Stat153 Project

Kenny Tran

May 10, 2021

1 Executive Statement

There is a lot of uncertainty surrounding Lots-of-Stuff Incorporated's (LOSI) stock price forecast. Although the company's stock price has been increasing steadily over the past couple of years, the events of 2020 has resulted in a gloomy and uncertain outlook for the company heading into 2021. According to our parametric model with AR(10) noise, LOSI's stock price should increase during the trading days of January 11, 2021 to January 25, 2021.

2 Exploratory Data Analysis

An initial look at the Time Series plot of LOSI (Figure 1) reveals a generally upward trend in stock prices over time, which we will remove to obtain stationarity. There also appears to be some fluctuations in price change, indicating that we may be dealing with heteroscedasticity. In order to achieve homoscedasticity, we apply a log variable stabilizing function to account for the increasing variance.

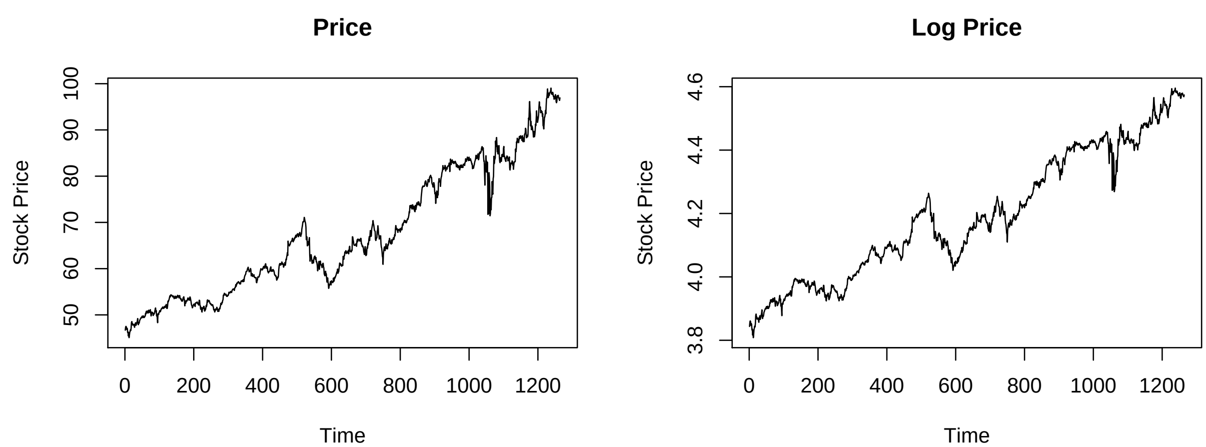

Figure 1: Original Time Series and Time Series of Log Prices

3 Models Considered

3.1 Second-order Differencing

Our first approach to achieve stationarity is to difference our data in order to eliminate trend. We apply a first order difference to our log transformed data. Figure 2 below shows the differenced data (left) and it's corresponding ACF (right).

