Analyzing the Swiss Market Index: Stochastic Processes and Time Series Forecasting Techniques

Glenn Anta Bucagu

November 2020

Contents

1 Introduction

Random Walk Hypothesis Efficient Market Hypothesis

1.1 Objectives

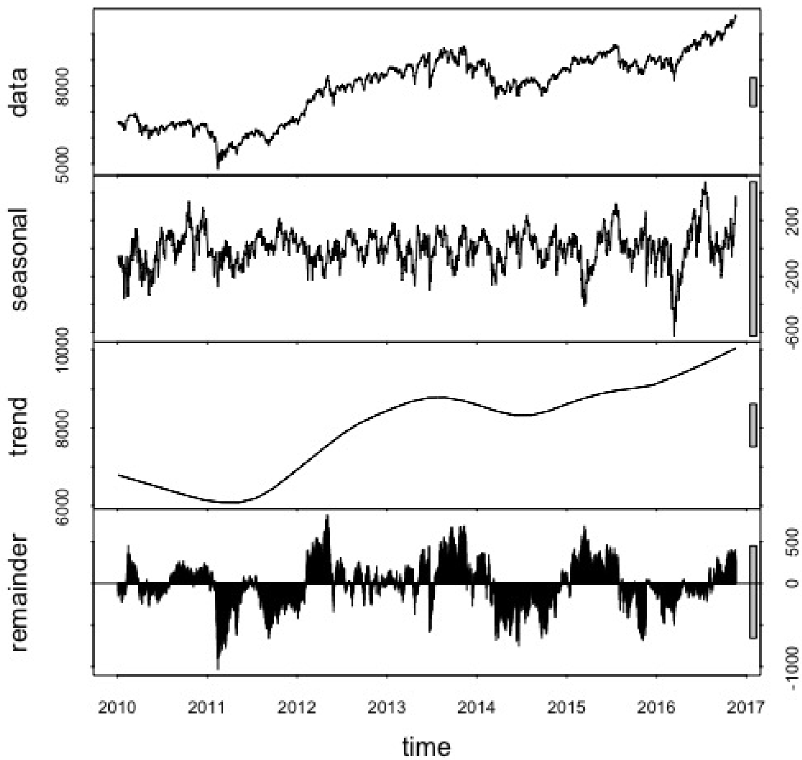
The accurate modelling and prediction of stock markets is a highly complex and intricate task. There are possibly millions of variables, conditions and events that may cause a given stock to move in a particular direction. The objective of this investigation is to compare and contrast discrete stochastic processes and machine learning methods in the prediction of the Swiss Market Index (SSMI). In particular, we will be looking at how Markov Chains and neural networks can be used to model the daily closing price of the SSMI from January 4th 2010 to December 30th 2019.

2 Exploratory Data Analysis

2.1 Time series Decomposition

Using an additive timeseries decomposition in R, we obtain the plot in figure 1

Figure 1: Additive Timeseries Decomposition of the SSMI Closing Stock Price



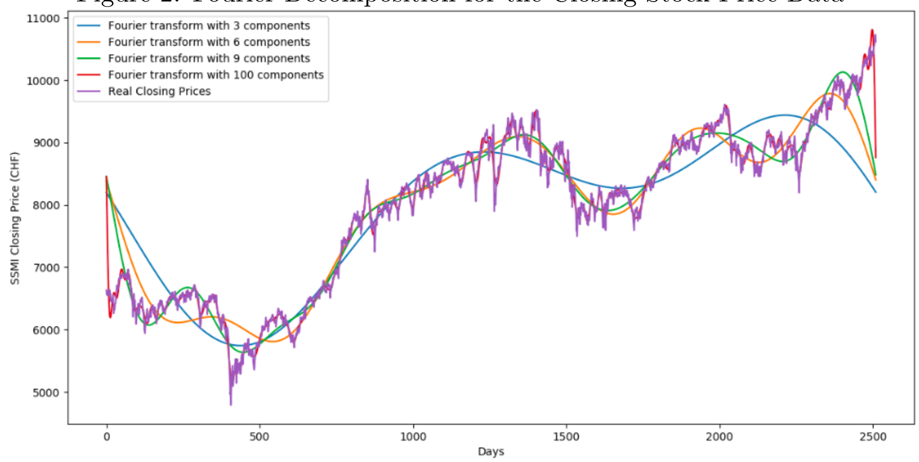
The additive time series decomposition assumes that the closing price can be expressed as the sum of a trend, seasonal and random noise component. The seasonal component is computed via loess smoothing the seasonal sub-series (every Monday etc). The seasonal component is subtracted from the data and the remainder is smoothed to compute the general trend. Iterating the process allows us to compute the remainder.
Looking at the grey vertical bars in figure 1, we see that huge variability in the seasonal component. This suggests that the seasonal component is not statistically significant in describing the data. In contrast, the general trend appears to be the most statistically significant component in describing the data.

