**Grayson County Sales Tax Forecasting**

ECMT 475 - Dr. Schulman

4 May 2022

**Authors:**

| William Hays  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | Joshlyn Javier  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | Kevin Pham  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| --- | --- | --- |

Nick Zekulin

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

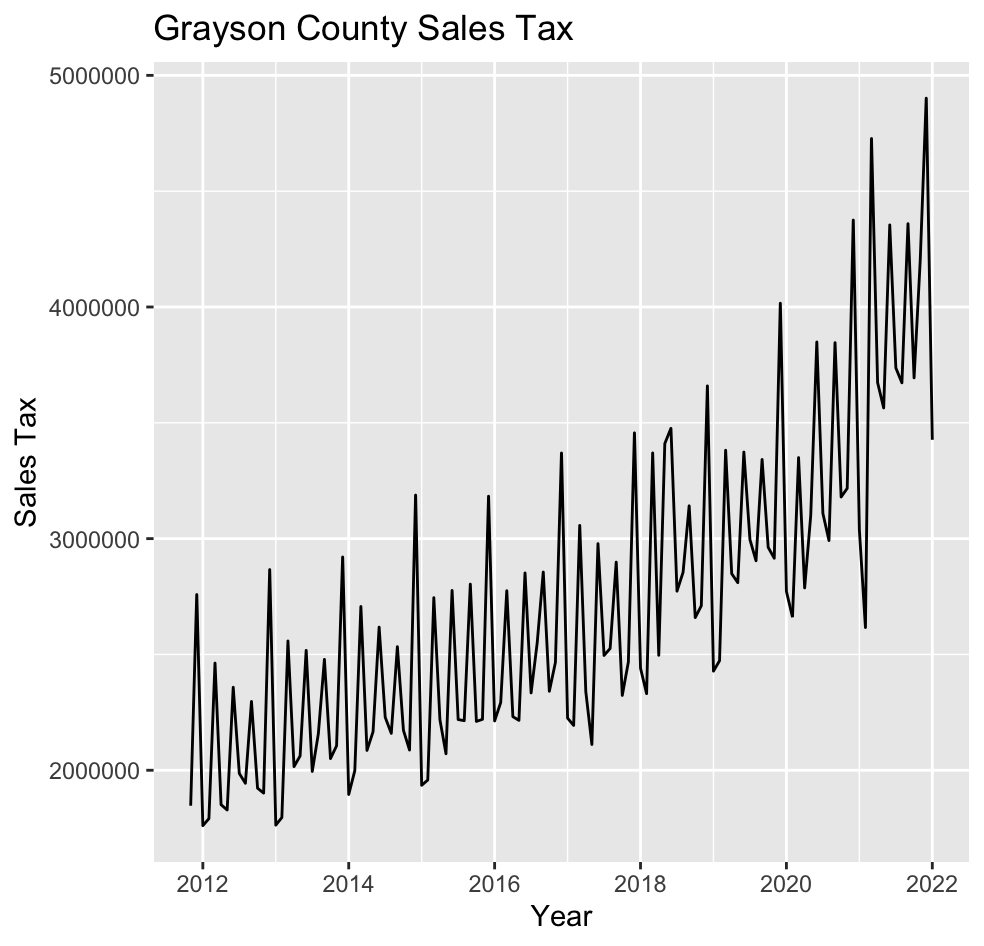
COUNTY PROFILE

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Grayson County is a county in Texas that is a northern suburb of Dallas. As of 2022 census data, the county has 135,543 people in total. The population has evolved over the last 12 years, growing by 18.55% since 2010. In 2015, there was a large boom in population, and the population grew by 27.38%. There has been diminishing upward growth over the last 12 years, but it has been consistently positive. The ethnic breakdown of the county is made up of roughly 74% white, 14% hispanic, 6% black, and 6% of other ethnicities. A majority of the population (52%) is between the ages of 18 and 65, and 24% of the population is under 18. The county is slightly under-educated, having 20.6% of their population earn a bachelor’s degree or higher, compared to the national average of 32.1%. Their largest industries by employment are healthcare/education, retail trade, manufacturing, and construction. As a result of these dominating industries, the major employers in the area include Texoma Medical Center and Tyson Foods, employing almost 4,000 and 1,750 respectively. The population is concentrated in the Sherman and Denison urban areas, and the county seat is located in Sherman. According to 2020 census data, Sherman has a population of 43,645, and Dennison follows with a population of 24,479. There are 2.66 persons per household on average and 129.6 persons per square mile.

