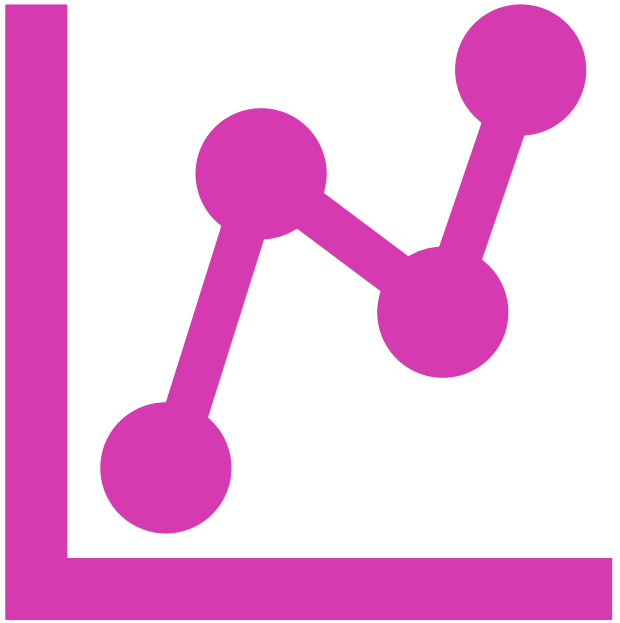
TECHNISCHE HOCHSCHULE DEGGENDORF THD

Stock Price Prediction using ARIMA Model

Allan Johns
Amritesh Vadakkedath Babu
Asin Sona Thomas
Snehagee Keshavan Saji

Introduction

- **Time series analysis** is a statistical technique that deals with time-ordered data points. It is widely used in various fields such as economics, finance, environmental studies, and more.
- In the **financial sector**, predicting stock prices is crucial for making informed investment decisions.
- Accurate forecasting models can provide significant advantages to traders and investors.



Objective and Data Source

Objective

- To predict Tesla's stock prices using the ARIMA model.

Data Source

- Yahoo Finance - Tesla - Historical Stock Price Data
- Period – Daily
- Start Date - 2020-01-01
- End Date - 2024-06-17

