Using time series to predict the change in the U.S. unemployment rate at covid-19

Jie Gao [1,†], Zhichao Li [2,*,†], Weifu Li [3,†], Boqing Shi [4,†]

[1] Erlianhot No.1 Middle School, Inner Mongolia, China

[2] University of Miami, Miami, US

[3] Tongji University, Shanghai, China

[4] Miami University of Oxford, US

* Corresponding Author Email: zxl823@miami.edu

[†] These authors contributed equally.

Abstract. To some extent, the unemployment rate can measure the economic situation of a country, generally speaking, the economic development is good during the period of low unemployment, and there are relatively few social contradictions and problems; Contrarily, the economy is weak, social conflicts are obvious, and the unemployment rate is high. Therefore, studying the unemployment rate can be used as one of the references for studying the good or bad economy of a country, and at the same time, predicting the future unemployment rate can also allow the government department to understand the future situation to prevent possible problems. The Covid-19 epidemic in 2020 had a great impact on the job market, this paper uses time series data of the US business rate, and uses the ETS model and the ARIMA model to predict and analyze the unemployment rate. in addition, The same methodology has been applied to states with the highest and lowest unemployment rates in the United States to predict unemployment rates and analyze the specific impact of the pandemic on employment.

Keywords: unemployment rate; ETS model; ARIMA model; forecast; Covid impact.

1. Introduction

The unemployment rate, a crucial lagging indicator of the capital market, calculates the percentage of the labor force over a certain length of time that is still jobless among those who are employed and fulfill all eligibility conditions for employment. An increase in the unemployment rate, on the other hand, will cause inflation, require the central bank to restrict the monetary base, and decrease monetary investment. The government may decide to relax the monetary base and encourage economic development if the unemployment rate rises as a symptom of economic fragility. It is published by the U.S. Department of Labor (USDL) on the first Friday of each month and is extensively read by market players. However, the Covid-19 epidemic raised unemployment rates over the world. The U.S. unemployment rate rose dramatically during the pandemic, and starting in April 2020, initial jobless claims immediately reached an all-time high of 14.7 percent (USDL,2022).

According to the Bureau of Labor Statistics, the U.S. unemployment rate saw a very rapid increase in the year 2020 after the COVID-19 outbreak. Why did the COVID-19 outbreak have such a large impact on the U.S. unemployment rate. Which occupations saw the fastest rise in unemployment in this scenario? Did unemployment rise sharply in all occupations, or did it vary depending on the different occupations that would be affected. For example, it is agriculture, industry and some tertiary occupations, such as services. In addition, the impact of government policies on the unemployment rate, where most occupations were in shutdown after the outbreak, also had an impact on the unemployment rate. Most companies had no income during this period, but fixed costs were not reduced, which led to severe losses in some industries. Did this have an impact on the unemployment rate?

1.1. Literature review

The CoreLogic home price index's county-level data on house prices was utilized by the local labor market in the United States in 2007 to examine how the unemployment rate and housing prices had changed in relation to one another in the two counties. The findings showed that the two variables had changed in different directions. Relationship [1]. Unemployment resources refer to the labor force. The blue line in the figure indicates that the ratio of unproduced goods to labor force can be called the unemployment rate. There are two types of unemployment, namely frictional unemployment and structural unemployment. They will be called the natural unemployment rate , are complementary to each other, leading to idle resources, other resources may also be idle [2]. The FRED figure shows that the rate of bank card delinquencies is rising in tandem with the recession's high unemployment rate. There is a negative association between both since the jobless rate increased significantly during the economic slowdown from February to April 2020, unlike in 2008 when the credit card delinquency rate declined [3]. Since COVID-19's emergence, it has had a significant negative influence on events all around the world. The COVID-19 outbreak and the government embargo have both had a significant impact on the tourist sector. There has also been a significant reduction in the number of jobs associated with it, leading to an increase in unemployment. In addition, agriculture and academic institutions has been affected by COVID-19. Because in education and agriculture, the mobility and socialization of students is inevitable. The impact of public health is unavoidable in a reopened face-to-face education system [4]. The COVID-19 outbreak has brought about public health and safety issues that have largely affected the agricultural and food industries. The spread of COVID-19 has caused many employees in the agricultural and food industries to lose their jobs, resulting in a significant increase in unemployment. In addition, the demand for food will change in this situation. The cessation of face-to-face classes in some public places, such as schools, will lead to a decrease in the demand for food needed in cafeterias. But the demand for some other special food, such as storable food, will increase greatly again [5]. In 2022, Fangyuan Hu et, al has done a research about the relationship between unemployment rate and the vaccination rate, they found out that getting vaccination would help to decrease the unemployment rate, when the vaccination rate increase by 1%, the unemployment rate went down by 0.09% [6]. Hanlai et, al have done a research of prediction of unemployment rate in Asian countries in 2021 by using hybrid ARIMA-ARNN, ARIMA-ANN and hybrid ARIMA-SVM model, and they find out hybrid ARIMA-ANN have a better accuracy when predicting developing countries. They also claims that in the future, the unemployment rate for developing countries in Asia is three times higher than the developed countries in Asia [7]. This book investigates the causes of the high unemployment rate and its extreme fluctuations. It illustrates the relationship between unemployment and inflation and explores the possibility of achieving both price stability and full employment. It queries why some demographic groupings experience higher unemployment rates than others [8].