SALES TAX VARIABLE OVERVIEW

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_



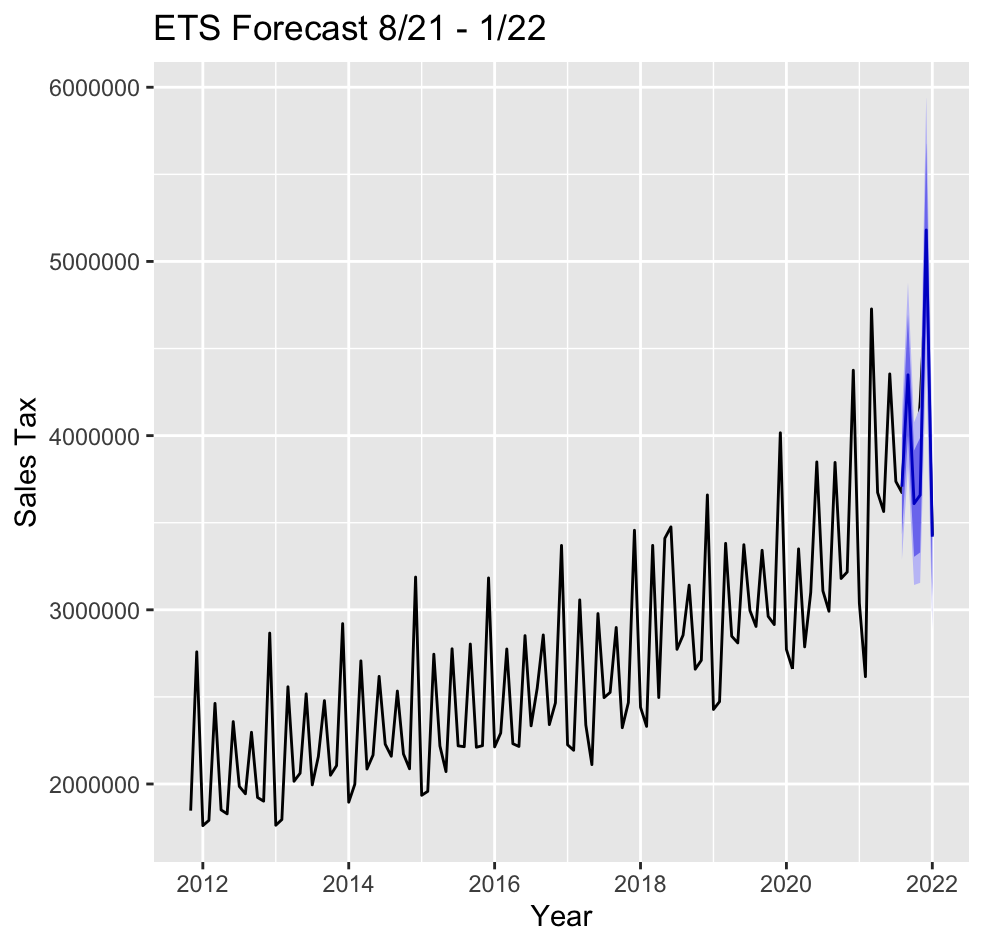
*Figure 1.1*

As displayed in Figure 1.1 above, there is a steady upward trend of the Grayson County Sales Tax, beginning from 2012 and lasting until 2022. It appears that the trend increases around 2020. There is a relatively large trough in January of 2021, but the upward trend continues. Seasonally, there are consistent peaks in the following months: March, June, September, and December, with December being the highest peak. Similarly, there are consistent troughs in the following months: January, February, April, August, and November, with April being the lowest trough. There is a cyclical pattern that can be observed in the figure, as the seasonal pattern repeats each year.

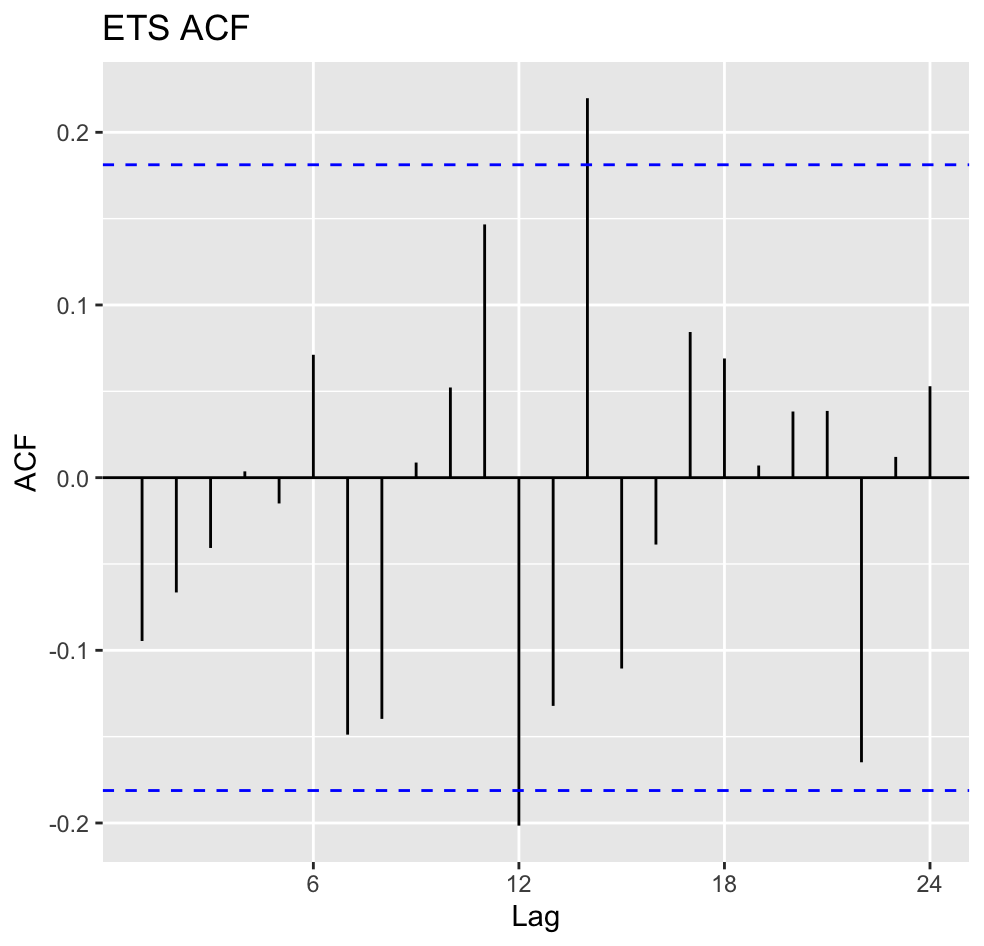
UNIVARIATE FORECASTING MODEL

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

In order to properly forecast the future Sales Tax for Grayson County, two univariate models were created. The first model was an ETS (M, Ad, M) model that was constructed by forecasting into the last six months of the data. That is, a training set of data was created from November 2011 to July 2021. Then the ETS model would be used to forecast the final six months of known data and compared to the actual values. The six-month ahead forecast for the ETS model is shown below in *Figure 2.1*.