ARIMA – An Overview

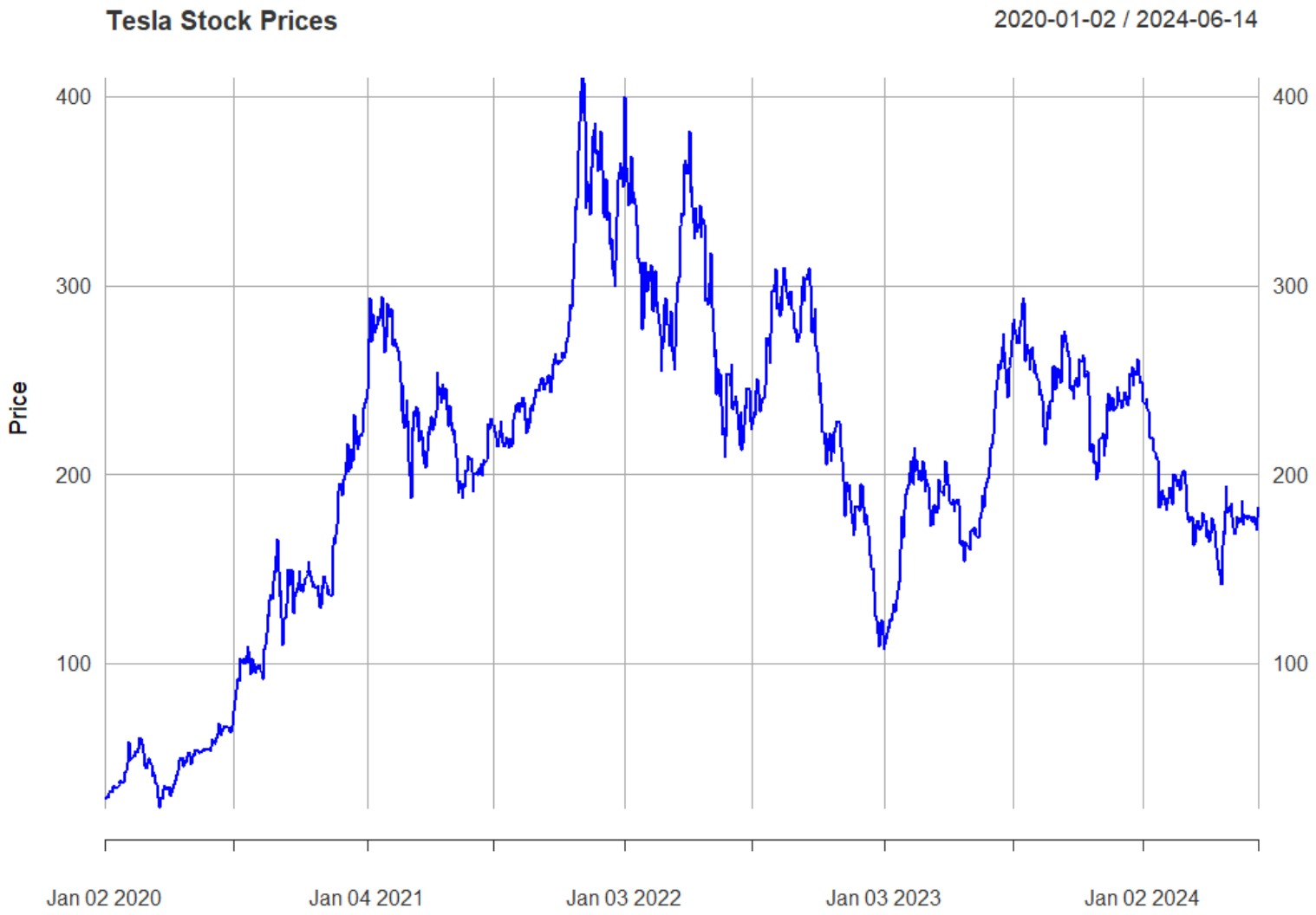
- Auto-Regressive Integrated Moving Average.
- It is a popular and widely used statistical method for time series forecasting.
- ARIMA models are capable of capturing different standard temporal structures in time series data.

Components of ARIMA

The ARIMA model is characterized by three parameters: ***p***, ***d***, and ***q***:

- ***p*** : The number of lag observations included in the model (autoregressive part).
- ***d*** : The number of times that the raw observations are differenced to make the time series stationary (integrated part).
- ***q*** : The size of the moving average window (moving average part).

Visualize the Time Series



Visualize the Time Series

- Significant growth early in the period, peaking in early 2021.
- High volatility with notable peaks and troughs.
- Major peak in early 2021 (~$400), significant troughs mid-2022 and late 2023.
- Frequent fluctuations indicate market reactions to various events.
- Despite volatility, the stock price remained higher than the start.

Stationarize the Series

- **What is Stationarizing ?**

➢ Stationarizing a time series involves transforming it to make its statistical properties (mean, variance, autocorrelation) constant over time.

- **Why is Stationarizing Important?**

➢ Many time series models, including ARIMA, assume the series is stationary for accurate modeling and forecasting.

- Stationary series are easier to model and predict because their behavior is consistent over time.