2.2 Fourier Transform for Trend Analysis

Another interesting decomposition for the timeseries is one that uses Fourier analysis. The oscillating nature of the stock price justifies the use of Fourier analysis. Using Fourier transforms in this context allows us to extract any global or local patterns in the data.

2

Figure 2: Fourier Decomposition for the Closing Stock Price Data



We can vary the number of components to adapt our Fourier curves to the true data. As we can see from figure 2, the higher the number of components, the better the approximation. The 3 component Fourier transform essentially describes the general trend in the data, similarly to the trend component in the "stl" decomposition. While such a model lacks in explanatory power, it can have predictive power. Given a Fourier series for our dataset, we can extend the Fourier series and therefore make predictions for the stock price at a future date. The issue is that we have no statistical measure of how accurate the prediction is.

3 Stochastic Modelling

3.1 Two State Markov Chain

We model the closing price of the SSMI as a two state Markov Chain. We assume that the daily closing price is randomly distributed, and that the future closing price is only dependent on the current closing price. This is the Markov property.

3.2 Derivation and Classification of the Two State Transition Probability Matrix

For the purpose of Markov Chain modelling, we need to define a set of states. We are considering the daily closing price of the Swiss Market Index. From one day to the next, this price can either increase, decrease or decrease. Careful analysis shows that throughout the time period considered, the SSMI never stayed constant; it always either increased or decreased. Hence, for this context, we define two states: state I (increase) and state D (decrease)

$$\mathbf{P}_{simple} = \begin{array}{c c} & \begin{array}{cc} I & D \end{array} \\ \begin{array}{c} I \\ D \end{array} & \begin{pmatrix} 0.4764353 & 0.5235647 \\ 0.4556301 & 0.5443699 \end{pmatrix} \end{array}$$

Note that by the way we define the transition matrix entries, the first and last day in the dataset are unaccounted for. We only have 2508 days for the Markov Chain. We will assume the Markov Chain is stationary, meaning that the n-th step probability matrix is simply the 1-step transition probability matrix raised to the n-th power, for $n \in \mathbb{N}$.
By inspection, both states communicate with each other. Therefore, this Markov Chain contains a single communication class, namely the class {I, C}. We can therefore conclude that:

- The Markov Chain is irreducible
- The communication class {I, C} is positive recurrent
- The Markov Chain is aperiodic

3.3 Defining the Initial State Vector for the Two State Markov Chain

We can define the initial state vector $\vec{\alpha}_0 = [\alpha_{0,I}, \alpha_{0,D}]$.
$\alpha_{0,I} = \frac{1167}{2508}$
$\alpha_{0,D} = \frac{1341}{2508}$
It follows that $\vec{\alpha}_0 = [0.46531, 0.53468]$.
In practice we can find the data for January 3rd 2010, but for the purpose of the model, it is useful to define an initial state vector.

3.4 Studying Long-term behavior of the Two State Markov Chain

Forecasting the longterm behavior of the SSMI is vital for various stakeholders such as investors, bankers and various government bodies.
We begin with a few sample calculations:

$$\mathbf{P}^2_{simple} = \begin{array}{c c} & \begin{array}{cc} I & D \end{array} \\ \begin{array}{c} I \\ D \end{array} & \begin{pmatrix} 0.46554 & 0.53445 \\ 0.46510 & 0.53489 \end{pmatrix} \end{array}$$

$$\mathbf{P}^4_{simple} = \begin{array}{c} \\ I \\ D \end{array} \begin{array}{c} \begin{array}{cc} I & D \end{array} \\ \begin{pmatrix} 0.46531 & 0.53468 \\ 0.46531 & 0.53468 \end{pmatrix} \end{array}$$

$$\mathbf{P}^8_{simple} = \begin{array}{c} \\ I \\ D \end{array} \begin{array}{c} \begin{array}{cc} I & D \end{array} \\ \begin{pmatrix} 0.46531 & 0.53468 \\ 0.46531 & 0.53468 \end{pmatrix} \end{array}$$

We begin to see a pattern. After 4 trading days, the transition probability matrix seems to stabilize. We can predict the long-term behavior of this Markov Chain by computing the steady-state probability distribution. For a formal argument, we apply the Basic Limit Theorem.
Recall that the Markov Chain is aperiodic, recurrent and irreducible. It follows that a unique steady-state probability distribution $\vec{\pi} = [\pi_I, \pi_D]$ must exist and satisfy the following constraints:
$\vec{\pi} = \vec{\pi}\mathbf{P}_{simple}$
$\pi_I + \pi_D = 1$

