Predicting Rail Freight by Carload: An Investigation into Time Series Model Construction

David Fecht

March 2023

1 Abstract

In this analysis, ARIMA model construction is explored on a data set analyzing rail freight carloads. Recorded on a monthly basis, and taken between the months of January 2000 through December 2022, 276 observations were recorded. With the model displaying seasonal trends, the data was respectively massaged for both seasonal and non-seasonal first differences. These procedures assisted in making the model stationary, with some outlying variance concerns still existing around the year 2020. When assessing the ACF and PACF plots of the differenced data, there showed no significant non-seasonal lags; however, there were significant seasonal lags at lag(12) in the ACF plot, and lag(12), lag(24), and lag(36) in the PACF plot. As such, applying parsimony to our model exploration, models with seasonal term Q = 1 were considered first before introducing additional terms P $\in$ {1,2,3}. For our non-seasonal terms, different combinations were explored that gave strong diagnostic results for normality of residuals, and also strong point estimates for parameters. Bringing in the best candidate models, six were chosen and reviewed for selection based on Akaike's Information Criterion, Akaike's Information Criterion (small sample adjusted), and Bayesian Information Criterion. In the end, an ARIMA $(1,1,1)$ x $(0,1,1)_{12}$ model was chosen based on the strength of its small number of variables and lowest BIC criterion amongst models with seasonal and non-seasonal terms. A forecast 12-months into the future from January 2023 through December 2023 was created. Layering the model over historical data, the fit was determined to be decent. Observations of the model showed that volatility around economic disruptions (pandemic, recession, monetary contraction, etc..), may recommend a different model framework using GARCH or "Long Memory" ARIMA.

2 Introduction

The dataset examines the frequency of train rail carloads incremented on a monthly basis that were transited from one destination to another across the United States. This data set was published by the United States Federal Reserve, and was available through their data repository, FRED (Federal Reserve Economic Data)[1].

Considering the impact of this data, samples of this sort tend to be consequential in predicting future economic events. Specifically, when we consider manufacturing and economic output, the delivery of raw and intermediate goods, and the frequency of which they are delivered via transit can help us make inference around the future of economic health. As a microcosm applied to the rail industry, this information can also be useful as a non-balance sheet metric for the demand of rail services, and a different perspective on how that can reflect around the profitability of firms in industry. Overall, there are multiplicative reasons as to how a time series investigation can help make strong decisions about how to allocate capital.

3 Stationary Procedures

Looking at the undifferenced data, we first explore the general trends of the data set to consider if any transformations are needed:

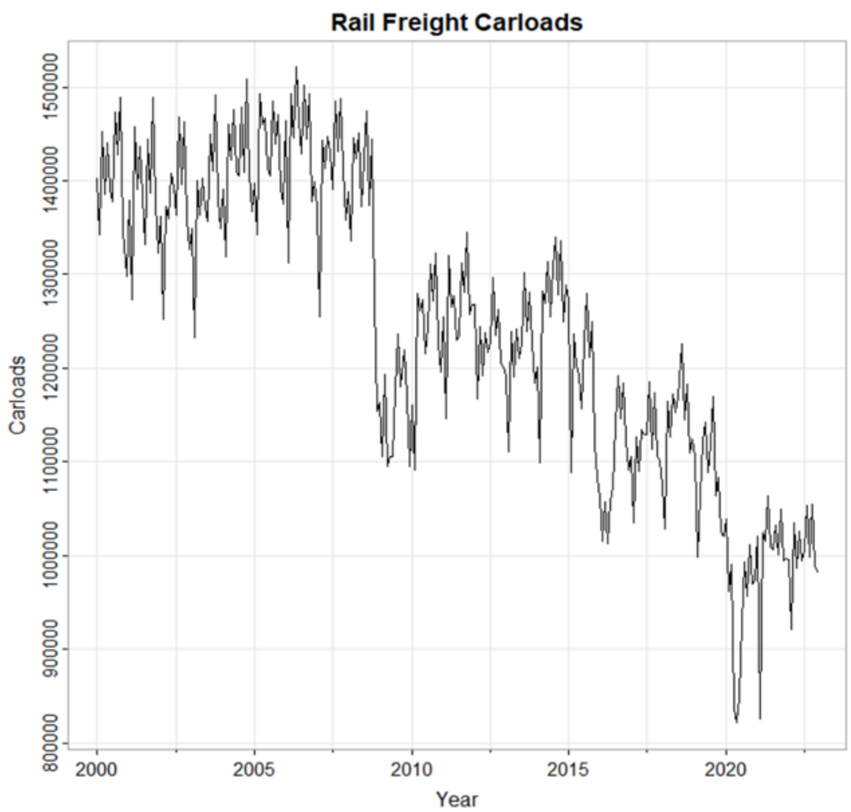

Figure 1: (Undifferenced Data)

The continuous trend follows a seasonal pattern with major disruptions occurring around 2008 (Great Recession) and 2020 (COVID-19 Pandemic). Outside of the seasonal trend, the data over the interval resembles what could be described as a decreasing step-wise trend; where outside the years of considerable volatility, the seasonal pattern looks constant. As such, the data does not resemble any similarities to higher order functions where transformations would be recommended. Later, the periods of volatility may allow an opportunity to try a different modeling framework with the desire to improve constant variance.

With no manipulation of the data needed, differencing of both the seasonal and non-seasonal portions of the data can be completed to ensure the model is stationary. This is achieved by ensuring a constant mean and constant variance of the data to build the ARIMA model.

Taking the first difference of the seasonal part of the data, $(1 - B^{12})X_t$, we have:

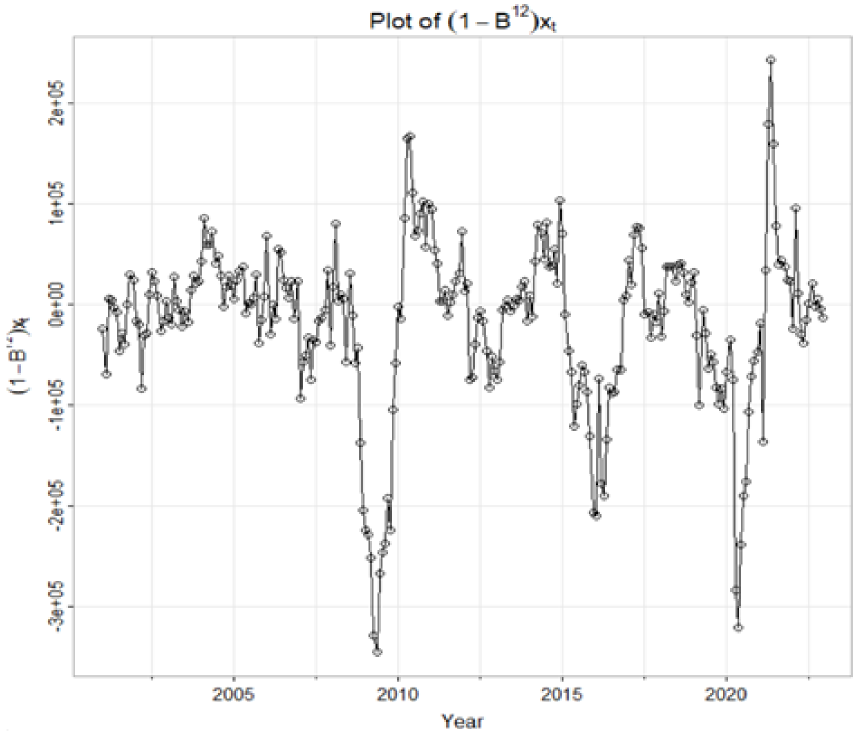

Figure 2: Seasonal First-Differenced Data

Introducing another difference to the non-seasonal part of the collected data, $(1-B)(1-B^{12})X_t$, we have:

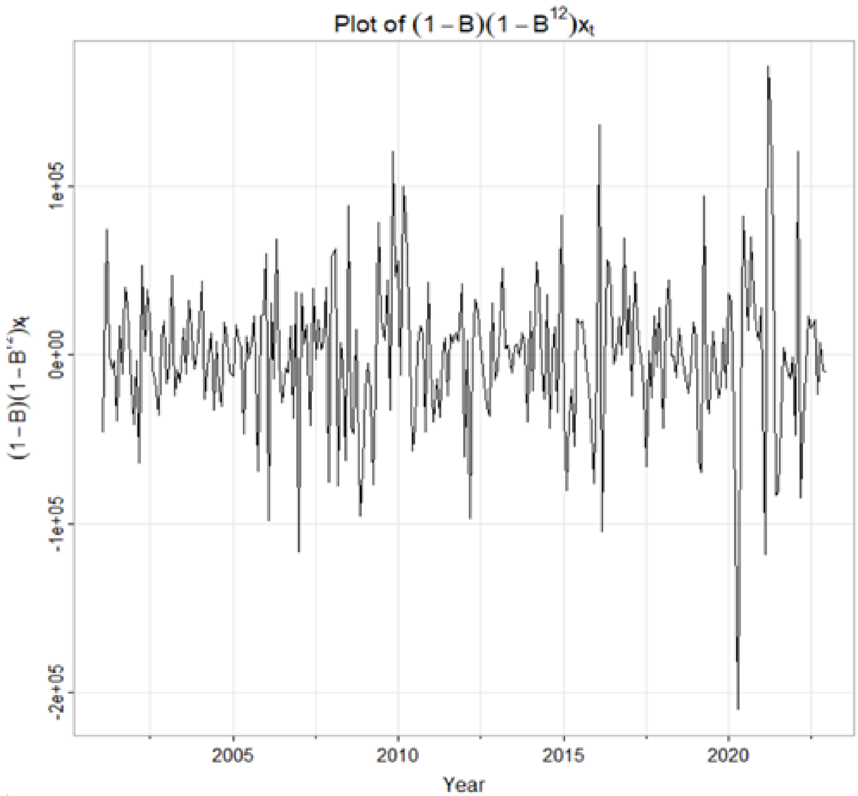

Figure 3: First Seasonal and Non-Seasonal Differences Applied Data

After taking the first differences of both the seasonal and non-seasonal parts of the data, we are able to determine from (Figure 3) that the data is now stationary. The mean is now constant, and the variance is roughly constant with some volatility prevailing around 2020.