Test for Stationarity

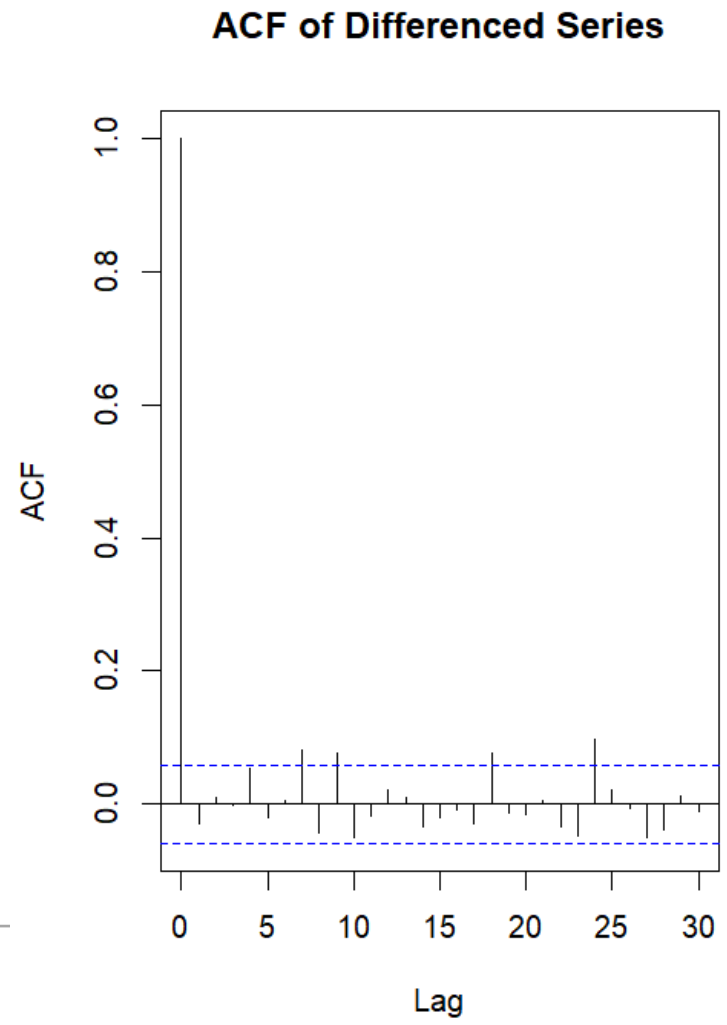
- **Augmented Dickey-Fuller (ADF) Test :** To statistically determine if a time series is stationary.

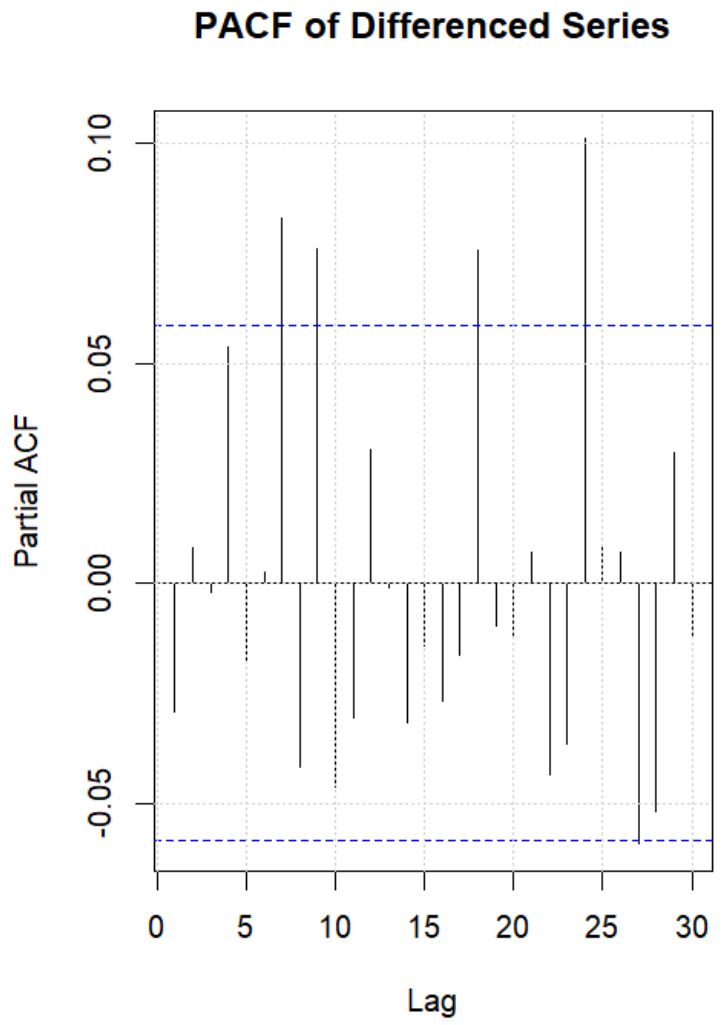
Null Hypothesis	**Alternative Hypothesis**
The series has a unit root (i.e., it is non-stationary).	The series does not have a unit root (i.e., it is stationary).

P-value (Threshold) : Typically, a significance level (α) of 0.05 is used.