The unique steady-state probability distribution is $\vec{\pi} = [0.46531, 0.53468]$. It follows that,

$$\lim_{n\to\infty} \mathbf{P}^n_{simple} = \begin{array}{c} \\ I \\ D \end{array} \begin{array}{c} \begin{array}{cc} I & D \end{array} \\ \begin{pmatrix} 0.46531 & 0.53468 \\ 0.46531 & 0.53468 \end{pmatrix} \end{array}$$

The mean recurrent time m_s for each state s can be computed as follows: $m_I = \frac{1}{\pi_I} \approx 2.149$ and $m_D = \frac{1}{\pi_D} \approx 1.870$. This suggests that on average, it takes about two trading days for the SSMI closing price to decrease and also two trading days for it to increase.

3.5 Five State Markov Chain

In the financial market, interest lies in predicting future stock price. The two state Markov model is limited because we only have an indication on the stock price movement. It provides no indication on the absolute change in the stock price. Defining a five state Markov chain may provide a better insight on the change in stock price. Note that the largest consecutive day-to-day decrease in the SSMI closing stock price is CHF -779.59 and the largest consecutive day-to-day increase in the SSMI closing stock price is CHF 289.90. We split this range of values into 5 equal intervals to define our states as follows:

- State 1 - Day-to-day closing stock price change in [-798.00, -580.40]

- State 2 - Day-to-day closing stock price change in (-580.40, -362.80]
- State 3 - Day-to-day closing stock price change in (-362.80, -145.72]
- State 4 - Day-to-day closing stock price change in (-145.72, 72.40]
- State 5 - Day-to-day closing stock price change in (72.40, 290.00]

Given the above states, we obtain the following transition probability matrix:

$$\mathbf{P}_{complex} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0.10290 & 0.5882 & 0.3088 \\ 0.000471 & 0 & 0.02496 & 0.8587 & 0.1159 \\ 0 & 0 & 0.02540 & 0.8254 & 0.1492 \end{pmatrix} \end{array}$$

The set 3, 4, 5 forms a communication class. The sets 1 and 2 form communication classes. By inspection, all three communication classes must be recursive.

3.6 Long Term Behavior of the Five State Markov Chain Model

Our Markov chain is not irreducible, so we cannot apply the Basic Limit theorem to study the long term behavior. By diagonalizing the matrix, it can be shown that two of the eigenvalues have non-zero imaginary parts. Taking the limit of the power of the matrix of eigenvalues, it can be shown that there does not exist a steady-state probability distribution for the five state transition probability matrix. However, by inspecting the transition probability matrix, we can deduce that the general trend in the SSMI price index over the given time period is increasing. While there is strong explanatory power in this model, there is weak predictive power. We still cannot make statements about future SSMI closing prices.

The modelling of stock prices as Discrete Time Markov Chain is a valuable academic exercise. However, the model is inherently limited for several reasons. Firstly, the underlying assumptions may not hold true in the real world. In reality, it can be argued that the SSMI index is entirely deterministic (not random), but it depends on millions of features and variates. Examples include, associated indices, investor confidence, and geopolitical factors. Due to this, the Markov property may not hold either.

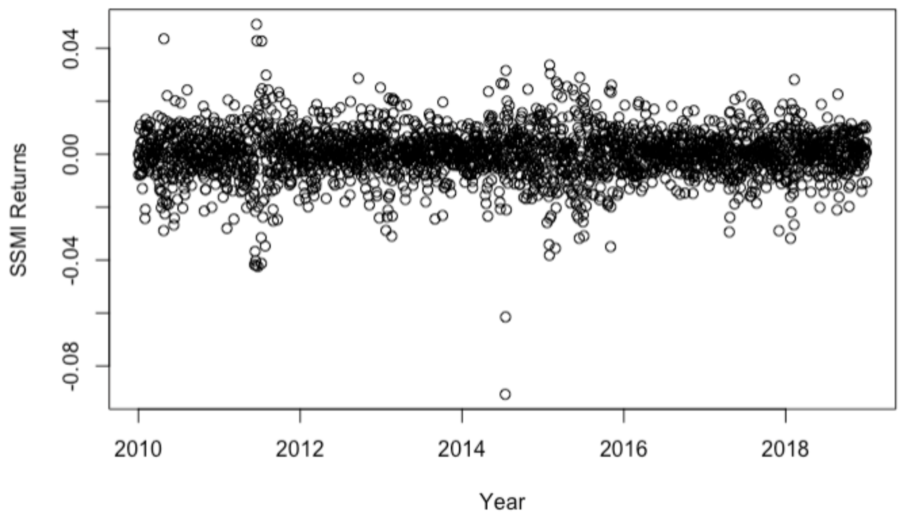
3.7 Defining the Geometric Brownian Model

Define $S(t_i)$ as the closing SSMI stock price (in CHF) on day t_i. Define $r_i = \log(\frac{S(t_i)}{S(t_{i-1})})$ to be the log return of the SSMI closing stock price. We will assume the our model follows the following stochastic differential equation:

$$dS_t = \mu S_t dt + \sigma S_t dW_t \tag{1}$$

where μ denotes the percentage drift, σ denotes the percentage volatility and W_t is a probability distribution function associated with Brownian motion or a Wiener process.

