Stat 153 Final Project

Noah Kaminer

5/6/2021

1 Executive Summary

Lots-of-stuff Incorporated's (LOSI) success in producing various consumer goods, including paper, pens and water bottles has allowed its stock price to, on average, perform well since 2016. However, recent events in 2020, particularly the COVID-19 pandemic, have increased business uncertainty and turmoil for the company at large. While the company has rebounded significantly, investors are concerned that over-anticipation for escaping 2020 will force the stock price down slightly over the next 10 days of 2021. A parametric model with ARMA(3,2)x(1,0)[5] noise was calculated to fit the data best, and further predicts that investors concerns are correct and the stock price will continue to trend down, on average, for the next 10 days of 2021.

2 Exploratory Data Analysis

The price of one share of LOSI stock has increased, on average from 2016 to 2021, which is displayed by the upward trend in Figure 1. Nonetheless, the stock price has fluctuated significantly within this timeframe. The time series also suggests some degree of seasonality within the data, specifically with a period of 1 year, which is approximately 253 days of trading for the NYSE and NASDAQ. This seasonality is approximately seen through the successive dips in the stock price following the start to each year. The strength of this seasonality will be evaluated in the following sections to determine if investors are right to be concerned with a consistent stock price below the 97 dollar mark for the next 10 days of 2021. Figure 1 also appears to suggest heteroskedasticity, considering the progressive increase in fluctuations in the stock price over time. In essence, the observed presence of an upward trend, annual and weekly seasonality and heteroskedasticity indicate the dataset's deviation from stationarity, and importance of modeling these features to better evaluate and forecast the data. It is important to note that 2016 and 2020 was a leap year, which will complicate modeling methods, and will thus be overlooked for the purpose of this analysis.

3 Models Considered

To fully evaluate the signal in the stock price data, two different approaches will be used: a parametric signal model and a differencing model. Each model will fit the signal, but will also be complemented with an ARMA model for the residual noise after accounting for the signal.

3.1 Parametric Signal Model

To begin, a parametric model is employed to model the signal in the data. Exploratory data analysis suggests that the data increases over time and that there may be some degree of annual and weekly seasonality within the data. Thus, it follows to create a sinusoid that increases with time and has a period of 253. Accordingly, an indicator for the day of the week was used because it has long been expected that a stock moves differently
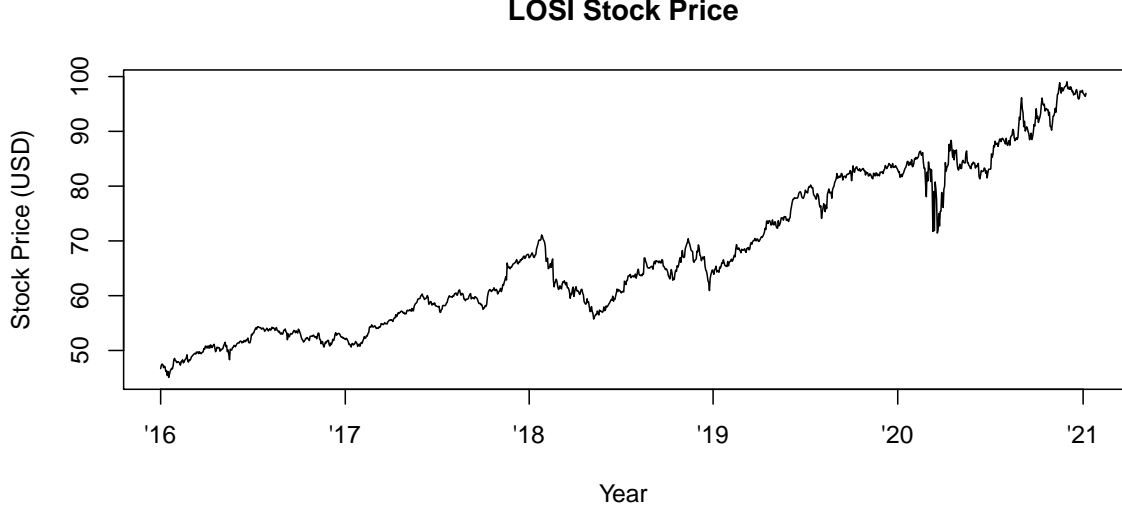
LOSI Stock Price