- **If p-value > 0.05**: Fail to reject the null hypothesis. The series is likely non-stationary.
- **If p-value ≤ 0.05**: Reject the null hypothesis. The series is likely stationary.

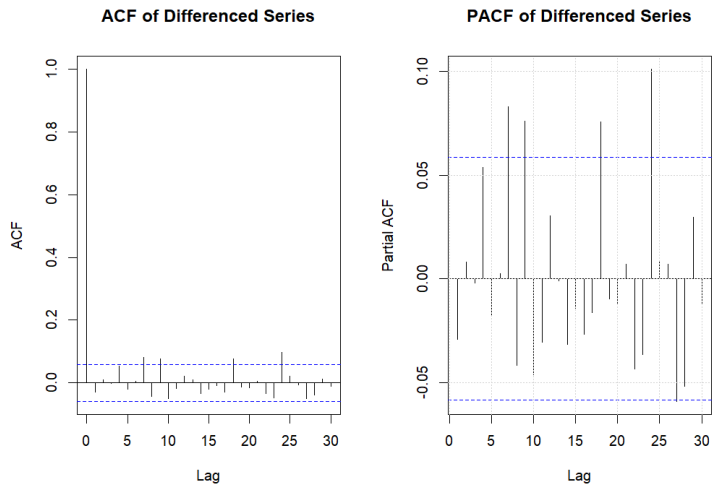
ACF/PACF Charts





ACF Plot

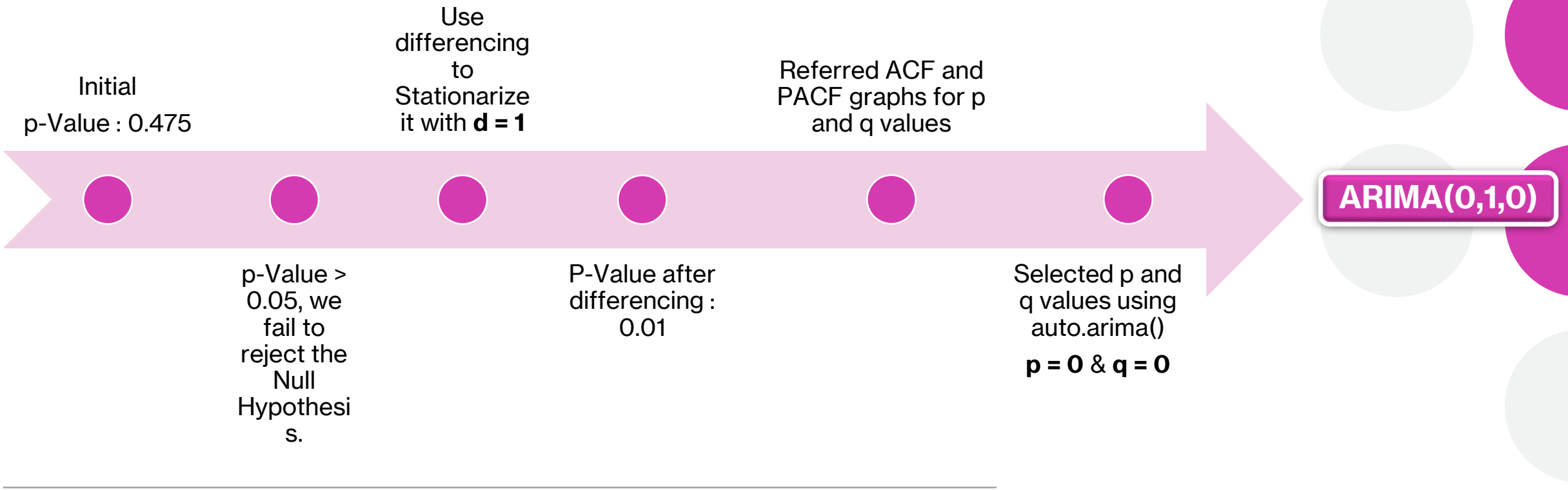
- The ACF plot of the differenced series shows no significant spikes beyond lag 0.
- This indicates that there is no substantial autocorrelation left in the differenced series.
- This suggests that the moving average component (q) is not needed, i.e., q = 0.