The data is now ready for model construction determined by the ACF and PACF plots below:

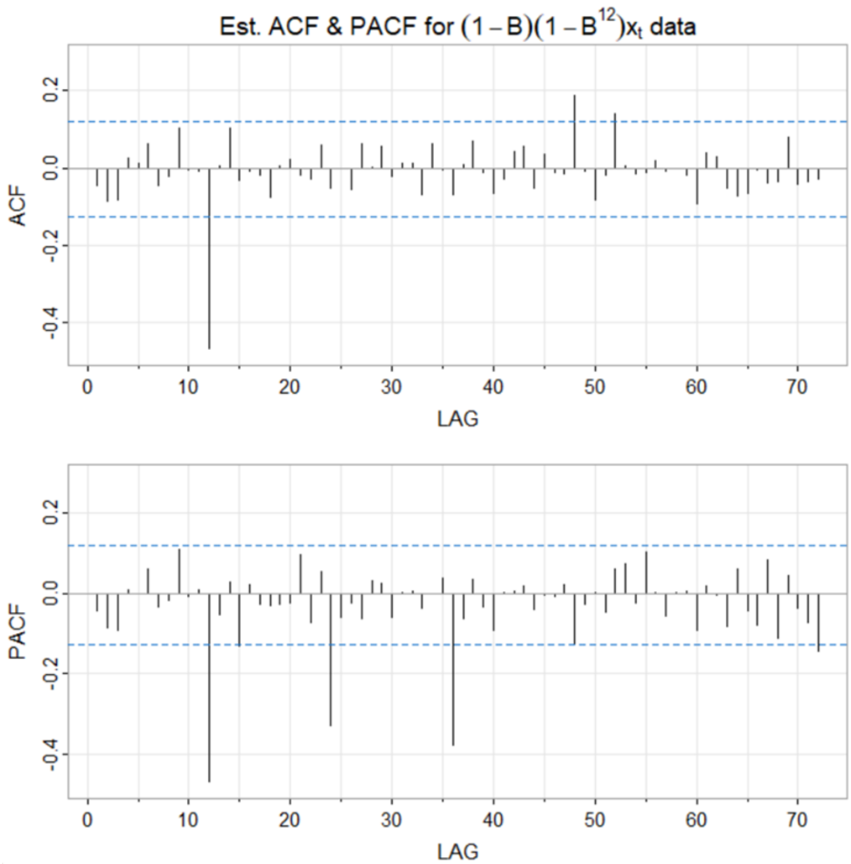

Figure 4: ACF and PACF of Seasonal and Non-Seasonal First Differences

As we assess the ACF and PACF plots in (Figure 4), there are no significant non-seasonal correlations at any lag. Specifically, both the ACF and PACF have no significance at lag(1) where there is a cut off or tail off convergence to 0. Examining the seasonal correlations at different lags, we can see that the ACF shows significant correlation at lag(12), and the PACF shows significant correlation at lag (12), lag(24), and lag (36). While the ACF also has a significant correlation at lag(48) and the PACF at lag(72), these occurrences have preceding lags($n\times12$) that are not significant; hence lacking applicability in model construction. Since all significant usable lags appear to be multiples of 12, we can use this consistency as a basis to determine likely candidates for model construction.

Applying the principal of parsimony, it would make sense to choose a model with the fewest parameters possible. Since we know there are no significant non-seasonal lags, we will have to try different combinations to see if any non-seasonal terms yield statistical significance in their parameters. When considering seasonal terms, we know with the ACF having significance at lag(12) that we could use the term Q=1. Similarly, the PACF showed significance at lag(12), lag(24), and lag(36). This would allow potential terms in the set P $\in$ [1,2,3].

To start with the least terms, model constructions in the form of ARIMA $(p,1,q)$ x $(0,1,1)_{12}$ will be considered first. From there, to improve the pool of candidates, we can introduce strong seasonal AR terms where we can try ARIMA $(p,1,q)$ x $(1,1,0)_{12}$; ARIMA $(p,1,q)$ x $(2,1,0)_{12}$; and ARIMA $(p,1,q)$ x $(3,1,0)_{12}$; ARIMA $(p,1,q)$ x $(1,1,1)_{12}$; ARIMA $(p,1,q)$ x $(2,1,1)_{12}$; and ARIMA $(p,1,q)$ x $(3,1,1)_{12}$.

4 Model Specification - $(p,1,q)$ x $(0,1,1)_{12}$

Starting with ARIMA $(0,1,0)$ X $(0,1,1)_{12}$, we have the following diagnostics:

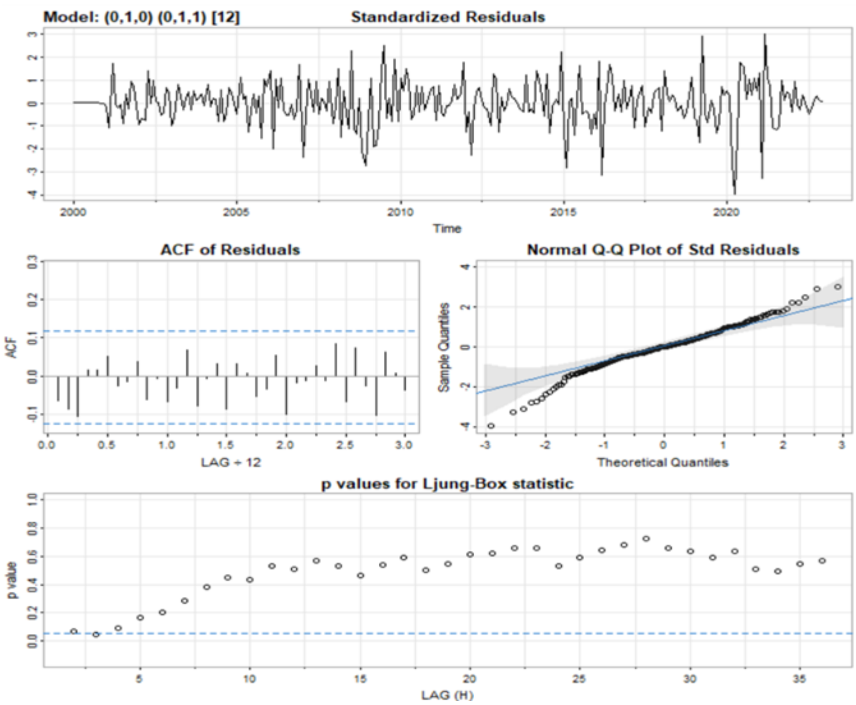

Figure 5: ARIMA $(0,1,0)$ x $(0,1,1)_{12}$ Diagnostics

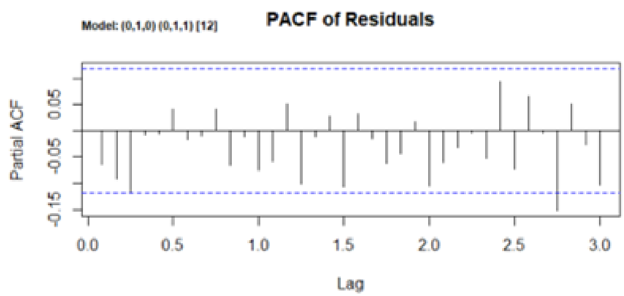

Figure 6: ARIMA $(0, 1, 0)$ x $(0, 1, 1)_{12}$ PACF of Residuals

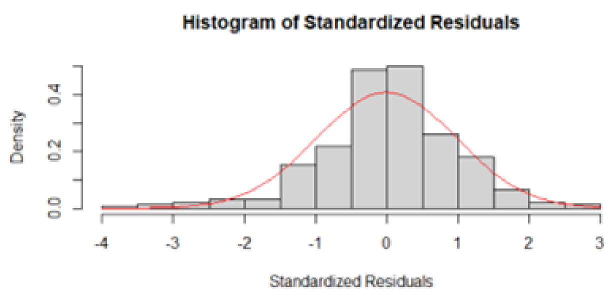

Figure 7: ARIMA $(0, 1, 0)$ x $(0, 1, 1)_{12}$ Standardized Residuals

```
     Estimate     SE t.value p.value
sma1  -0.8253 0.0364 -22.655       0
```

Figure 8: ARIMA $(0, 1, 0)$ x $(0, 1, 1)_{12}$ Parameter Estimates

Examining the first model in our selection process (Figures 5-8), we choose a model that has no non-seasonal terms. Reviewing residual diagnostics, majority of standardized residuals look to be within 3 standard deviations; the Ljung-Box Statistic shows independence of residual autocorrelations with the exception of lag(3); the ACF of residuals shows no significant correlation; and the PACF of residuals shows no significant correlation except at lag(28) (not in model). For normality properties, the histogram of standardized residuals shows normal approximation, and the Q-Q-Plot looks strong with the exception of a few outliers.

The parameter estimate of this model, Θ_1, has a p-value ≈ 0 at -0.8253.

Overall, this model could be considered a candidate for final selection given the strong diagnostics and parameter estimate.

