

Modeling Gewgle's Stock Price

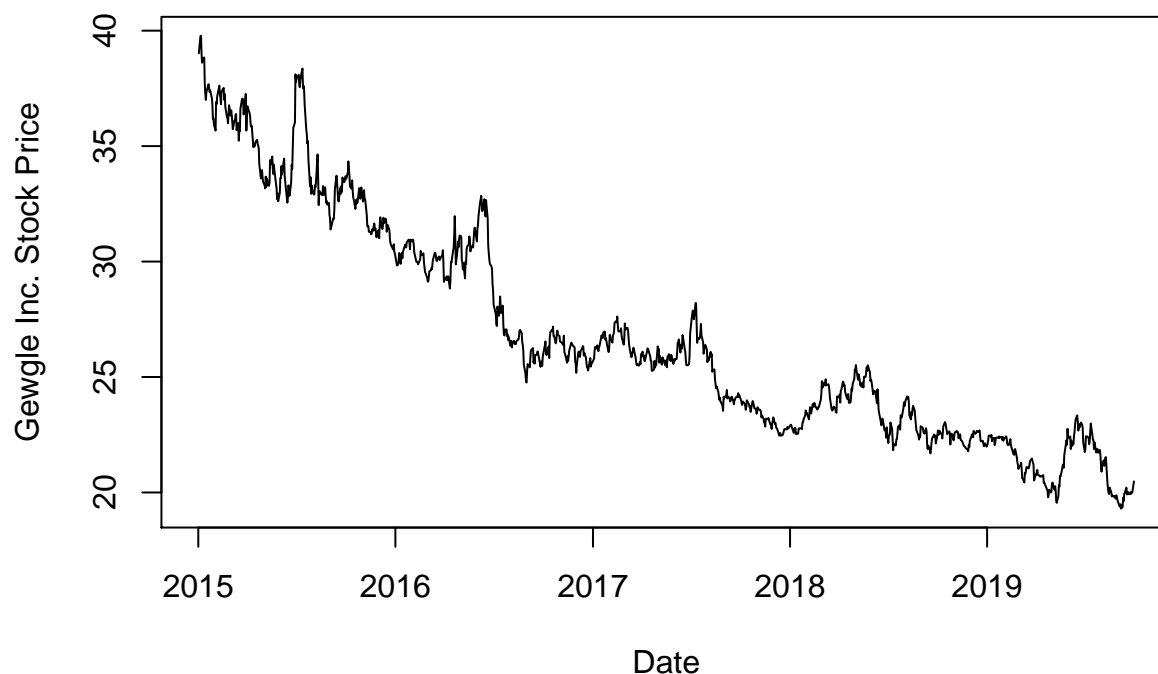
Raymond Chang

Assignment: Gewgle Incorporated, a fictional search engine company, has been struggling for a few years as the company faces a multitude of challenges. From high-level scandals to an increase in competition, Gewgle hopes to turn things around and begin Q4 on a good note. Given their stock prices from January 2015 to September 2019, forecast Gewgle's stock prices for the first 10 trading days of October 2019.

1 Exploratory Data Analysis

1.1 First Glance Observations

Figure 1: Gewgle Inc. Stock Price



After plotting Gewgle's stock prices, I noticed 3 unique features about the dataset:

1. There is a overall downwards trend.
2. The variance or "spread" is decreasing year after year.
3. There are seasonal spikes occuring around the middle of each year.

1.2 Modeling

I will be using both ARIMA (Autoregressive Integrated Moving Average) and SARIMA (Seasonal Autoregressive Integrated Moving Average) models for this project.

1.2.1 ARIMA Model: denoted as $\text{ARIMA}(p,d,q)$

The components of ARIMA are:

AR: Autoregression - a model that uses the dependent relationship between an observation and some number of lagged observations

I: Integrated - the use of differencing of individual observations (such as subtracting an observation from an observation at the previous time step) in order to make the time series stationary

MA: Moving Average - a model that uses the dependency between an observation and a residual error from a moving average model applied to lagged observations

The parameters of the ARIMA model are:

p: the number of lag observations included in the model

d: the number of times that individual observations are differenced

q: the size of the moving average “window”