2. Organization of the Text

2.1. Background

During the Covid-19 pandemic, unemployment rates increased in most countries around the world, with a sharp climb in the United States in particular. Starting in April 2020, the number of initial claims for unemployment benefits reached an all-time high of 14.7 percent directly after the Covid-19 outbreak, and the number of confirmed cases in the United States peaked in 2020, according to the Bureau of Labor Statistics. Therefore, we focus on analyzing the trends between confirmed cases of covid-19 and unemployment rates in the face of different patterns of covid-19.

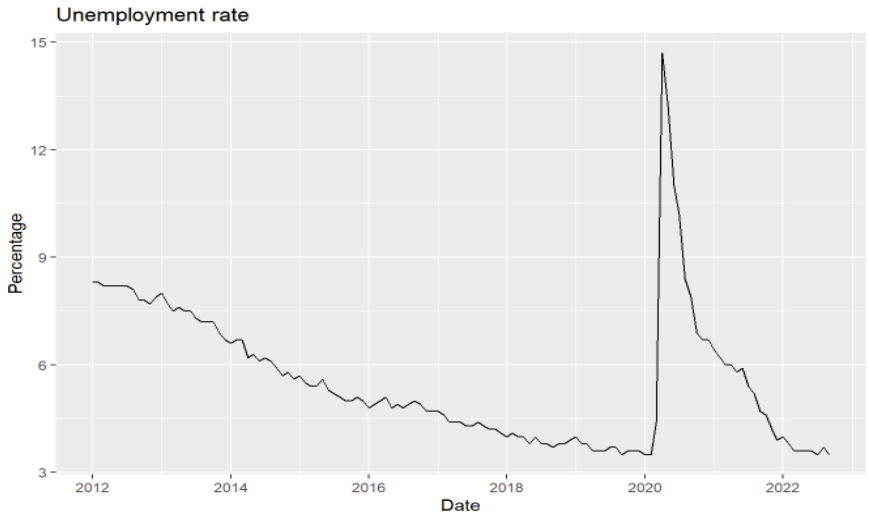


Fig 1. U.S. Unemployment Statistics

In Figure 1&2, by applying different models to predict how covid-19 affects the unemployment rate and why there is such a large variation in the unemployment rate across states. During covid-19, the unemployment claims rate in each state rose substantially. Likewise, the unemployment rate changed significantly.

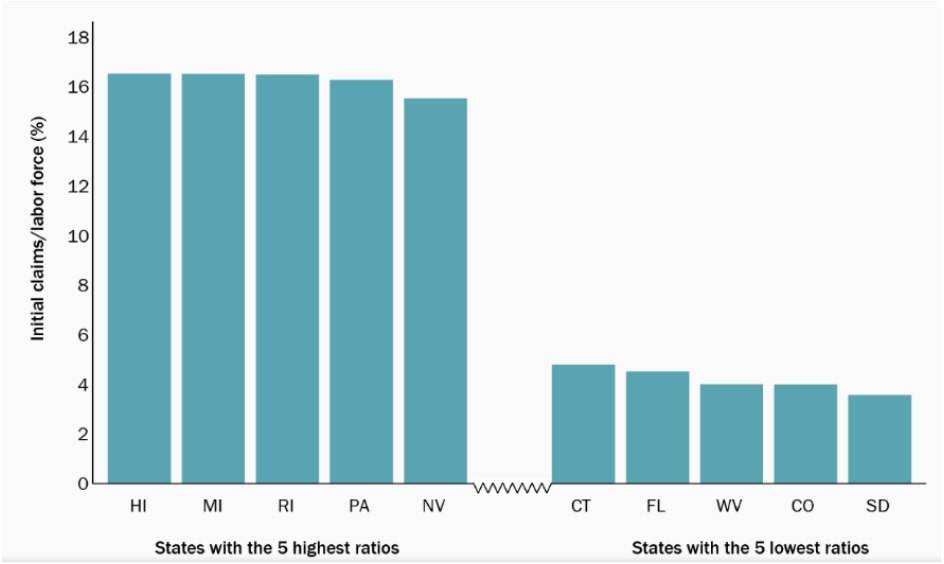


Fig 2. Initial Unemployment Insurance Claims to Labor Force Ratio, by State

Source: Economic Studies at BROOKINGS

2.2. Methods

The number of confirmed cases during the 2020 epidemic in the United States and the fall in unemployment rates at that time are the main topics of this study. Additionally, a state with the greatest unemployment rate and a state with the lowest unemployment rate are chosen in order to do the study and forecasts. The selection of these two representative states allows for a valid analysis of whether unemployment rates are affected by the government's management of the different states.

The main forecasting models used in the article are ETS and ARIMA. finally, these two forecasting models are compared with the actual situation to find out which model is closer to the actual situation.

2.2.1 ETS Model

Time series univariate forecasting is a specialty of the ETS (Error, Trend, Seasonal) method. This ETS model emphasizes trend and seasonal factors. The ETS model's adaptability lies. Regarding the trend and seasonal elements for many parameters.[9]

2.2.2 ARIMA Model

There is no movement in any ARIMA models. ARIMA stands for Auto Regressive Integrated Moving Average Model. ARIMA's goal is to stabilize the data.

