Forecasting Load to Truck Ratios in the Los Angeles Market

STA 5250: Time Series Analysis

Erik Doan and Samantha Secor

Fall Semester 2019

Abstract

In this paper, we will analyze a transportation market indicator metric called the load-to-truck ratio, LTTR, specifically for the Los Angeles market. We will determine if gas prices are a good predictor of this figure and come up with methods to create the best model for our data.

Introduction

The transportation industry is very different than many other categories of business. For one, there is a huge amount of spend. The entire U.S. logistics sector (which includes various modes of transportation such as air, water, rail, and trucks) was worth roughly $1 trillion in 2018, according to Plunkett Research.

Another difference is that the transportation industry is highly fragmented. There are many freight companies of ranging sizes of scope and scale that are profitable today with the strategies they implement. Because of this, competition between freight carriers is fierce.

Another point to note is the lack of innovation in logistics. While companies like Uber Freight and Convoy are dramatically changing the way the U.S. operates in freight, we are just seeing these innovations after many years of stagnation. When other industries flourished with new technologies and modernism, transportation remained relatively unchanged.

Because of these differences, many people in the transportation industry are now looking into big data to come up with insights that can help answer their many questions. The biggest question anyone in transportation can ask is: What will the market look like in 2020? Market predictions remain to be a phenomenon that's very challenging to predict, however, market indicators such as the load-to-truck ratio may be easier.

The load-to-truck ratio, referred in this paper as LTTR, is the ratio of the number of loads (shipments needing trucks) to the number of trucks in a certain market. If there are 100 loads that need to ship in an area but only one truck available, the market is "tight" and the rate for that truck is high. If there is one loads that need to ship in an area where 100 trucks are available, the market is "soft" and the rate for those trucks are low.

Using the LTTR to predict market softness vs. tightness is a incredible tool for transportation. Within this paper, we will try to do just that for the Los Angeles market.

Correlating Load to Truck Ratios with Gas Prices

Many individuals in transportation claim that there is a distinct correlation between gas prices and the LTTR. Our knowledge aquired in Time Series Analysis helps us determine whether these experts are correct.

We have acquired both Los Angeles gas price data (from the U.S. Energy Information Administration) and LTTR data (from DAT Solutions) from January 5th, 2015 to July 29th, 2019. Both data sets are recorded on a weekly basis and there are 239 weeks spanning this time period. We have plotted that data in Figure 1 and Figure 2, respectively.

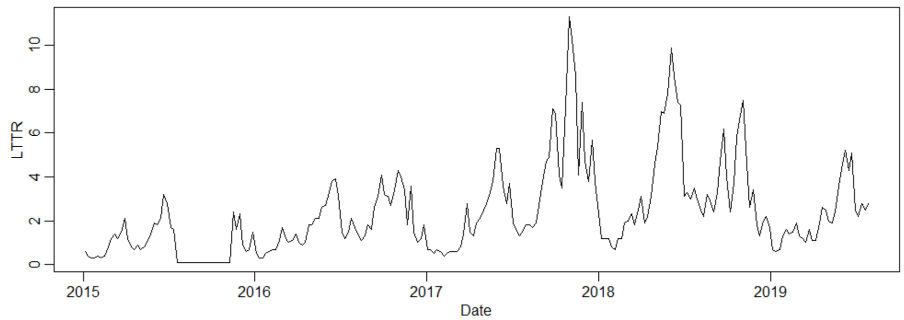

Figure 1: LTTR data in the Los Angeles area from January 5, 2015 to July 29, 2019

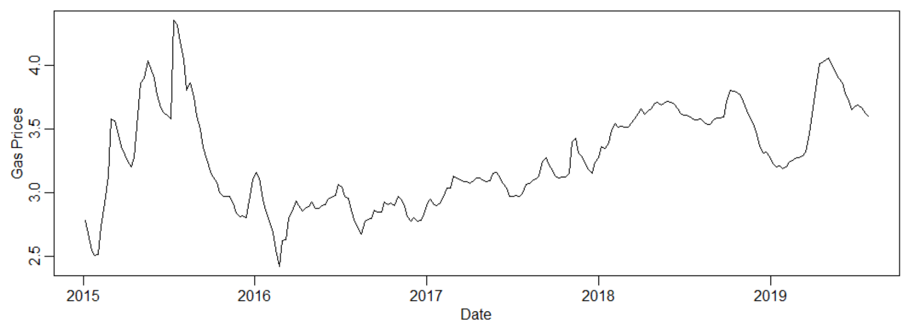

Figure 2: Gas Prices in the Los Angeles area from January 5, 2015 to July 29, 2019

In Figure 1, for LTTR data, note that from July, 2015 to November, 2015, the LTTR was 0.1. It is unlikely that the LTTR was exactly 0.1 for this entire time period but more likely that the data is unclean. As such, we will not be taking log transforms on this data set. However, due to the slight upward trend, we will be differencing the data.

