PHINEAS PHAM
OCT 22nd, 2022
MATH 422 - Dr. White

Mid Semester Project

I. Introduction:

Earth's sky has always been a mystery of human interest throughout history. We put our dreams and science to satisfy our curiosity. But not until 1957 did humanity first launch an artificial satellite into space. The USSR kicked off the Cold War's "Space Race" with Sputnik, a basketball-sized satellite (Oct 4, 1957 CE: USSR launches Sputnik). Since that event, we have been launching space missions from all over the world to satisfy our curiosity. In this project, I tried to see the trend of the number of space missions monthly from 1957 to 2022, fit models to this time series, and try to forecast the number of space missions monthly in the near future. After trying out some models, I find out that the number of space missions now is correlated with the noise of two months ago, and there is a seasonal trend with the same month last year.

II. Methods:

I used the space missions data set from Maven Analytics to provide information about all the missions that happened between 1957 to 2022. I did some wrangling of the data set by grouping the dates into months., resulting in a new data frame called 'sp' which consists of a list of months and the number of space missions launched that month. As the data set is originally from Next Spaceflight, I strongly believe the data is the true record from history. Thus, it is hardly biased and reflected the true space missions in the world.

I assumed that the number of missions is not constant and is increasing at a faster speed (not linearly) over time. I reasoned that because it is clear that the technology and space industry has been developing day by day, especially in these recent years (Ludwig). Also, I believe that seasonality impacts the number of missions. Like how the weather may impact the launch of Artemis I, scientists may have to calculate and forecast the best weather to make the most efficient launch, and seasonality impacts weather (Ritz, Weather could interfere with the Artemis I launch).

In this project, I tried 3 different approaches to fit the 'sp' time series: the non-parametric trend, the function of time model, and the SARIMA model. I used seasonality as a parameter in those models. And Figure 1 shows what the time series of the number of space missions looks like. From that figure, I can hardly identify any obvious and useful trend, except for an unsteady increase.

I performed non-parametric trends with Loess and Ksmooth. However, both 3 smoothing trends (Figures 2, 3, 4) for monthly, quarterly, and yearly failed to identify any significant or useful trend. They cannot capture the fact that there is a significant increase lately, thus going forward with these trends will lead to prone-error results.

The function of time also did not give useful findings. Figure 5 shows the summary of the function of time I got with significant p-values. I found that the seasonal means have the p-values for all the months less than 0.05. This proves that the seasonality of months helps explain the trend. However, the adjusted R-squared of 3.18% argued that the model is not so helpful, as only about 3% of the observations can be explained by this model, which would not perform well for forecasting and other uses.

Last but not least, I found the SARIMA model to perform well with this time series. The SARIMA model with one-time differencing, MA(2), SAR(1), and SMA(1) performs very well with model utility tests. As a result, I used this model for forecasting.

III. Results and Conclusions:

After fitting various models, ARIMA(0,1,2) and the seasonality of (1,0,1) with a period of 12 (months) is the best model. Figure 6 shows the residual plots of this SARIMA model. It looks not too good to be stationary, but thanks to the results from Figures 9 and 10 of the 4 stationary tests, I am confident that the residuals are stationary. The ACF test only has 2 bad lags at 7 and 27, and the PACF test has 1 bad lag at 7. These bad lags seem random, as it is unreasonable to see that the number of space missions has a relationship with the number 7 months ago. Thus, I called it a pass for ACF and PACF tests. Next, I check the model utility tests. From figure 11, we can see that the Q-Q plot of the residuals looks normal, and the p-values for the Ljung-Box statistic show no sign of concern. Figure 12 shows the density plot of the residuals, and they look good with no significant skew.

Because all the model utility tests look good, I built a forecasting model in figure 13. The forecast looks reasonable as it captures my opinion on the trend of increasing the number of space missions in the near future. The time series looks like it is increasingly volatilely, and the 95% confidence interval of the forecast seems to capture this trend.

IV. Conclusion:

This project is fun to play with. It was challenging to attempt to find the best models to fit this time series. SARIMA models are the cornerstone of fitting this time series, as there is no useful trend that I can identify with bare eyes. Using auto.arima() function helped me find the best model, but I was able to make a better one with a lower AICc value and better ACF/PACF lags. I learned that we can always find better models than one recommended by auto.arima() function, but the function seems to be a good guide to finding the model we want. Based on the model I presented, the number of space missions in the world is correlated with the number from 2 months ago and 12 months ago.

