Time series analysis on the stock price of Tesla Inc.

Wenhao Pan (3034946058), Ruojia Zhang, Mengzhu Sun, Xiangxi Wang, Mingmao Sun

November 13, 2021

Contents

1 Abstract

2 Introduction

3 Data Description

4 Exploratory Data Analysis

To obtain a comprehensive understanding of the data, we conduct explanatory data analysis (EDA) first. Figure 1(a) is the time series plot of all the given time points. We observe that the stock prices of Tesla before 2020 are averagely and considerably lower than those after 2020. The significantly different scales of different parts of the time series make it hard to visually examine the trend and seasonality pattern of the time series. Moreover, since we are majorly interested in the recent activities of Tesla, it is unnecessary for us to analyze all the available data. Therefore, for the sake of interest and convenience, we decide to only analyze the last 300 time points, which cover the period from `2020-08-26` to `2021-11-02` excluding weekends. Thus, whenever we use the word "data" in the following analysis, we implicitly mean the time series of last three hundred time points.

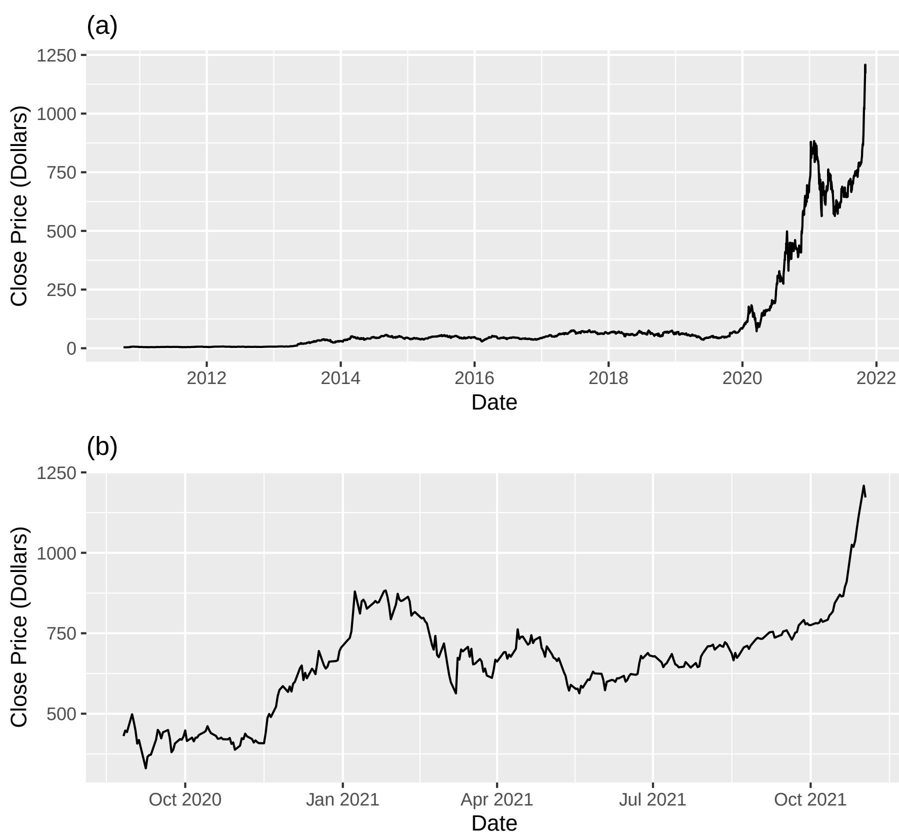

Figure 1: (a) Time series plot of all available trading days. (b) Time series plot of last 300 trading days

Figure 1(b) is the time series plot of the close prices of Tesla in last three hundred trading days before and including `2021-11-02`. We first observe that our data is roughly homoscedastic based on Figure 1(b). To verify our observation, we try the square root and natural log transformations and see whether they effectively stabilize the variance of the time series. Their plots are below in Figure 2.

