S. (2004), ”Time series forecasting using holt-winters exponential smoothing,” *Kanwal Rekhi Sch. Inf. Technol*, 4329008, 1–13.

Kane, M.J., Price, N., Scotch, M. and Rabinowitz, P. (2014) ”Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks,” *BMC Bioinform*, 15, 276.

Lu, H., Stratton, C.W. and Tang, Y.W. (2020), ”Outbreak of Pneumonia of Unknown Etilogy in Wuhan China: The Mystery and the Miracle,” *J. Med Virol.*, 92, 401-402.

McCloskey, B., Zumla, A., Ippolito, G., Blumberg, L., Arbon, P., Cicero, A., Endericks, T., Lim, P.L., Borodina, M. and M.G.E. Group. (2020), ”Mass gathering events and reducing further global spread of COVID-19: A political and public health dilemma,” *Lancet*, 395, 1096.

McKenzie, E. (1984), ”General exponential smoothing and the equivalent ARMA process,” *J. Forecast*, 3, 333–344.

Oreshkin, B.N., Carpov, D., Chapados, N. and Bengio, Y. (2019), ”N-BEATS: Neural basis expansion analysis for interpretable time series forecasting,” arXiv:1905.10437.

Petropoulos, F. and Makridakis, S. (2020), ”Forecasting the novel coronavirus COVID-19,” *PLoS ONE*, 15, e0231236.

Roosa, K., Lee, Y., Luo, R., Kirpich, A., Rothenberg, R., Hyman, J., Yan, P. and Chowell, G. (2020), ”Real-time forecasts of the COVID-19 epidemic in China from 5 February to 24 February 2020,” *Infect. Dis. Model.*, 5, 256–263.

Salinas, D., Flunkert, V., Gasthaus, J. and Januschowski, T. (2019), ”DeepAR: Probabilistic forecasting with autoregressive recurrent networks,” *Int. J. Forecast*, doi:10.1016/j.ijforecast.2019.07.001.

Soebiyanto, R.P., Adimi, F. and Kiang, R.K. (2010), ”Modeling and predicting seasonal influenza transmission in warm regions using climatological parameters,” *PLoS ONE*, 5, e9450.

Song, X., Xiao, J., Deng, J., Kang, Q., Zhang, Y. and Xu, J. (2016), ”Time series analysis of influenza incidence in Chinese provinces from 2004 to 2011,” *Medicine*, 95, e3929.

Taylor, S.J. and Letham, B. (2018), ”Forecasting at scale,” *Am. Stat*, 72, 37–45.

Wold, H. (1939), ”A Study in Analysis of Stationary Time Series. *J. R. Stat. Soc.*, 102, 295.

Yang, Z., Zeng, Z., Wang, K., Wong, S.S., Liang,W., Zanin, M., Liu, P., Cao, X., Gao, Z., Mai, Z., et al. (2020), ”Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions,” *J. Thorac. Dis.*, 12, 165.

Yin, R., Luusua, E., Dabrowski, J., Zhang, Y. and Kwoh, C.K. (2020), ”Tempel: time-series mutation prediction of influenza A viruses via attention-based recurrent neural networks,” *Bioinformatics*, 36, 2697–2704.

Yule, G.U. (1926), ”Why do we sometimes get nonsense-correlations between Time-Series?-a study in sampling and the nature of time-series,” *J.R. Stat. Soc.*, 89, 1-63.

Zhang, Y., Yakob, L., Bonsall, M.B. and Hu, W. (2019), ”Predicting seasonal influenza epidemics using cross-hemisphere influenza surveillance data and local Internet query data,” *Sci. Rep.*, 9, 1–7.

Zhang, S., Diao, M., Yu, W., Pei, L., Lin, Z. and Chen, D. (2020),”Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis,” *Int. J. Infect. Dis.*, 93, 201–204.

Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. (2020), China Novel Coronavirus Investigating and Research Team. A novel coronavirus from patients with pneumonia in China, 2019,” *N Engl J Med.*, 382:727–33.