There are two categories of ARIMA models: seasonal and non-seasonal. The non-seasonal model is a hybrid of differencing and autoregression, as well as a moving average model.[10]

2.3. Data collection and variables

The Federal Reserve Economic Data (FRED), which includes routinely updated US macro and regional economic time series with annual, quarterly, monthly, weekly, and daily frequencies, was used to compile the information for the US rate of unemployment (UNRATE), Hawaii's unemployment rate (HIUR), and South Dakota's unemployment rate (SDUR). We chose Hawaii and South Dakota as our data sources because they have the greatest and lowest unemployment rates, respectively. The time series have been set to two period in case to avoid the outliers, the first period is from January 2000 to January 2020, the second period is from April 2020 to September 2022.

2.4. Result

2.4.1 ETS Model

The ETS model was used to forecast the unemployment rate in the future, at first the time was set from January, 2012 to September 2022. However the prediction is obviously influenced by the outliers as the plot shown below (Figure3). In this plot the RMSE is greater than 1 which indicate that this model is not quite fit to this data set, and the MAPE value is really high which shows ETS model wasn't fit well with the data set with outliers.

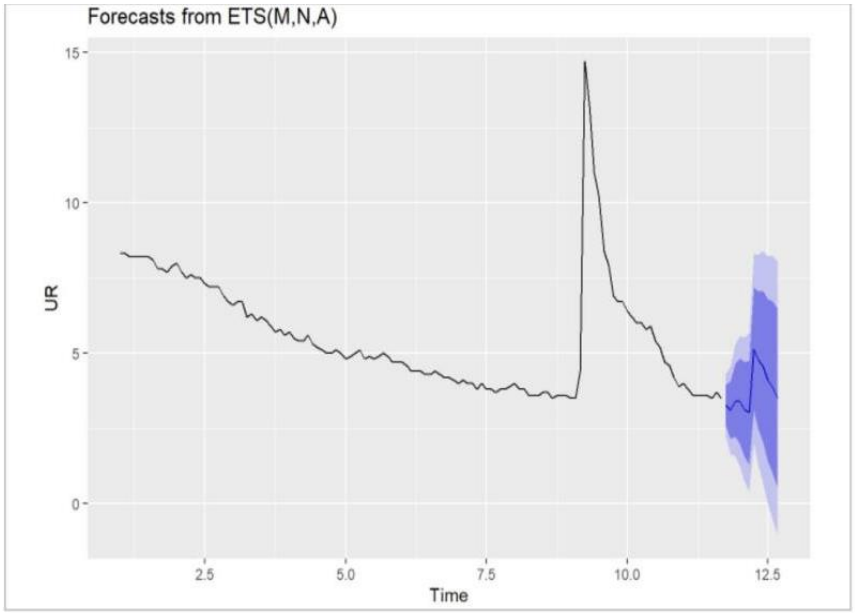


Fig 3. Forecasting UR using ETS model from 2012 to 2022

Table 1. Training set error measures from ETS (M, N, A)

Values	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	-0.03016022	1.012522	-0.9297542	0.4838652	8.498773	0.3266718

In order to fix this model, the time slot has been set to two different period as mentioned in the method section, after this modification, the prediction became more accurate, as the graph shown the RMSE value for the second time slot has lowered to 0.27 which shows that the ETS model fit this model pretty good (Figure4). The prediction of the pre-pandemic time period, which is the first period also tend to have a RMSE value of 0.145(table3).

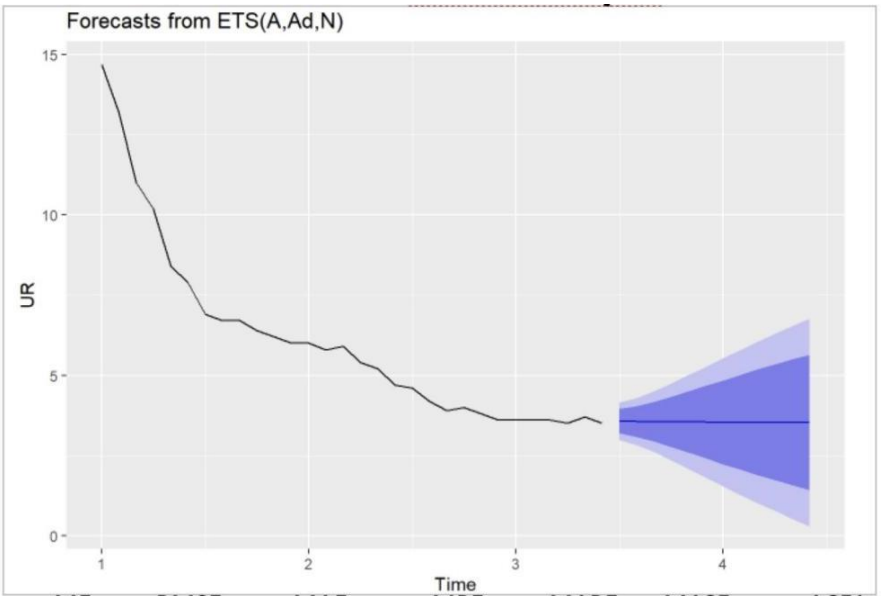


Fig 4. Forecasting UNRATE using ETS model from 2020 to 2022

Table 2. Training set error measures from ETS (A, Ad, N)

Values	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	-0.01345135	0.2703473	0.2144979	0.2886417	3.517489	0.0657745