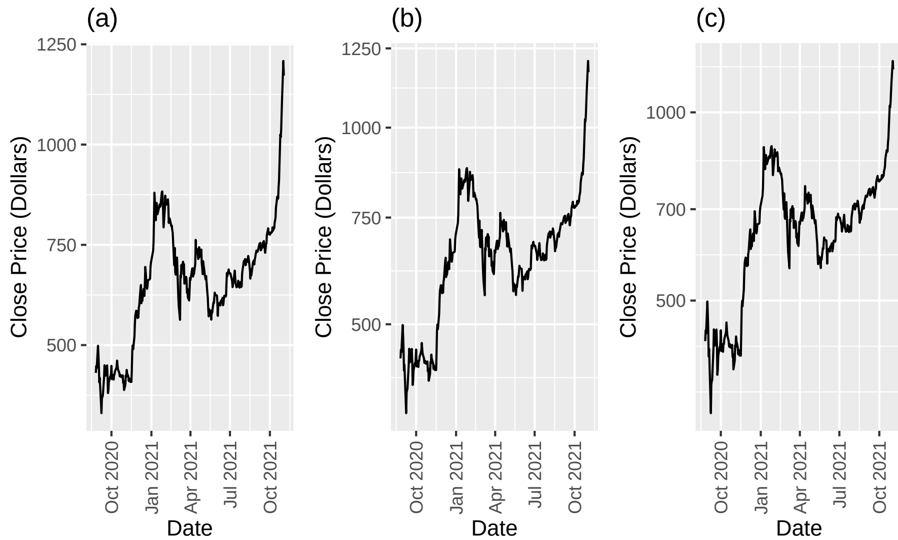

Figure 2: (a): Original time series. (b): Square root transfromed time seires. (c): Natural log transformed time series.

We can see that both transformations unnecessarily increase the variance of the time series before mid-November in 2020 and do not change the variance of other time series data. Although both transformations shorten the vertical distance between the maximum and minimum of the time series after Oct. 2021, the spike after Oct. 2021 is more like an increasing trend rather than a huge fluctuation. In short, both transformations are redundant, and we do not need to use any variance stabilizing transformation.

Back to Figure 1(b), intuitively, the data is not stationary because there exists a nonlinear and generally increasing trend. The trend first increases until around Feb. 2021 and then decreases until around Mid-May. 2021. Finally, the trend increases again until the end of the time series. Nonetheless, we do not observe an obvious or significant seasonality pattern. It matches the intuition since the granularity of our data is day, and the structure of stock price data is too complicated to have a seasonality pattern.

In conclusion, based on all the previous discussion in EDA, we decide to construct possible models on the original time series data including only the last three hundred time points.

5 Model Construction

With a comprehensive understanding of our data, we start to experiment and construct different time series model. We choose and build two non-parametric signal models of the trend and seasonality in our data. We aim to make the residuals approximately weekly stationary. We do not consider any parametric trend model because we think the trend of the stock price data is too complicated to be modeled by a parametric model, such as a high-order polynomial. Certainly, we could use a 15 or 20 order polynomial, but it may overfit the training data and produce imprecise predictions. We do not consider a parametric seasonality model either by the analysis at the end of the EDA section. Finally, based on each signal model, we provide two ARMA models or its extension, such as SARMA or ARIMA, to whiten the residuals of the signal model. Thus, we have four candidate models, and we will explain how we select a final model among them in the next section.

5.1 Non-parametric Signal Model: exponential smoothing

In this model, we choose exponential smoothing with weight $\alpha = 0.8$ and lag $k = 10$ and a seasonal differencing with period $d = 5$.