Figure 3: SSMI Log Return



For our model, whose log returns are illustrated in figure 3, we assume that $r_i = \log(\frac{S(t_i)}{S(t_{i-1})}) = \mu + \epsilon_i$, with $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$. This model is inherently stationary, and as such, does not necessarily reflect reality. A more complex model could relax the stationary assumption and provide a time-dependent log return mean.

3.8 Evaluating the Normality of the SSMI Return

We now perform a goodness-of-fit hypothesis test to evaluate the normality assumption. The null hypothesis is that the SSMI log returns are normally distributed. The alternate hypothesis is that the SSMI log returns are not normally distributed. We compute the Jarque-Bera Test statistic (in R with the tseries package) for this problem to be 4501.306 with a p-value less than $2 \cdot 10^{-16}$. Therefore, at the 1% significance level, there is insufficient evidence to support the normality of log returns assumption.

3.9 Estimating Drift and Volatility

Per our model, $\{r_1, r_2, ..., r_n\}$ is a set of independent and identically distributed random variables. Our previous discussion suggests that it is not realistic to assume that drift and volatility are not constant throughout our 10-year time period. This leads to:

$$E[r_i] = (\mu - \frac{1}{2}\sigma^2)\Delta t \tag{2}$$

$$Var[r_i] = \sigma^2 \Delta t \tag{3}$$

The drift estimate $\hat{\mu}$ corresponds to the sample mean:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} r_i \tag{4}$$

The sample variance S_μ^2 is expressed as:

$$S_\mu^2 = \frac{1}{n-1}\sum_{i=1}^{n}(r_i - \mu)^2 \tag{5}$$

The volatility estimate $\hat{\sigma}$ is:

$$\hat{\sigma} = \frac{S_\mu}{\sqrt{\Delta t}} \tag{6}$$

Estimating the location and scale parameters can be reduced to maximum likelihood estimation problems, given that we have a reasonably accurate distribution for the log returns.

3.10 Non-parametric Estimation Via Bootstrap Methods

One way to estimate statistics such as the drift for this model can also be achieved by generating Bootstrap simulations. We know that using the sample mean is not likely to be accurate since our previous hypothesis test suggests the log returns for the SSMI are not normally distributed. To reduce estimation errors, using large number of data points (including re-sampling) is preferred. We begin by assuming $r_i = \log(\frac{S(t_i)}{S(t_{i-1})}) \sim F$ where F denotes an arbitrary probability distribution. Assuming all observations are identically and independently distributed, a sample of observations $S = \{r_1, ..., r_n\}$ can be used to estimate a probability distribution for the estimator $\tilde{\theta}(S_i)$. This is achieved by randomly sampling the data with replacement, drawing random re-samples S_i of size n from S. This process is repeated N times. By the Central Limit Theorem, the distributions of means will approach normality. Letting $\hat{\theta}_i$ denote the arithmetic mean of each S_i.

Using the two techniques outlined (sampling and bootstrap) with $N = 5000$ re-samples, we obtain estimates for out drift μ and 95% confidence intervals. The results are summarized below:

$\hat{\mu}_{\mathbf{sample}}$	$\hat{\mu}_{\mathbf{bootstrap}}$	$\mathbf{95\%CI_{sample}}$	$\mathbf{95\%CI_{bootstrap}}$
0.0001875329	0.0001875329	[0.0001831941, 0.0001918717]	[-0.0002, 0.0005]

For the volatility, we have:

$\hat{\sigma}_{\mathbf{sample}}$	$\hat{\sigma}_{\mathbf{bootstrap}}$	$\mathbf{95\%CI_{sample}}$	$\mathbf{95\%CI_{bootstrap}}$
0.009346813	0.009346813	[0.00073, 0.02]	[0.0089, 0.0100]

3.11 Solving the Stochastic Differential Equation

Recall, our stochastic differential equation:

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t) \tag{7}$$

Letting $f = \log(S(t))$, we apply Ito's lemma to obtain:

$$df = d(\log(S(t))) = \left(\frac{\partial f}{\partial t} + \frac{\partial f}{\partial S}\mu S(t) + 0.5\frac{\partial^2 f}{\partial S^2}\sigma^2 S(t)^2\right) dt + \sigma S(t)dW(t) \tag{8}$$