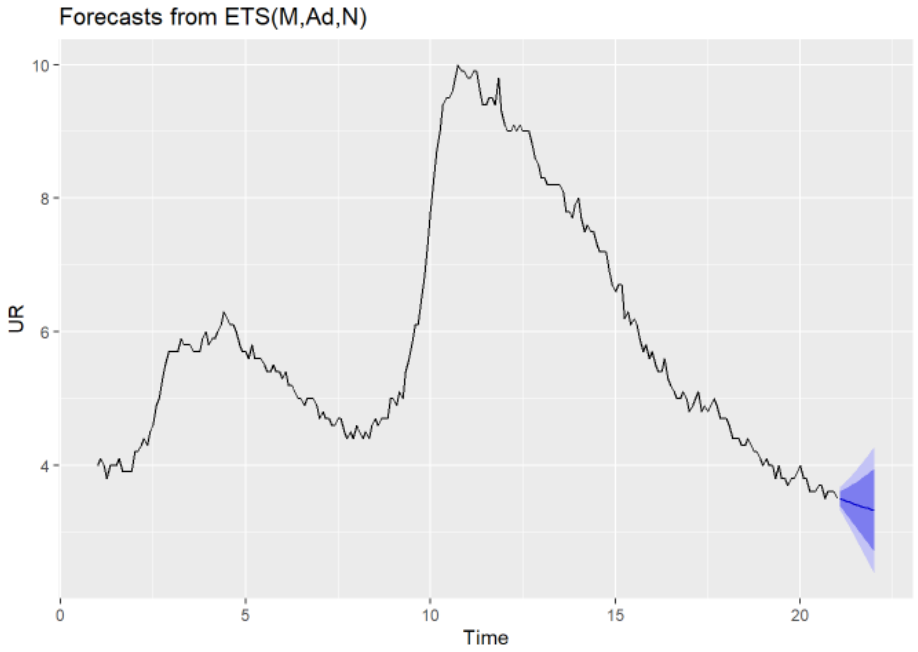


Fig 5. Forecasting UNRATE using ETS model from 2000 to 2020

Table 3. Training set error measures from ETS (M, Ad, N)

Values	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	-0.000933051	0.1446368	0.1154503	0.01978948	2.040511	0.1472056

Furthermore, the same ETS model were used on the state with the highest unemployment-Hawaii and the state with the lowest unemployment rate-South Dakota. The accuracy test for the first period of South Dakota prediction has a RMSE of 0.059, and the second time period have a RMSE of 0.169.(Table4)

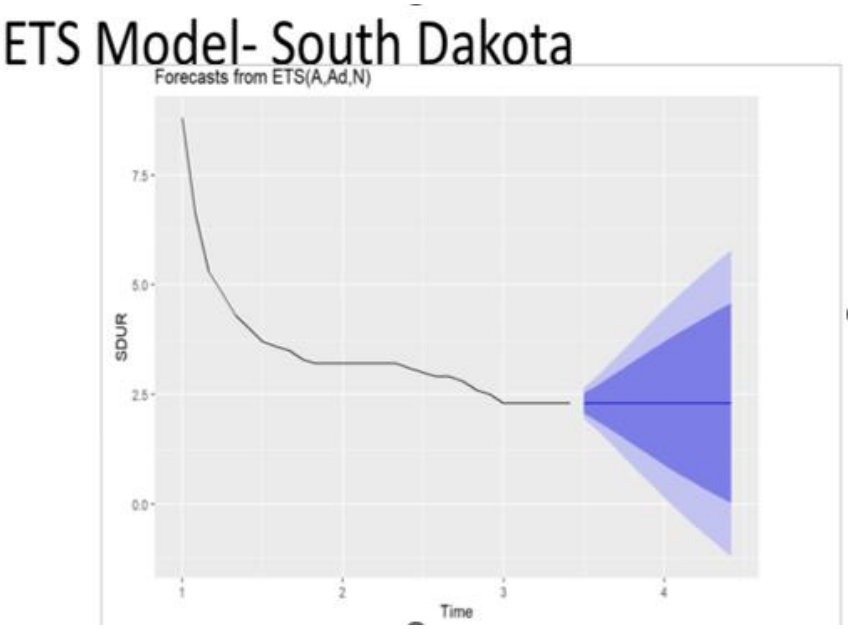


Fig 6. Forecasting SDUR using ETS model from 2020 to2022

Table 4. Training set error measures from ETS (A, Ad, N)

Values	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	0.04075151	0.168565	0.09404701	0.7246966	2.481355	0.07112799