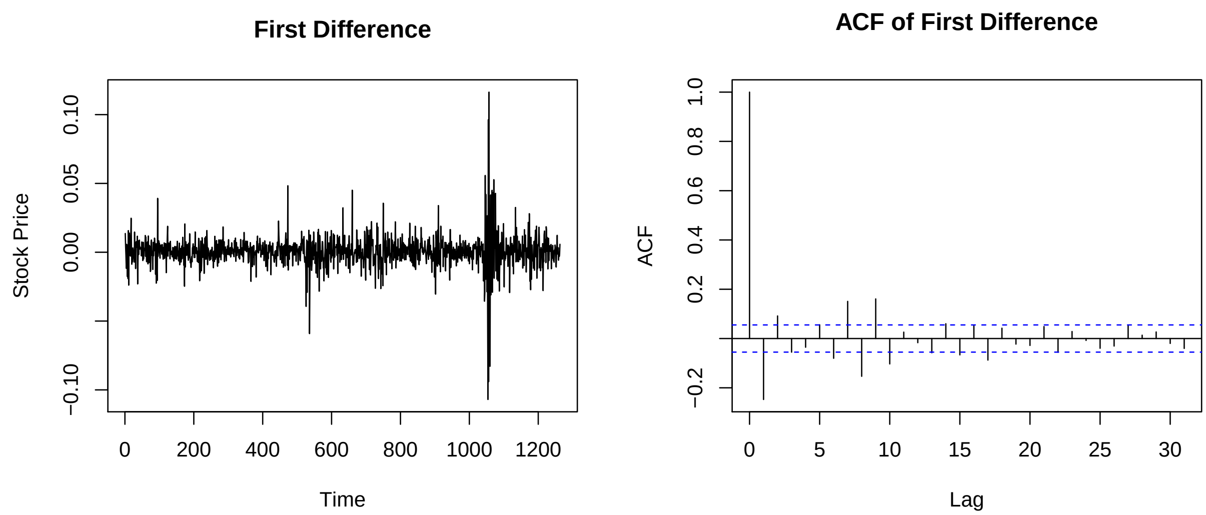

Figure 2: First Difference of Log Stock Prices and it's ACF

The mean and variance are relatively constant over time and our data looks roughly stationary (with the exception of time = ~1050). There are some large magnitude ACF values at lags 1, 7, 8, and 9. My initial attempt was an AR(1) model because the ACF plot (Figure 2, Right) shows a decreasing trend after Lag 1.

3.1.1 First Difference with AR(1)

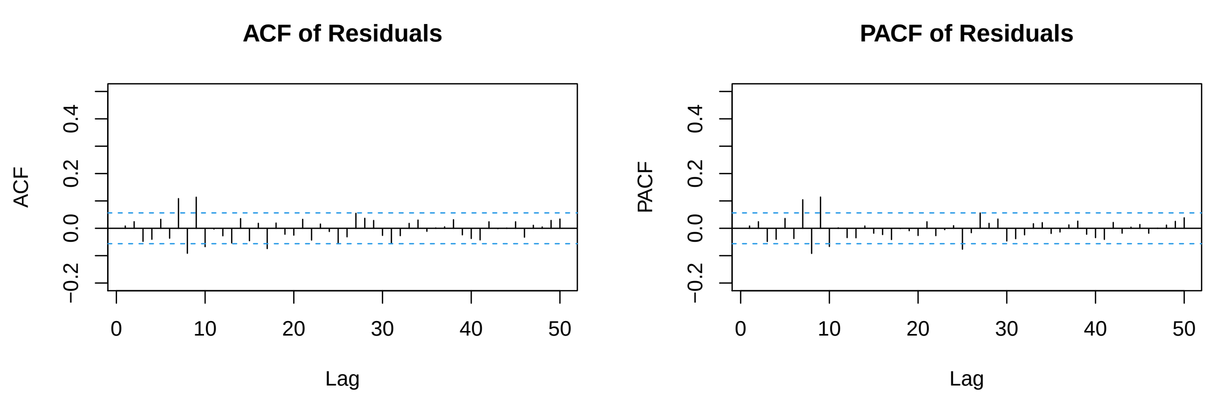

Figure 3: ACF and PACF of Residuals

The residuals of the first difference with AR(1) noise look reasonably stable based off of the ACF and PACF of Residuals plot (Figure 3). Most of the values are within the blue 95% confidence bands or at least close enough to being within the blue bands. As a result, we can say that the residuals look like white noise.

3.1.2 First Difference with AR(9)

My next attempt was an AR(9) model because lag 9 was the last lag with large magnitude ACF values as depicted in Figure 2 (Right).

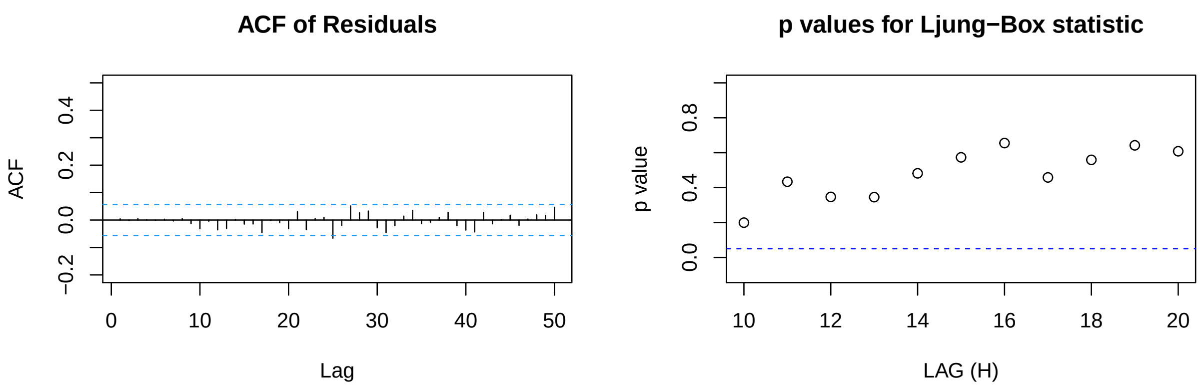

Figure 4: ACF of Residuals and p-values of Ljung-Box test

As we can see from Figure 4 (Left), the residuals of the first difference with AR(9) noise model look stable and like white noise as most of the values fall under the blue 95% confidence bands. In addition, the Ljung-Box test (Figure 4, Right) shows us that there are no significant p-values across all lags. This means that we reject the null hypothesis that the data comes from the model, indicating that this model is an acceptable fit.