Letting $\frac{\partial f}{\partial t} = \frac{\partial \log(S(t))}{\partial t} = 0$. Applying the chain rule to all remaining partial derivatives, we obtain:

$$d(\log(S(t))) = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma dW(t) \tag{9}$$

Integrating both sides of this equation with respect to t, we get:

$$\log(S(t)) - \log(S(0)) = \left(\mu - \frac{\sigma^2}{2}\right) t + \sigma(W(t) - W(0)) \tag{10}$$

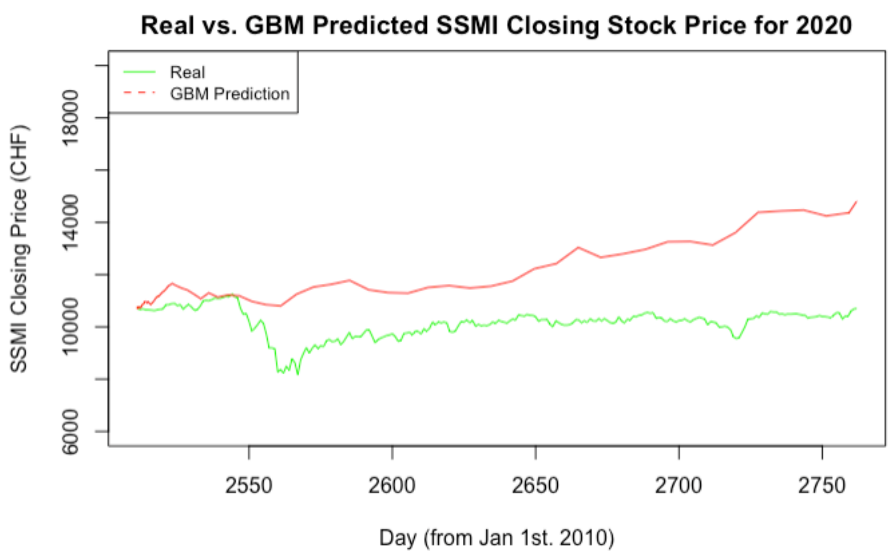
Given that we have simple geometric Brownian motion in the movement of the stock price $S(t)$, it follows that $W(t) \sim \mathcal{N}(0, t)$ for $t \geq 0$. Our solution for the initial stochastic differential equation becomes:

$$S(t) = S(0)\exp\left((\mu - \frac{\sigma^2}{2})t + \sigma W(t)\right) \tag{11}$$

3.12 Predicting SSMI Stock Prices for 2020

We have an exact solution to our stochastic differential equation. Given parameter estimates for the drift and velocity, we can predict future (January to December 2020) SSMI closing stock prices and see how they compare to the true SSMI closing stock prices.

Figure 4: Real vs. GBM Predicted SSMI Closing Stock Price for 2020



Observe in figure 4, we have the real and GBM-predicted SSMI closing stock price for the year 2020. We observe several qualitative differences. Firstly, the predicted price show an overall general rising trend from January to December 2020. In reality, we know that most global were affected by the outbreak of Covid-19. For European countries like Switzerland, this happened around March 2020, which is why we observe a noticeable drop in the SSMI closing price. Before the outbreak of Covid-19, we see that the real and predicted stock prices are almost equal (difference of less than 300 CHF). The prediction appears more smooth, and that is due to the idyllic nature of our model. Including time-dependent drift and volatility might better reflect the movement of the real SSMI stock price. However, the sheer number of variables to account for such as global context and related stocks, calls for more sophisticated prediction techniques.