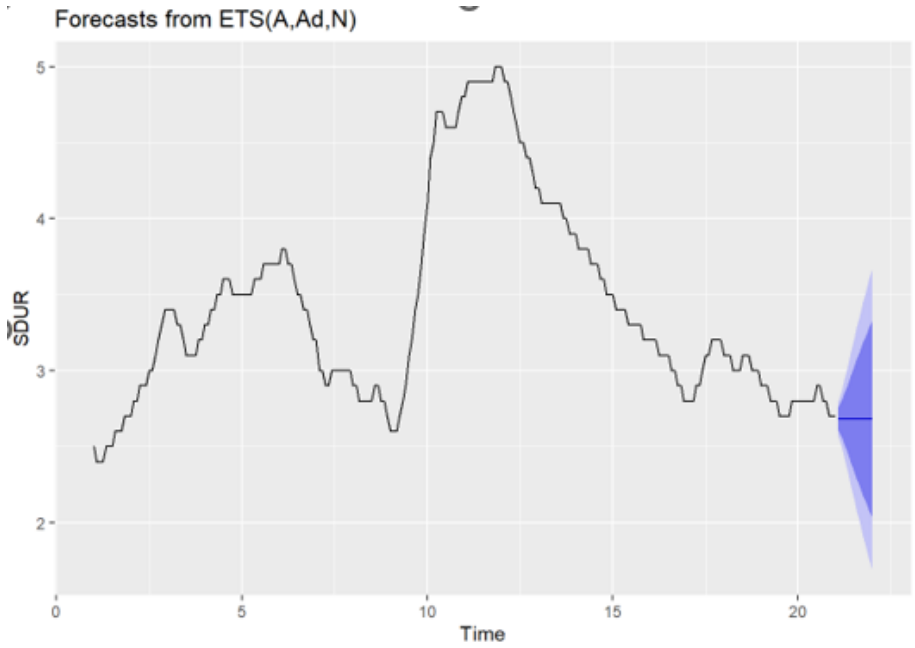


Fig 7. Forecasting SDUR using ETS model from 2000 to 2020

Table 5. Training set error measures from ETS (A, Ad, N)

Values	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	0.0004679585	0.05878456	0.04727754	0.034316	1.436196	0.1284289

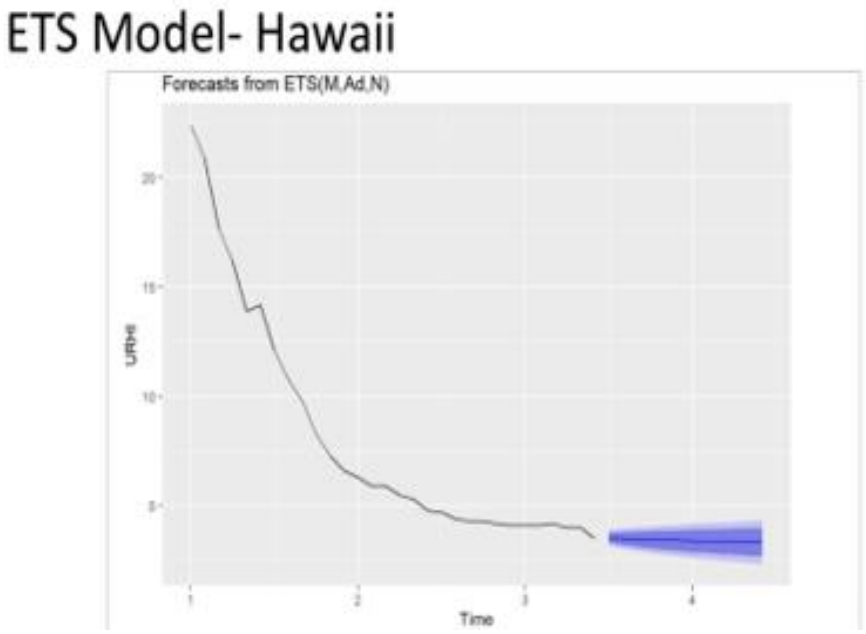


Fig 8. Forecasting HIUR using ETS model from 2020 to 2022

Table 6. Training set error measures from ETS (M, Ad, N)

Values	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	-0.03291082	0.4762697	0.3222496	-0.4272754	3.636696	0.05258833

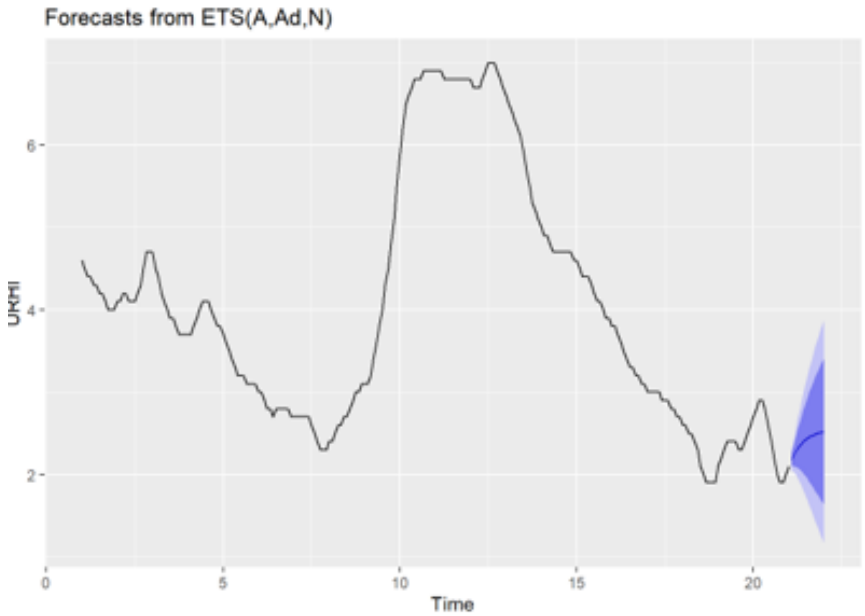


Fig 9. Forecasting HIUR using ETS model from 2000 to 2020

Table 7. Training set error measures from ETS (A, Ad, N)

Values	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	-0.001431099	0.06600303	0.05326915	0.004844521	1.543066	0.08137849

The ETS model also shows a good performance on the prediction of Hawaii (Table 6), the test shows the RMSE for the second period is 0.476, and the first period have a RMSE of 0.066

2.4.2 ARIMA Models

In addition, the ARIMA model can also be used to predict future unemployment rate, and similar to the previous ETS model, when forecasting using the ARIMA model, the forecast results are clearly affected by outliers in the time series (2020 year's unemployment rate).

