Term Project

Gabe Alpar, Margaret Kemp, Lucy Neat, Ilan Tijerina

Professor Schulman

English 475

30 April 2024

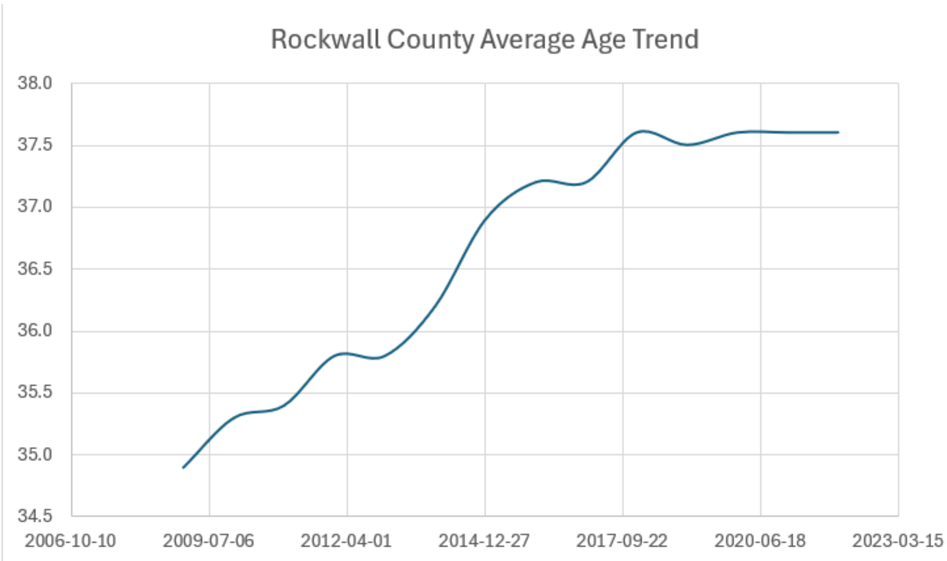
ECMT 475 TERM PROJECT - Total Alcohol Sales in Rockwall County

1. County Socio-Economic Profile:

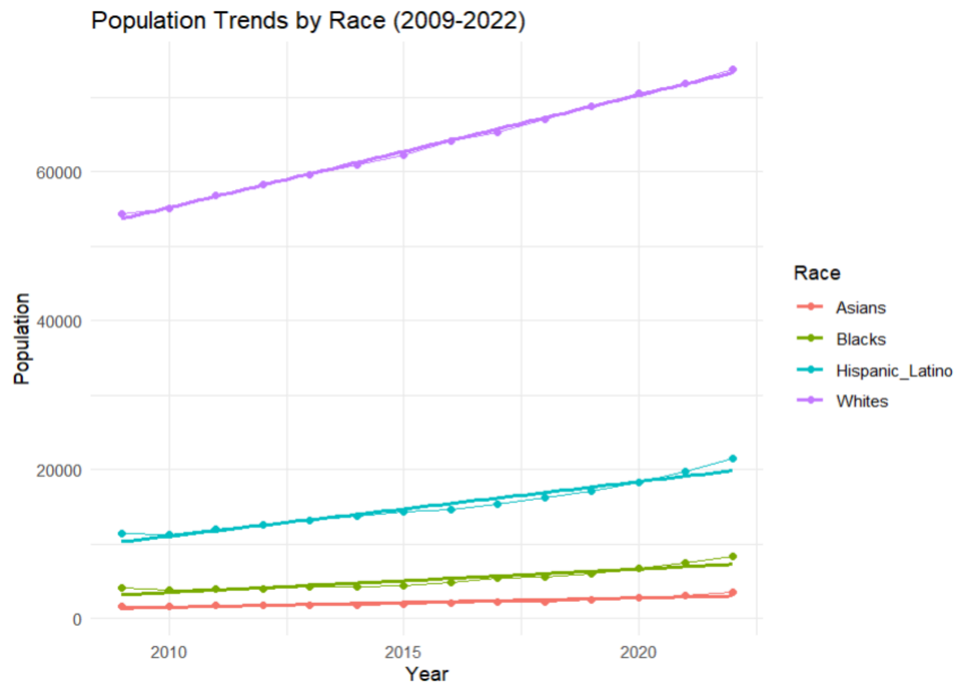
Despite being the smallest county in Texas, Rockwall is one of the state's most rapidly growing regions. Over the past decade and a half, Rockwall County has experienced significant population growth. In 2022, the population stood at 123,208, a staggering 56% higher than in 2010, which stood at 78,971. Moreover, within the last year, the population had grown an additional 6.5%, reaching 131,307 residents. This remarkable growth is driven by suburbanization from the nearby Dallas-Fort Worth area. Many families are drawn to Rockwall County for its superior educational opportunities and suburban lifestyle while maintaining proximity to the Dallas-Fort Worth metropolitan area.