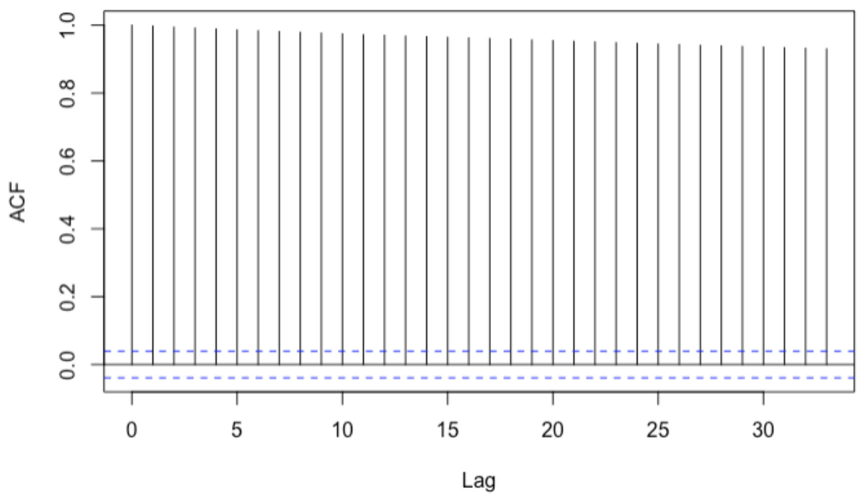
4 Time series Forecasting Techniques

4.1 Testing for Stationarity

If we think of the stock price $S(t) : t \in \mathbf{Z}^+$, one property that makes forecasting more convenient, is the stationarity of the timeseries. Generally speaking, we think of stationarity as regularity in the timeseries regardless of the window of data observed. In general, there are two levels of stationarity. A time series $S(t)$ is strongly stationary if $S(t_1), ..., S(t_k) \stackrel{D}{=} S(t_1 + h), ..., S(t_k + h)$, for

$t_1, ..., t_k, h \in \mathbf{Z}^+$. A time series $S(t)$ is weakly stationary if $\mathbb{E}[S(t)]$ is independent of t, and $\mathbb{E}[(S(t) - \mathbb{E}[S(t)])(S(s) - \mathbb{E}[S(s)])] = f(|s - t|)$ for some function f, for any $s, t \in \mathbf{Z}^+$. Let us inspect the autocorrelation function for the stock price data:

Figure 5: Autocorrelation Function for the SSMI Daily Closing Price



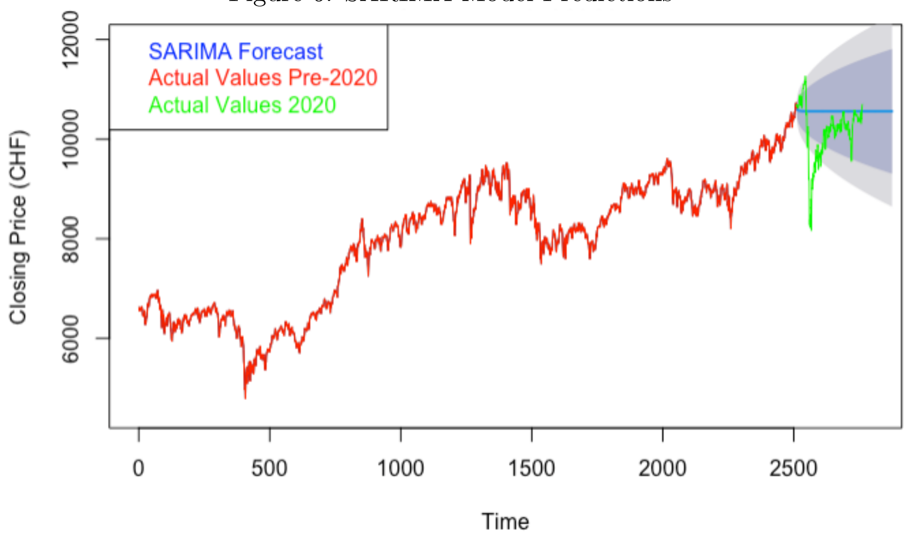
Notice in figure 5 that the ACF is quite high for all lags up to and including 30. We also note that the ACF remains outside the blue bands, which represent the 95% ACF prediction interval for a strong white noise. This is indicative of high serial dependence in the daily closing stock price. For a relatively stable stock such as the SSMI, this is intuitive. This is corroborated by the relatively low stock volatility we computed in the idealized GBM model. In summary, there is strong evidence to suggest that the SSMI daily closing stock price is neither strongly nor weakly stationary.

Several operations can be performed on the observed timeseries to render it more stationary. Two popular choices are de-trending and timeseries differencing.