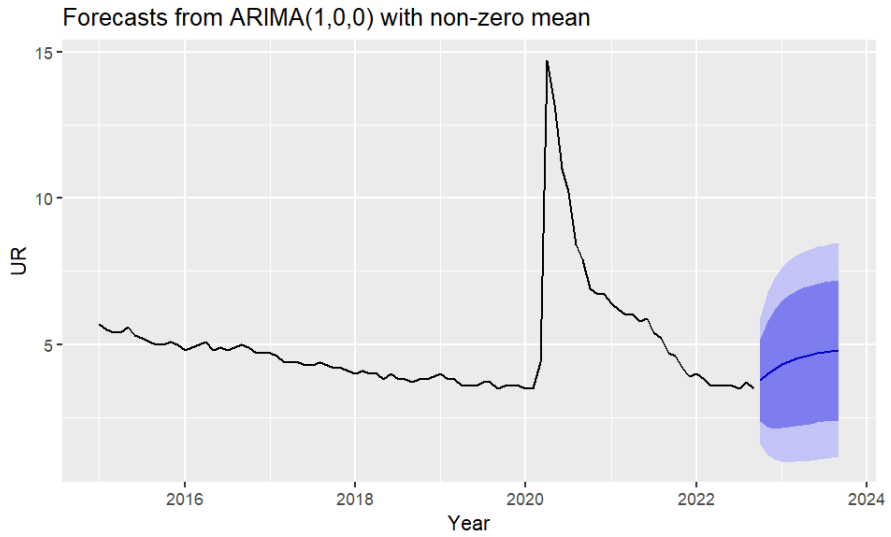


Fig 10. Forecast from ARIMA (1,0,0)

In terms of overall results, the rapid increase in unemployment after 2020 is empirically unreasonable and is most likely the result of extreme values in 2020.

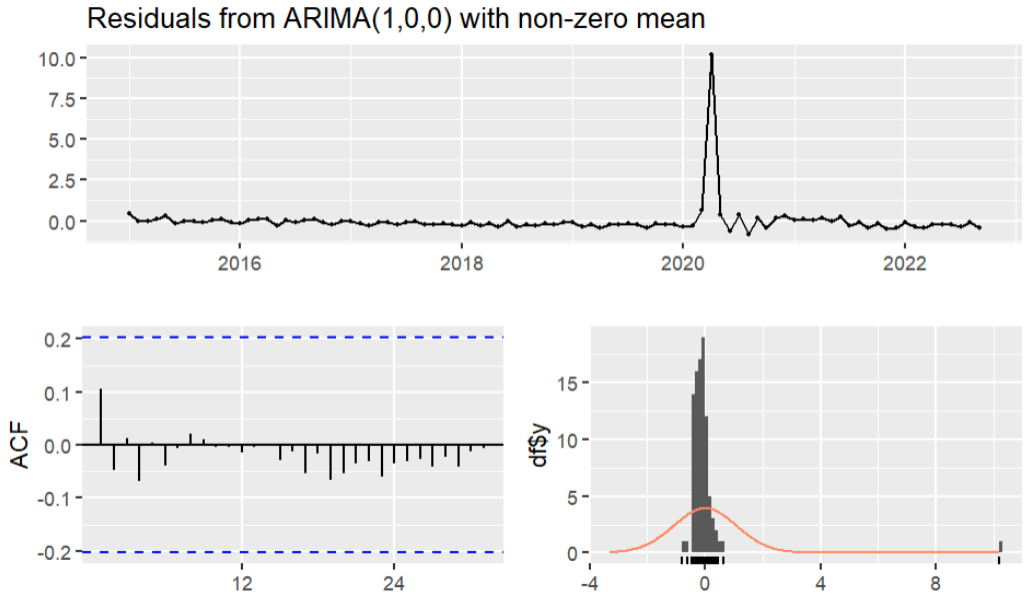


Fig 11. Residuals from ARIMA (1,0,0)

Of course, the overall image is deceptive, so it needs to be analyzed from the perspective of residuals. As can be seen from the residual image (Figure11), the residual values show a stable trend overall, but the corresponding time series of unemployment rate (Figure10) is not stable. In combination with the three parameters of ARIMA (p, d, q), the value of d reflects the number of times the difference is required, but the corresponding d value in the model is 0, indicating that the original data is not differentiated. The choice of the next model begins without a difference in nonstationary data, which is obviously unreasonable. At the same time, there are no lines in the ACF image that cross the critical value, but the residuals in the distributed image do not resemble the normal distribution, so the residuals cannot be completely considered white noise, so this model is unreasonable from the perspective of residuals.

Table 8. Training set error measures from ARIMA (1,0,0)

Values	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	-0.0103682	1.089835	0.32167	-2.428691	5.4437	0.1865087

Moreover, the RMSE value in the training set error measures is already greater than 1, which can indicate that the model cannot predict the unemployment rate time series well.

However, unlike the previous ETS model, the ARIMA model is a more complex predictive model, which means that it is not necessary to use the ARIMA model when there are few observations. The unemployment rate time series after 2020 has fewer observations, and dividing the time series into two segments may raise new problems. However, the ARIMA model also has the advantage that ETS does not have, that is, it will diverge data to eliminate some parts which are not stationary and seasonal factors, and at the same time, the impact of outliers on a period will be much reduced after more observational data. Therefore, the solution is to expand the time series and then make forecasts, and this article selects data from the year 2000.