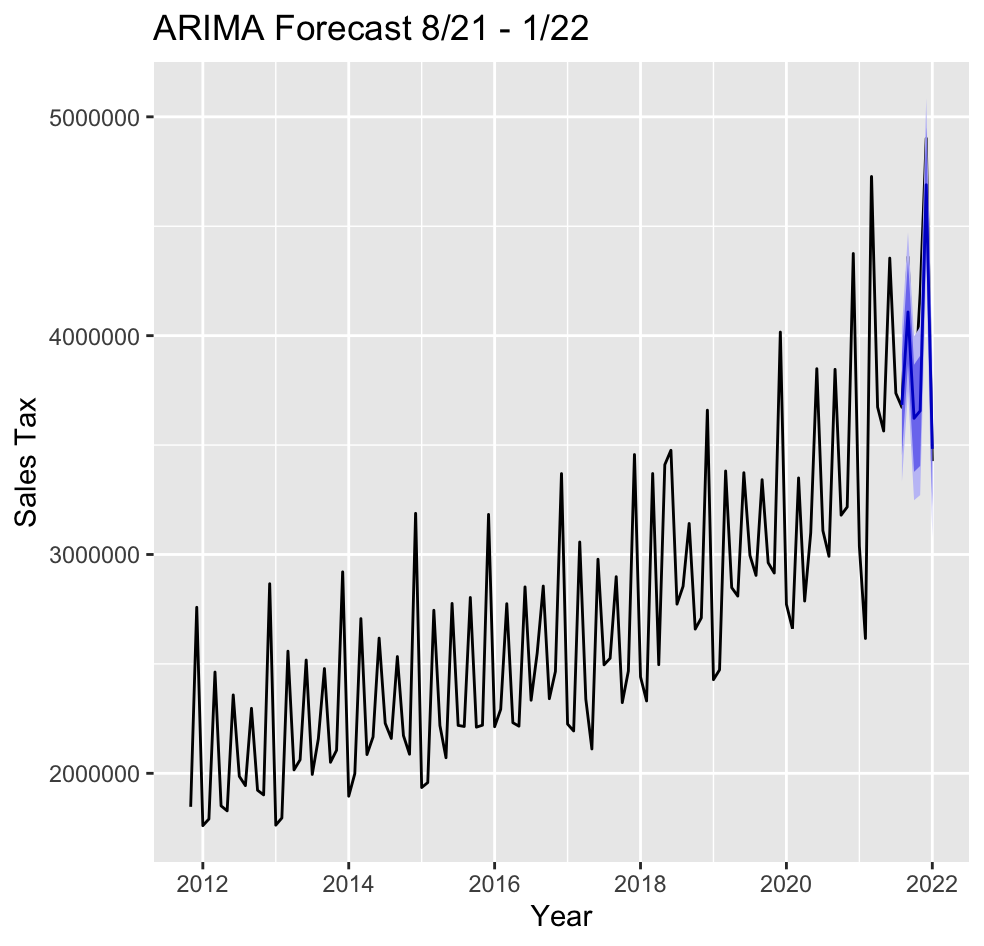
*Figure 2.1*



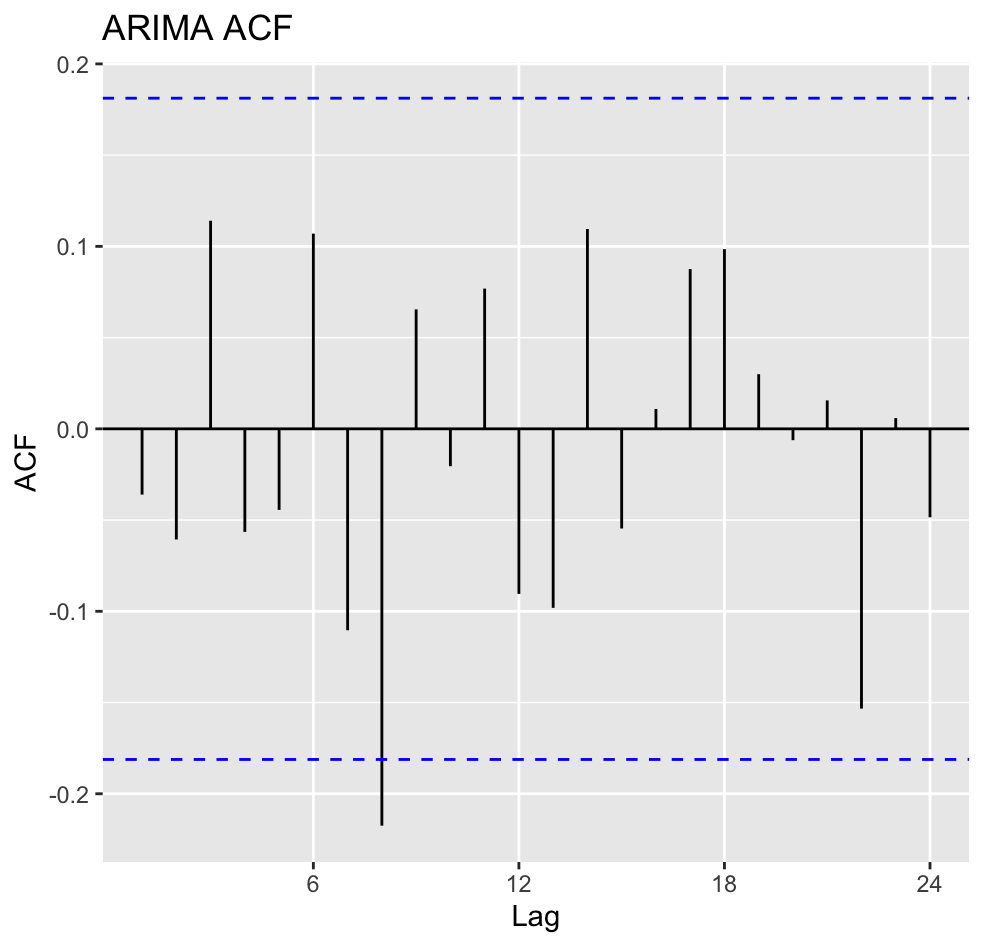
*Figure 2.2*

*Figure 2.2* above shows the residual diagnostics for the ETS model based on the training series. The plot shows a mean that is close to zero and no extreme correlation in the series.

In addition to the ETS model, an ARIMA(0,1,1)(0,1,2)[12] model was created from the training set and a six-month forecast was created. The forecast was then used to compare the accuracy of actual vs predicted values between the ETS model and the ARIMA model. The six-month ARIMA forecast into the final part of the data is shown below in *Figure 2.3*.



*Figure 2.3*



*Figure 2.4*

In the residual diagnostics given by the ACF plot in *Figure 2.4*, we can see a mean of the residuals near zero with no correlation and only one extreme outlier. Compared to the ACF from the ETS model, the ARIMA model appears to be a more accurate fit of the data, though this does not necessarily mean that it will be a better predictor of future values.

*Figure 2.5* below shows a comparison of the forecasted values from each model for the final six months of data compared to their actual values.

| **Date** | **08/21** | **09/21** | **10/21** | **11/21** | **12/21** | **01/22** |
| --- | --- | --- | --- | --- | --- | --- |
| True Value | 3,672,647 | 4,359,435 | 3,694,128 | 4,192,345 | 4,901,413 | 3,426,738 |
| ETS | 3,706,881 | 4,349,016 | 3,609,599 | 3,658,786 | 5,180,501 | 3,419,488 |
| ARIMA | 3,683,706 | 4,108,083 | 3,622,333 | 3,656,340 | 4,689,506 | 3,482,668 |

*Figure 2.5*

| **Training Set** | **ME** | **RMSE\*** | **MAE** | **MPE** | **MAPE** | **MASE** | **ACF1** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ETS: | 20364.5 | 155282.6 | 96995.39 | 0.4661299 | 3.45926 | 0.4839946 | -0.07388445 |
| ARIMA: | 20839.45 | 166706.4 | 98752.88 | 0.311657 | 3.434121 | 0.4927642 | -0.03603916 |