3.2 Parametric Signal Model

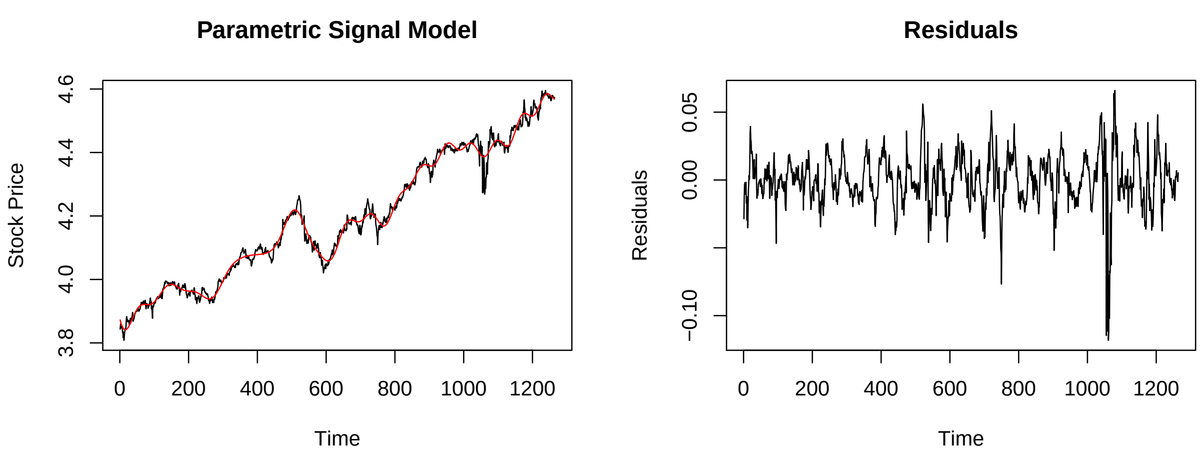

Figure 5: Parametric Signal Model and it's Residual Plot

Next, we consider a parametric model. The fitted line seems to capture the time series pattern reasonably well. In addition, the residual plot (Figure 5, Right) seems reasonably stationary. However the ACF and PACF plots (Figure 6) tell us that this is not white noise as there are too many values sticking out of the confidence bands. Judging by the PACF plot (Figure 6, Right), we will first try an AR(10) model as the graph seems to cut off after lag 10.