1.2.2 SARIMA Model: denoted as $\text{SARIMA}(p,d,q) \times (P,D,Q)_s$

The components of SARIMA is the same as ARIMA with the addition of:

S: Seasonal - seasonal component used to model univariate data

The parameters of SARIMA is the same as ARIMA with the addition of:

P: the *seasonal* number of lag observations included in the model

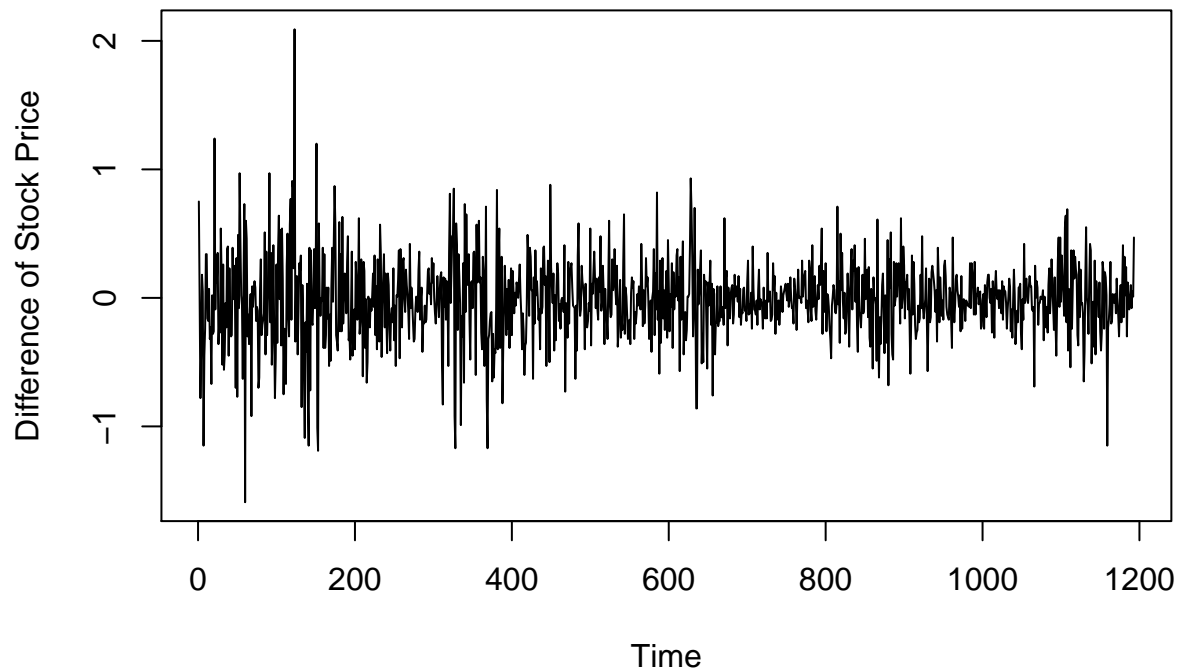
D: the *seasonal* number of times that individual observations are differenced

Q: the *seasonal* size of the moving average “window”