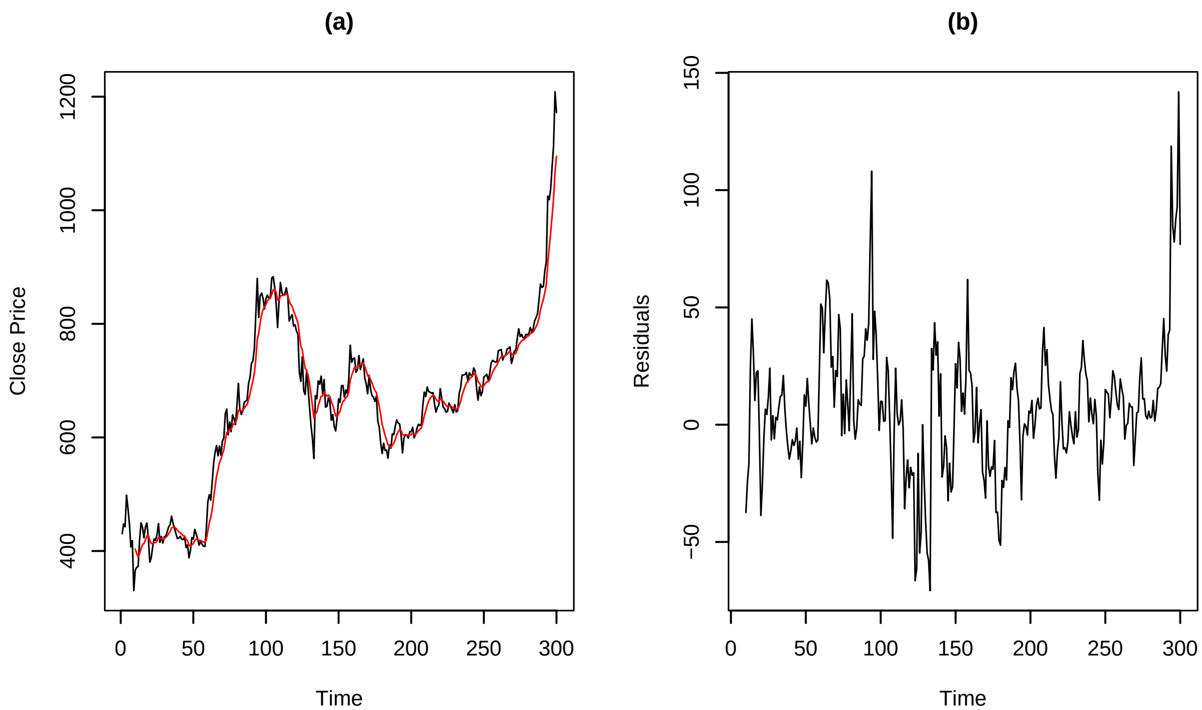

Figure 3: (a): Time series plot of the original data and fitted values. (b): The residual plot of exponential smoothing.

We experiment with different combinations of α and k with a careful consideration of overfitting issue. we choose $k = 10$ as the final value because we want to only use past two weeks, which are ten days in our data, to forecast. We choose $\alpha = 0.8$ as the final value because it best balances the smoothing effect and the capture of trend pattern. Indeed, the smoothing line in Figure 4(a) fits the data in the way that we want. Note that we lose the first nine time points due to the computation process of the smoothing filter.

However, the residual plot Figure 4(b) is fairly non-stationary, as it has cycling fluctuation pattern. We use the seasonal differencing with period $d = 5$, which is one week in our data, on the residuals to remove the pattern.

We believe that the time series of the differenced residuals shown in Figure 5(a) is sufficiently stationary.

5.2 Non-parametric Signal Model: second-order differencing

In this model, we choose a second-order differencing. We observe that after the first-order differencing, there is still some trend pattern, such as the increasing one between 270 and 300, as shown by Figure TODO. Thus, we take another differencing and acquire the second-order differencing data shown in Figure TODO.

The second-order differenced data looks more stationary than the first-order differenced data. Moreover, we are satisfied with the approximate stationarity of the second-order data, so we do not need to consider the seasonality any further.