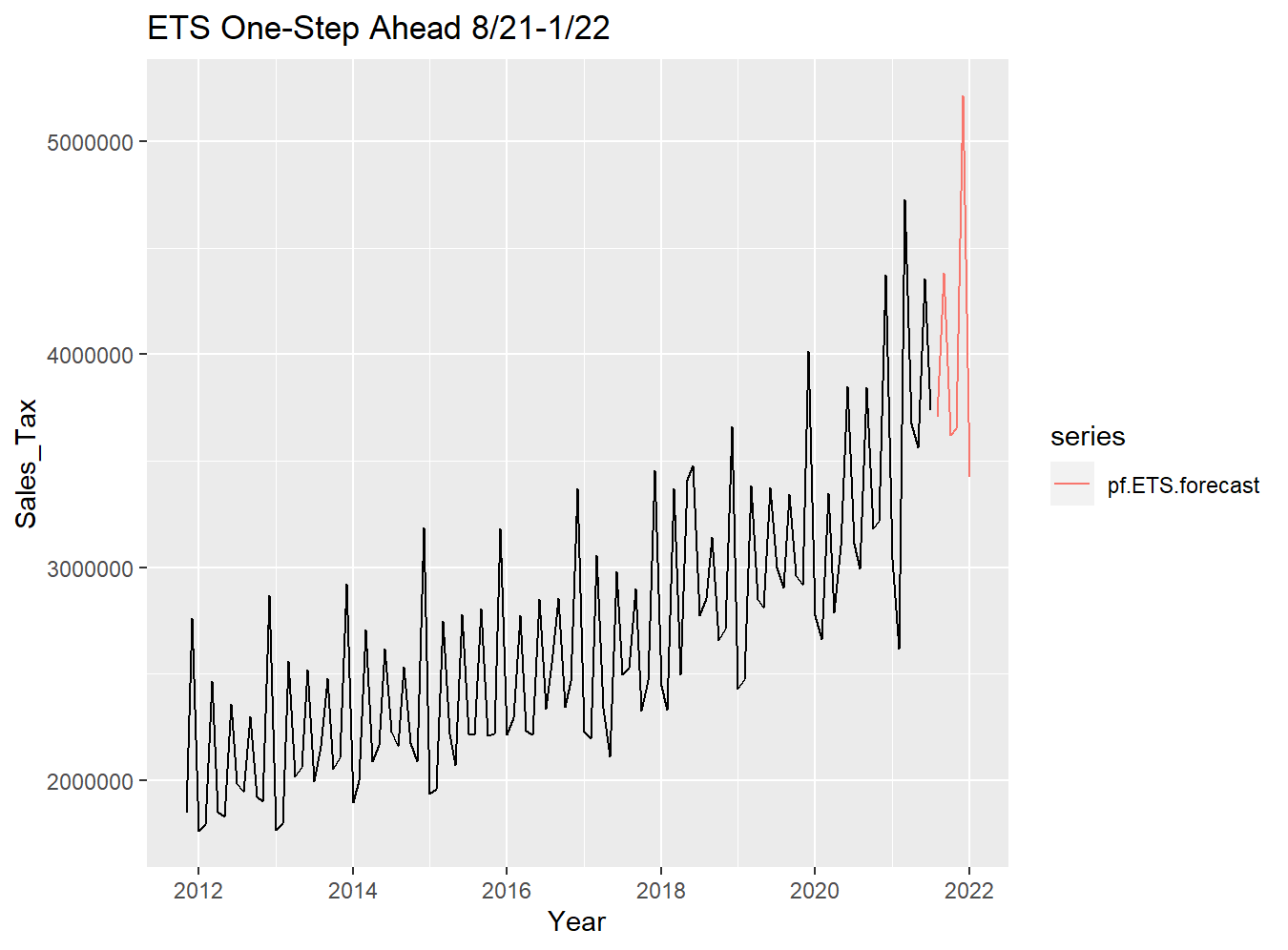
*Figure 2.6*

From the given predicted values in *Figure 2.5* and the accuracy measures of each model in *Figure 2.6* it is apparent that the ETS model provided a better forecast of the values of the last six months of data within the series. From the table we can see that the ETS model has a lower Mean Error, Residual Mean Squared Error, Mean Average Error, and Mean Absolute Scaled Error.

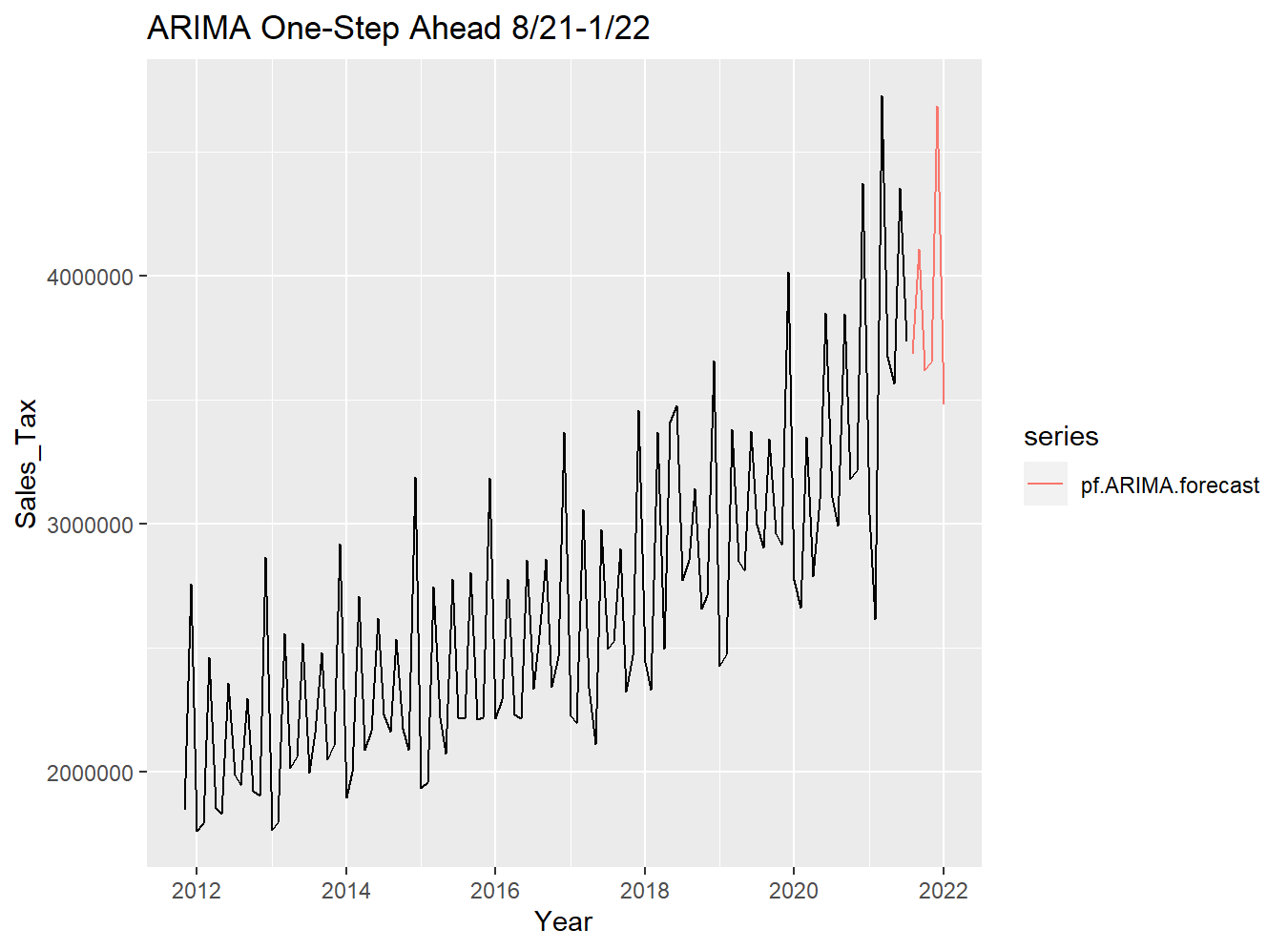
Before beginning to construct a full 12 month forecast for the future, two additional forecasts were created based on the training data. Using both the ETS and ARIMA models again, two series of 6 one-step-ahead forecasts were created for the last six months of given data. The predicted values obtained from these series are compared to the actual values below in *Figure 2.7*. The plots of each one-step-ahead forecast are shown in *Figure 2.8* and *Figure 2.9* below.