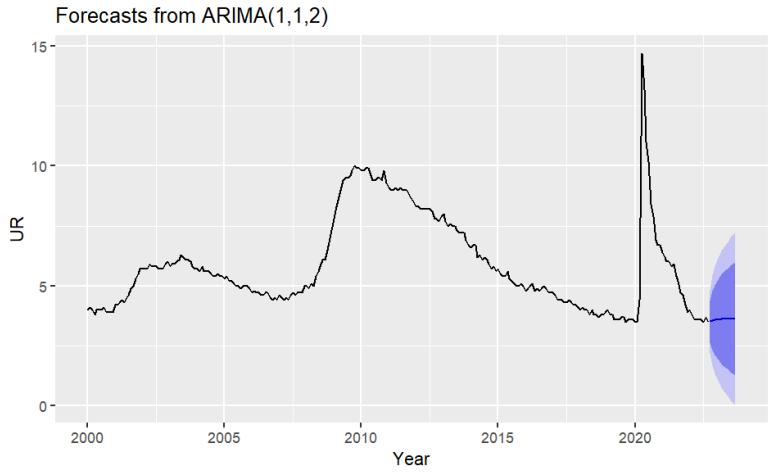


Fig 12. Forecast from ARIMA (1,1,2)

Table 9. Training set error measures from ARIMA (1,1,2)

Values	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	-0.00203875	0.6718146	0.2027113	--0.4647546	3.093097	0.1765943

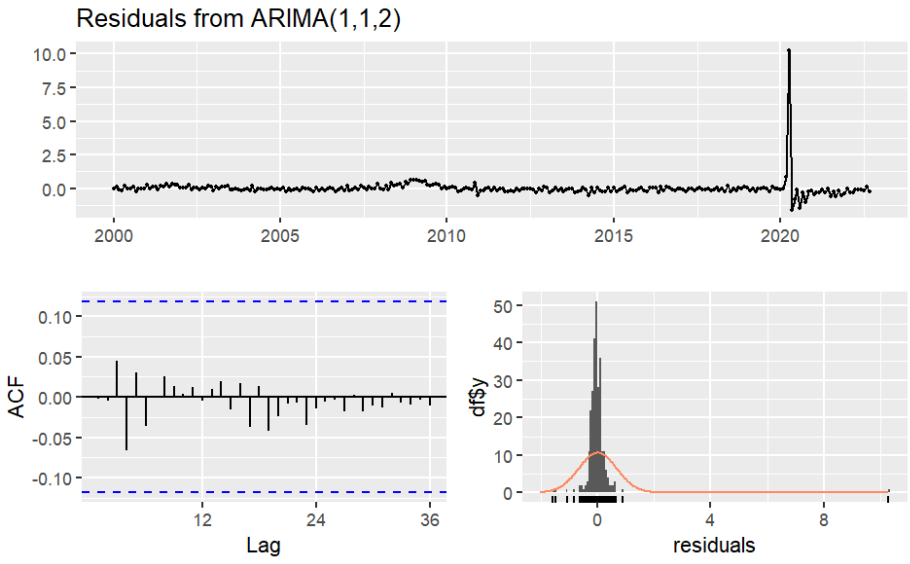


Fig 13. Residuals from ARIMA (1,1,2)

As can be seen from the overall (Figure 14), the forecast trend receives images of extreme data for 2020. In the residual plot (Figure 15), the time series after differentiation can also be regarded as stationary, while the residual ACF image has no lines crossing the critical value, the distribution is approximately normal, and the residual can be regarded as white noise. At the same time, the values in the training set error measures are small, where the RMSE value is less than 1. In summary, the model used in this prediction is appropriate.

Same as the previous ETS model, the ARIMA model were used on the state with the highest unemployment-Hawaii and the state with the lowest unemployment rate-South Dakota.

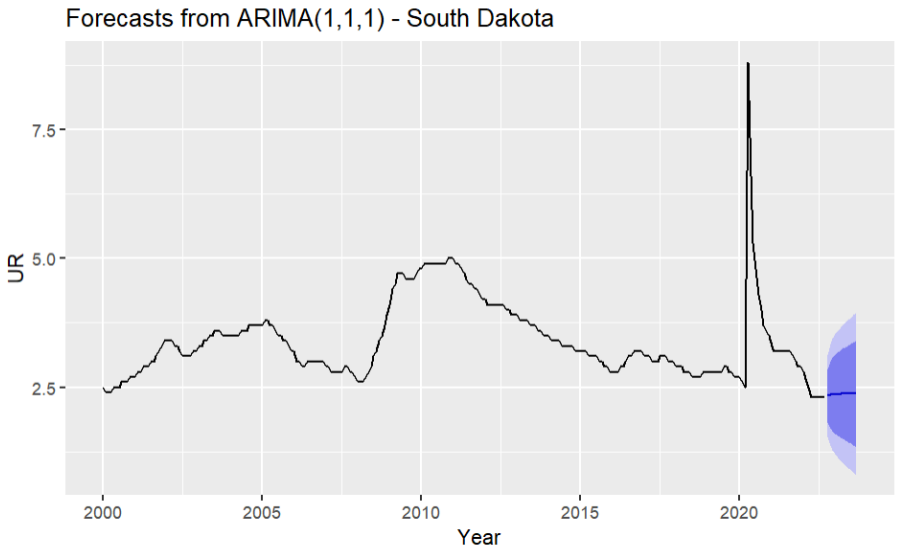


Fig 14. Forecast from ARIMA (1,1,1)- South Dakota

Table 10. Training set error measures from ARIMA (1,1,1)- South Dakota

Values	ME	RMSE	MAE	MPE	MAPE
Training set	-0.000879324	0.4014149	0.1156695	--0.5075163	2.973074