Figure 1: A time series of Lots-of-stuff Incorporated's (LOSI) stock price, 1/4/2016 - 1/8/2021.

on different days of the week. There is also an indicator variable for 2018 in this model because cursory analysis of the stock price time series displays that 2018 does not appear to fit perfectly into the rest of the time series – though the reasoning is unknown. While the deterministic model described above appears to model the stock trend relatively well and helps to make the residuals more stationary, it certainly is not perfect.

$$Price_t = \beta_0 + \beta_1 t + \sum_{j=1}^{4} \beta_{1+j} I_{\text{weekday}_{jt}} + \beta_6 I_{2018_t} + \sum_{i=1}^{6} \beta_{6+i} \sin(\frac{2\pi it}{253}) + \sum_{i=1}^{6} \beta_{12+i} \cos(\frac{2\pi it}{253}) + X_t$$

Figure 2 illustrates the parametric signal model discussed above. As shown in the left panel, the deterministic model appears to fit the stock price time series relatively well. However, taking a closer look at the right panel of Figure 2 indicates that the variance of the data, even after a parametric fit, changes over time, and thus the data is still slightly heteroskedastic. Nonetheless, this model offers an interesting and valuable analysis of the signal component to the LOSI stock price time series.

3.1.1 Parametric Signal with ARMA(3,2)x(1,0)[5]

The ACF and PACF plots for the parametric signal model's residuals are displayed in Figure 3. Comparing the ACF and PACF plots, it is quite apparent that there are much more significant lags on the ACF than the PACF. Moreover, the most significant values on the ACF are at lags of 1 and 2. This lends to the interpretation that possible parameters are p=2 and P=1. However, a closer look at the ACF and PACF plots, particularly their shapes suggest that q may also be non-zero. After trial and error, a ARMA(3,2)x(1,0)[5] depicted the best fit to the model. The red circles on Figure 3 display this model's fit to the sample autocorrelations.

3.1.2 Parametric Signal with ARMA(1,3)x(0,2)[5]

The auto.arima() function in R predicted that a ARMA(1,3)x(0,2)[5] model would be a good fit to the data. While somewhat different from the first ARMA model used and discussed above, it seems reasonable that this model could fit the data quite well. The results of this model are displayed by the blue dots on Figure 3, and as shown on the ACF and PACF plots, this model actually appears to fit the model less well than the first ARMA model displayed in red, despite being produced from an R function.
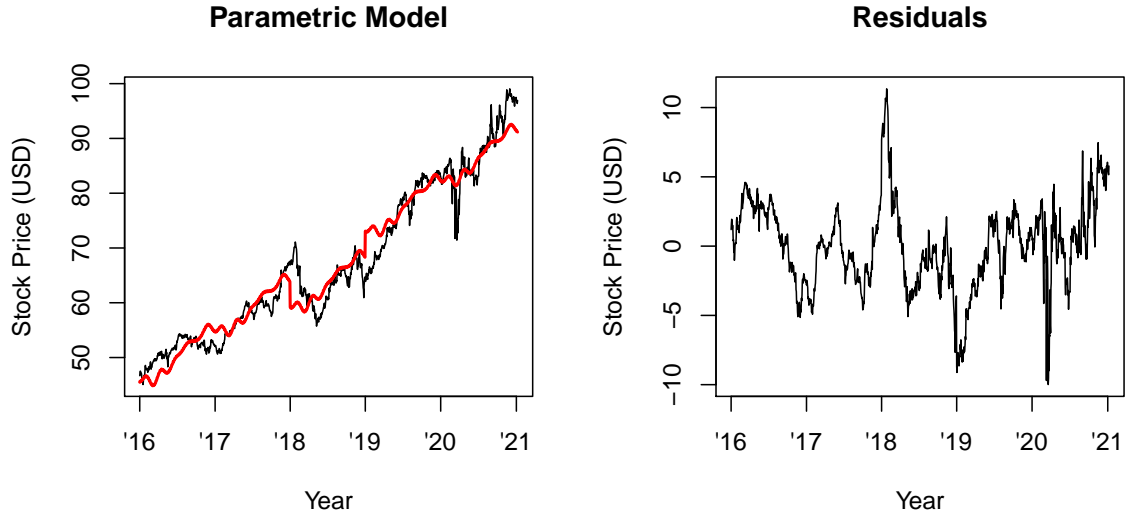
2

Parametric Model

Residuals

Figure 2: Parametric model for LOSI stock price signal. The left panel displays the model's fitted values in red and the stock price data in black. The right panel illustrates the residuals of this signal model.