PACF Plot

- The PACF plot of the differenced series also shows no significant spikes beyond lag 0.
- This indicates that there is no substantial partial autocorrelation left in the differenced series.
- This suggests that the autoregressive component (p) is not needed, i.e., p = 0.

ARIMA Model Development



Predictions

- Stock Price Prediction for the next 30 days

Predictions

•The forecasted prices are represented by a continuation of the line plot extending to the right.

•There is a visible blue shaded area around the forecasted prices, which represents the confidence interval of the forecast.

•The blue shaded area indicates the range within which the actual stock price is expected to lie with a certain level of confidence (typically 95% in most forecasts).

•A wider shaded area suggests greater uncertainty in the forecast, while a narrower area indicates higher confidence in the predicted values.

•In this plot, the confidence interval widens over time, indicating increasing uncertainty as the forecast horizon extends.

Evaluation

RMSE	MAE	MAPE
8.325	5.798	2.998

- An RMSE of 8.32 indicates that, on average, the forecasted stock prices deviate from the actual prices by about 8.32 units. For stock prices, this can be considered relatively low if the stock price is in the range of hundreds of dollars, suggesting that the model has a reasonable level of accuracy.
- An MAE of 5.79 means that, on average, the forecasted stock prices are off by about 5.79 units.
- A MAPE of 2.99% indicates that the forecasted prices deviate from the actual prices by about 2.99% on average. This is a relatively low percentage, suggesting that the model performs well in terms of percentage accuracy.

Dive into the Code

```
> print(forecast_df)
         Date Forecast     Lo80     Hi80      Lo95     Hi95
1  2024-06-18 185.1247 174.4613 195.7881 168.81641 201.4330
2  2024-06-19 184.9719 170.1052 199.8385 162.23533 207.7084
3  2024-06-20 185.0601 166.8393 203.2809 157.19380 212.9264
4  2024-06-21 185.0092 164.0127 206.0056 152.89789 217.1204
5  2024-06-22 185.0386 161.5667 208.5105 149.14143 220.9357
6  2024-06-23 185.0216 159.3253 210.7179 145.72251 224.3207
7  2024-06-24 185.0314 157.2811 212.7817 142.59099 227.4718
8  2024-06-25 185.0257 155.3673 214.6842 139.66711 230.3844
9  2024-06-26 185.0290 153.5759 216.4822 136.92556 233.1325
10 2024-06-27 185.0271 151.8774 218.1768 134.32906 235.7252
11 2024-06-28 185.0282 150.2640 219.7924 131.86100 238.1954
12 2024-06-29 185.0276 148.7210 221.3342 129.50147 240.5537
13 2024-06-30 185.0279 147.2417 222.8142 127.23883 242.8171
14 2024-07-01 185.0277 145.8177 224.2378 125.06113 244.9943
15 2024-07-02 185.0279 144.4439 225.6118 122.96001 247.0957
16 2024-07-03 185.0278 143.1149 226.9406 120.92760 249.1280
17 2024-07-04 185.0278 141.8269 228.2287 118.95775 251.0979
18 2024-07-05 185.0278 140.5762 229.4794 117.04491 253.0107
19 2024-07-06 185.0278 139.3597 230.6959 115.18447 254.8712
20 2024-07-07 185.0278 138.1748 231.8808 113.37230 256.6833
21 2024-07-08 185.0278 137.0191 233.0365 111.60485 258.4508
22 2024-07-09 185.0278 135.8906 234.1650 109.87895 260.1767
23 2024-07-10 185.0278 134.7875 235.2682 108.19181 261.8638
24 2024-07-11 185.0278 133.7080 236.3476 106.54093 263.5147
25 2024-07-12 185.0278 132.6508 237.4048 104.92407 265.1316
26 2024-07-13 185.0278 131.6145 238.4411 103.33920 266.7164
27 2024-07-14 185.0278 130.5979 239.4577 101.78450 268.2711
28 2024-07-15 185.0278 129.6000 240.4556 100.25831 269.7973
29 2024-07-16 185.0278 128.6197 241.4359  98.75912 271.2965
30 2024-07-17 185.0278 127.6562 242.3994  97.28554 272.7701
```