Next in our analysis, we consider ARIMA $(1, 1, 0)$ x $(0, 1, 1)_{12}$:

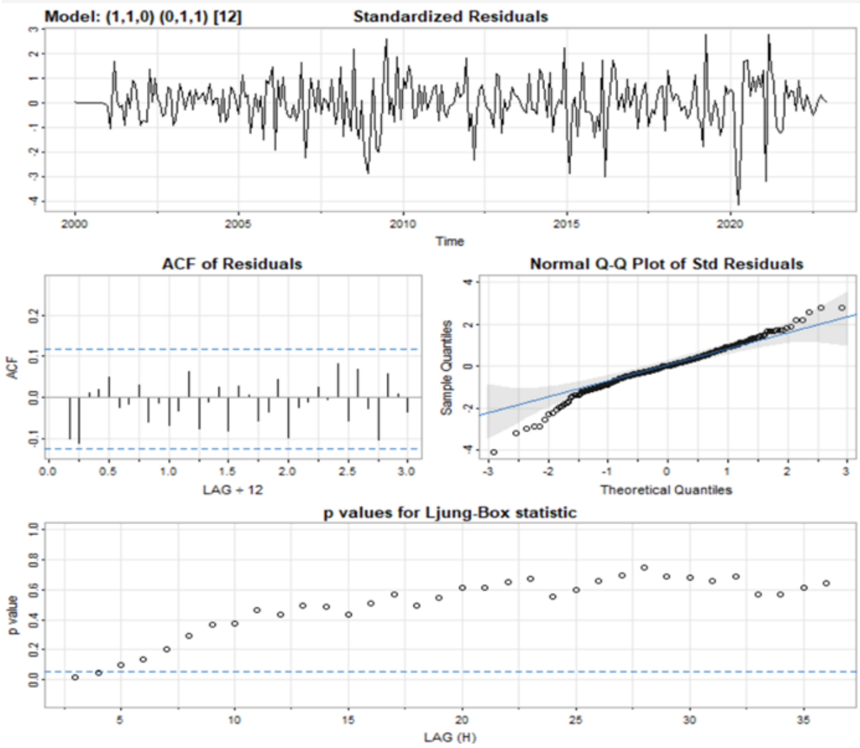

Figure 9: ARIMA $(1, 1, 0)$ x $(0, 1, 1)_{12}$ Diagnostics

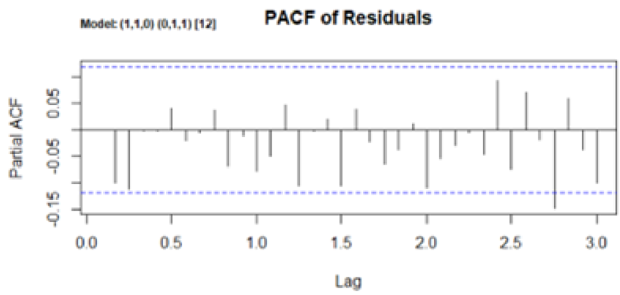

Figure 10: ARIMA $(1,1,0)$ x $(0,1,1)_{12}$ PACF of Residuals

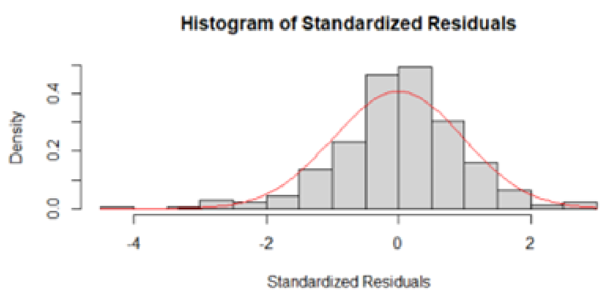

Figure 11: ARIMA $(1,1,0)$ x $(0,1,1)_{12}$ Standardized Residuals

	Estimate	SE	t.value	p.value
ar1	-0.0696	0.0616	-1.1302	0.2594
sma1	-0.8265	0.0364	-22.7186	0.0000

Figure 12: ARIMA $(1,1,0)$ x $(0,1,1)_{12}$ Parameter Estimates

Reviewing the second model in our selection process (Figures 9-12), we choose a model that has a single non-seasonal AR term while keeping our $(0,1,1)_{12}$ seasonal component. Reviewing residual diagnostics, majority of standardized residuals look to be within 3 standard deviations; the Ljung-Box Statistic shows dependence of residual autocorrelations at the earlier lags; the ACF of residuals shows no significant correlation; and the PACF of residuals shows no significant correlation except at lag(28) (not in model). For normality properties, the histogram of standardized residuals shows normal approximation, and the Q-Q-Plot looks strong with the exception of a few outliers.

The seasonal parameter estimate of this model, Θ_1, has a p-value ≈ 0 at -0.8265. The non-seasonal parameter estimate, ϕ_1, has p-value ≈ 0.2594 at -0.0696.

In summary, this model would not be a good candidate since all parameter estimates are not significant and the Ljung-Box Statistic shows instances of residual dependence.

Moving on to the third model, ARIMA $(0, 1, 1)$ x $(0, 1, 1)_{12}$, we have:

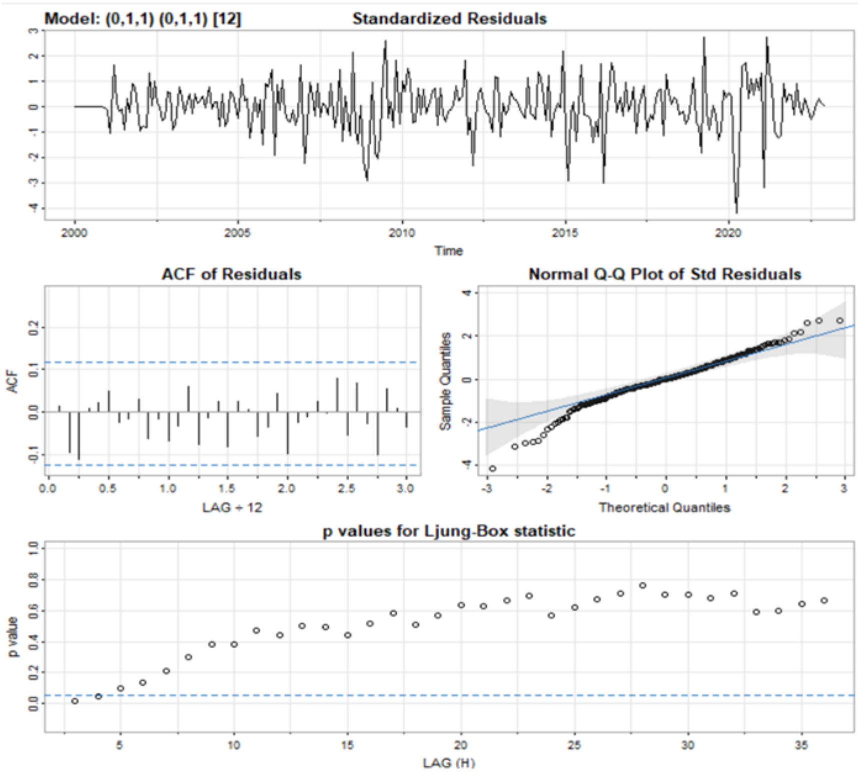

Figure 13: ARIMA $(0, 1, 1)$ x $(0, 1, 1)_{12}$ Diagnostics

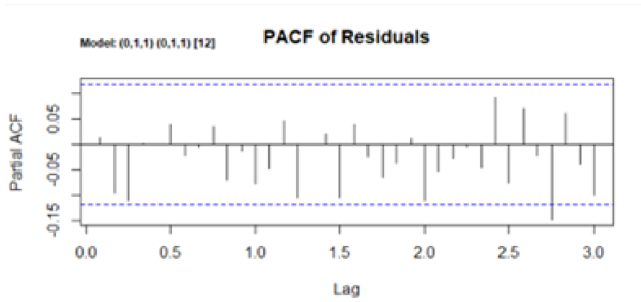

Figure 14: ARIMA $(0, 1, 1)$ x $(0, 1, 1)_{12}$ PACF of Residuals

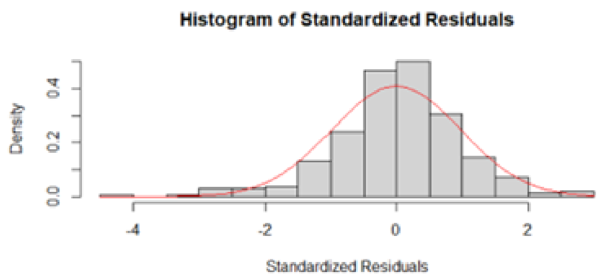

Figure 15: ARIMA $(0, 1, 1)$ x $(0, 1, 1)_{12}$ Standardized Residuals

	Estimate	SE	t.value	p.value
ma1	-0.0858	0.0691	-1.2411	0.2157
sma1	-0.8269	0.0364	-22.7265	0.0000

Figure 16: ARIMA $(0, 1, 1)$ x $(0, 1, 1)_{12}$ Parameter Estimates

Considering the third model in our selection process (Figures 13-16), we choose a model that has a single non-seasonal MA term while keeping our $(0, 1, 1)_{12}$ seasonal component. Reviewing residual diagnostics, many of the features are similar to the previous model. Majority of standardized residuals look to be within 3 standard deviations; the Ljung-Box Statistic shows dependence of residual autocorrelations at the earlier lags; the ACF of residuals shows no significant correlation; and the PACF of residuals shows no significant correlation except at lag(28) (not in model). For normality properties, the histogram of standardized residuals shows normal approximation, and the Q-Q-Plot looks strong with the exception of a few outliers.

The seasonal parameter estimate of this model, Θ_1, has a p-value ≈ 0 at -0.8269. The non-seasonal parameter estimate, θ_1, has p-value ≈ 0.2157 at -0.0858.

Determining validity, this model would not be a good candidate since all parameter estimates are not significant and the Ljung-Box Statistic shows instances of residual dependence.