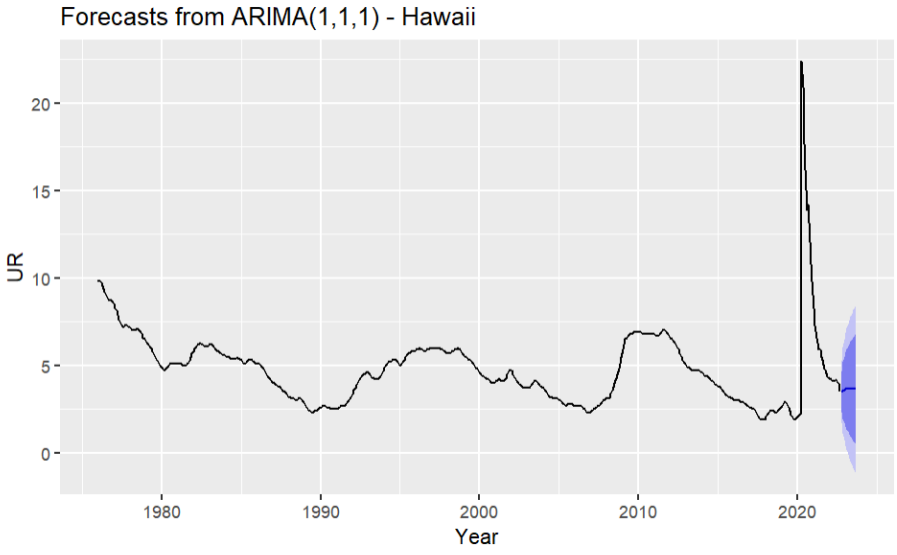


Fig 15. Forecast from ARIMA (1,1,1)- Hawaii

Table 11. Training set error measures from ARIMA (1,1,1)- Hawaii

Values	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	-0.01476605	0.8833378	0.1537388	-0.6930585	2.612023	0.1479969

3. Conclusion

At the level of data analysis, first of all, the unemployment rate in the United States has gradually declined since the peak of the unemployment rate in 2020, and it is also expected to gradually stabilize in the future. The main reason for the unemployment outlier is the pandemic in 2020. From the perspective of the two states under the subdivision, Hawaii has the highest unemployment rate in the United States, mainly because many local industries in Hawaii are related to tourism, and at the same time, when people want to go to Hawaii, they need to take planes or ships, and the epidemic has hit tourism and public transportation such as aviation, so Hawaii's local income will decline and unemployment will be high. As for South Dakota, the reason for the low unemployment rate may be that the population is small and the tourism industry is relatively underdeveloped, so people will have relatively stable jobs, so the unemployment rate will be low.

Furthermore, in terms of model selection, when using the ETS model, 2020 Observations are limited after the year, so it is not necessary to use a complex model, and you can choose to use a relatively simple model to make predictions instead. At the same time, because the model prediction is too affected by outliers, which will cause the forecast to be inaccurate, it is necessary to divide the time series into two segments for analysis. Therefore, "dynamic regression models" that can be added to other time series may be a better choice. Therefore, the next step in the study can use dynamic regression models to try to predict the unemployment rate.

References

[1] Dvorkin, M. A., & Shell, H. G. (n.d.). The recent evolution of U.S. local labor markets. Economic Research - Federal Reserve Bank of St. Louis. Retrieved November 16, 2022.

[2] Yetter, E. A. (n.d.). Why scarce resources are sometimes unemployed. Economic Research - Federal Reserve Bank of St. Louis. Retrieved November 16, 2022.

[3] FRED. (2021, December 27). On the relationship between unemployment and late credit card payments. FRED Blog. Retrieved November 16, 2022.

[4] Shrestha, N., Shad, M. Y., Ulvi, O., Khan, M. H., Karamehic-Muratovic, A., Nguyen, U.-S. D. T., Baghbanzadeh, M., Wardrup, R., Aghamohammadi, N., Cervantes, D., Nahiduzzaman, K. M., Zaki, R. A., & Haque, U. (2020, October 13). The impact of covid-19 on Globalization. One Health. Retrieved November 21, 2022.

[5] Larue, B. (2020). Labor issues and Covid-19. Canadian Journal of Agricultural Economics/Revue Canadienne D'agroeconomie, 68(2), 231–237.

[6] Hu, F., Pang, J., & Sun, H. (n.d.). Research on the relationship between unemployment rate and ... - JOEBM. Retrieved November 21, 2022.

[7] Lai, H., Khan, Y. A., Thaljaoui, A., Chammam, W., & Abbas, S. Z. (2021). COVID-19 pandemic and unemployment rate: A hybrid unemployment rate prediction approach for developed and developing countries of Asia. Soft computing, 1–16. Advance online publication.

[8] Layard R, Nickell S, Jackman R. Unemployment: Macroeconomic Performance and the Labour Market. Oxford University Press, 1991.

[9] Hyndman, R. J., & Athanasopoulos, G. (2021). Forecasting: Principles and practice. Otexts.

[10] Jofipasi, C. A., Miftahuddin, & Hizir. (2018). Selection for the best ETS (error, trend, seasonal) model to forecast weather in the Aceh Besar District. IOP Conference Series: Materials Science and Engineering, 352, 012055.