In Figure 2 we see an upward trend beginning in 2016 which continues through 2019. Because of this, we will be differencing this data set as well. We have plotted the differenced data in Figure 3 and Figure 4

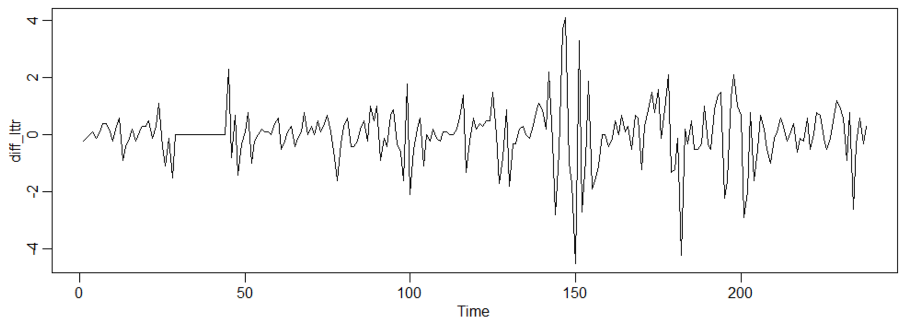

Figure 3: Differenced LTTR data in the Los Angeles area from January 5, 2015 to July 29, 2019

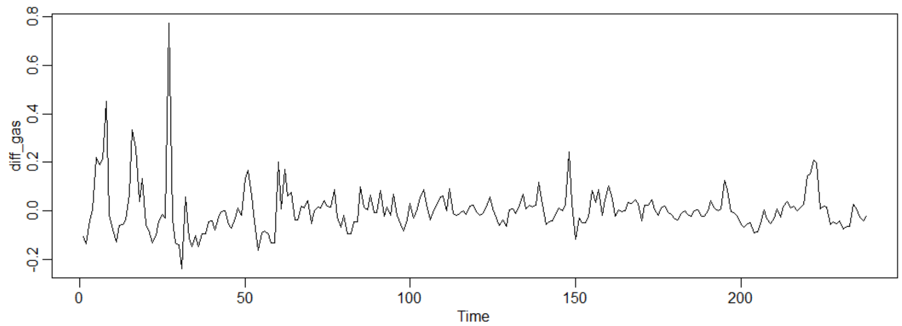

Figure 4: Differenced Gas Prices in the Los Angeles area from January 5, 2015 to July 29, 2019

In both time series, there appears to be no trend and the variance seems to be relatively constant. We determined that both differenced time series are stationary.

Now that we have stationary data, we attempt to correlate the data on the differenced time series.

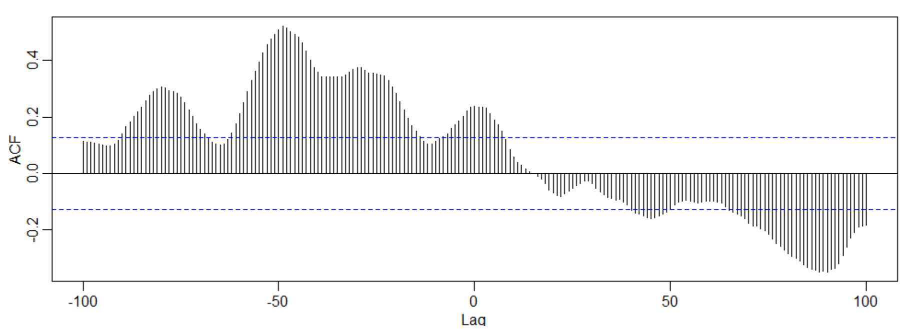

Figure 5: Cross-correlation of the two univariate series, differenced LTTR and Gas Prices data