4.2 SARIMA Model

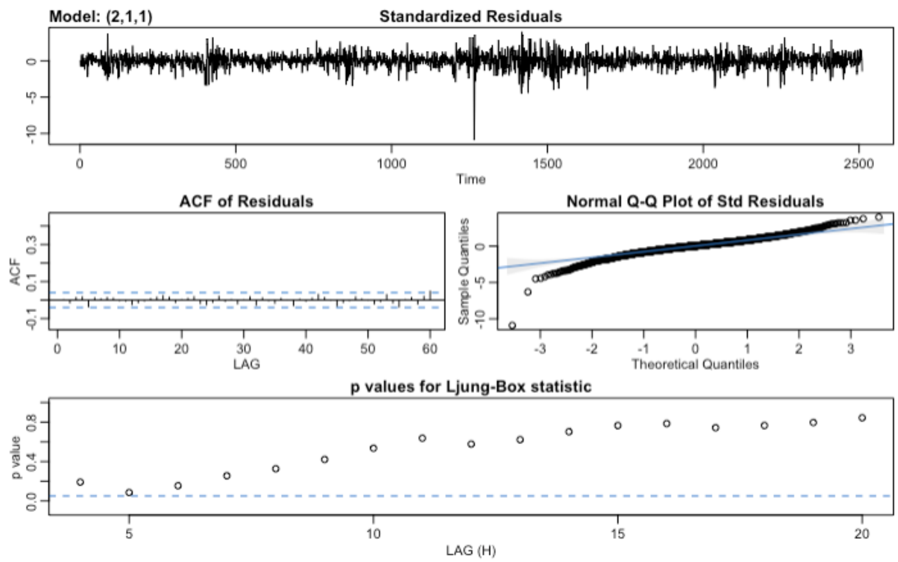
One of the most common and highest performing timeseries modelling techniques is the SARIMA model. Such models are generally parsimonious, computationally efficient and provide easy to compute prediction intervals. For our data, we will automatically generate a SARIMA model which minimizes the AIC over the parameter space where p, d, q, P, D, Q are within [0, 10]. We will use maximum likelihood estimation to estimate each parameter. We obtain the plot in figure 6:

Figure 6: SARIMA Model Predictions



Notice that the prediction for 2020 appears more or less constant. The SARIMA forecast does not quite capture the stock price volatility, nor does it capture the generally increasing trend we see over the first 2500 days of stock data. The 2020 stock price generally stays within the SARIMA 95% prediction interval. In terms of its predictive power, the SARIMA model appears to have a similar performance to the GBM model. Nonetheless, we run some diagnostics to thoroughly evaluate the strength of this model.

Figure 7: SARIMA Model Diagnostics

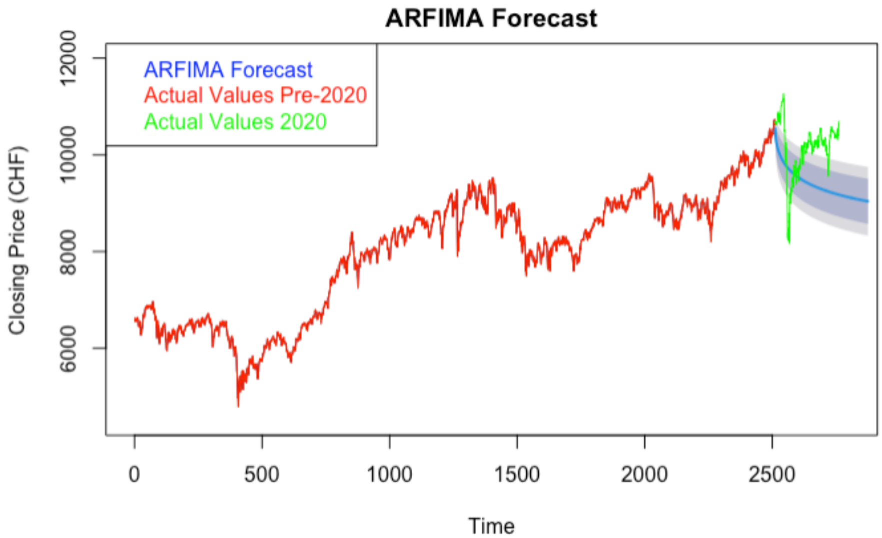


Looking at figure 7, we see the SARIMA residuals are approximately Gaussian, although the extremities of the QQplot suggest there are heavy tails in the residual empirical CDF. We see that the ACF values all remain within the 95% prediction interval for a strong white noise, which suggests our model is likely stationary. This is corroborated in the relatively high p-values for the Ljung-Box statistic. Ultimately, the SARIMA model seems to fit the data relatively well, although can be improved. We also note that 2020 is an abnormal year in terms of financial data, so that may account for at least some of the prediction inaccuracy.

4.3 ARFIMA Model

Recall that the ACF for our stock price data decays very slowly. We may even posit that the ACF is not absolutely summable over all integer lag values. In this case, we say our stock price data exhibits long-range time dependence. This may explain why a standard SARIMA model is unable to truly capture the patterns in this data. An alternative would be to use a fractional ARIMA model, where the differencing parameter $d \in (0, 0.5)$. These values of d have been shown to accurately portray timeseries with long-range time dependence all while being stationary. Using maximum likelihood estimation for parameter estimation and AIC minimizing for model selection, we obtain an ARFIMA model whose forecast plot is depicted in figure 8.