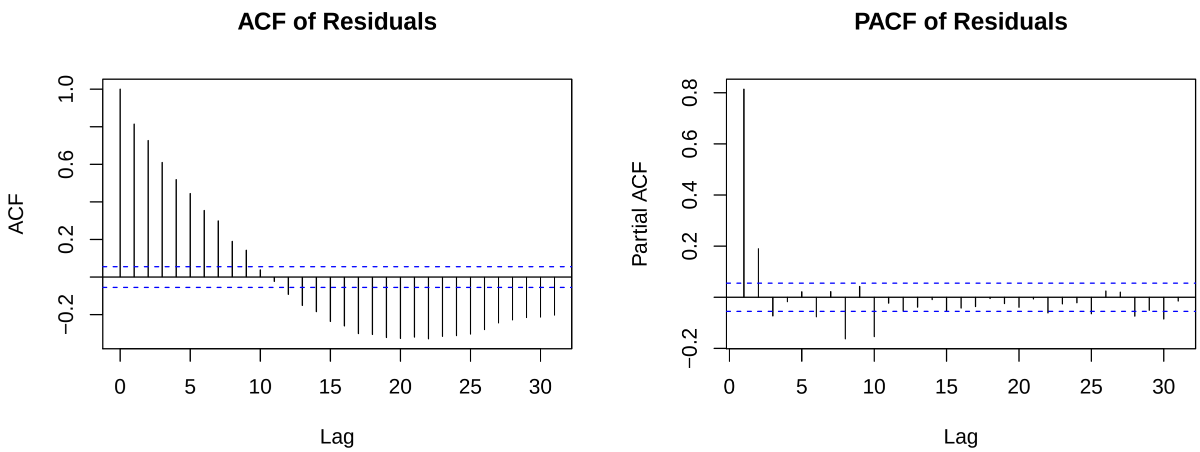

Figure 6: ACF and PACF of Residuals

3.2.1 Parametric Signal Model + AR(10)

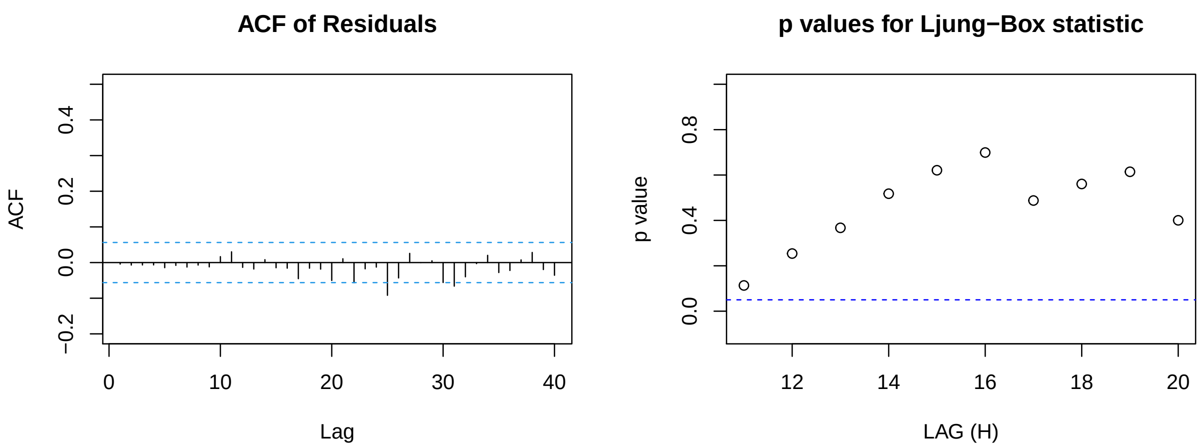

Figure 7: ACF of Residuals and p-values of Ljung-Box test

Judging by the ACF of Residuals (Figure 7, Left), the residuals of the Parametric Model + AR(10) look stable and resemble white noise. In addition, all of the p-values in the Ljung-Box statistic (Figure 7, Right) are non significant, indicating that the model is an acceptable fit.

3.2.2 Parametric Signal Model + ARMA(1,2)

Lastly, we use R's auto.arima() function for our final model. The result is an ARMA(p = 1, q = 2) model. This seems plausible as the residual ACF (Figure 6, Left) seems to tail off after lag 1, indicating p = 1. In addition, we can make an argument that the PACF plot (Figure 6, Right) tails off after lag 2, indicating q = 2. Thus, we get an ARMA(1,2) model. Judging from the ACF and PACF of residuals plot (Figure 8), most of the values are either within the blue bands or are very close to being within the blue bands so we can make an argument that it fits reasonably well.

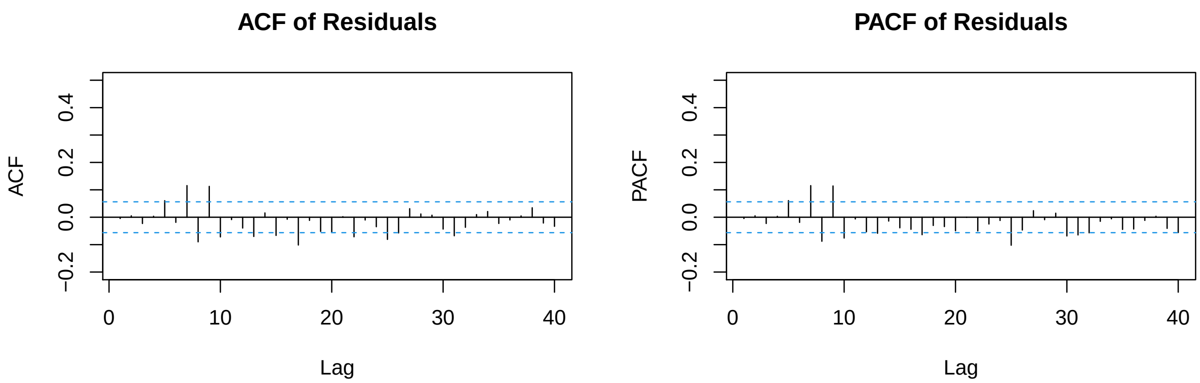

Figure 8: ACF and PACF of Residuals of Parametric + ARMA(1,2)

4 Model Comparison and Selection

The four models are compared through time series cross validation. The data was split into a training and testing set with a rolling window of 10 points for the test set. We compare each of the forecast performances by root mean square error (RMSE). The model with the lowest RMSE will be chosen to forecast stock prices for the 10 trading days between January 11, 2021 and January 25, 2021.