| **Date** | **08/21** | **09/21** | **10/21** | **11/21** | **12/21** | **01/22** |
| --- | --- | --- | --- | --- | --- | --- |
| True Value | 3,672,647 | 4,359,435 | 3,694,128 | 4,192,345 | 4,901,413 | 3,426,738 |
| ETS | 3,706,881 | 4,382,581 | 3,622,601 | 3,654,394 | 5,216,131 | 3,421,914 |
| ARIMA | 3,683,706 | 4,107,363 | 3,621,443 | 3,655,719 | 4,687,413 | 3,482,186 |

*Figure 2.7*



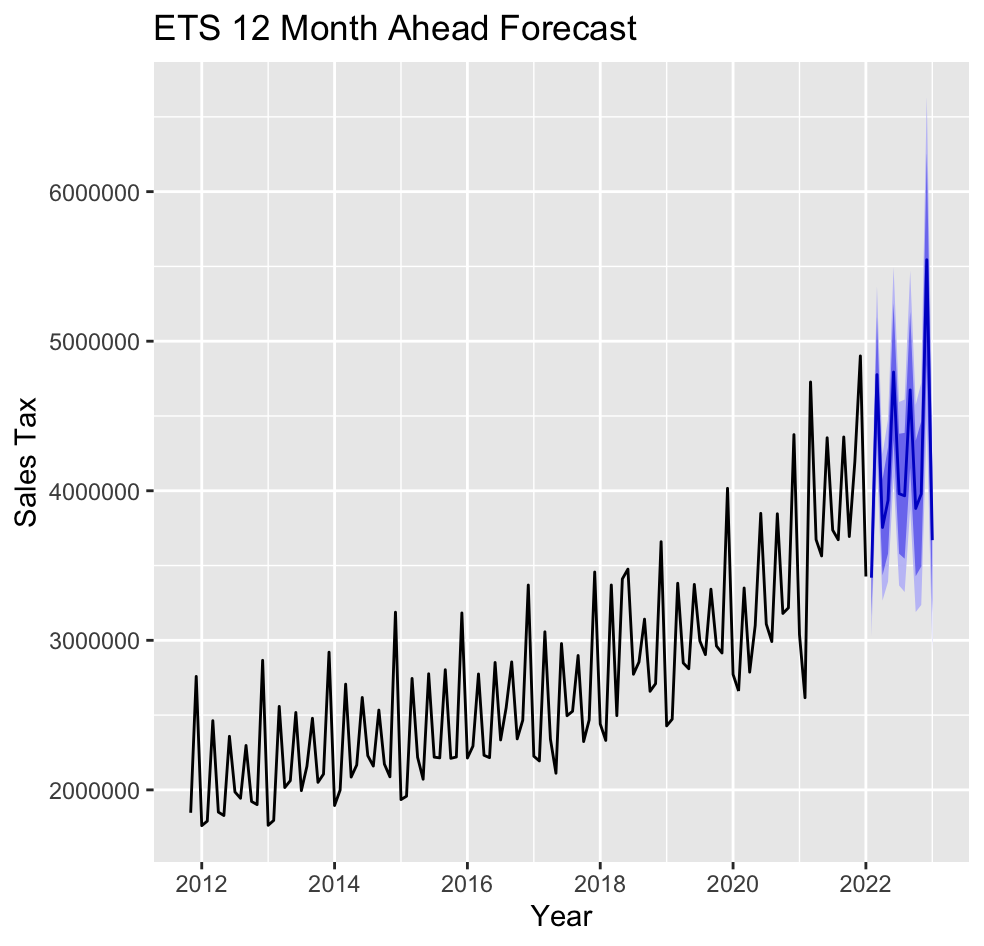
*Figure 2.8*



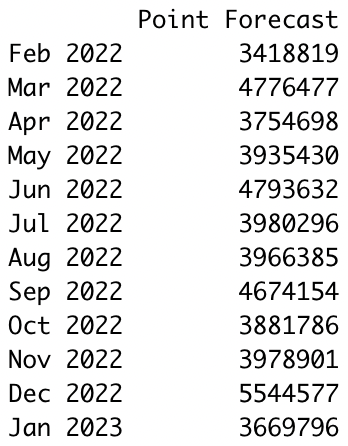
*Figure 2.9*

After comparing both the ETS and ARIMA models created from the training set of data, the ETS model was selected to be used to create a full 12 month forecast of the Sales Tax data. The ETS and ARIMA models both closely fit the data for the first 117 months of the series. Additionally, both models contained no significant correlation in their residuals. Given the accuracy measures provided in *Figure 2.6*, the ETS model appears to be a more accurate model with which to forecast the data and will be used for future predicted values. While both models should provide predictions that are fairly similar in terms of accuracy, the ETS model is slightly better than the ARIMA model.

After selecting the ETS model to forecast our series, a 12 month forecast was constructed to predict the Sales Tax for Grayson County over the next year. The forecast begins in February of 2022 and ends in January of 2023. The plot of the forecast is provided below in *Figure 2.10* along with the predicted values in *Figure 2.11*.