The draw to Rockwall extends beyond its educational, location, and lifestyle offerings; its affordability is an additional significant attraction. Mayor Kevin Fowler notes that Rockwall's housing expenses are notably lower than those of neighboring states, enhancing its appeal to people moving from other states, such as California. With approximately 147 square miles of flat and gently rolling prairies, the county's landscape complements its booming population.

Economically, Rockwall County employs a workforce of around sixty thousand people: the key industries in the county revolve around health care, retail trade, and professional, scientific, and technical services, according to the American Community Survey conducted by the US Census Bureau. However, the county's strategic proximity to the Dallas-Fort Worth area is a significant catalyst for its growth, with many residents drawn to the ease of access to employment opportunities. Because of this, there are no disproportionately large employers located in Rockwall. The average commute from Rockwall to Dallas spans 32.7 minutes. The county is well connected with reliable bus and rail networks provided by the Dallas Area Rapid Transit (DART), which further enhances the county's connectivity and attractiveness to residents and businesses.



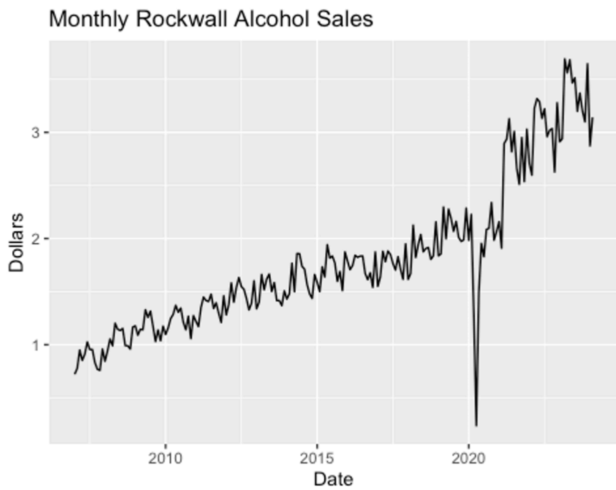
Reflecting on the graph above, we can see the direct correlation of "Average Age" and a chronologically forward sequence. From the beginning of 2006 to the finishing year of 2023, we see a starkly consistent increase in the average age of citizens in Rockwell County. The graph starts with an average age of 35.3, and within the next 17 years, it climbs to 37.6 as the average age per citizen within the county. This signifies an aging population and community settling down elderly, revealing a lower influx of younger generations moving into the county. This could result from its charming suburban lifestyle and distance from the city, which does not attract the younger populations.

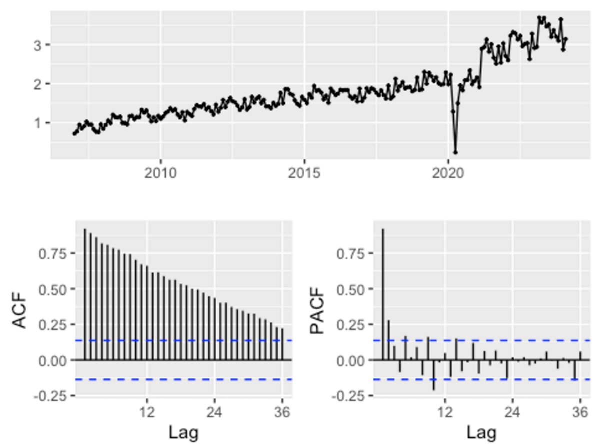


The figure above displays racial demographic trends from 2009-2022 in Rockwall County. It's clear that all races have had steady increases over the past several years, with the white population accounting for nearly 70% of Rockwall County's population. However, there is a rapidly growing Hispanic population in the county, as shown in the figure. A notable uptick in the Hispanic population growth rate between 2020-2022 has occurred. This can be attributed to the rise in the Hispanic population not just in Rockwall County but across the entire state of Texas. Hispanic citizens now maintain the largest race population in Texas. Released in 2023 by the Census, data revealed the shocking rate of 40.2% Hispanic citizens in Texas, surpassing the historically overpowered rates of White citizens, coming in at 39.8%.