2 Modeling a Deterministic Function of Time

2.1 Detrending

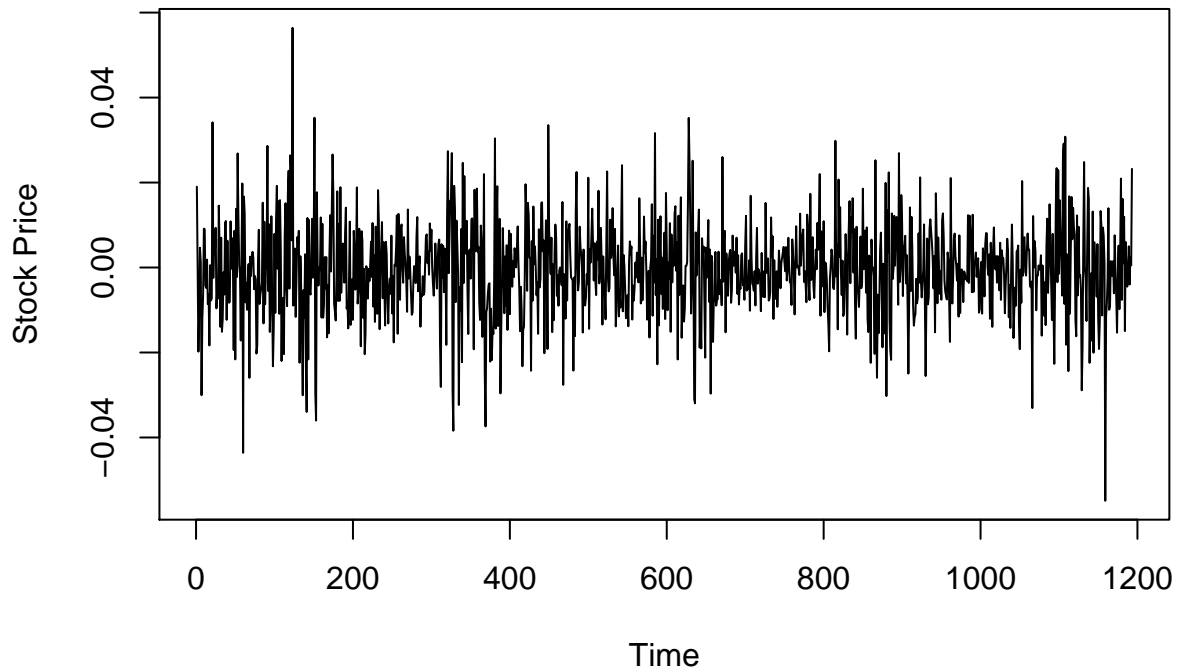
Figure 2: Differenced Stock Prices



As part of the I in ARIMA, I will take the difference of the time series to remove the downwards trend. The differenced stock prices show heteroscedasticity (a fluctuation of variance or spread of the data over time) centered at mean zero. This indicates that a variance stabilizing transform should be applied to the time series to achieve homoscedasticity (constant variance over time) and stationarity (consistent mean, variance, and covariance over time).

2.2 Variance Stabilizing Transformation

Figure 3: Difference of Log of Stock Prices



To stabilize the heteroscedasticity from the differenced data, I applied a log transformation because they are effective at removing exponential variance.

Note: The seasonality component will be accounted for in the SARIMA model.