V. References:

1. Oct 4, 1957 CE: USSR launches Sputnik. National Geographic Society. (2022, May). Retrieved October 23, 2022, from https://education.nationalgeographic.org/resource/ussr-launches-sputnik
2. Ludwig, S. (2022, September 15). How the space industry is taking off in 2022. U.S. Chamber of Commerce. Retrieved October 22, 2022, from https://www.uschamber.com/space/how-the-space-industry-is-taking-off-in-2022
3. Ritz, B. (2022, August 26). Weather could interfere with the Artemis I launch. CNN. Retrieved October 23, 2022, from https://www.cnn.com/2022/08/26/world/nasa-launches-forecaster-artemis-i-scn/index.html

VI. Appendix:

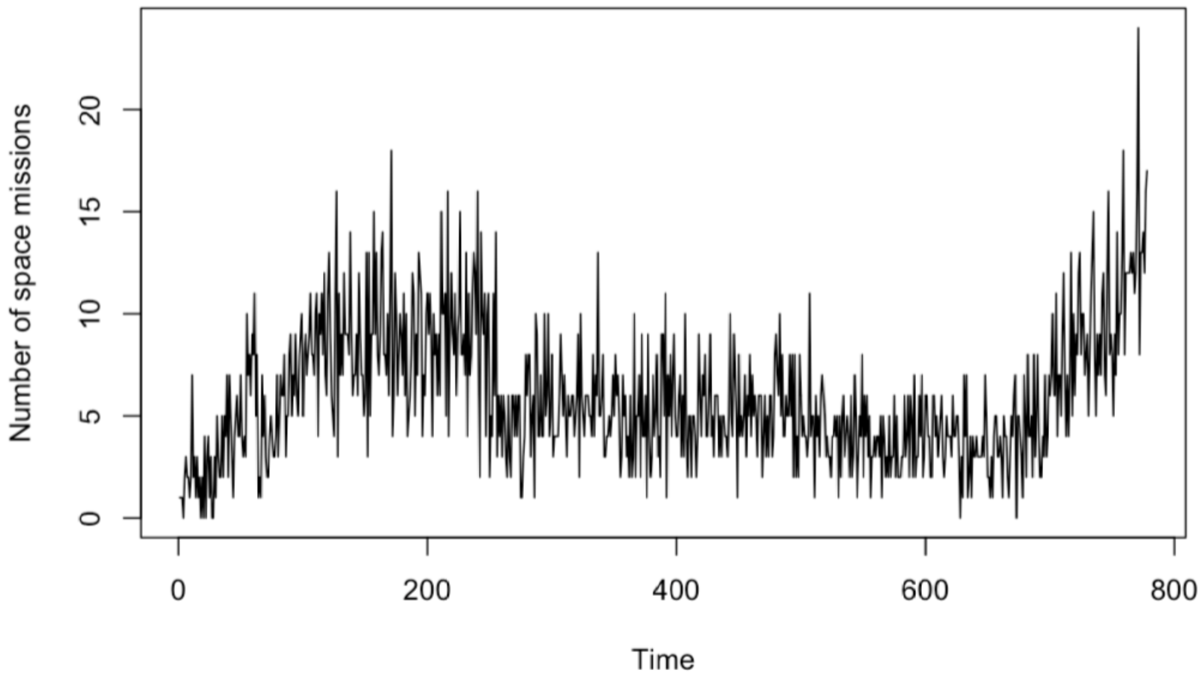

Figure 1: Time series of the number of space missions

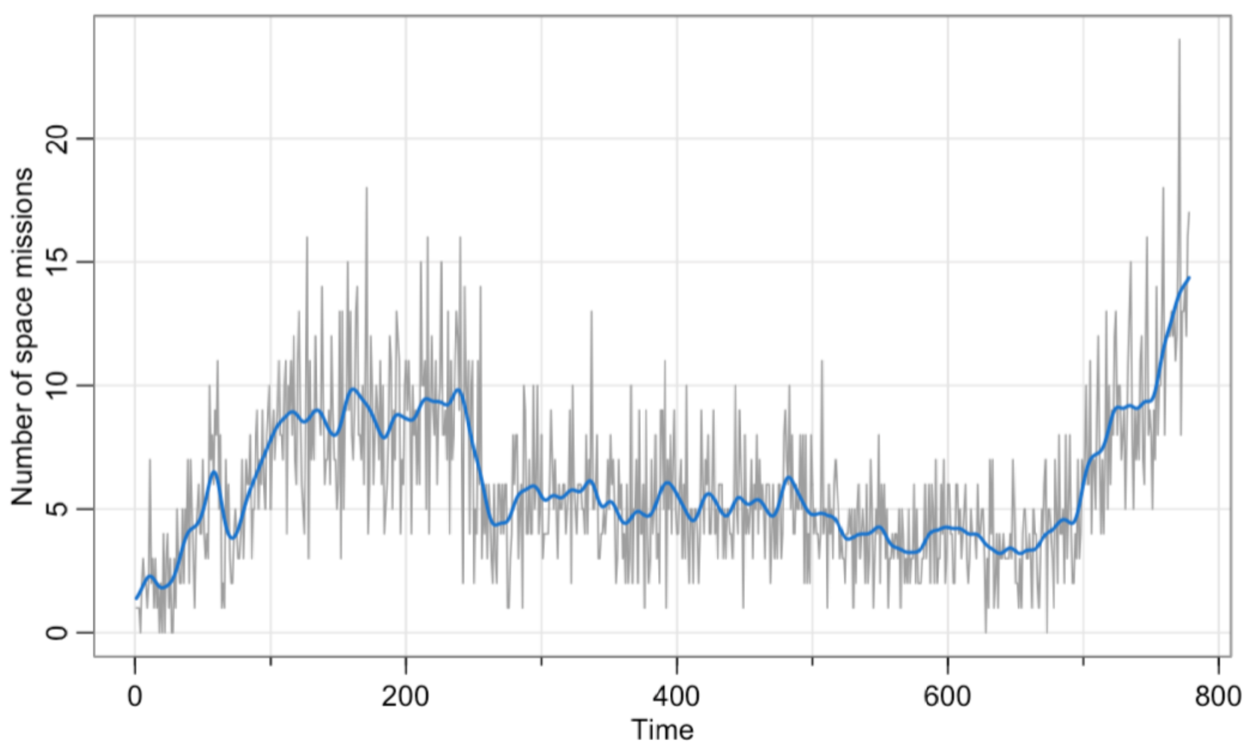

Figure 2: Monthly smoothing

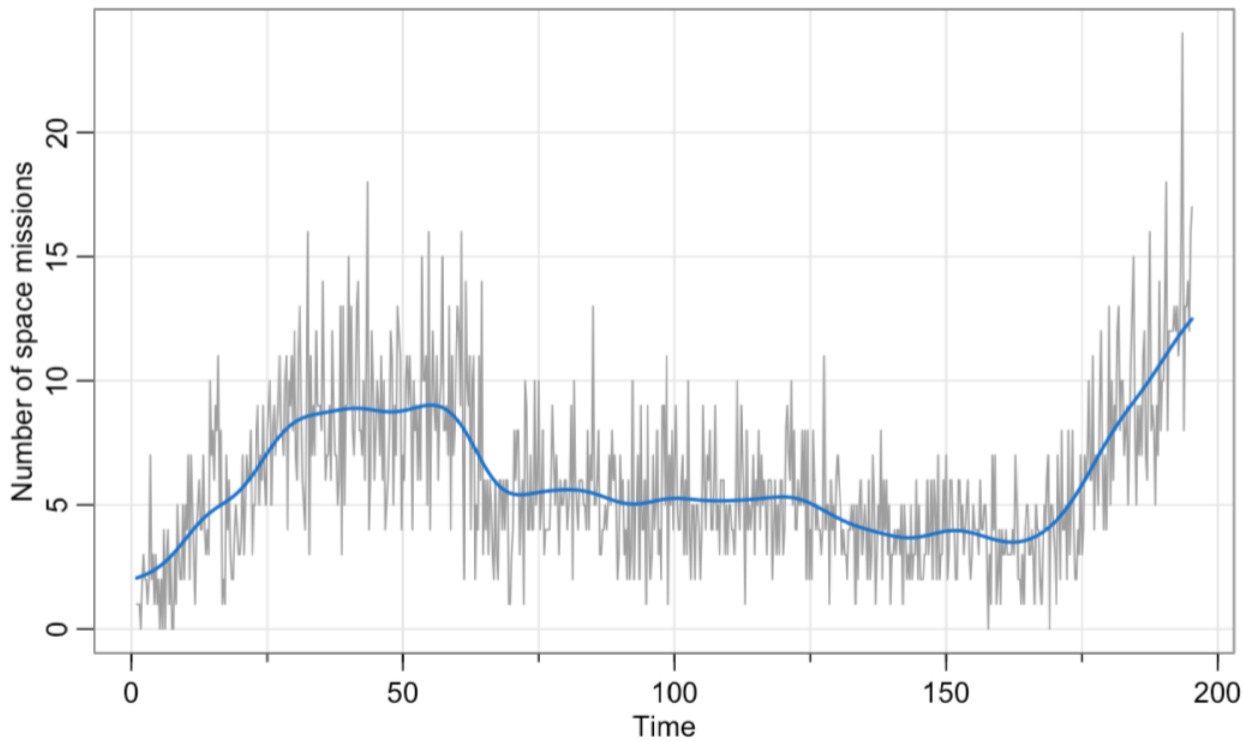

Figure 3: Quarterly smoothing

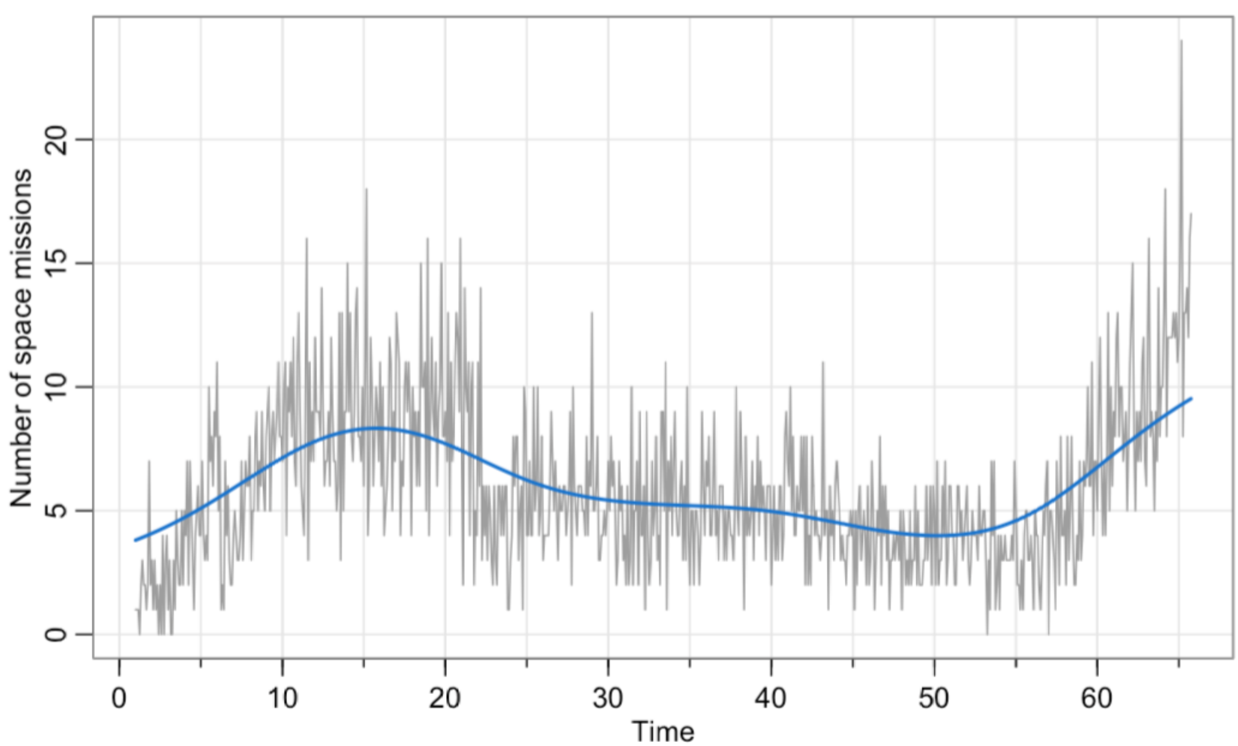

Figure 4: Yearly smoothing

```
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)          4.3692     0.4092  10.677  < 2e-16 ***
as.factor(sp$M)2     1.1846     0.5787   2.047 0.040998 *
as.factor(sp$M)3     1.4462     0.5787   2.499 0.012665 *
as.factor(sp$M)4     1.9231     0.5787   3.323 0.000933 ***
as.factor(sp$M)5     1.0154     0.5787   1.755 0.079731 .
as.factor(sp$M)6     2.2615     0.5787   3.908 0.000101 ***
as.factor(sp$M)7     1.4769     0.5787   2.552 0.010900 *
as.factor(sp$M)8     1.7401     0.5810   2.995 0.002830 **
as.factor(sp$M)9     1.6620     0.5810   2.861 0.004341 **
as.factor(sp$M)10    1.8308     0.5787   3.164 0.001620 **
as.factor(sp$M)11    1.2462     0.5787   2.153 0.031603 *
as.factor(sp$M)12    3.2000     0.5787   5.530 4.41e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.299 on 766 degrees of freedom
Multiple R-squared:  0.04762,	Adjusted R-squared:  0.03394
F-statistic: 3.482 on 11 and 766 DF,  p-value: 9.177e-05
```