2. Overview of The Total Sales Variable:

To effectively capture and analyze the fluctuations in alcohol sales in Rockwall County, we constructed a time series plot. The plot shows various elements, including trends, seasonality, and cyclical patterns. The dataset provides information on "liquor by the drink" sales from establishments such as bars and restaurants, thus excluding liquor or grocery store sales in Rockwall County. Spanning from January 2007 to February 2024, the data comprises data points for individual beer, wine, and liquor sales and a data point for the total sales. Our primary analysis focuses on the total sales variable.





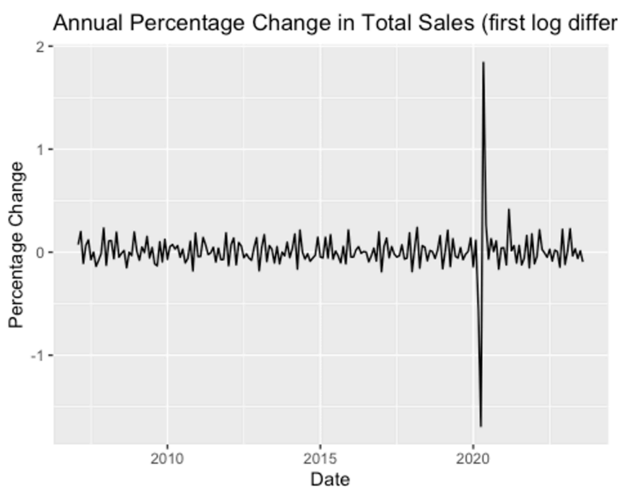
When looking at the time series plot, it is evident that there is an upward trajectory in the total alcohol sales within Rockwall County. However, this trend encountered a significant deviation in 2020. This is attributed to the impact of the COVID-19 pandemic, which caused the temporary closure of a considerable amount of restaurants and bars; thus, alcohol sales were very low. This highlights the strong influence of external factors on alcohol consumption patterns. There is also seasonality in the time series plot. Total alcohol sales have seasonal troughs and peaks. In addition to the trend and seasonal pattern, there is a cyclical pattern throughout the time series. The autocorrelation function decreases slowly, meaning Rockwall's total alcohol sales are not stationary. When looking at the partial autocorrelation function, we see significant lags that indicate autocorrelation, which explains the cyclical pattern.

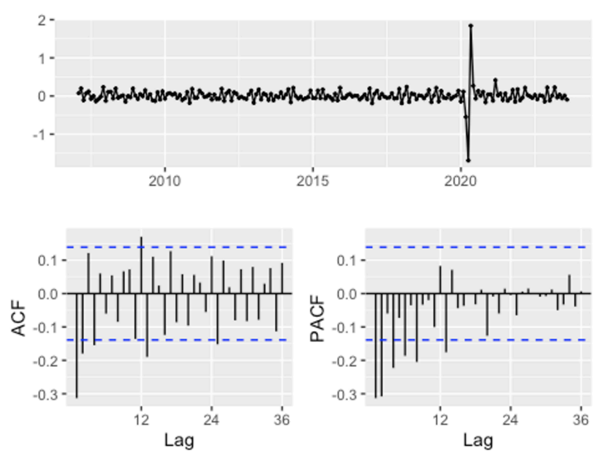
3. Univariate Forecasting Models:

I.Background Information

To accurately forecast the total alcohol sales in Rockwall County, it is imperative to apply a log-difference transformation to the variable. The results from the ndiffs() and nsdiffs() functions confirm that a single log difference is sufficient.

ndiffs(totalsales)	## [1] 1
ndiffs(**log**(totalsales))	## [1] 1
nsdiffs(**log**(totalsales))	## [1] 0
ndiffs(**diff**(**log**(totalsales)))	## [1] 0