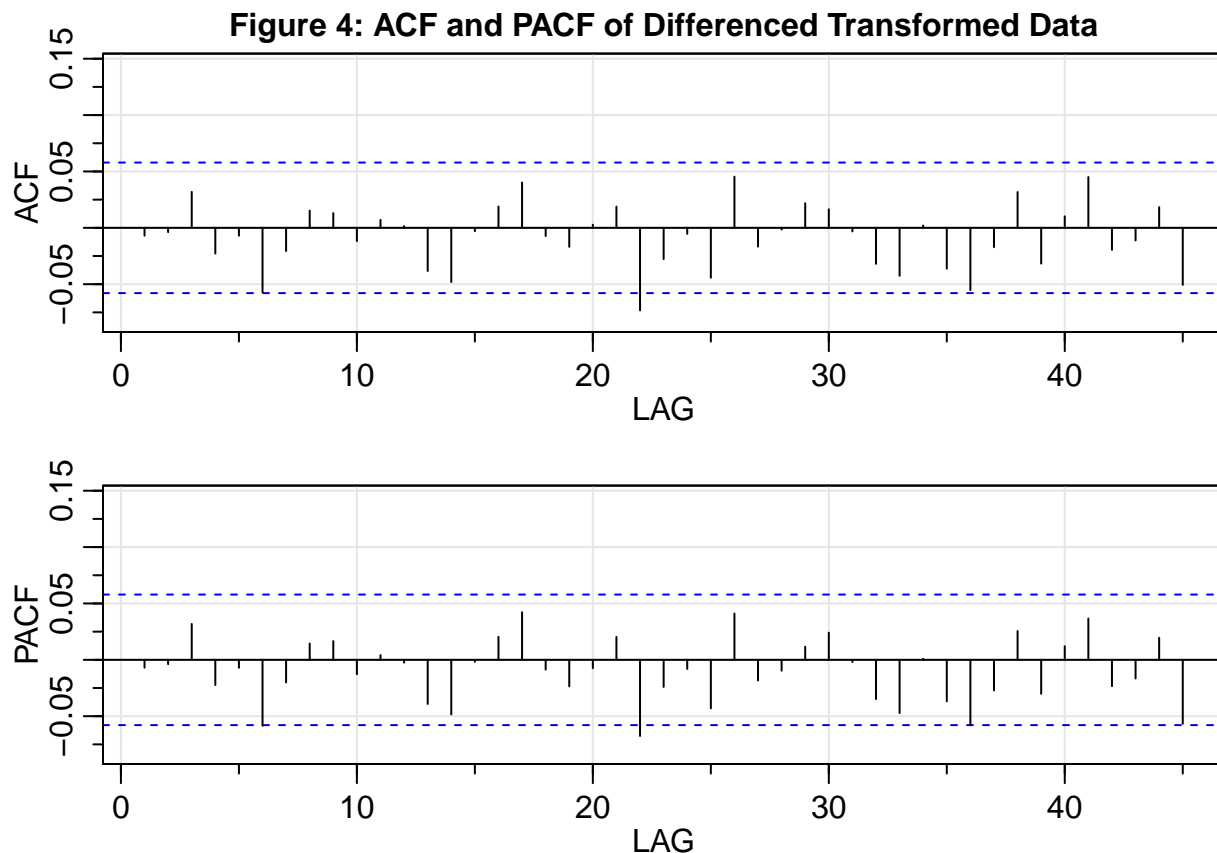
3 ARIMA Model Selection

3.1 ACF & PACF

To determine possible candidate models, I will plot the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the differenced logged data.

The ACF is a measure of the correlation between observations that are separated by k time units. The PACF is a measure of the correlation between observations that are separated by k time units after adjusting for the presence of all the other terms of shorter lag. Here is a table showing the rule of thumb for choosing p and q values from ACF and PACF plots.

Model	ACF	PACF
AR(p)	Tails off slowly	Cuts off after lag p
MA(q)	Cuts off after lag q	Tails off slowly
ARMA(p, q)	Tails off slowly	Tails off slowly



`acf2` shows that the ACF is tailing off slowly. The PACF cuts off after lag $p=1$ which indicates that an AR(1) model might be feasible for this dataset.

3.2 Generating Models

Another approach is to use `auto.arima` which gives us more precise candidates for p , d , and q .

From `auto.arima`, we get values of $p = 1$, $d = 1$, and $q = 0$.

`sarima` calculates the Akaike information criterion (AIC), Bayesian information criterion (BIC), and AICc which are all Information Criterion (IC). IC judges out-of-sample prediction errors. It also quantifies the quality of a statistical model by taking into account how well the model fits the dataset and penalizes the model based on its complexity.

A model is considered better if it has a lower information criteria than another model. AIC penalizes complex models less which indicates that it may put more emphasis on model performance on the training dataset. This in turn makes AIC prone to selecting more complex models. BIC penalizes model complexity more heavily. AICc is similar to AIC with an extra penalty term added for the number of parameters in the model.

All of the models are created using the logged data, and all except Model 3 are differenced once. Models 1, 4, and 5 have a seasonality component of period 12, 3, and 4 respectively. $s = 12$ accounts for yearly seasonality, $s = 3$ accounts for quarterly seasonality, and $s = 4$ accounts for seasonality for each third of the year.

3.3 Cross Validation Model Comparison

The AIC, BIC, and AICc values for the 5 models are:

Model	Notation	AIC	BIC	AICc
1	ARIMA(1,1,0)	-6.140505	-6.127720	-6.140496
2	ARIMA(1,1,1)	-6.138829	-6.121782	-6.138812
3	AR(2)	-6.131122	-6.114086	-6.131105
4	SARIMA(1,1,1) x (1,0,0) ₃	-6.138162	-6.116853	-6.138133
5	SARIMA(1,1,2) x (1,1,0) ₆	-5.689036	-5.667731	-5.689008

Table 1: AIC, BIC, AICc of the 5 models.

Because the Information Criteria for these 5 models are very close in value, another metric, mean squared error (MSE), will be used to pick the best model. MSE measures the deviation between the fitted values with the actual time series.

Model	Notation	MSE
1	ARIMA(1,1,0)	3.352261
2	ARIMA(1,1,1)	3.352281
3	AR(2)	3.423817
4	SARIMA(1,1,1) x (1,0,0) ₃	3.349121
5	SARIMA(1,1,2) x (1,1,0) ₆	3.945073

Table 2: Out-of-sample MSE's for our models of interest.