ACF of Residuals
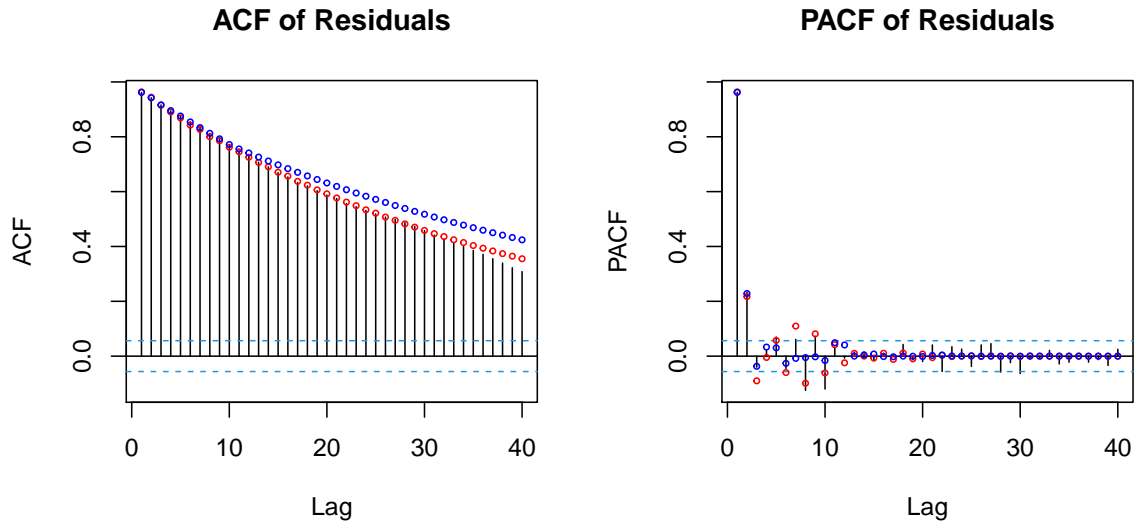
PACF of Residuals

Figure 3: Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) values for the parametric signal model's residuals. Red circles illustrate the ARMA(3,2)x(1,0)[5] model and the blue circles display the ARMA(1,3)x(0,2)[5] model.

3.2 Annual and Weekly Differencing

As discussed in the beginning sections, there appears to be yearly seasonality within the time series data. More specifically, the stock price appears to increase at the end of each year, and dip slightly at the beginning of each year. Given this predicted seasonality and the fact that the stock market is open 253 days a year, differencing with a lag of 253 makes the most sense. Moreover, the stock market is also only open on weekdays, and it is expected that the day of the week has some degree of impact on the stock price. Thus, a subsequent lag of 5 will help in this analysis. This twice-differenced result will account for any linear and quadratic trend that may exist. As Professor Fisher mentioned in lecture, it is sufficient to ignore the increase in variance in 2020 data for simplicity. Ignoring the aforementioned heteroskedasticity in 2020 produces the conclusion that Figure 4 appears to illustrate weak stationarity.
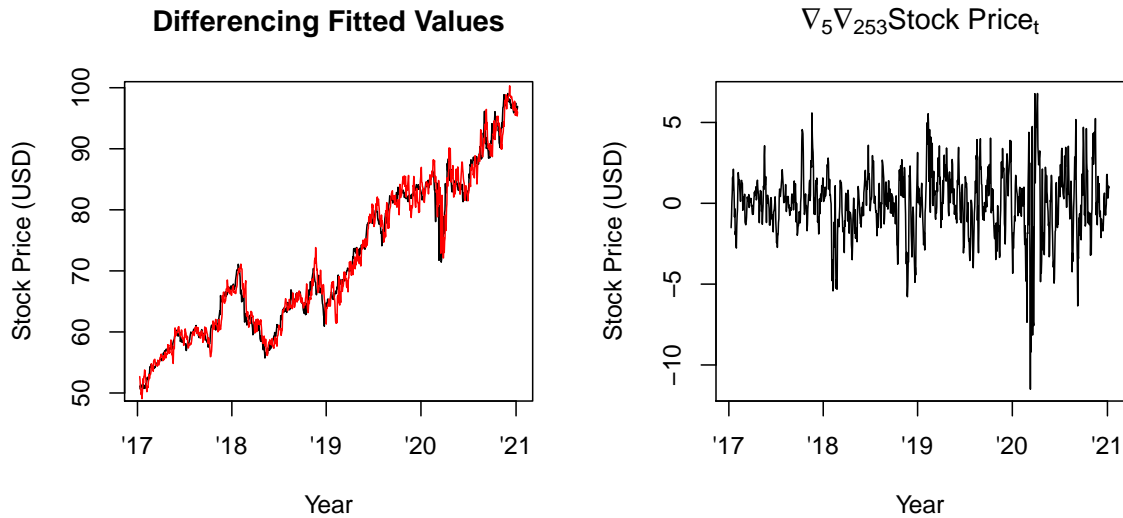