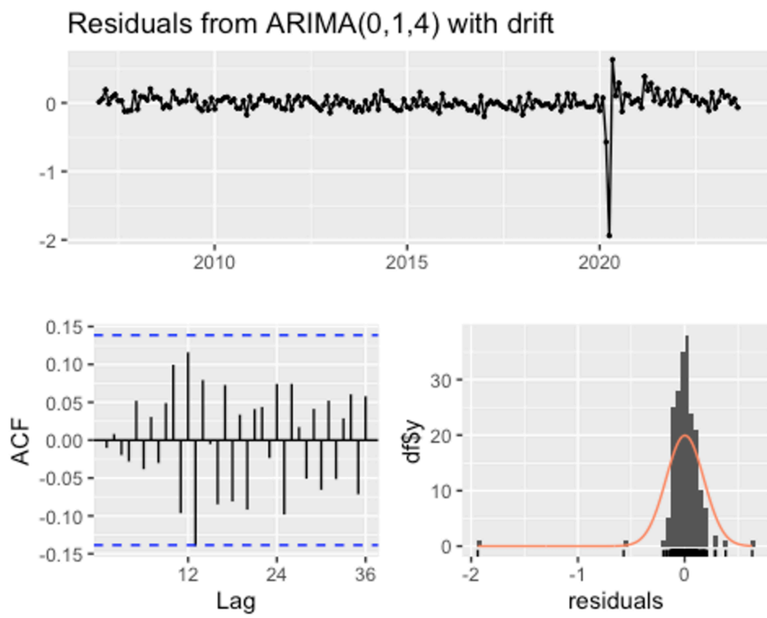




You can see in the above time series graph that the log difference of the variable is much more stationary. The logarithm helped to stabilize the variance of the time series, while the differencing helped stabilize the mean of the time series by removing changes in the level of the time series, thus reducing trend and seasonality. Even though we saw seasonality in our original time series plot, taking the seasonal difference is unnecessary; the first-log difference has already eliminated it. In addition to looking at the time plot of the data, the ACF plot is also helpful in identifying if the time series is stationary. In this case, the ACF drops to zero relatively quickly and resembles a white noise. There are very few autocorrelations lying outside the 95% limits. This suggests that the annual percent change in total sales of alcohol in Rockwall County is essentially a random amount that is uncorrelated with that of previous years.

II. Preferred ARIMA model:

```
## ARIMA(0,1,4) with drift
## Box Cox transformation: lambda= 0
##
## Coefficients:
##           ma1      ma2     ma3      ma4   drift
##       -0.5183  -0.2929  0.1192  -0.1837  0.0067
## s.e.   0.0704   0.0784  0.0798   0.0681  0.0017
##
## sigma^2 = 0.03233:  log likelihood = 60.98
## AIC=-109.96   AICc=-109.52   BIC=-90.2
```

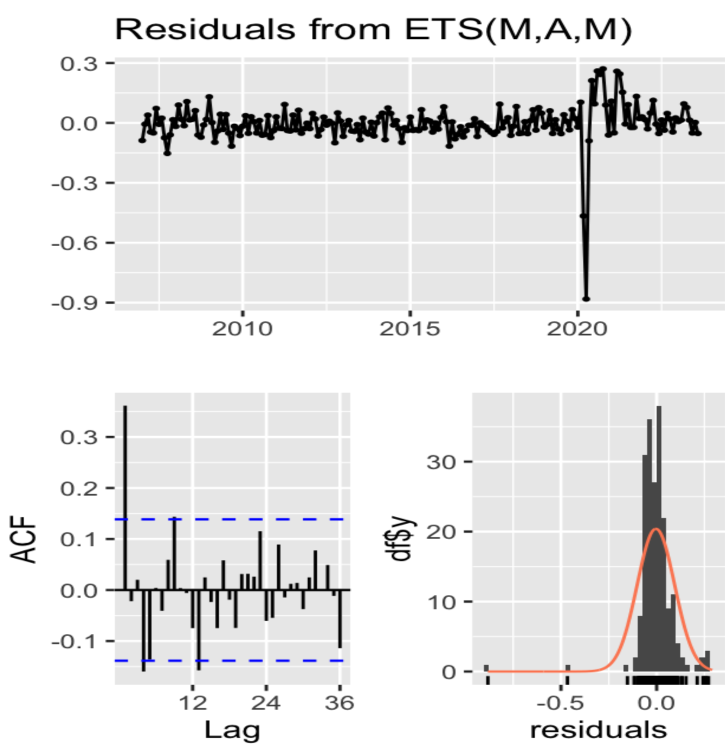


The preferred ARIMA model of the total sales variable is based on the six-month "hold out" sample of our data, where we forecast into the final six months of our data. We identified ARIMA(0,1,4) with drift as the model that best fits our data, which has the smallest AICc value. This specification indicated an autoregressive order of 0, a first differencing degree of 1, and a moving average order of 4.