Model 4 will be used to predict the stock prices of Gewgle Inc.'s next 10 trading days because it has the lowest MSE out of the 5 models.

4 Results

The ARIMA model chosen for prediction is a SARIMA(1,1,1) x (1,0,0)₃ model, defined as:

$$(1 - \Phi B^3)(1 - \phi B)\nabla X_t = (1 - \theta B)Z_t \quad (1)$$

4.1 Estimation of Model Parameters

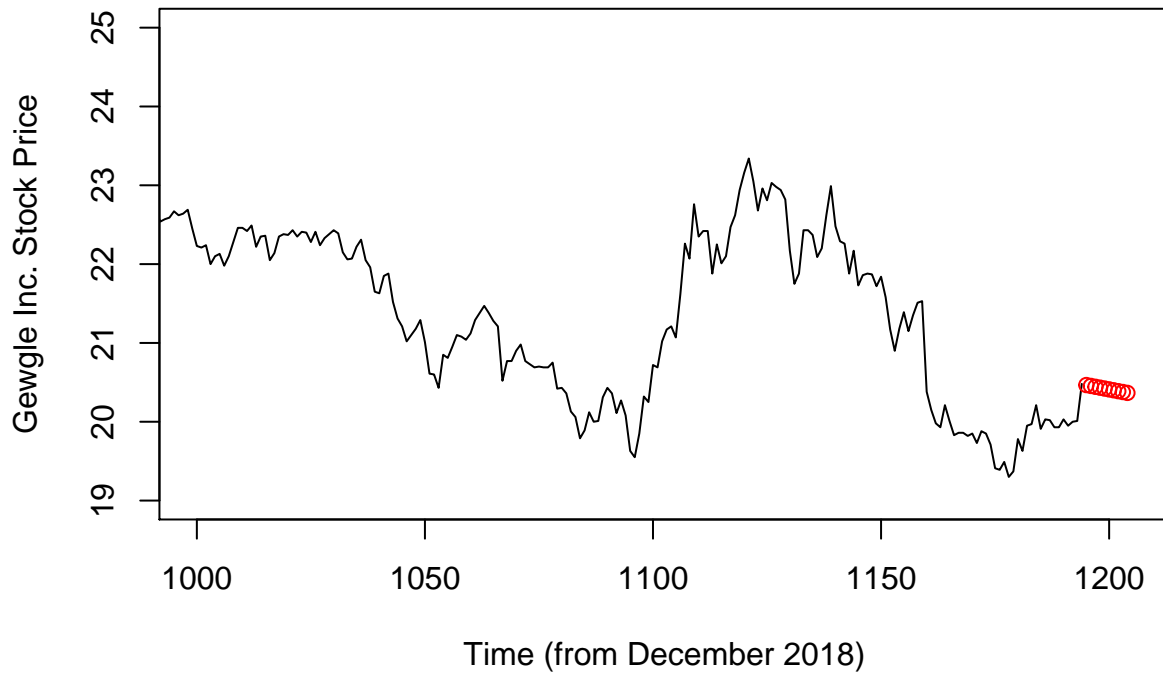
The estimates of the model parameters from **sarima** are:

Parameter	Estimate
Φ	0.0013
ϕ	-0.004
θ	-0.0031

Table 3: Model parameter estimates corresponding to variables in equation (1).

4.2 Final Prediction

Figure 5: Plot of Predicted Prices in Red



Day	Stock Price
1	20.46552
2	20.45448
3	20.44343
4	20.43238
5	20.42134
6	20.41030
7	20.39927
8	20.38824
9	20.37723
10	20.36621

Table 4: Next 10 days of stock prices projected using `sarima.for`.

4.3 Conclusion

The stock price data for Gewgle Inc. shows a downward trend with a seasonal component and heteroscedasticity. After applying a variance-stabilizing transformation and pursuing stationarity, I decided upon a SARIMA(1,1,1) \times (1,0,0)₃ model. My prediction shows a decline of about \$0.10 in stock price over the first 10 days of October, indicating that the stock is not turning around as Gewgle hoped it would.