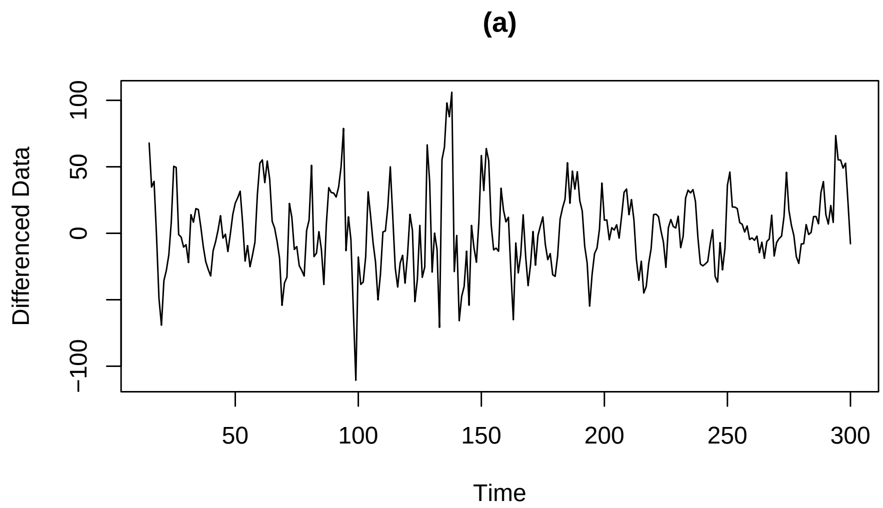

Figure 4: (a): Time series plot of the seasonal differenced (d = 5) residuals from the previous smoothing.

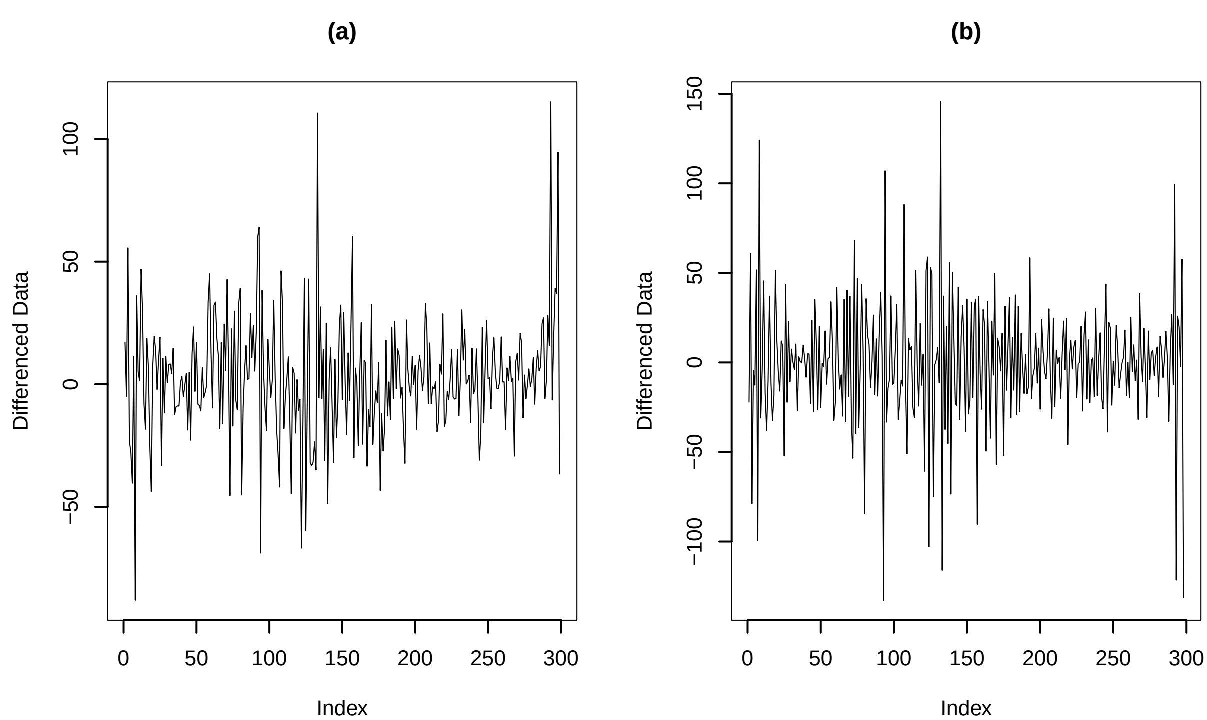

Figure 5: (a): The first-order differenced data. (b): The second-order differenced data.

6 Model Comparision and Selection

7 Final Model

7.1 Model interpretation

7.2 Prediction

8 Conclusion