*Figure 2.10*



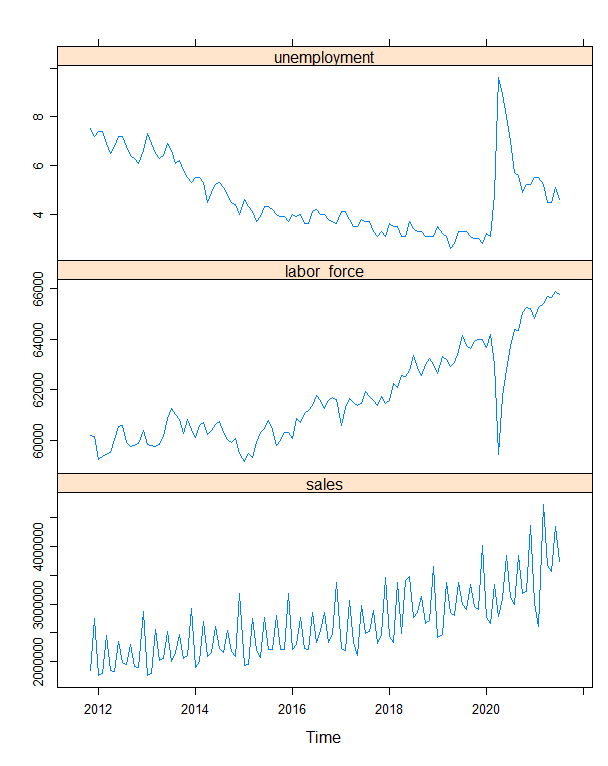
*Figure 2.11*

From the forecast, the upward trend is expected to continue although apparently slightly less steep than previously. A similar large spike in Sales Tax collection is expected in December of 2022, predicted to be around $5.5m. The seasonality and trend of the series are forecasted to continue.

MULTIVARIATE FORECASTING MODEL

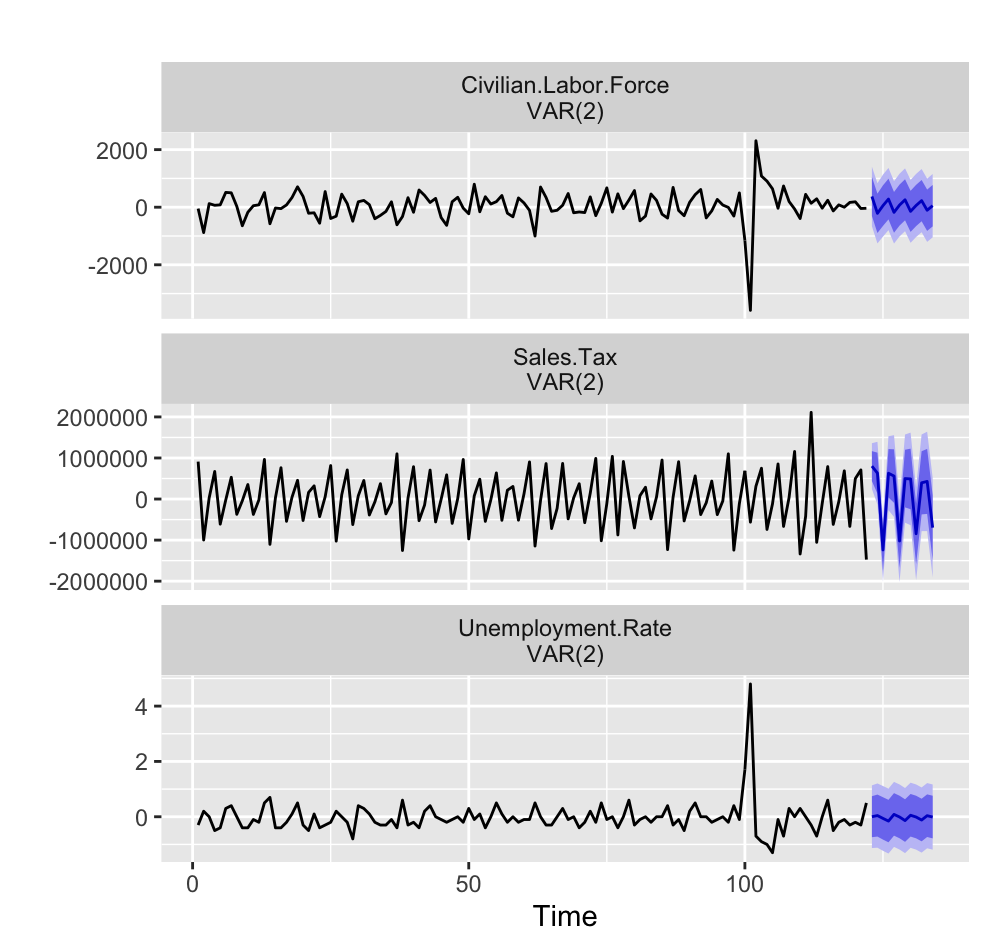
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

In addition to the ETS and ARIMA models, a VAR model was constructed to forecast the future 12 months of Sales Tax collections in Grayson county. The VAR contained three time series variables: the Sales Tax collections, the unemployment rate, and the civilian labor force population. All three of these variables went monthly from November 2011 to January 2022, providing 123 data points. These two additional time series data were chosen for their hypothesized relation to the Sales Tax collections. Economically, it would make theoretical sense that the civilian labor force and the Sales Tax collections would be positively correlated while the unemployment rate and Sales Tax collections would be negatively correlated. As more people enter the workforce, more income is generated. Most of this income will be spent within Grayson County leading to higher collections. Oppositely, as the unemployment rate increases, there is a larger amount of citizens that are spending a decreased wage from being out of a job. This decreased spending should be negatively correlated with the Sales Tax collections. In *Figure 3.1* below are the time series plots of each of the three variables used to construct the VAR model. An obvious point to note in the plots of both the unemployment rate and civilian labor force are the large spikes in 2020, as a result of the COVID-19 pandemic.