Figure 5: Summary of fitting seasonal means

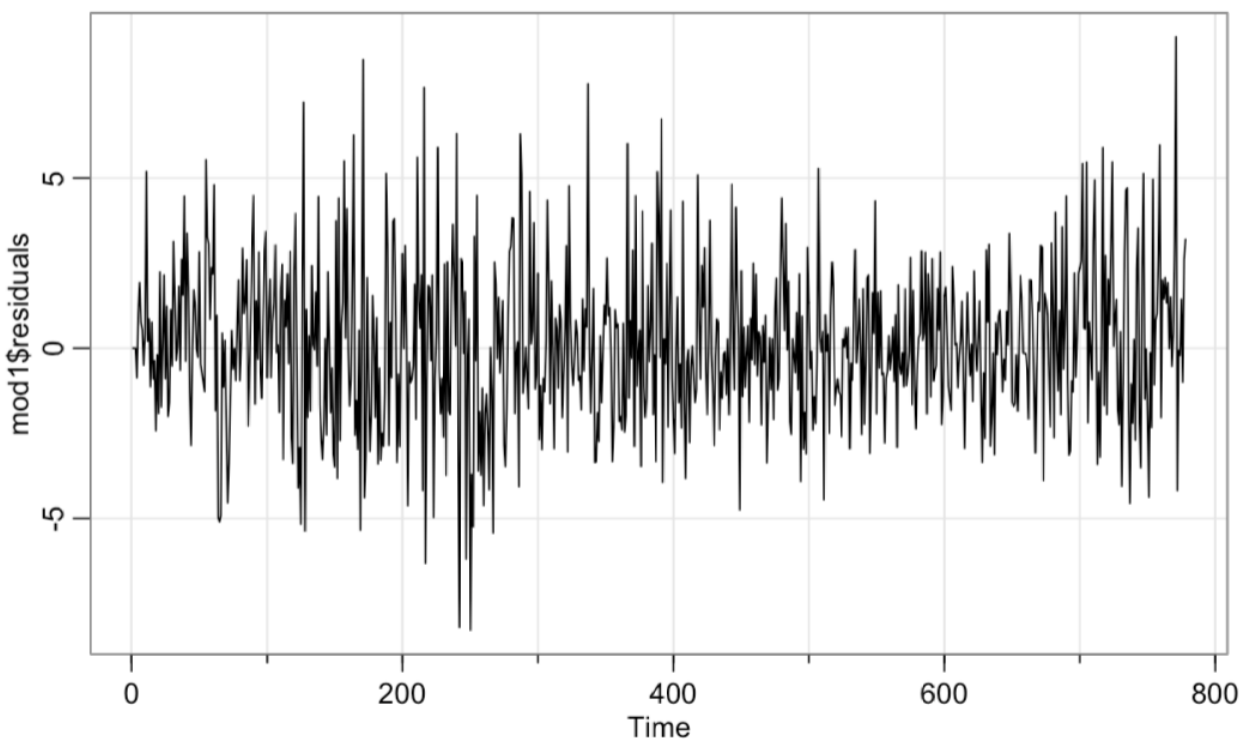

Figure 6: Model residuals

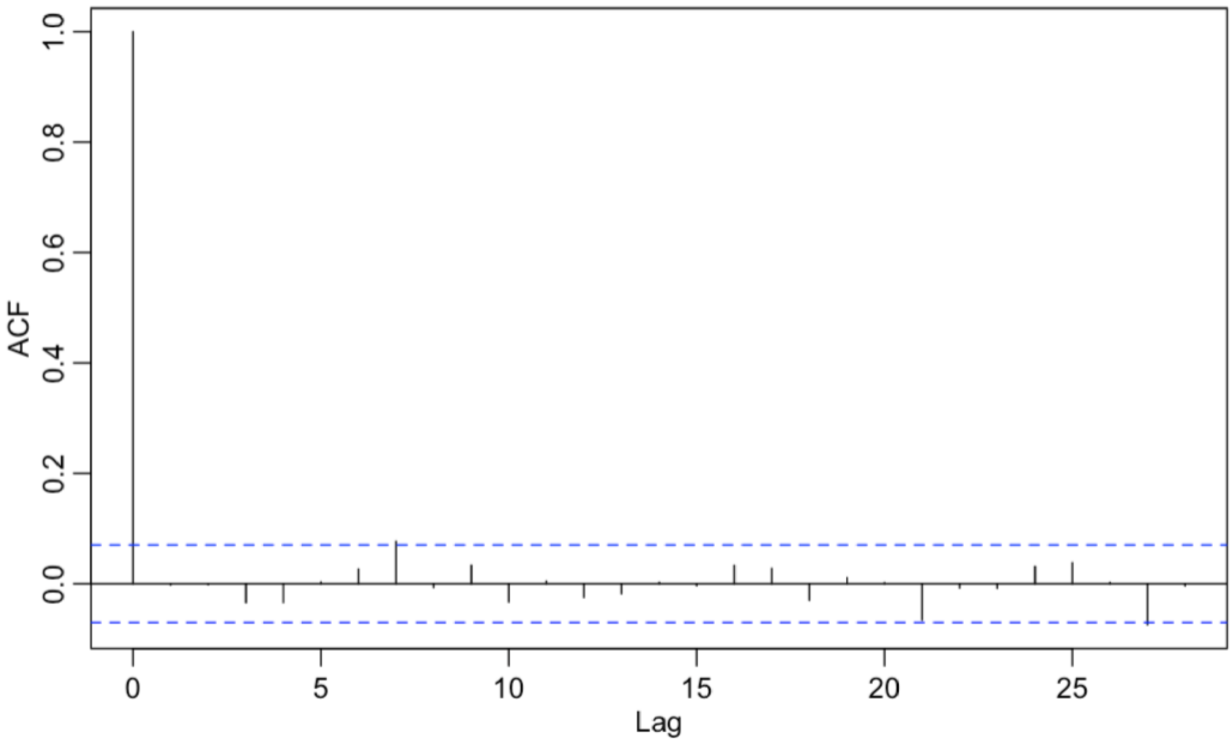

Figure 7: ACF test

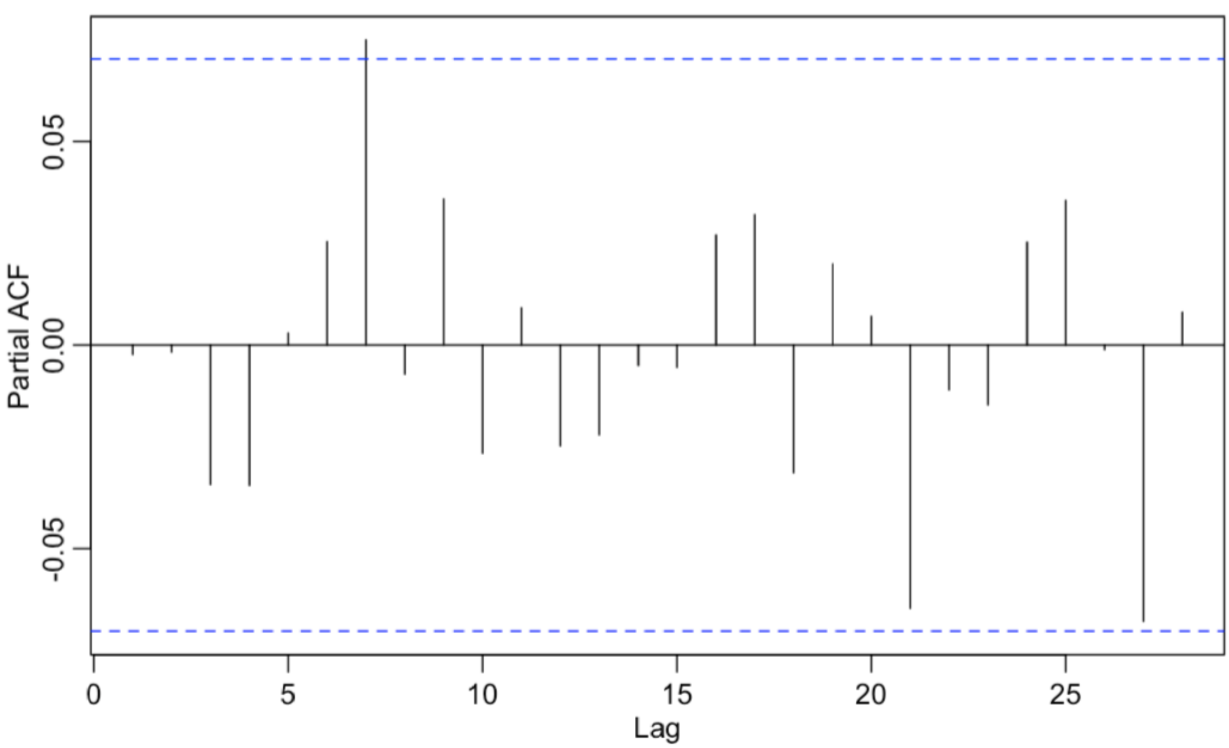

Figure 8: PACF test

```
Warning: p-value smaller than printed p-value
	Augmented Dickey-Fuller Test

data:  mod1$residuals
Dickey-Fuller = -8.3179, Lag order = 9, p-value = 0.01
alternative hypothesis: stationary

Warning: p-value smaller than printed p-value
	Phillips-Perron Unit Root Test

data:  mod1$residuals
Dickey-Fuller Z(alpha) = -753.87, Truncation lag parameter = 6, p-value = 0.01
alternative hypothesis: stationary

Warning: p-value greater than printed p-value
	KPSS Test for Level Stationarity

data:  mod1$residuals
KPSS Level = 0.25193, Truncation lag parameter = 6, p-value = 0.1
```

Figure 9: Stationary tests (1)

```
Value of test-statistic is: -12.9101

Critical values of DF-GLS are:
                 1pct  5pct 10pct
critical values -2.57 -1.94 -1.62
```

Figure 10: Stationary tests (2)