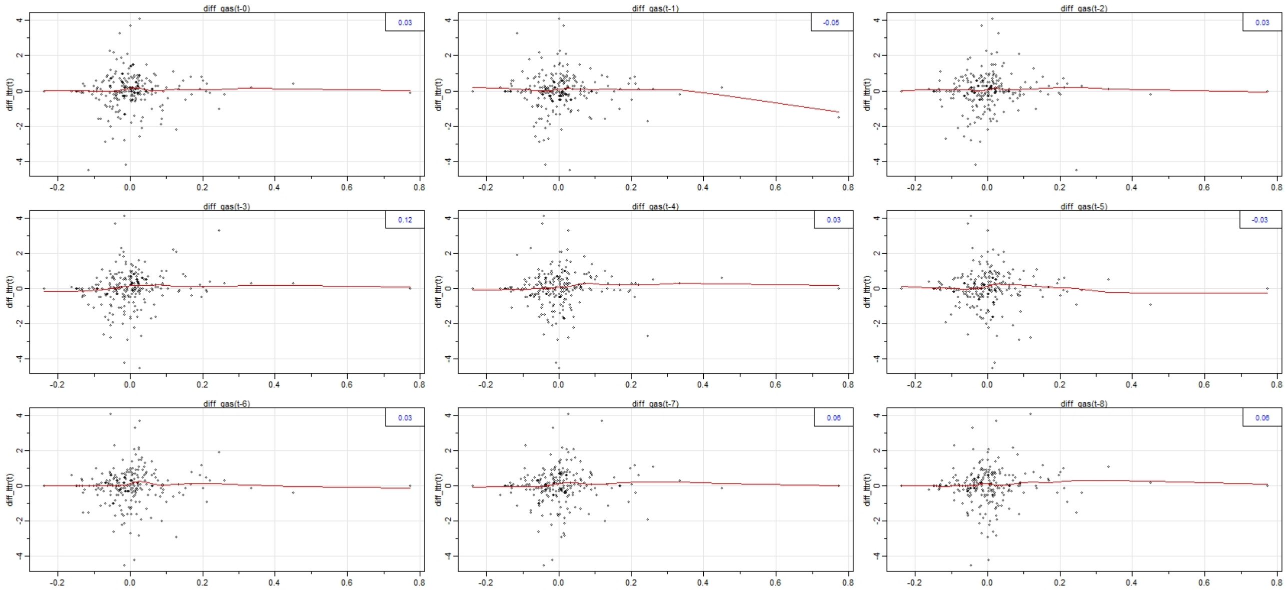

Figure 6: Lag Plot of the two time series, LTTR and Gas Prices

Note that the CCF in Figure 5 has a lot of varying data and it's very hard to take anything away from the plot. The Lag Plot in Figure 6, explicitly shows that there is actually no correlation going on between the two data sets. From these two charts, we are able to ascern that gas prices are actually not a good useful predictor for LTTRs for the Los Angeles area. Therefore, we will continue our analysis of LTTRs without using gas prices as a predictor.

Building an ARIMA Model

Now that we are able to focus solely on the LTTR data, we can utilize the obvious periodic cycles that are affecting our data. We converted our data into a seasonal time series with a period of 52. Figure 7 provides a plot of the LTTR time series data with a period of 52.

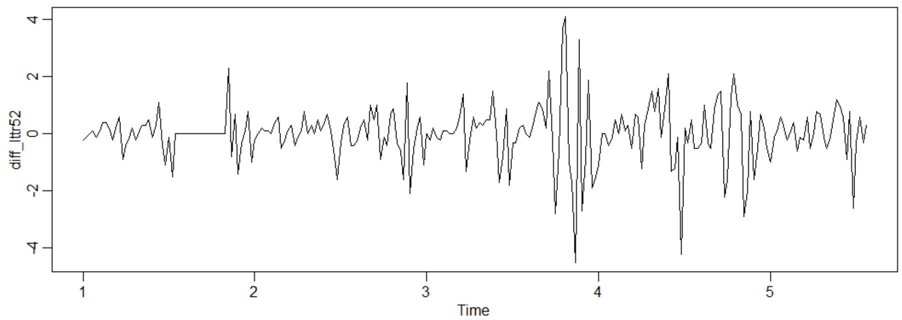

Figure 7: Time series plot of differenced LTTR data with a period of 52

To determine if there is a correlation between yearly data, we needed to plot the ACF and PACF functions for our data set.