Arriving at the fourth model, ARIMA $(1,1,1)$ x $(0,1,1)_{12}$, we have:

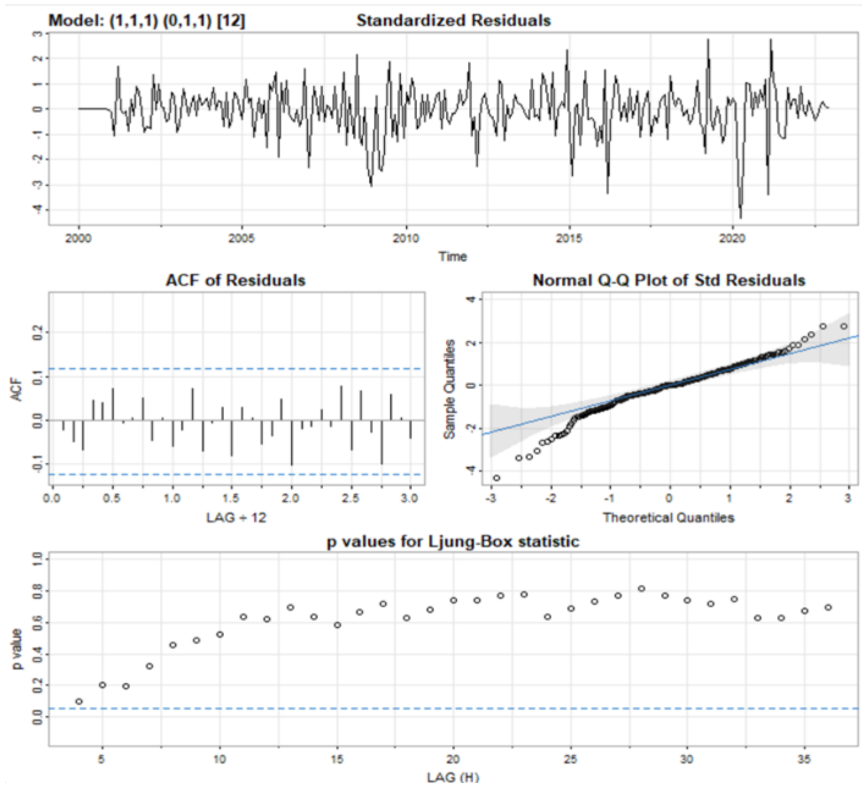

Figure 17: ARIMA $(1,1,1)$ x $(0,1,1)_{12}$ Diagnostics

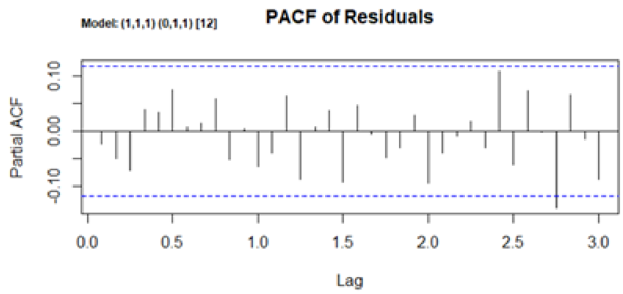

Figure 18: ARIMA $(1, 1, 1)$ x $(0, 1, 1)_{12}$ PACF of Residuals

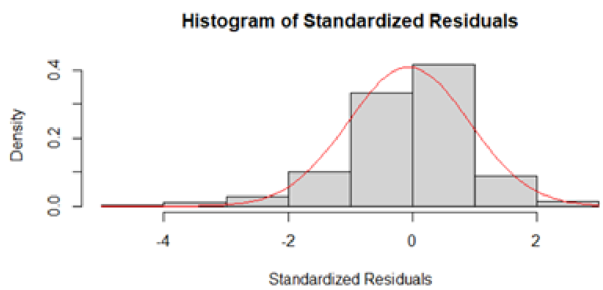

Figure 19: ARIMA $(1, 1, 1)$ x $(0, 1, 1)_{12}$ Standardized Residuals

	Estimate	SE	t.value	p.value
ar1	0.9060	0.0350	25.8563	0
ma1	-0.9955	0.0441	-22.5788	0
sma1	-0.8203	0.0378	-21.6977	0

Figure 20: ARIMA $(1, 1, 1)$ x $(0, 1, 1)_{12}$ Parameter Estimates

Studying the fourth model in our selection process (Figures 17-20), we choose a model that has two non-seasonal terms, ϕ_1 and θ_1, while keeping our $(0, 1, 1)_{12}$ seasonal component. Reviewing residual diagnostics, majority of standardized residuals look to be within 3 standard deviations; the Ljung-Box Statistic shows independence of residual autocorrelations; the ACF of residuals shows no significant correlation; and the PACF of residuals shows no significant correlation except at lag(28) (not in model). For normality properties, the histogram of standardized residuals shows normal approximation; however, the distribution is slightly skewed to the right, and the Q-Q-Plot looks strong with the exception of a few outliers.

The seasonal parameter estimate of this model, Θ_1, has a p-value ≈ 0 at -0.8203. For the non-seasonal parameter estimate, θ_1, we have a p-value $\approx$ 0.0 at -0.9955. For non-seasonal parameter estimate, ϕ_1, we have a p-value ≈ 0 at 0.9060.

All-around, this model is the strongest candidate model so far with the seasonal component $(0, 1, 1)_{12}$. The diagnostics are strong, and the parameter estimates are all significant.

Exploring other models keeping the $(0, 1, 1)_{12}$ seasonal component in place, we can examine other diagnostics by adding additional non-seasonal terms for analysis. First, by increasing our p term to 2 or 3, we have the following:

	Estimate	SE	t.value	p.value
ar1	0.4773	0.3022	1.5796	0.1154
ar2	-0.0652	0.0764	-0.8541	0.3939
ma1	-0.5628	0.2993	-1.8802	0.0612
sma1	-0.8322	0.0366	-22.7068	0.0000

Figure 21: ARIMA $(2, 1, 1)$ x $(0, 1, 1)_{12}$ Diagnostics

	Estimate	SE	t.value	p.value
ar1	0.0419	0.4823	0.0870	0.9308
ar2	-0.0836	0.0713	-1.1723	0.2421
ar3	-0.1034	0.0777	-1.3302	0.1846
ma1	-0.1286	0.4846	-0.2655	0.7909
sma1	-0.8351	0.0369	-22.6307	0.0000

Figure 22: ARIMA $(3, 1, 1)$ x $(0, 1, 1)_{12}$ Diagnostics

As seen in both figures above, the additional AR terms do not yield any significant p-values that would be useful for model construction. Like we have seen before; however, the seasonal MA term still has a low p-value ≈ 0. Since not all estimates are significant, we cannot use these models as candidates for our final model.

Similar to increasing our AR values, we can simulate higher MA values by increasing our q term to 2 or 3. The results of which are seen below:

	Estimate	SE	t.value	p.value
ar1	0.4684	0.3934	1.1907	0.2349
ma1	-0.5570	0.3909	-1.4249	0.1554
ma2	-0.0553	0.0826	-0.6694	0.5038
sma1	-0.8320	0.0366	-22.7176	0.0000

Figure 23: ARIMA $(1, 1, 2)$ x $(0, 1, 1)_{12}$ Diagnostics

	Estimate	SE	t.value	p.value
ar1	-0.1721	0.8585	-0.2005	0.8412
ma1	0.0861	0.8546	0.1007	0.9199
ma2	-0.1023	0.1077	-0.9499	0.3431
ma3	-0.1051	0.0863	-1.2177	0.2244
sma1	-0.8333	0.0374	-22.3089	0.0000

Figure 24: ARIMA $(1, 1, 3)$ x $(0, 1, 1)_{12}$ Diagnostics

Similar to the previous results increasing the AR values, increasing the MA values do not lead to any significant results. None of the models can be used for model construction given the high p-values of the non-seasonal MA estimates.

After further manipulation of non-seasonal terms, there was one other model that was able to produce decent diagnostics and strong parameter estimates.

In the fifth model, ARIMA $(4,1,1)$ x $(0,1,1)_{12}$, we have:

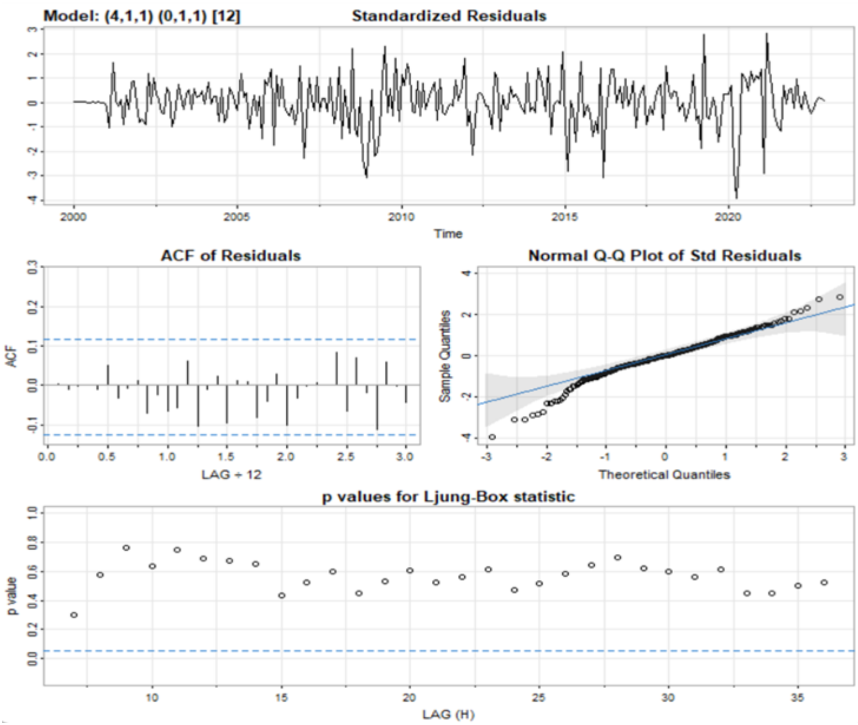

Figure 25: ARIMA $(4,1,1)$ x $(0,1,1)_{12}$ Diagnostics

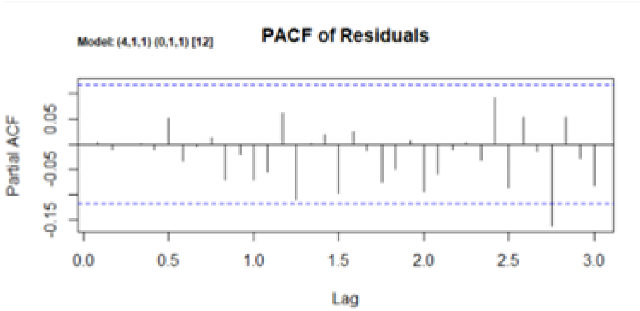

Figure 26: ARIMA $(4, 1, 1)$ x $(0, 1, 1)_{12}$ PACF of Residuals

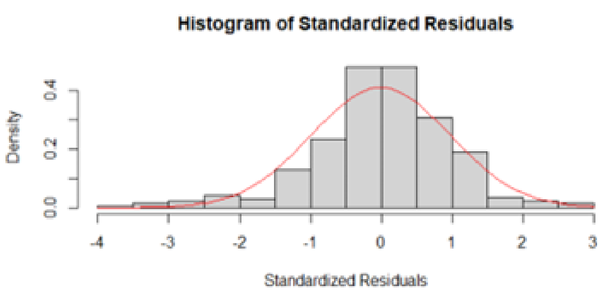

Figure 27: ARIMA $(4, 1, 1)$ x $(0, 1, 1)_{12}$ Standardized Residuals

	Estimate	SE	t.value	p.value
ar1	-1.0684	0.0613	-17.4162	0.0000
ar2	-0.1775	0.0890	-1.9936	0.0472
ar3	-0.2056	0.0894	-2.3010	0.0222
ar4	-0.1221	0.0614	-1.9896	0.0477
ma1	1.0000	0.0153	65.2873	0.0000
sma1	-0.8307	0.0374	-22.2271	0.0000

Figure 28: ARIMA $(4,1,1)$ x $(0,1,1)_{12}$ Parameter Estimates

Surveying the fifth model in our selection process (Figures 25-28), we choose a model that has 5 non-seasonal terms, ϕ_1, ϕ_2, ϕ_3, ϕ_4 and θ_1, while keeping our $(0,1,1)_{12}$ seasonal component. Reviewing residual diagnostics, majority of standardized residuals look to be within 3 standard deviations; the Ljung-Box Statistic shows independence of residual auto-correlations; the ACF of residuals shows no significant correlation; and the PACF of residuals shows no significant correlation except at lag(28) (not in model). For normality properties, the histogram of standardized residuals shows normal approximation with no skewness, and the Q-Q-Plot looks strong with the exception of a few outliers.

The seasonal parameter estimate of this model, Θ_1, has a p-value ≈ 0 at -0.8307. For the non-seasonal parameter estimates: θ_1 has a p-value ≈ 0 at 1.0; ϕ_1 has a p-value ≈ 0 at -1.0684; ϕ_2 has a p-value ≈ 0.0472 at -0.1775; ϕ_3 has a p-value ≈ 0.0222 at -0.2056; and ϕ_4 has a p-value ≈ 0.0477 at -0.1221.

Compared to the other candidates, this model is one of the strongest candidates with ideal diagnostics and significant parameter estimates.

5 Model Specification - Other Model Formats

Up to this point, we have been exclusively examining models in the form of ARIMA $(p,1,q)$ x $(0,1,1)_{12}$. While this step was initially chosen to render the simplest model construction given the recommended seasonal correlations in our ACF plot, we also know that we have strong seasonal correlations in the PACF plot that would allow us to try models with higher seasonal AR terms. Specifically, keeping our $Q \in$ [0,1], we can also try $P \in$ [1,2,3]. This can be substantiated given we had significance at lag(12), lag(24), and lag(36) in our PACF plot.

Examining a few different models, we are given the following parameter estimates:

	Estimate	SE	t.value	p.value
ar1	-0.065	0.0616	-1.0546	0.2926
sar1	-0.484	0.0540	-8.9667	0.0000

Figure 29: ARIMA $(1,1,0)$ x $(1,1,0)_{12}$ Parameter Estimates

	Estimate	SE	t.value	p.value
ma1	-0.0776	0.0678	-1.1453	0.2531
sar1	-0.4841	0.0539	-8.9746	0.0000

Figure 30: ARIMA $(0,1,1)$ x $(1,1,0)_{12}$ Parameter Estimates

	Estimate	SE	t.value	p.value
ar1	0.9098	0.0347	26.1870	0.0000
ma1	-0.9953	0.0397	-25.0706	0.0000
sar1	-0.0562	0.0723	-0.7770	0.4379
sma1	-0.8036	0.0447	-17.9661	0.0000

Figure 31: ARIMA $(1,1,1)$ x $(1,1,1)_{12}$ Parameter Estimates

Analyzing the results above, while the ARIMA $(1,1,0)$ x $(1,1,0)_{12}$ and ARIMA $(0,1,1)$ x $(1,1,0)_{12}$ had no significant non-seasonal parameter estimates, both models had a significant seasonal AR term where $P = 1$. When analyzing the ARIMA $(1,1,1)$ x $(1,1,1)_{12}$ model, the seasonal AR term where $P = 1$ was not significant. In this investigation, it was determined that by adding higher order seasonal AR terms that it may be possible to find stronger model candidates.

This was examined below by introducing such terms in the models below:

	Estimate	SE	t.value	p.value
ar1	0.7049	0.4724	1.4921	0.1369
ma1	-0.7959	0.4142	-1.9216	0.0558
sar1	-0.8435	0.0654	-12.9015	0.0000
sar2	-0.6946	0.0779	-8.9193	0.0000
sar3	-0.5226	0.0775	-6.7430	0.0000
sar4	-0.1091	0.0682	-1.5991	0.1110

Figure 32: ARIMA $(1,1,1)$ x $(4,1,0)_{12}$ Parameter Estimates

	Estimate	SE	t.value	p.value
ma1	-0.0823	0.0680	-1.2106	0.2271
sar1	-0.7975	0.0558	-14.3050	0.0000
sar2	-0.6217	0.0645	-9.6369	0.0000
sar3	-0.4444	0.0598	-7.4361	0.0000

Figure 33: ARIMA $(1,1,1)$ x $(3,1,0)_{12}$ Parameter Estimates

As seen in (Figure 32) and (Figure 33), increasing our P value does indeed render more significant parameter estimates. While some of the non-seasonal terms had no significant parameter estimates, the results of this may change when removing and adding non-seasonal variables. With this discovery, the following models were selected as final candidates:

For our sixth model, ARIMA $(1, 1, 1)$ x $(2, 1, 0)_{12}$, we have:

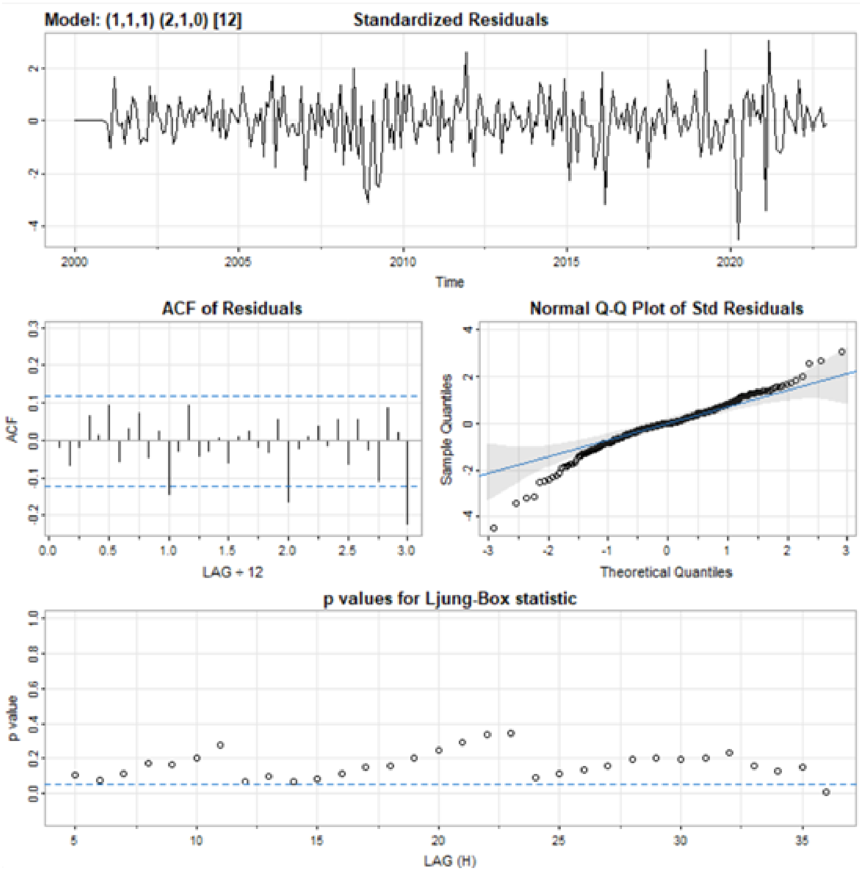

Figure 34: ARIMA $(1, 1, 1)$ x $(2, 1, 0)_{12}$ Diagnostics

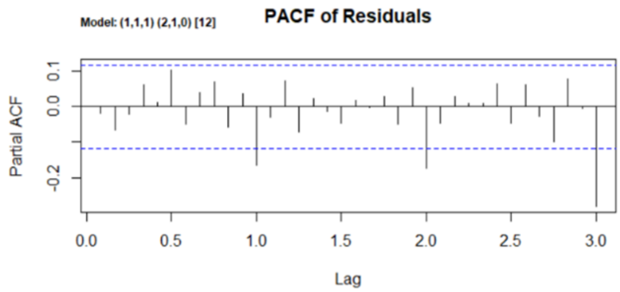

Figure 35: ARIMA $(1, 1, 1)$ x $(2, 1, 0)_{12}$ PACF of Residuals

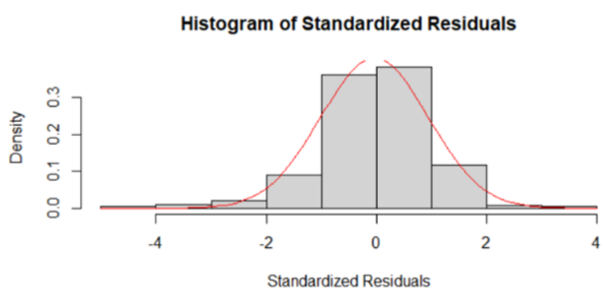

Figure 36: ARIMA $(1, 1, 1)$ x $(2, 1, 0)_{12}$ Standardized Residuals

	Estimate	SE	t.value	p.value
ar1	0.8999	0.0287	31.3508	0
ma1	-1.0000	0.0137	-72.8110	0
sar1	-0.6295	0.0583	-10.7928	0
sar2	-0.3639	0.0613	-5.9343	0