Figure 4: Diagnostics for twice-differenced signal model, which illustrates the actual difference to evaluate the stationarity through both an annual and weekly difference. The left panel displays the actual stock price data in black and the fitted values in red. The right panel shows the differences to help evalaute stationarity.

3.2.1 Seasonal and Weekly Differencing with ARMA(3,2)x(1,0)[5]

The sample ACF and PACF of the differenced data is shown in Figure 5. The ACF appears to display a pattern similar to that of an AR model with p=3, but with some variations at later lags. Moreover, the ACF seems to indicate that P is not 0, so a value of 1 was used. As for the PACF, there appears to large value every 5 lags, so S=5. The aforementioned ARMA(3,2)x(1,0)[5] model is displayed with red circles on Figure 5, which appears to illustrate a relatively good fit to the data.

3.2.2 Seasonal and Weekly Differencing with ARMA(2,2)x(2,0)[5]

To keep consistency with the Parametric Signal Model used previously, the differencing model's second ARMA noise model employed utilized the auto.arima() function in R. This function recommends ARM(2,2),(2,0)[5], thus suggesting paramters: p=2, q=2, P=2, Q=2, and S=5. Clearly this model is very similar to the first noise model used, but simply with different values of p and P. This suggested model from the auto.arima() function is displayed on Figure 5 with blue circles. This noise model appears to be a good fit to the data, but by eyeballing it, it seems to have a worse fit than the first noise model used for the differencing approach. However, this comparison will be analyzed further in the following section.
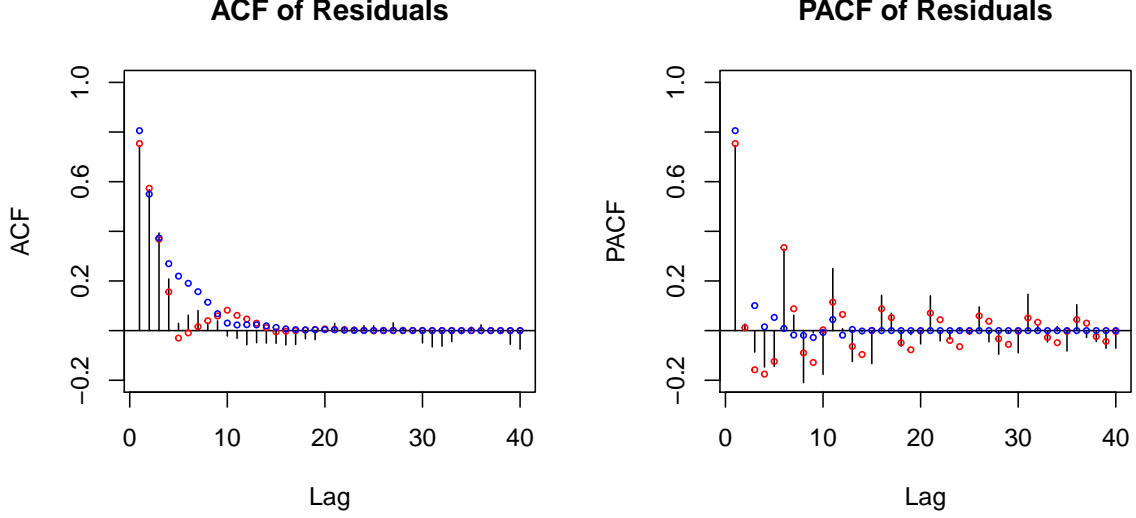
Figure 5: Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) values for the seasonal differencing model. Red circles display the ARMA(3,2)x(1,0)[5] model, while the blue circles illustrate the ARMA(2,2)x(2,0)[5].