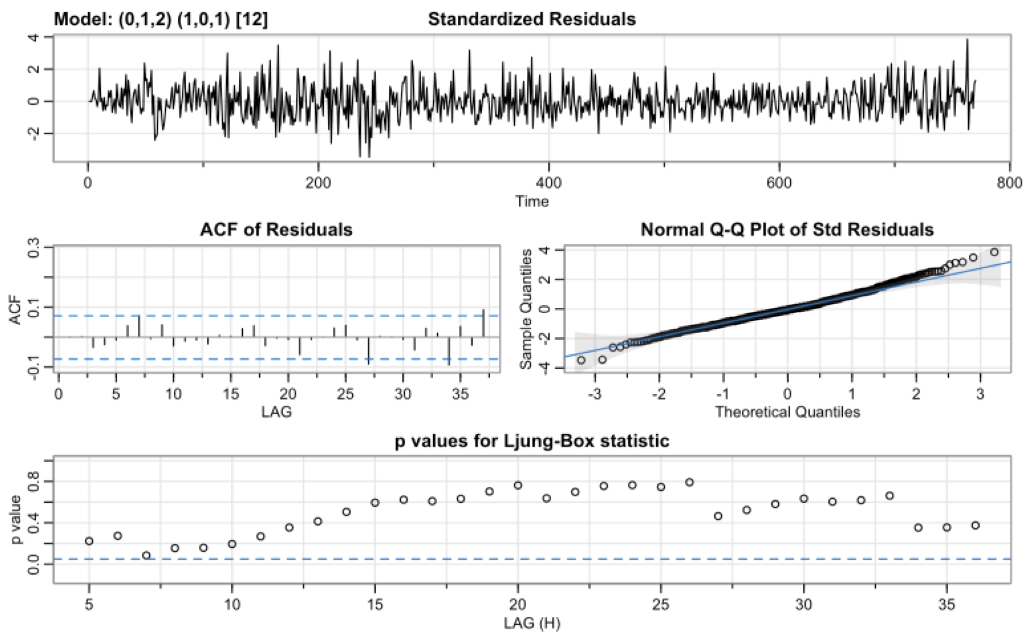

Figure 11: Model Utility tests

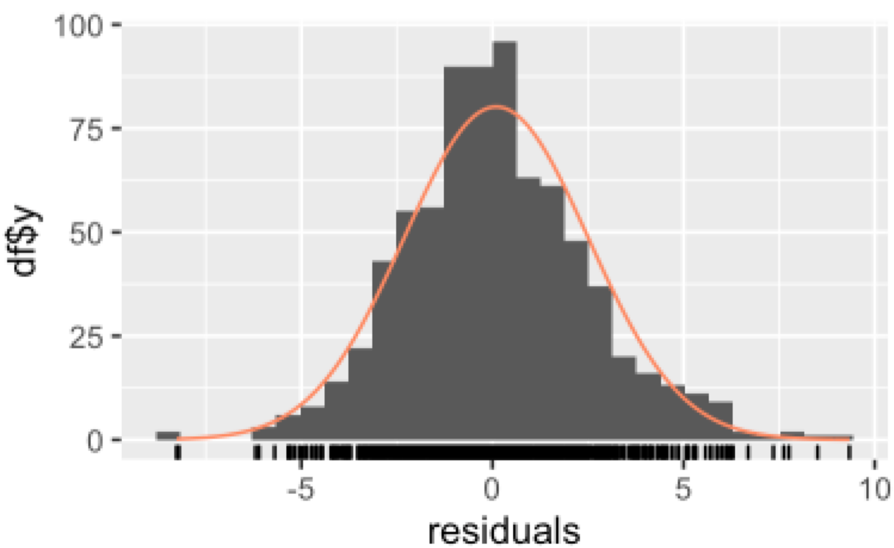