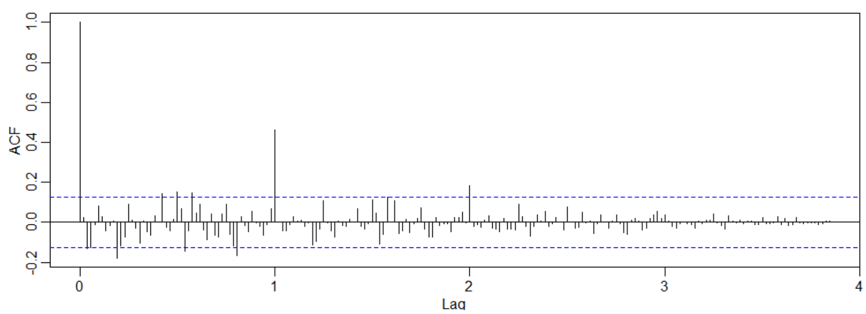

Figure 8: Auto Correlation Function of LTTR

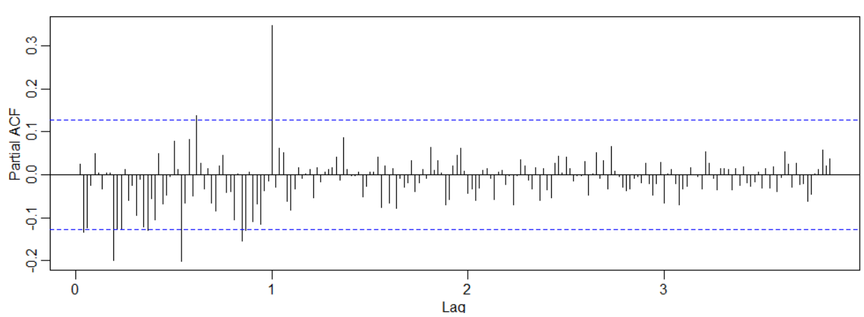

Figure 9: Partial Auto Correlation Function of LTTR

Inspecting the ACF in Figure 8, we find a strong correlation at lag 1 year and 2 years. In the PACF in Figure 9, we find a strong correlation at lag 1.

There are many ways to interpret this pair of ACF and PACF plots. Initial reactions indicate the ACF cutting off at lag 1 or 2 and the PACF tailing off, suggesting that the differenced data is an MA(1) or MA(2). We might also assume that the PACF cuts off at lag 1 and that the ACF is trailing off, suggesting that the raw data is an ARIMA(1,1,0). An alternative hypothesis is that the data is ARIMA(0,1,1) or ARIMA(0,1,2).

To help us determine the correct model, we use the `auto.arima` function in R on the differenced data (i.e. $I = 1$). Below in Table 1, we compare models based on their AIC values.

ARIMA	(p,d,q)	Seasonality	AIC
ARMA(1,1)	(1,0,1)	(0,0,1)	621.1477
ARMA(1,2)	(1,0,2)	(0,0,1)	621.9560
MA(1)	(0,0,1)	(0,0,1)	631.7362
MA(2)	(0,0,2)	(0,0,1)	626.9407
AR(1)	(1,0,0)	(0,0,1)	632.8942

Table 1: Possible ARIMA models with seasonality and AIC values

From Table 1, we can see that our initial guesses of MA(1), MA(2) and AR(1) have a much higher AIC value as compared to the ARMA(1,1) and ARMA(1,2) models. The ARMA(1,1) was the best model based on the AIC.

We now use `sarima` to fit our ARMA(1,1) and ARMA(1,2) models and we will determine which of the two models has the best fit. Below are the outputs of each model.