Figure 37: ARIMA $(1,1,1)$ x $(2,1,0)_{12}$ Parameter Estimates

Investigating the sixth model in our selection process (Figures 34-37), we choose a model that has 2 non-seasonal terms, ϕ_1 and θ_1, and 2 seasonal terms, Φ_1 and Φ_2. Reviewing residual diagnostics, majority of standardized residuals look to be within 3 standard deviations; the Ljung-Box Statistic shows independence of residual autocorrelations with the exception of some borderline cases, and a dependency at lag(36); the ACF of residuals shows no significant correlation with the exceptions of lag(10), lag(20) and lag(30); and the PACF of residuals shows no significant correlation except at lag(10), lag(20), and lag(30) as well. For normality properties, the histogram of standardized residuals shows normal approximation with skewness to the right, and the Q-Q-Plot looks strong with the exception of a few outliers.

The seasonal parameter estimates, we have the following: Φ_1 has a p-value ≈ 0 at -0.6295, and Φ_2 has a p-value ≈ 0 at -0.3639. For non-seasonal parameter estimates: ϕ_1 has a p-value ≈ 0 at 0.8999, and θ_1 has a p-value ≈ 0 at -1.0.

Juxtaposed against other candidates, this model is decent in terms of its parameter estimates; however, it is borderline in its residual diagnostics given some dependency issues.

Inspecting the seventh model, ARIMA $(1, 1, 1)$ x $(3, 1, 0)_{12}$, we have:

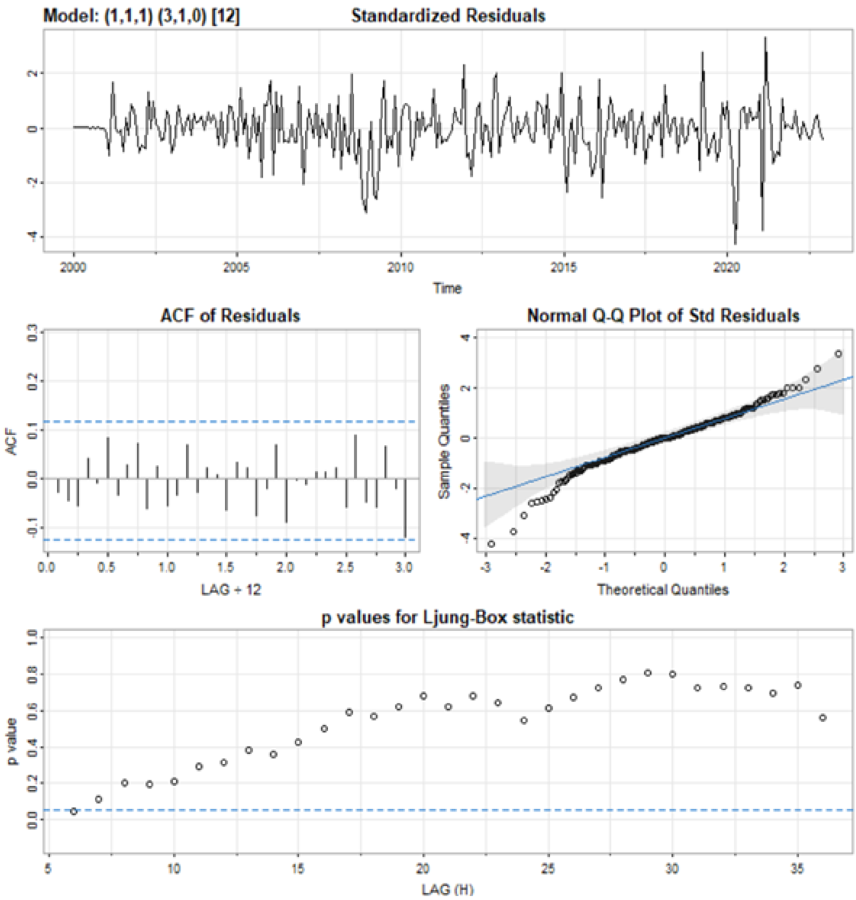

Figure 38: ARIMA $(1, 1, 1)$ x $(3, 1, 0)_{12}$ Diagnostics

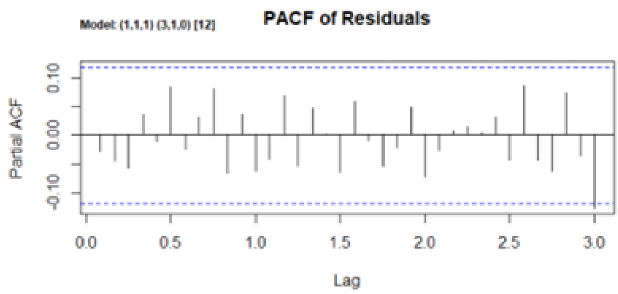

Figure 39: ARIMA $(1,1,1)$ x $(3,1,0)_{12}$ PACF of Residuals

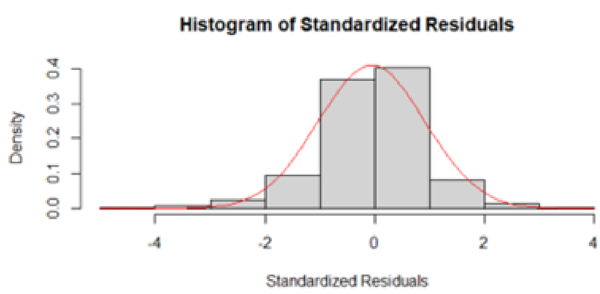

Figure 40: ARIMA $(1,1,1)$ x $(3,1,0)_{12}$ Standardized Residuals

	Estimate	SE	t.value	p.value
ar1	0.9207	0.0265	34.6866	0
ma1	-0.9999	0.0282	-35.3972	0
sar1	-0.7798	0.0566	-13.7840	0
sar2	-0.6057	0.0650	-9.3175	0
sar3	-0.4328	0.0603	-7.1804	0

Figure 41: ARIMA $(1, 1, 1)$ x $(3, 1, 0)_{12}$ Parameter Estimates

Qualifying the seventh model in our selection process (Figures 38-41), we choose a model that has 2 non-seasonal terms, ϕ_1 and θ_1, and 3 seasonal terms, Φ_1, Φ_2, and Φ_3. This model looks to have slightly better diagnostics than its predecessor. Reviewing residual diagnostics, majority of standardized residuals look to be within 3 standard deviations; the Ljung-Box Statistic shows independence of residual autocorrelations with the exception of lag(6); the ACF of residuals shows no significant correlations; and the PACF of residuals shows no significant correlation with the exception of lag(30) . For normality properties, the histogram of standardized residuals shows normal approximation with skewness to the right, and the Q-Q-Plot looks strong with the exception of a few outliers.

The seasonal parameter estimates, we have the following: Φ_1 has a p-value ≈ 0 at -0.7798, Φ_2 has a p-value ≈ 0 at -0.6057, and Φ_3 has a p-value ≈ 0 at -0.4328 . For non-seasonal parameter estimates: ϕ_1 has a p-value ≈ 0 at 0.9207, and θ_1 has a p-value ≈ 0 at -0.9999.

All factors considered, this model is decent in terms of its parameter estimates, and has stronger residual diagnostics than the preceding model. With that said, there diagnostics are not ideal, but better than most.

Concluding with the last model before selection, ARIMA $(1, 1, 1)$ x $(3, 1, 1)_{12}$, we have:

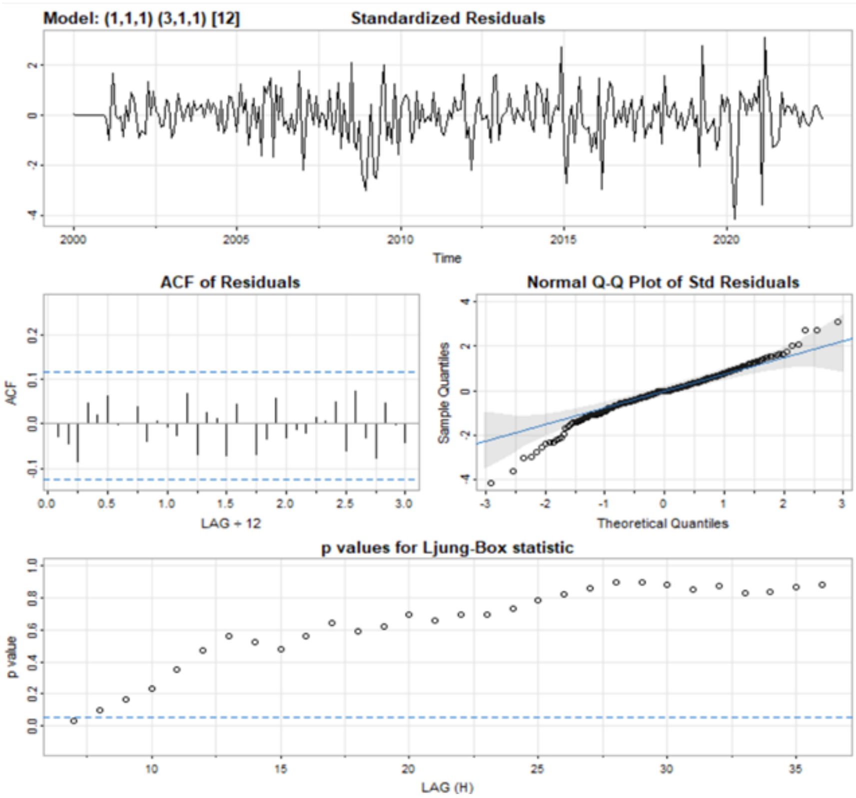

Figure 42: ARIMA $(1, 1, 1)$ x $(3, 1, 1)_{12}$ Diagnostics

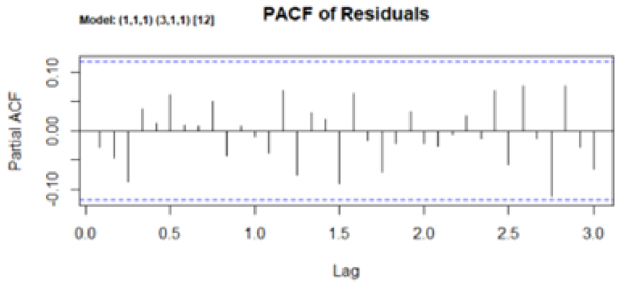

Figure 43: ARIMA $(1,1,1)$ x $(3,1,1)_{12}$ PACF of Residuals

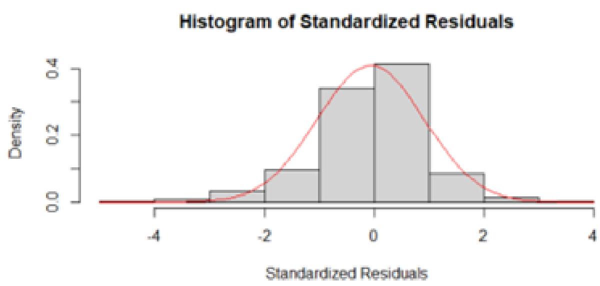

Figure 44: ARIMA $(1,1,1)$ x $(3,1,1)_{12}$ Standardized Residuals

	Estimate	SE	t.value	p.value
ar1	0.9232	0.0354	26.0628	0.0000
ma1	-0.9920	0.0254	-39.0226	0.0000
sar1	-0.3206	0.1176	-2.7262	0.0068
sar2	-0.3106	0.0952	-3.2616	0.0013
sar3	-0.2415	0.0890	-2.7142	0.0071
sma1	-0.5693	0.1161	-4.9016	0.0000

Figure 45: ARIMA $(1,1,1)$ x $(3,1,1)_{12}$ Parameter Estimates

Vetting the eighth model in our selection process (Figures 42-45), we choose a model that has 2 non-seasonal terms, ϕ_1 and θ_1, and 4 seasonal terms, Φ_1, Φ_2, Φ_3 and Θ_1. Reviewing residual diagnostics, majority of standardized residuals look to be within 3 standard deviations; the Ljung-Box Statistic shows independence of residual autocorrelations with the exception of lag(6); the ACF of residuals shows no significant correlations; and the PACF of residuals shows no significant correlations either. For normality properties, the histogram of standardized residuals shows normal approximation with skewness to the right, and the Q-Q-Plot looks strong with the exception of a few outliers.

The seasonal parameter estimates, we have the following: Φ_1 has a p-value $\approx$ 0.0068 at -0.3206, Φ_2 has a p-value $\approx$ 0.0013 at -0.3106, Φ_3 has a p-value $\approx$ 0.0071 at -0.2415, and Θ_1 has a p-value ≈ 0 at -0.5693 . For non-seasonal parameter estimates: ϕ_1 has a p-value ≈ 0 at 0.9232, and θ_1 has a p-value ≈ 0 at -0.9920.

Rating comparative strength, this model is decent in terms of its parameter estimates, and has decent residual diagnostics. With that said, there diagnostics are not ideal, but better than the preceding two models.

6 Fitting and Model Selection

After completing our analysis of residuals and parameter estimates, we are ready to verify our final model based on the strongest candidates explored. They are as follows:

$$\begin{gathered}\text{ARIMA } (0,1,0) \text{ x } (0,1,1)_{12}\\ \text{ARIMA } (1,1,1) \text{ x } (0,1,1)_{12}\\ \text{ARIMA } (4,1,1) \text{ x } (0,1,1)_{12}\\ \text{ARIMA } (1,1,1) \text{ x } (2,1,0)_{12}\\ \text{ARIMA } (1,1,1) \text{ x } (3,1,0)_{12}\\ \text{ARIMA } (1,1,1) \text{ x } (3,1,1)_{12}\end{gathered}$$

With six potential models listed, we now compare their estimates of prediction error using AIC, AICc, and BIC. We know that the ideal models will have the lowest values, and if possible, the lowest number of parameters. The results yielded the following:

	mod.name	AIC	AICc	BIC
1	ARIMA(0,1,0)x(0,1,1)_12	23.72799	23.72805	23.75516
2	ARIMA(1,1,1)x(0,1,1)_12	23.71166	23.71201	23.76598
3	ARIMA(4,1,1)x(0,1,1)_12	23.73504	23.73629	23.83012
4	ARIMA(1,1,1)x(2,1,0)_12	23.86590	23.86649	23.93381
5	ARIMA(1,1,1)x(3,1,0)_12	23.70701	23.70790	23.78851
6	ARIMA(1,1,1)x(3,1,1)_12	23.68614	23.68738	23.78121

Figure 46: Criterion Selection Process

Looking at the results above, ARIMA $(1,1,1)$ x $(3,1,1)_{12}$ has the lowest AIC and AICc values. Similarly, both ARIMA $(0,1,0)$ x $(0,1,1)_{12}$ and ARIMA $(1,1,1)$ x $(0,1,1)_{12}$ have the lowest BIC values. Given that the models with lowest BIC values have fewer parameters than ARIMA $(1,1,1)$ x $(3,1,1)_{12}$, and the fact that this model had issues with its Ljung-Box independence, we will apply the principal of parsimony by narrowing down our model selection to ARIMA $(0,1,0)$ x $(0,1,1)_{12}$ and ARIMA $(1,1,1)$ x $(0,1,1)_{12}$.

With the decision now between two models, and while ARIMA $(0,1,0)$ x $(0,1,1)_{12}$ has the lower BIC values, ARIMA $(1,1,1)$ x $(0,1,1)_{12}$ has the lower AIC and AICc values. In addition, ARIMA $(1,1,1)$ x $(0,1,1)_{12}$ also has stronger independence of residuals compared with ARIMA $(0,1,0)$ x $(0,1,1)_{12}$.

Therefore, we will choose ARIMA $(1,1,1)$ x $(0,1,1)_{12}$ as our final model.

The final model equation is iterated below:

$$(1 - B^{12})(1 - B)X_t = \frac{(1 + \theta B)(1 + \Theta B^{12})}{(1 - \phi B)} W_t$$

$$(1 - \phi B)(1 - B^{12})(1 - B)X_t = (1 + \theta B)(1 + \Theta B^{12})W_t$$

$$(1 - \phi B)(1 - B - B^{12} + B^{13})X_t = (1 + \theta B)(1 + \Theta B^{12})W_t$$

$$(1 - B - B^{12} + B^{13} - \phi B + \phi B^2 + \phi B^{13} - \phi B^{14})X_t = (1 + \theta B + \Theta B^{12} + \theta\Theta B^{13})W_t$$

$$X_t - (1 + \phi)X_{t-1} + \phi X_{t-2} - X_{t-12} + (1 + \phi)X_{t-13} - \phi X_{t-14} = W_t + \theta W_{t-1} + \Theta W_{t-12} + \theta\Theta W_{t-13}$$

$$X_t = (1 + \phi)X_{t-1} - \phi X_{t-2} + X_{t-12} - (1 + \phi)X_{t-13} + \phi X_{t-14} + W_t + \theta W_{t-1} + \Theta W_{t-12} + \theta\Theta W_{t-13}$$

Here, $\phi = 0.9060$, $\theta = -0.9955$, and $\Theta = -0.8203$

With the model chosen, we will move to create our forecast. Since we have 23 years of data in our sample, and 276 observations in total, we want to be careful about how far into the future we predict frequency of rail freight. Specifically, it will make sense for our prediction interval to increase as our forecast goes deeper into the future, since there will be more uncertainty; but moreover, we want to ensure that the amount of time predicted is not unrealistic. Therefore, we will use a seasonal cycle and predict at most 12-months into the future to make a reasonable prediction.

7 Forecasting Rail Freight Carloads

The forecast of the ARIMA $(1,1,1)$ x $(0,1,1)_{12}$ is seen below:

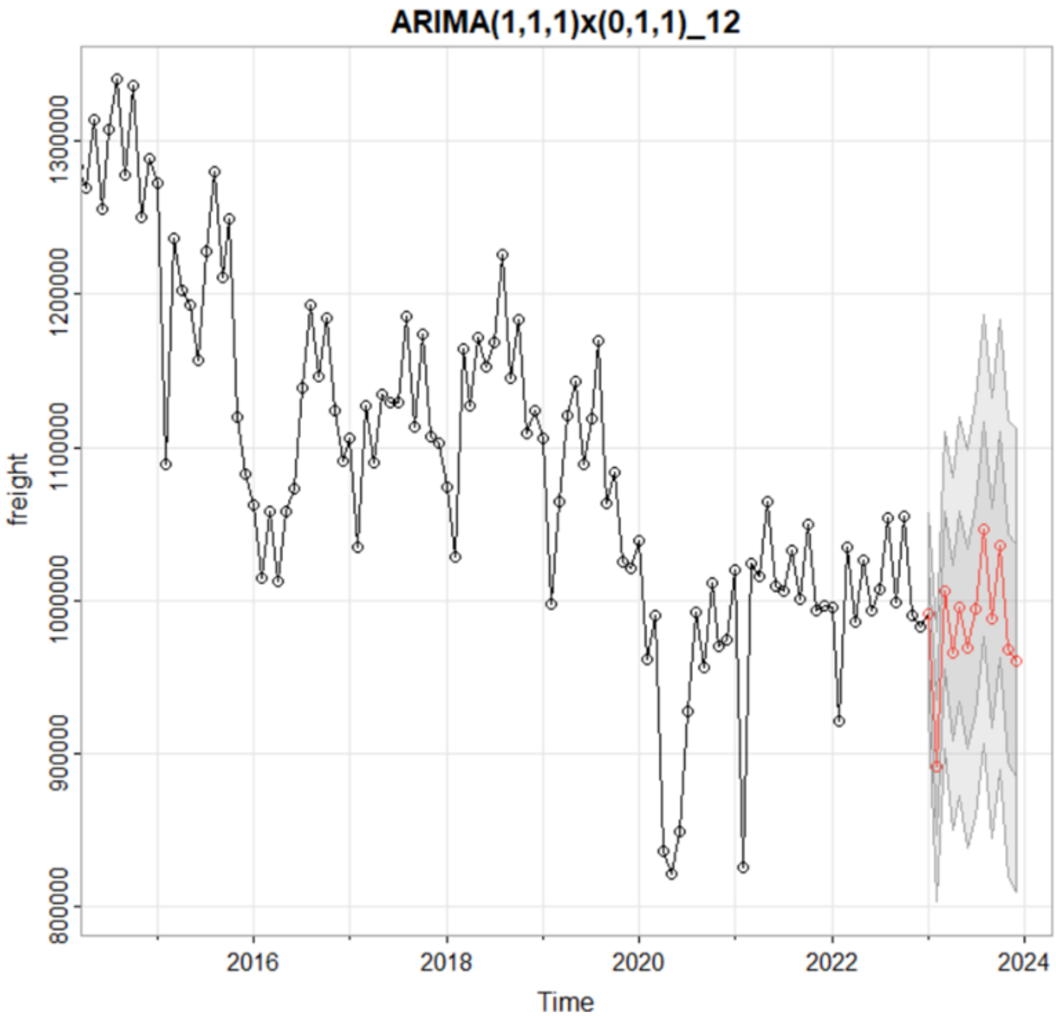

Figure 47: ARIMA $(1,1,1)$ x $(0,1,1)_{12}$ Forecast 12-Months

Overall, as seen in the forecast above (Figure 47), the predicted values follow a reasonable seasonal trend. It is noteworthy; however, to point out that that second predicted value dips considerably in the month of February 2023. No doubt, in part, this is due to the volatility trends we saw earlier in our data exploration around 2008 and 2020. The full listing of predicted values and 95% prediction intervals are iterated in the table below:

Month	Predicted Value	Standard Error	95% Prediction Interval
Jan-23	991,566.00	32,525.08	[927,816.84, 1,055,315.16]
Feb-23	891,276.10	44,035.15	[804,967.21, 977,584.99]
Mar-23	1,006,622.60	51,712.15	[905,266.79, 1,107,978.41]
Apr-23	965,549.20	57,342.59	[853,157.72, 1,077,940.68]
May-23	996,045.80	61,659.59	[875,193, 1,116,898.6]
Jun-23	968,697.00	65,057.29	[841,184.71, 1,096,209.29]
Jul-23	994,589.80	67,778.58	[861,743.78, 1,127,435.82]
Aug-23	1,046,905.10	69,985.91	[909,732.72, 1,184,077.48]
Sep-23	988,653.10	71,793.97	[847,936.92, 1,129,369.28]
Oct-23	1,036,392.70	73,286.78	[892,750.61, 1,180,034.79]
Nov-23	968,551.40	74,527.63	[822,477.25, 1,114,625.55]
Dec-23	960,947.80	75,565.16	[812,840.09, 1,109,055.51]

The seasonal trend is validated by the prediction numbers above. Despite the dip observed in February 2023, the model seems as one would expect over a year. The forecast predicts that in the months of March, August, and October the frequency of freight cars will be the highest, while they will be at their lowest in February and April.

The prediction intervals seem to remain roughly constant as the prediction goes further out. This is what we would want to see, given that this is a seasonal ARIMA model where seasonal differencing has been incremented by month. If we predicted further out, we would run the risk of the prediction interval growing.

Lastly, it makes sense to fit our model over historical data to see how well the prediction model observes the historical pattern.

Backfitting of the model is seen below:

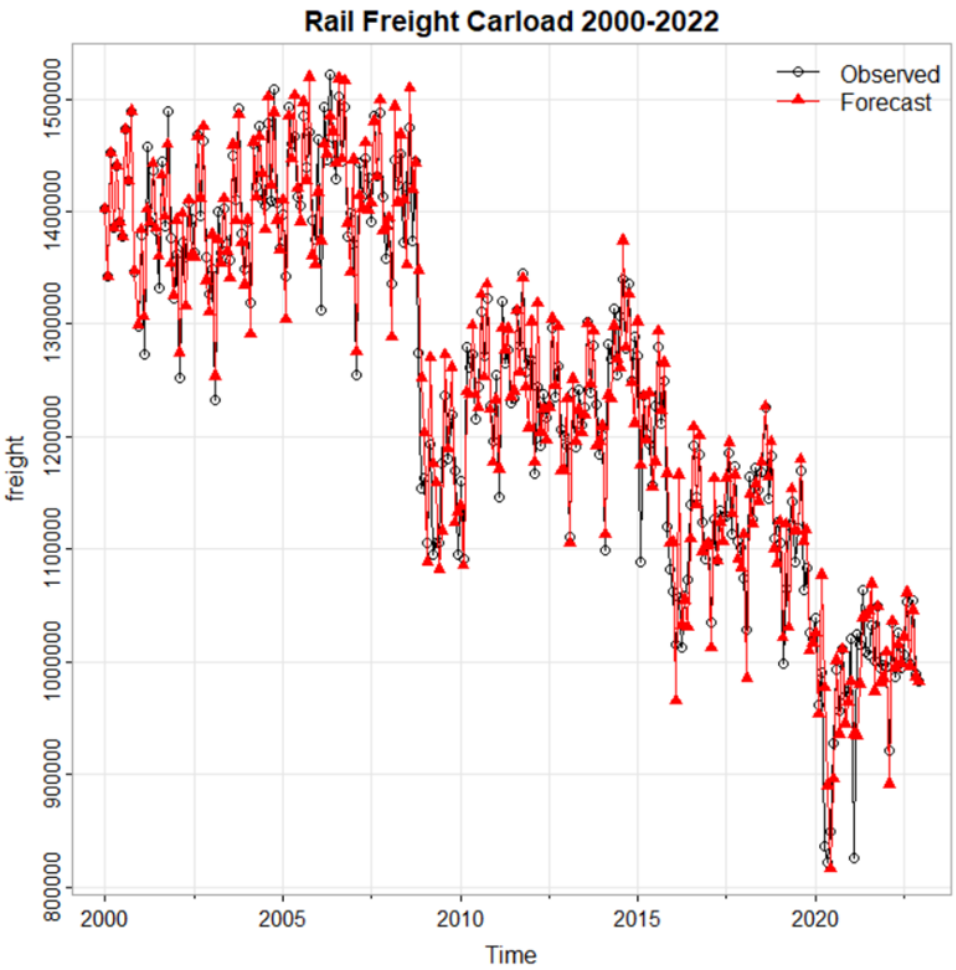

Figure 48: ARIMA $(1, 1, 1)$ x $(0, 1, 1)_{12}$ Forecast backfitted

As see above (Figure 48), the predicted model fits reasonably well over the historical data. While there may have been some accuracy issues around periods of strong volatility, it follows the seasonal trends in a fairly accurate fashion. Without the places of strong volatility, 2008 and 2020, the model would probably fit over the historical data better as there would be more predictability around monthly seasonal trends.

8 Discussion

As we can see from our model construction process, we we're successfully able to take the data set, make it stationary, and apply diagnostics of residuals and parameter estimates to create our 12-month rail freight forecast.

Applying information of the forecast to industry, I believe that our model gives some helpful insights that may be useful for financial planning. Specifically, given the drop-off in February, despite the fact that the predication may be inflated, it would be reasonable that with the frequency of rail freight travel decreasing, there may be supply-chain issues or a potential contraction in manufacturing output over Q1 of 2023. As a result, this would be useful in predicting the equity price and derivative contracts of major rail freight companies over that quarter.

Similarly, if one worked for a rail company, this model could give insight into capital budgeting. With less freight moving earlier in the year, and reaching its highest frequency in August and October, Q1 and Q2 budgets for labor expense can be lowered, and more money can be budgeted for investment in procurement or rail yard projects.

9 Recommendations

Ultimately, the ARIMA model did an adequate job of predicting rail freight frequency for 2023. The primary issues that remained with the model largely had to do with the major volatility we observed in 2008 and 2020. With these anomalies being tied to economic events, it built uncertainty around the constant variance assumption of our stationary model.

One way volatility could have been mitigated in an ARIMA model framework, would be if there was more historical data in the sample. Specifically, if the number of observations roughly doubled to include January 1980 through December 1999, this would allow two decades of observations to be brought in. With this additional data, we could observe the economic shocks that may have caused volatility in earlier year's rail freight frequency. With these observations, this may have normalized the economic shocks in the model as a whole.

Another issue the volatility caused was the skewed normality of residuals in our final model. While again, this may have been helped with more data in the sample, the ARIMA model framework may not have been the best methodology given the data that we had.

The GARCH (Generalized Autoregressive Conditional Heteroskedasticity) approach may have been a more forgiving and more accurate choice in the modeling. Given that there is volatility clustering around major economic events, this may have assisted in interpreting stronger parameter estimates where they are constrained to allow time dependent variance, $\sigma_t^2 > 0$.

Lastly, "Long Memory" ARIMA could be a last resort if GARCH did not show any improvement from the initial ARIMA model.

References

[1] U.S. Bureau of Transportation Statistics. *Rail Freight Carloads*. 2000-2022. URL: `https://fred.stlouisfed.org/series/RAILFRTCARLOADS` (visited on 03/02/2023).