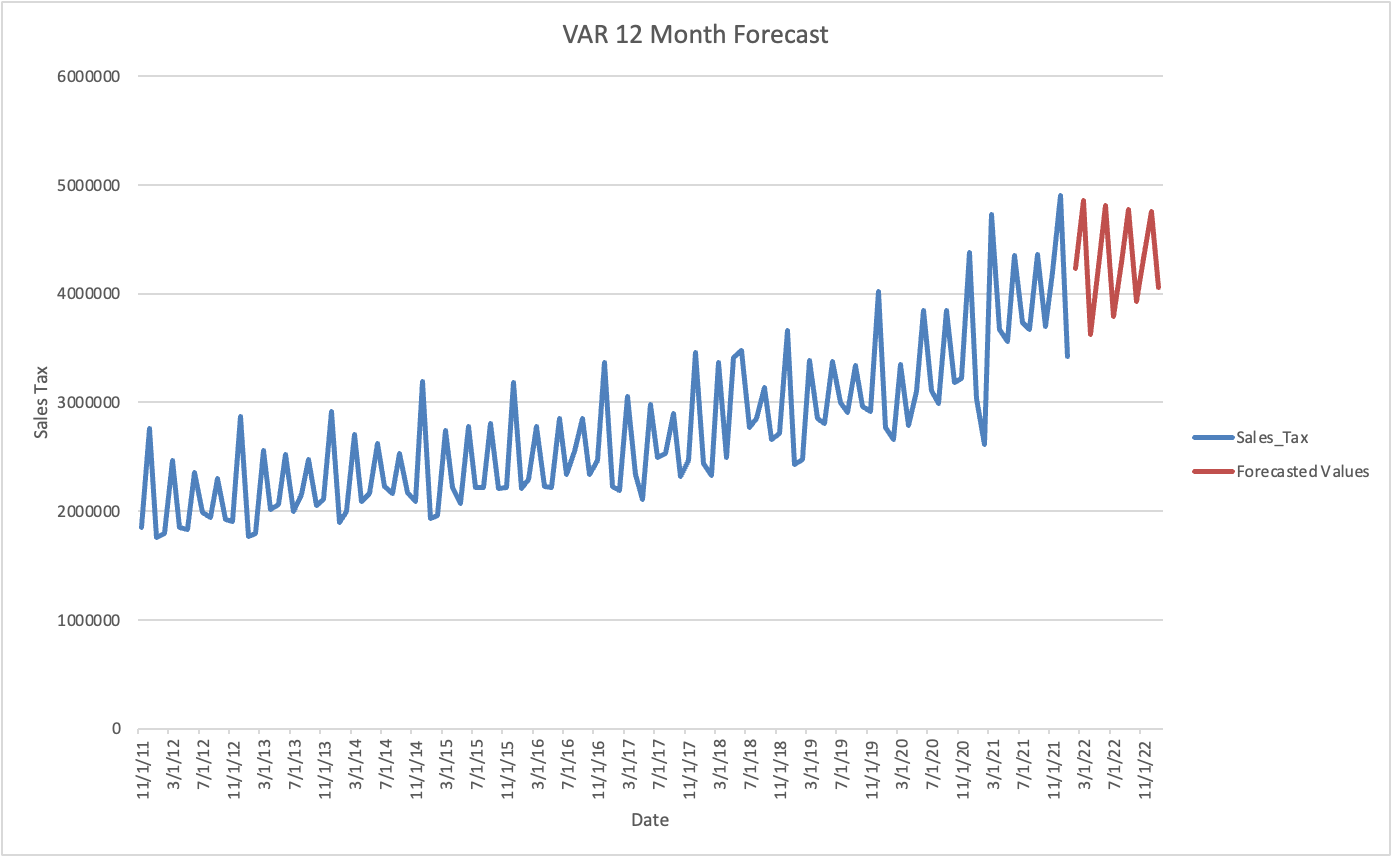


*Figure 3.1*

Each variable was differenced once to make them stationary and a VAR model was constructed with two lags. A 12-month ahead forecast was made from February 2022 to January 2023. *Figure 3.2* below shows the stationary forecast plots for each of the variables. *Figure 3.3* shows the undifferenced forecast for the next 12 months of Sales Tax collections based on the VAR model.



*Figure 3.2*



*Figure 3.3*

As compared to the ETS model, the VAR model indicates a lesser continuation of the trend from 2022 onward along with decreasing seasonality. Reasoning for these differences could be that accounting for our two labor statistics allows the VAR model to forecast using the lagged and current values of each variable, which in turn creates less extreme forecasts than the univariate models. The results of the VAR model were not significant (p = 0.146). Therefore, the ETS model is the preferred model to be used to predict the next 12 months of data.

BIBLIOGRAPHY

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

“Civilian Labor Force in Grayson County, TX.” *FRED*, US Bureau of Labor Statistics, 27 Apr. 2022, https://fred.stlouisfed.org/series/TXGRAY0LFN.

“Grayson County, TX.” *Data USA*, Data USA, https://datausa.io/profile/geo/grayson-county-tx#economy.

Hyndman, Rob J, and George Athanasopoulos. *Forecasting: Principles and Practice*. 2nd Edition, 2nd ed., OTexts, https://otexts.com/fpp2/, Accessed 1 May 2022.

“U.S. Census Bureau Quickfacts: Grayson County, Texas.” *US Census Bureau*, United States Government, https://www.census.gov/quickfacts/graysoncountytexas.

“Unemployment Rate in Grayson County, TX.” *FRED*, US Bureau of Labor Statistics, 27 Apr. 2022, https://fred.stlouisfed.org/series/TXGRAY0URN.