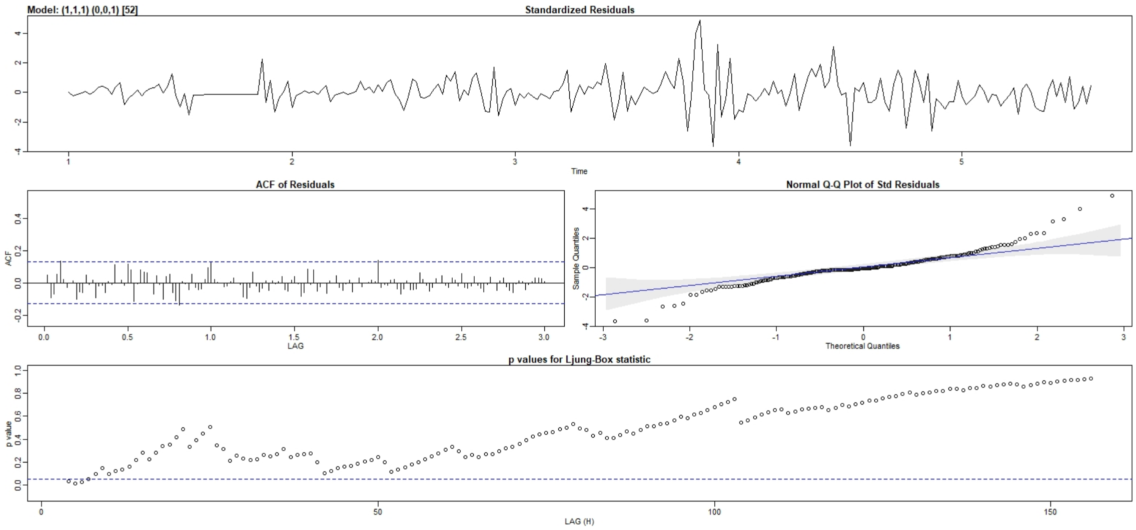

Figure 10: SARIMA outputs for ARIMA(1,1,1)

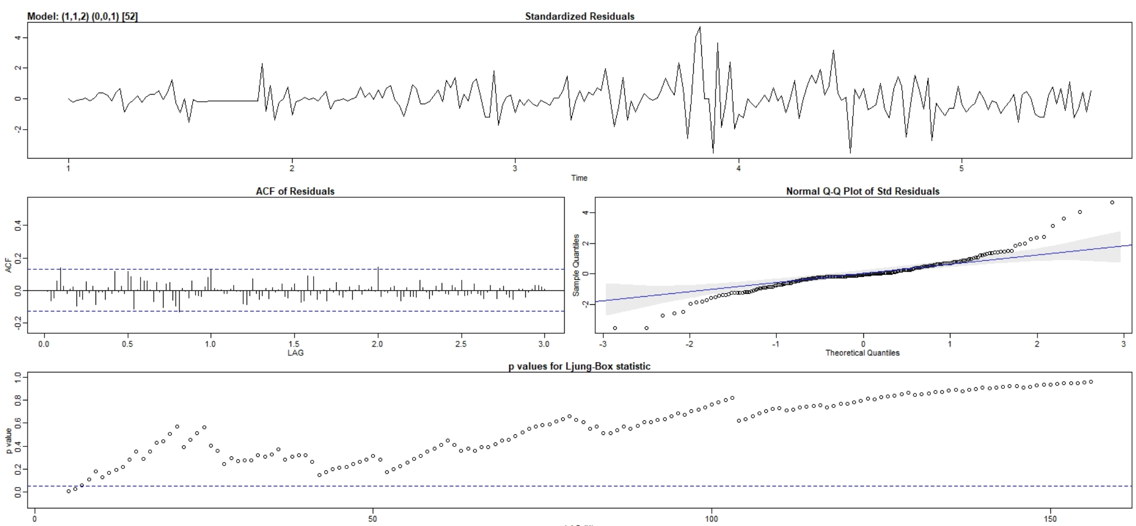

Figure 11: SARIMA outputs for ARIMA(1,1,2)

Based on the p-values for the Ljung-Box statistic in Figure 11, we determined that the ARIMA(1,1,2)(0,0,1)[52] had the best fit by a slight margin. We can also see from our Q-Q plot that this data is normally distributed and analyzing the standarized residuals shows no obvious patterns. The ACF shows no obvious deviations from our model assumptions. Finally, the Q statistic is not significant at the lags shown. We can conclude that this model is a good fit for LTTR data.

Now that we've realized the best fit, we can now forecast the results using our ARIMA(1,1,2)(0,0,1)[52] model to help understand how Los Angeles LTTR will play out in the next year. The results of this prediction are displayed in Figure 12.

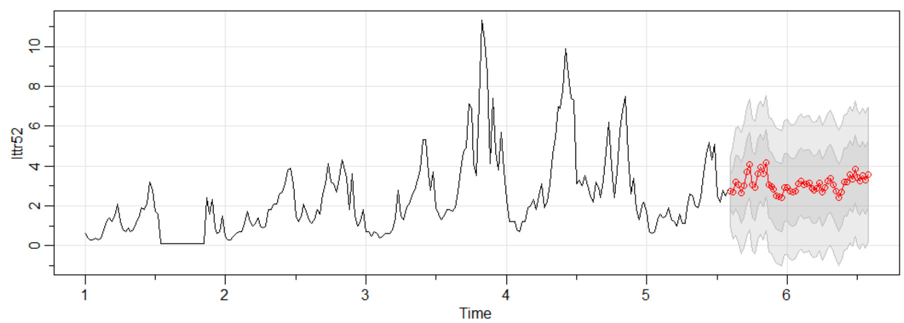

Figure 12: Fifty-two week forecast using ARIMA(1,1,2)x$(0,0,1)_{52}$ model on the LTTR data