Figure 12: Density plot

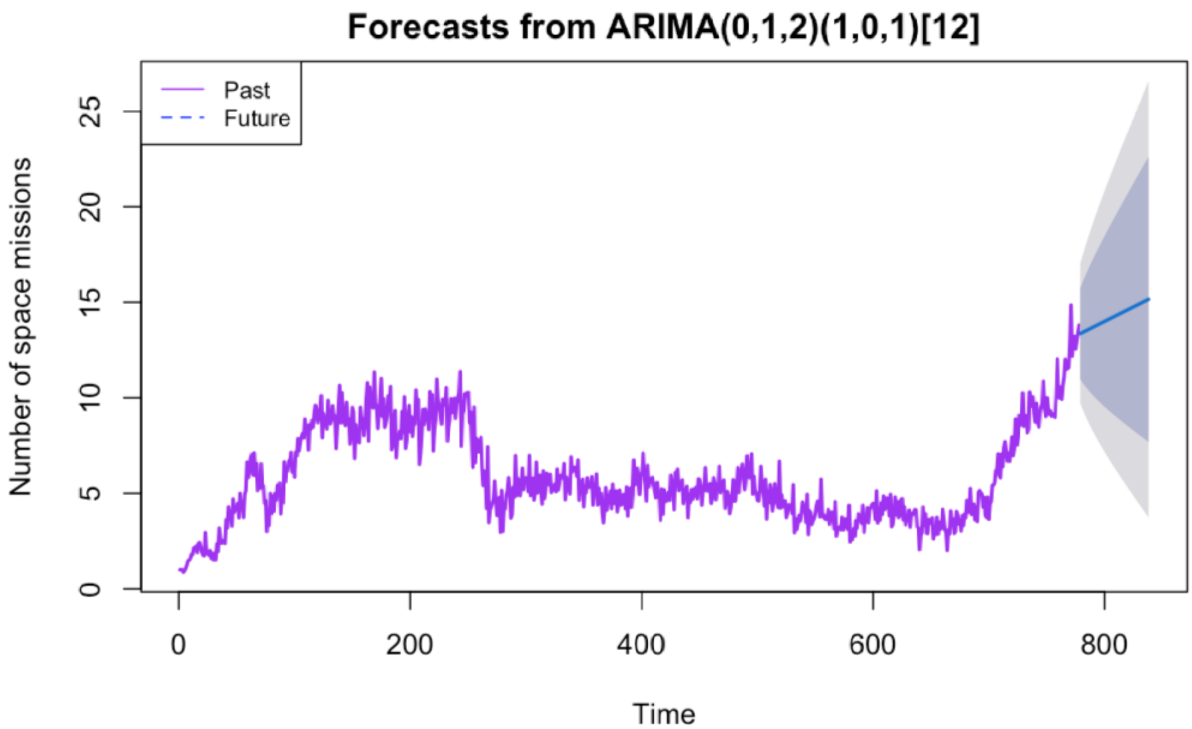

Figure 13: Forecasting