Table 1: Cross-Validated Out-of-Sample Root-Mean Squared Prediction Error for the Four Models Under Consideration.

	RMSPE
Parametric Model + ARMA(3,2)x(1,0)[5]	1.879338
Parametric Model + ARMA(1,3)x(0,2)[5]	1.889602
Annual Differencing + Weekly Differencing + ARMA(3,2)x(1,0)[5]	2.210771
Annual Differencing + Weekly Differencing + ARMA(2,2)x(2,0)[5]	2.223684

4 Model Comparison and Selection

Table 1 displays the results of comparing the four aforementioned models using time series cross validation. The test sets utilize the last 180 days of the data, on a rolling basis, in 10 day increments. This produces 180 predicted values, or forecasted data points, over the 10 day windows previously discussed. To compare the results of each model's forecasted data points, the root-mean squared prediction error (RMSPE) will be calculated for each model. The model of best fit will the be chosen out of the four models tested, based on the lowest RMSPE and that model will ultimately be used to predict the next 10 days of 2021.

Table 1 clearly illustrates that the Parametric Model with ARMA(3,2)x(1,0)[5] noise produces the lowest RMSPE among all of the models tested. While all four models have very similar RMSPE, the following sections will employ the Parametric Model with ARMA(3,2)x(1,0)[5] noise to predict the next 10 days of 2021.

5 Results

Per the request and concern of the LOSI's Board of Directors, the model with the lowest RMSPE will be used to predict the next 10 days of LOSI stock price. This model was a parametric model with ARMA(3,2)x(1,0)[5] noise. Let $Price_t$ denote LOSI stock price on day t with a noise term X_t. The parametric model with the best fit is shown again below. The addition to the equation below assumes stationarity via the ARMA(3,2)x(1,0)[5] noise model, given that W_t is white noise with a variance σ_W^2.

$$Price_t = \beta_0 + \beta_1 t + \sum_{j=1}^{4} \beta_{1+j} I_{\text{weekday}_{jt}} + \beta_6 I_{2018_t} + \sum_{i=1}^{6} \beta_{6+i} \sin\left(\frac{2\pi i t}{253}\right) + \sum_{i=1}^{6} \beta_{12+i} \cos\left(\frac{2\pi i t}{253}\right) + X_t$$

$$X_t = \sum_{i=1}^{3} \phi_i X_{t-i} + \Phi X_{t-5} - \sum_{i=1}^{3} \phi_i X_{t-i} \Phi X_{t-6} + W_t + \sum_{j=1}^{2} \theta_j W_{t-j}$$

The model contains a number of binary indicators to improve the given model. Specifically, $I_{\text{weekday}_{jt}}$ indicates if day t is the jth day of the week. Moreover, I_{2018} indicates if day t is in 2018. ϕ, Φ, θ and each of the βs are coefficients estimated in the following section of the time series analysis. While it is typically assumed that a linear model employs X with a mean of 0, the function written to create and analyze the models in this paper, notably referred to as "SARIMA", more accurately estimates the mean of X in this model to be 0.1237818.

5.1 Estimation of Model Parameters

The estimates of the model parameters are illustrated in Table 2 in Appendix 1. For the weekday indicator variables, it is important to understand that Friday is used as the base day of the week (the day of the week that all other days are compared to). From this, it is interesting to point out that Friday has a significantly larger impact on stock price than any other day of the week.