Interpreting these results shows us that there might be some fluctuations with LTTR coming up on the holidays to finish 2019, however the predictions look relatively flat for 2020

GARCH Model

Viewing our differenced time series LTTR data in Figure 3, it looks like there are some spikes in variance and a GARCH model might be appropriate. We can fit a GARCH model using the R command `garchFit`. We use the argument `cond.dist = "std"` since our Q-Q plot suggest a we should turn on residuals into a t-distribution. We can check the AIC of our GARCH model with the equation

$$AIC = -2 * loglik + 2p. \tag{1}$$

Running `garchFit` on ARIMA(1,1,2)(0,0,1)[52] + GARCH(1,1), we obtain a log likelihood value of -270.8929. Plugging this value into Equation 1, we obtain an AIC value of 555.7858. This AIC value is lower than those obtained in Table 1. Thus using an ARIMA(1,1,2)(0,0,1)[52] + GARCH(1,1) is an even better model fit solely based on the AIC.

Similarly with the SARIMA model, we'd like to see how the forecasted results look like with our ARIMA + GARCH model.

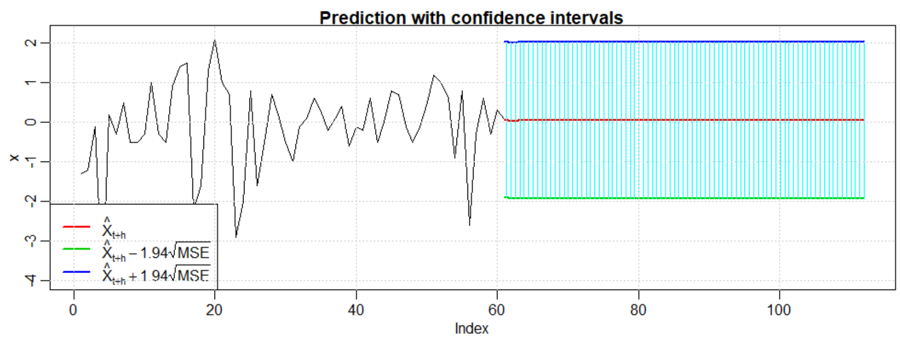

Figure 13: Fifty-two week forecast using ARIMA(1,1,2)x$(0,0,1)_{52}$+GARCH(1,1) model on the LTTR data

We can see that the ARIMA(1,1,2)x$(0,0,1)_{52}$+GARCH(1,1) performs poorly when predicting LTTR data. This might be because GARCH models do not take into account seasonal effects which we say had high significance in our previous models.

Conclusion

The transportation industry can be thought of as a well-established and also emerging field. Data is becoming more and more necessary as shippers and carriers commit to driving innovations.

This paper scratched the surface to demonstrate the importance of statistical analysis and predictions in logistics. Market indicators such as the LTTR are great datasets for analysts in transportation departments. Ultimately, we discovered that the ARIMA(1,1,2)x$(0,0,1)_5$2 was the best model forecast that we could provide for the Los Angeles market. It indicated that the LTTR would likely be relatively flat for 2020, which matches what many experts claim about the Los Angeles market.

While it is great to now have a model and forecast for the Los Angeles LTTR, there are many ways we can take this research farther. Los Angeles is just one market - and there is LTTR data for many different areas in the U.S.. We realized gas prices were not the best predictor, but we have yet to explore a market's GDP and consumer expectation, which experts also claim are a good forecasters for LTTR.

References

[1] Gasoline and Diesel Fuel Update,
https://www.eia.gov/petroleum/gasdiesel/

[2] DAT Data Analytics,
https://www.dat.com/industry-trends/data-analytics

[3] Rate Forecasting and Analytics in Practice,
https://www.dat.com/blog/post/rate-forecasting-and-analytics-in-practice

[4] Market Research:Business Industries and Consumers,
https://www.dat.com/blog/post/rate-forecasting-and-analytics-in-practice

[5] Shumway, Robert H and Stoffer, David S *Time Series Analysis and its Applications: with R examples*, 2017, Springer