Additionally, the autocorrelations are within the threshold limits, indicating that the residuals behave like white noise. Finally, the Ljung-Box test returns a significant p-value of .2792, reinforcing the assertion that the residuals conform to white noise characteristics.

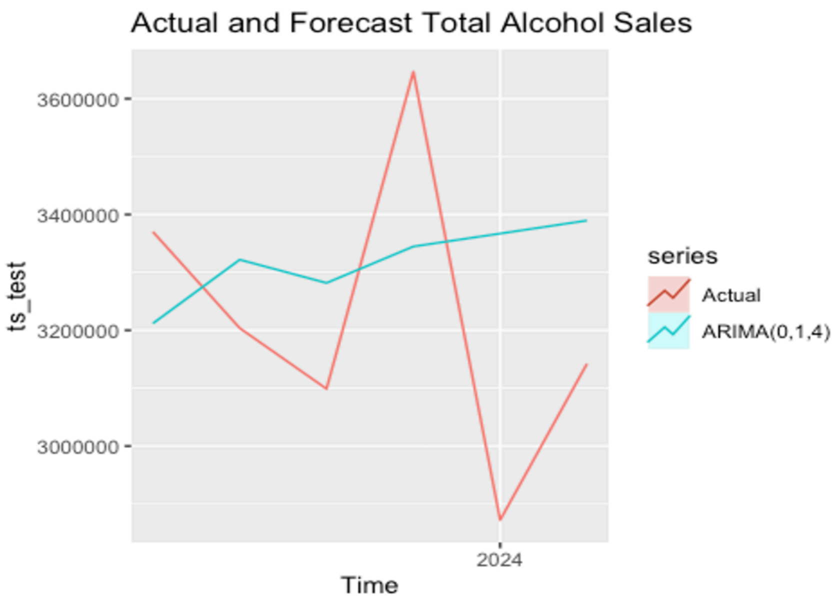
II. Preferred ETS Model

```
## ETS(M,A,M)
##
## Call:
##  ets(y = totalsales, model = "MAM", damped = NULL)
##
##   Smoothing parameters:
##     alpha = 0.2756
##     beta  = 1e-04
##     gamma = 2e-04
##
##   Initial states:
##     l = 840621.1857
##     b = 13148.5214
##     s = 1.037 0.8844 0.9668 0.9516 1.008 1.0512
##            1.0838 1.087 1.014 1.0655 0.9238 0.927
##
##   sigma:  0.102
##
##      AIC     AICc      BIC
## 5892.524 5895.887 5948.596
##
```



The ETS(M,A,M) model is determined as the best fit for the total sales variable based on the above output. This model of the total sales variable is based on the six-month “hold out” sample of our data, where we forecast into the last six months of our data. This model configuration incorporates exponential smoothing for the level (alpha), trend (beta), and seasonality (gamma). The model’s residual standard deviation is relatively low (.102), which indicates a good fit to the data. This model’s evaluation criteria, AIC, AICc, and BIC, support the suitability of this model as they are the lowest indicating the best fit. The ETS(M,A,M) model offers a holistic approach to capturing the underlying patterns in Rockwall County’s total alcohol sales data.

III. ARIMA Model six-month ahead forecast for September 2023 – February 2024.

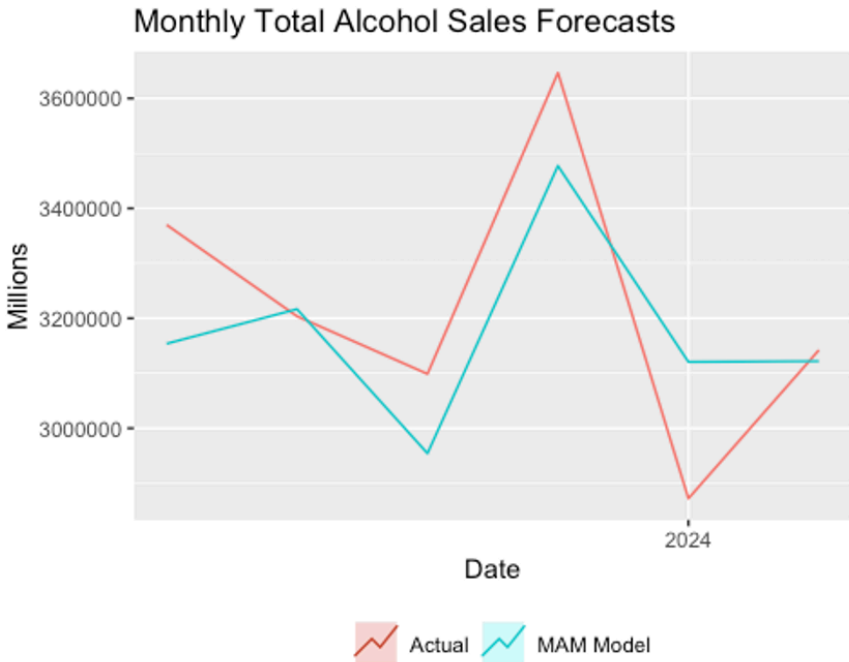



```
##                    ME     RMSE      MAE       MPE      MAPE      MASE
## Training set  24173.74 231163.7 149331.2 -2.348441 10.936123 0.7036504
## Test set     -97275.56 279751.1 250552.1 -3.622026  7.944185 1.1806046
```