5.2 Prediction

Figure 6 displays the forecasted values of LOSI's stock price for the next 10 days in January 2021. More specifically, the forecasted dates are 1/11/21 - 1/15/21 and 1/18/21 - 1/22/21. The model predicts that the stock price will continue to decrease, on average, over the next 10 days and hover around a share price of $96. Unfortunately, LOSI's board of directors and investors are rightfully concerned that the share price will stay below $97/share. However, when analyzing these predictions, it is important for the board of directors to keep in mind that there is certainly a margin of error and the stock price could very well increase over the next 10 days.
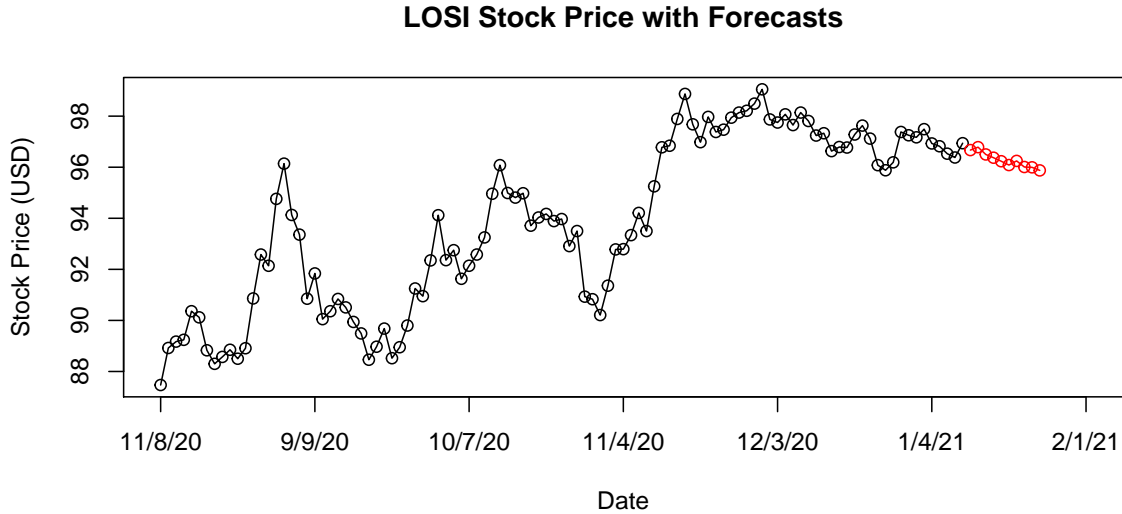
LOSI Stock Price with Forecasts

Figure 6: Forecasts of LOSI Stock Price. The black line indicates recent historical stock price data, beginning on 11/8/20. The red line indicates the forecasts for the next ten weekdays in Janaury 2021.

6 Appendix 1 - Table of Parameter Estimates

Table 2: Estimates of the forecasting model parameters with their standard errors (SE).

Parameter	Estimate	SE	Coefficient Description
β_0	40.817	0.312	Intercept
β_1	0.036	0.000	Time
β_2	-0.019	0.289	Monday
β_3	0.034	0.283	Thursday
β_4	0.164	0.283	Tuesday
β_5	0.047	0.283	Wednesday
β_6	-4.776	0.226	2018
β_7	0.289	0.127	$cos(2*pi*time*1/253)$
β_8	0.304	0.127	$cos(2*pi*time*2/253)$
β_9	0.135	0.127	$cos(2*pi*time*3/253)$
β_{10}	-0.120	0.127	$cos(2*pi*time*4/253)$
β_{11}	-0.491	0.127	$cos(2*pi*time*5/253)$
β_{12}	-0.185	0.127	$cos(2*pi*time*6/253)$
β_{13}	-1.417	0.129	$sin(2*pi*time*1/253)$
β_{14}	-0.108	0.128	$sin(2*pi*time*2/253)$
β_{15}	-0.345	0.127	$sin(2*pi*time*3/253)$
β_{16}	0.154	0.127	$sin(2*pi*time*4/253)$
β_{17}	-0.240	0.127	$sin(2*pi*time*5/253)$
β_{18}	-0.188	0.127	$sin(2*pi*time*6/253)$
ϕ_1	-0.734	.043	AR1 Coefficient
ϕ_2	0.827	0.018	AR2 Coefficient
ϕ_3	0.818	0.044	AR3 Coefficient
θ_1	1.546	0.059	MA1 Coefficient
θ_2	0.657	0.065	MA2 Coefficient
Φ	-0.038	0.033	Seasonal AR Coefficient
σ_W^2	.666		Variance of White Noise