Figure 8: ARFIMA Model Predictions

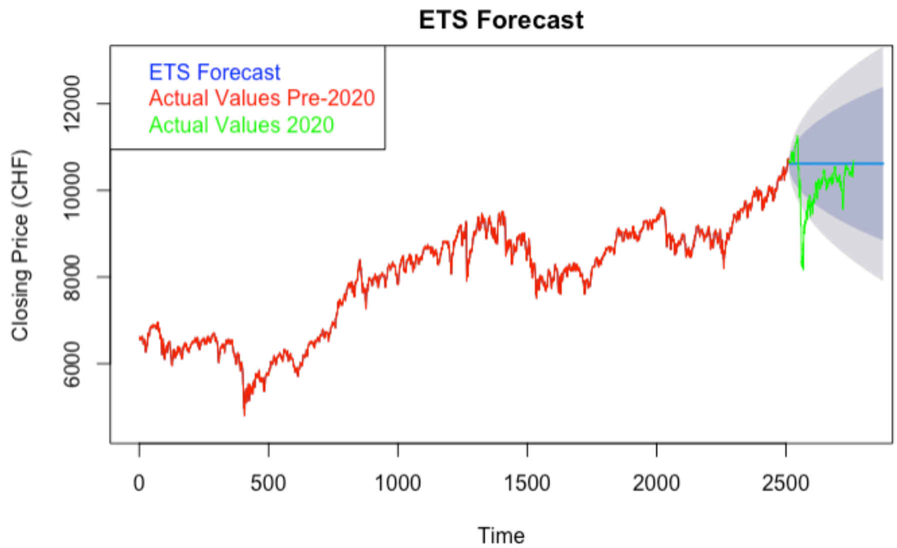


The ARFIMA model predicts a downward trend for the year 2020, which as we can see from the true 2020 values, is not generally true. Despite having smaller prediction intervals that the SARIMA model, the true 2020 values are not generally in the ARFIMA prediction intervals. This suggests the ARFIMA model does not have as much predictive power as the SARIMA model for our particular timeseries data. The underlying assumption of long-range time dependence in the stock data may therefore not be true.

4.4 ETS Model

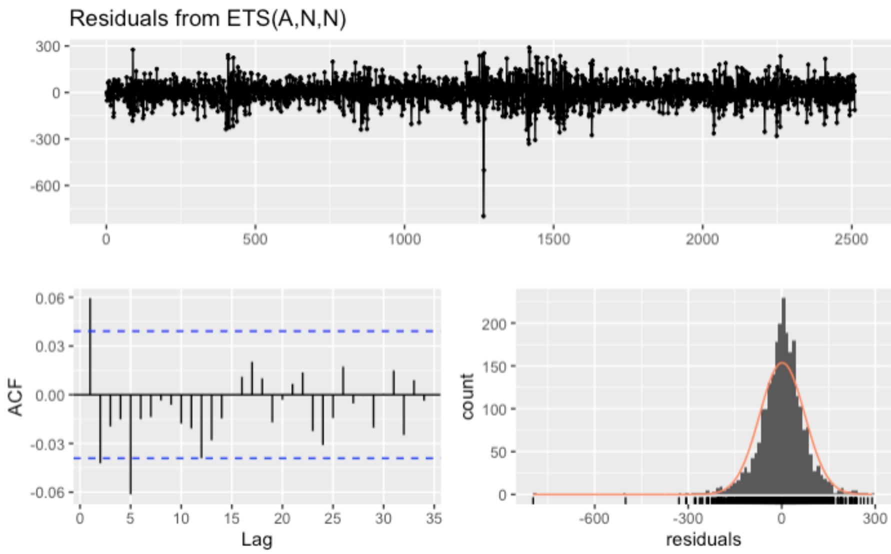
We can also explore state space models, which generally provide more flexibility than SARIMA (and ARFIMA) models. Exponential Smoothing Models (ETS) have been shown to be high performing in the M4 timeseries forecasting competitions. We therefore use an ETS model to forecast the SMI stock price. In particular, we minimize AIC for model selection. The ETS forecast is illustrated in figure 9.

Figure 9: ETS Model Predictions



Visually, the ETS forecast looks very similar to the SARIMA forecast. The true stock price is generally within the prediction interval, except for the first quarter of 2020. However, this is expected due to the impact of Covid-19 on global markets in early 2020. The actualy forecast does not seem to reflect the volatility or generally increasing trend that we seein the SMI stock price for the first 2500 days. We carry out model diagnostics for more thorough scrutiny of the ETS model.

Figure 10: ETS Model Diagnostics



Looking at figure 10, the ETS residuals are approximately Gaussian with a slightly higher peak than expected. Nonetheless residual ACF is generally within the prediction interval for a strong white noise. This suggests that the residuals are independent and indentically distributed Gaussian realizations. The ETS model is therefore a relatively good fit, compared to all previously explored models.

4.5 GARCH Model

4.6 Prophet Model

https://towardsdatascience.com/time-series-forecasting-predicting-stock-prices-using-facebooks-prophet-model-9ee1657132b5

5 Evaluation