How does the ARIMA(0,1,4) Forecast compare to the actual values for the total sales variable?

Based on the graph and reading the accuracy measures, it's clear that the ARIMA(0,1,4) forecast model is not entirely accurate. The blue line that represents the ARIMA forecast into our 6-month "hold out" period doesn't appear to have much sign of correlation with the actual data for this six-month period. Looking at the accuracy measures as well, the RMSE and MAPE are higher than those of the ETS forecast model, meaning more errors are indicating lower accuracy. That said, the MAPE for the test set is still under 10%, indicating a relatively accurate forecast. The forecast exhibits a more significant negative mean error (ME) on the test set compared to the training set, indicating a tendency to underpredict sales during this period. Overall, the ARIME(0,1,4) model captures some aspects of the underlying sales trends; its performance on the test set suggests room for improvement, especially in accurately predicting sales fluctuations.

III. ETS Model six-month ahead forecast for September 2023 – February 2024.



##	ME	RMSE	MAE	MPE	MAPE	MASE
## Training set	-2904.311	196591.1	100824.4	-4.311183	8.937046	0.4750860
## Test set	48072.547	162565.2	135240.9	1.216537	4.235843	0.6372566

How does the ETS MAM Model Forecast compare to the actual values for the total sales variable?

The ETS MAM model appears more accurate than the ARIMA 6-month ahead forecast. This may be due to the fact that the ETS model tends to focus more on trend and seasonality, both of which we see in our data. The graph shows that the MAM forecast aligns better with the actual data than the previous ARIMA 6-month ahead forecast. The lower accuracy measures confirm this interpretation. Looking specifically at the MAPE (mean absolute percent error) of the test set, it is only 4.24%, indicating that it's overall very accurate.

IIII. ARIMA Model: series of 6 one-step-ahead forecasts for September 2023 – February 2024


```
##                ME     RMSE      MAE       MPE     MAPE       ACF1 Theil's U
## Test set -62017.29 305662.9 243112.9 -2.575376 7.707189 -0.5921692 0.7106742
```

How do the six one-step ahead ARIMA model forecasts compare to the actual values for the total sales variable?

Compared to the six months ahead ARIMA forecast, this series of 6 one-step-ahead forecasts is more accurate. As we can see, the blue line representing the forecast aligns better with the actual data than it did in the 6-month ahead forecast, although it is still imperfect. With this type of ARIMA forecast, we actually see it following the trend line as it falls and rises around the same time as the actual data. The accuracy measures still indicate that this model is relatively accurate, as the MAPE is still under 10%. Like the 6-month ahead forecast, the model captures some aspects of the underlying sales trends; its performance on the test set suggests room for improvement.

IIII. ETS: MAM Model series of 6 one-step-ahead forecasts for September 2023 – February 2024


```
##                ME     RMSE      MAE         MPE    MAPE       ACF1 Theil's U
## Test set 9349.589 179938.2 143092.9 -0.02517683 4.56867 -0.2838367 0.3641097
```