Table 1: Cross-validated out-of-sample root mean squared error for the four models under consideration.

	RMSE	AIC
First Order Difference + AR(1)	5.715662	-6.134200
First Order Difference + AR(9)	5.735435	-6.164573
Parametric Model + AR(10)	4.196354	-6.265094
Parametric Model + ARMA(1,2)	4.196360	-6.217170

As we can see from Table 1, the model with the lowest RMSE is the Parametric model with AR(10). It is also the model with the lowest AIC value. This is the model that will be used for forecasting.

5 Results

The selected model that will be used for forecasting is described as follows:

$$Price_t = B_0 + B_1 t^2 + \sum_{j=1}^{10}[B_{2j} cos(\frac{\pi jt}{365}) + B_{2j+1} sin(\frac{\pi jt}{365})] + \sum_{j=1}^{5}[B_{2j+20} t^2 cos(\frac{\pi jt}{365}) + B_{2j+21} t^2 sin(\frac{\pi jt}{365})] + X_t$$

Where $X_t - \phi X_{t-1} - \phi X_{t-2} - \phi X_{t-3} - \phi X_{t-4} - \phi X_{t-5} - \phi X_{t-6} - \phi X_{t-7} - \phi X_{t-8} - \phi X_{t-9} - \phi X_{t-10} = W_t$

X_t is a stationary process defined as AR(10) and W_t is white noise with variance σ_W^2

5.1 Estimation of model parameters

Our final model contains 32 $\beta's$, 10 AR coefficients, and the variance of white noise. The Estimates of the model's parameters(Coefficients and Standard Errors) are given in Table 2 in Appendix 1.

5.2 Predictions

Figure 9 below shows the forecasted stock prices for Lots-Of-Stuff Incorporated (LOSI) for the trading period between January 11, 2021 to January 25, 2021. Based on the prediction of our selected model (Parametric + AR(10)), the stock price for Lots-of-Stuff Incorporated (LOSI) should increase linearly. After the events of 2020, this is a good sign for LOSI investors as this indicates growth for the company.

Figure 9: Forecast for LOSI Stock Prices, Historical Data in Black and 10 Day Forecast in Blue

6 Appendix 1 - Table of Parameter Estimates

Table 2: Estimates of the forecasting model parameters

Parameter	Estimate	SE
β_0	3.992	8.993e-03
β_1	3.890e-07	2.143e-08
β_2	-2.029e-02	1.332e-02
β_3	1.803e-04	4.879e-03
β_4	3.031e-03	5.282e-03
β_5	-1.831e-03	5.830e-03
β_6	-4.292e-02	4.485e-03
β_7	-2.210e-02	3.087e-03
β_8	-1.217e-02	2.098e-03
β_9	-1.594e-02	1.758e-03
β_{10}	-2.705e-03	1.504e-03
β_{11}	-9.721e-04	1.307e-03
β_{12}	-6.886e-02	1.078e-02
β_{13}	1.472e-03	1.396e-02
β_{14}	-3.916e-02	1.025e-02
β_{15}	-6.451e-02	5.675e-03
β_{16}	-1.613e-02	2.874e-03
β_{17}	3.534e-09	3.048e-08
β_{18}	-3.267e-08	1.008e-08
β_{19}	2.817e-08	1.484e-08
β_{20}	2.607e-08	1.427e-08
β_{21}	6.254e-08	9.739e-09
β_{22}	4.715e-08	6.237e-09
β_{23}	3.066e-08	4.029e-09
β_{24}	2.981e-08	3.419e-09
β_{25}	1.367e-08	2.814e-09
β_{26}	2.200e-08	2.135e-09
β_{27}	1.291e-07	2.703e-08
β_{28}	4.219e-08	3.344e-08
β_{29}	6.343e-08	2.238e-08
β_{30}	7.232e-08	1.161e-08
β_{31}	1.584e-08	5.652e-09
ϕ_1	0.6948	0.0278
ϕ_2	0.1899	0.0337
ϕ_3	-0.0418	0.0339
ϕ_4	-0.0247	0.0338
ϕ_5	0.0670	0.0337
ϕ_6	-0.0523	0.0337
ϕ_7	0.1114	0.0337
ϕ_8	-0.1567	0.0339
ϕ_9	0.1484	0.0337
ϕ_{10}	-0.1521	0.0278
σ^2_W	0.0001091	