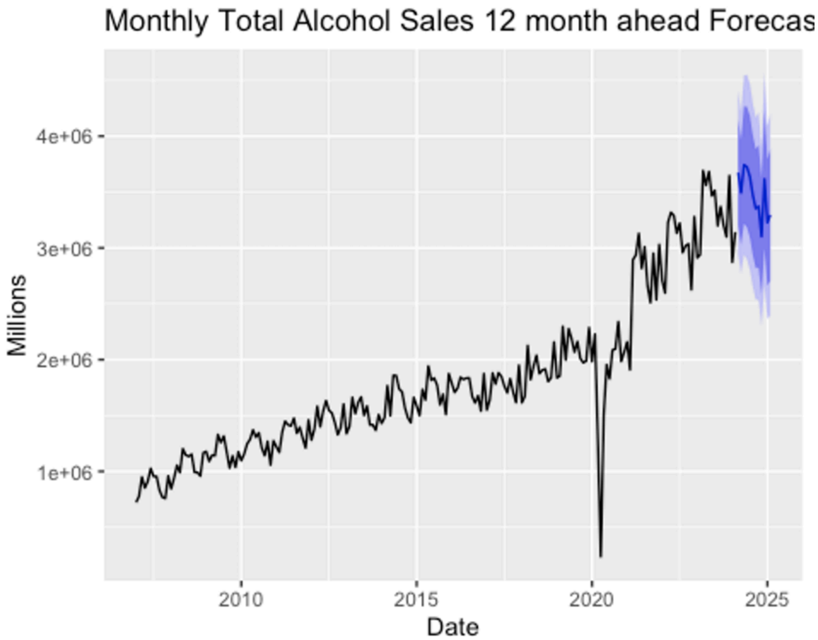
How do your forecasts compare to the actual values for the total sales variable?

Like the previous ETS model, this series of 6 one-step-ahead forecasts performed better than the ARIMA one-step-ahead forecasts. The blue line is aligned much better with the actual line, rising and falling very similarly at the same time. The accuracy measures are also lower than the previous ARIMA forecast, indicating that it is more accurate. Though this model proved to be more precise than both the ARIMA forecasts, the 6-month ahead ETS forecast still proves to be the best forecast.

VI. Based on your analysis in i-iii, what is your preferred forecasting model, ETS or ARIMA?

Although both ARIMA and ETS models are efficient in forecasting data, they differ in the components that they focus on. The ARIMA model focuses on autocorrelations within the data, whereas the ETS model focuses on trend and seasonality. The ETS model consistently demonstrated superior performance compared to the ARIMA model across various accuracy metrics, including Root Mean Square Error (RMSE), Mean absolute error (MAE), Mean Percentage Error (MPE), Mean Absolute Percentage Error (MAPE), Autocorrelation Function and Theil's U Statistics. The ETS model exhibits a more consistent performance, suggesting its robustness in capturing the underlying patterns in the data and providing accurate forecasts. Therefore, the ETS model is our preferred model for forecasting the total sales variable.

VII. Using the full dataset and the model you developed in 2.a., present a 12-month ahead path forecast for March 2024 – February 2025



Following the determination of the ETS model as our preferred choice for forecasting, we first recalibrated the model with the complete dataset encompassing each month of alcohol sales in Rockwall County, not just the training set our previous six-month forecasts were based on. We then projected the model 12 months ahead. The resulting time series plot shows historical data trends and the anticipated path for the next 12 months, offering valuable insights into the future trajectory of Alcohol Sales in Rockwall County.

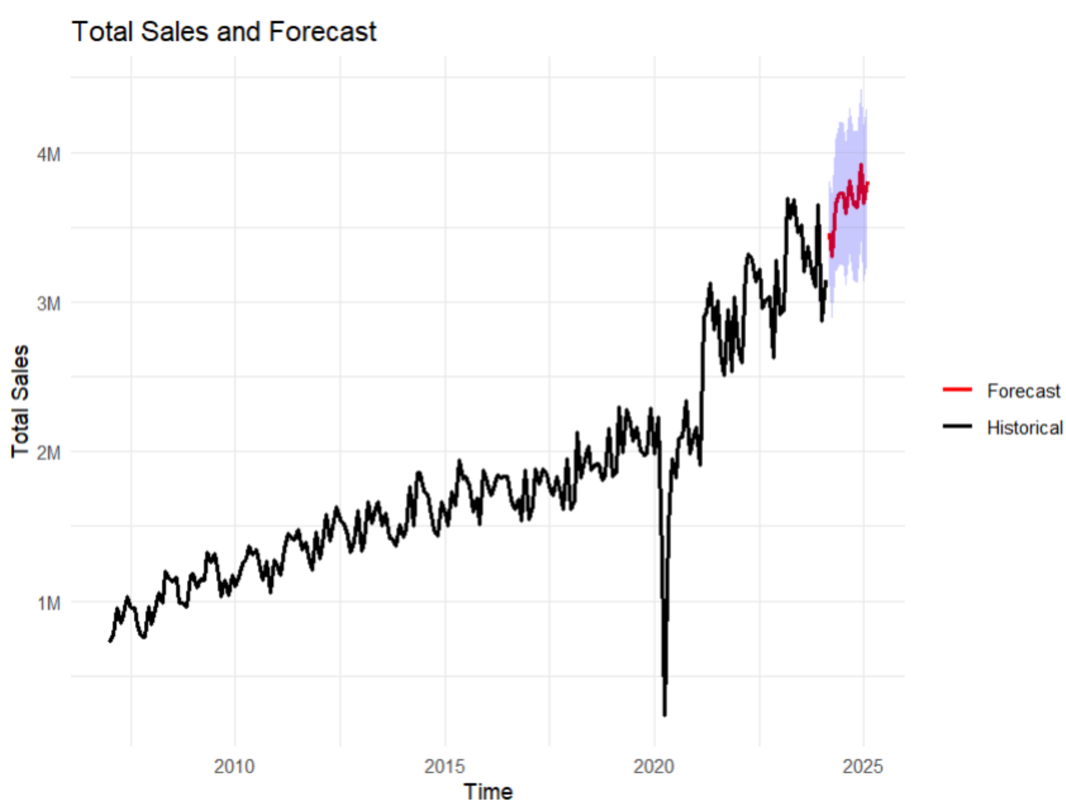
4. Multivariate Forecasting Models:

For our multivariate model, we needed to choose two variables that are most indicative of the purchase volume of Alcohol in Rockwall County. With that, we decided to go with the monthly unemployment rate in Rockwall County and the monthly national Consumer Price Index. Including these two variables is crucial for several reasons. First, economic conditions, reflected by unemployment rates, directly impact consumer purchasing power, potentially influencing a consumer's alcohol purchases. Second, changes in the national Consumer Price Index signal changes in overall price levels and inflation, which also affect consumer purchasing power and behavior. Thus, by incorporating these variables into our model, we can better understand the factors driving changes in alcohol sales, enabling better predictions and forecasts.

To create our multivariate VAR model, we first had to figure out how many lags were appropriate for our model after binding our variables. For this, we used lagselect with a max lag value of 12 and analyzed the AIC:

```
> alc.bv <- cbind(CPINationalts,TotalSales,UnemploymentRockwall)
> colnames(alc.bv) <- c("CPINationalts", "TotalSales", "UnemploymentRockwall")
> lagselect <- VARselect(alc.bv, lag.max = 12, type = "const")
> lagselect$selection
AIC(n)  HQ(n)  SC(n) FPE(n)
     9      2      2      9
```

Here, we can see that our appropriate lag value is nine by our AIC. After that, we created our multivariate VAR model with p=9. Once our model was created, we were able to develop a 12-month ahead forecast of our Total Sales variable:



Our VAR Total Sales forecast indicates an overall increasing trend in total sales. This can be compared to our univariate forecast, which showed a decreasing trend in total sales for the 12-month ahead forecast.

<u>Works Cited:</u>

Bureau, United States Census. "Rockwall County, Texas." Explore Census Data, data.census.gov/profile/Rockwall_County,_Texas?g=050XX00US48397. Accessed 22 Apr. 2024.

"Estimate, Median Age by Sex, Total Population (5-Year Estimate) in Rockwall County, TX." FRED, 7 Dec. 2023, fred.stlouisfed.org/series/B01002001E048397.

"Inflation, Consumer Prices for the United States." *FRED*, 19 Dec. 2023, fred.stlouisfed.org/series/FPCPITOTLZGUSA.

Mekouar, Dora. "Hispanic Residents Now Outnumber White Population in Texas." Voice of America, Voice of America (VOA News), 23 Oct. 2023, www.voanews.com/a/hispanics-now-outnumber-whites-in-texas/7316478.html#.

"Rockwall County, TX." Data USA, datausa.io/profile/geo/rockwall-county-tx#:~:text=In%202021%2C%20Rockwall%20County%2C%20TX,%24111%2C595%2C%20a%205.32%25%20increase. Accessed 22 Apr. 2024.

"Rockwall County: Workforce Solutions for North Central Texas: Workforce Solutions for North Central Texas." DFWJobs, www.dfwjobs.com/employers/labor-market-information/rockwall-county. Accessed 22 Apr. 2024.

"Unemployment Rate in Rockwall County, TX." *FRED*, 3 Apr. 2024, fred.stlouisfed.org/series/TXROCK7URN.