us_china_dependence

March 19, 2021

1 Dependence of US market on Chinese

1.1 Introduction

Hypothesis:

The dynamics of trading in Chinese markets has an impact on the US market. In this paper I am making a reasoned conclusion about the presence or absence of influence, statistically proving it and making a model model, according to which traders, seeing how China traded off at night, will be able to guess how America will behave.

Initial data: - S&P 500 - futures on the stock index S&P 500 (USA) - FTSE China A50 - Shanghai Stock Exchange Index Futures (China) - Coronavirus cases data (USA)
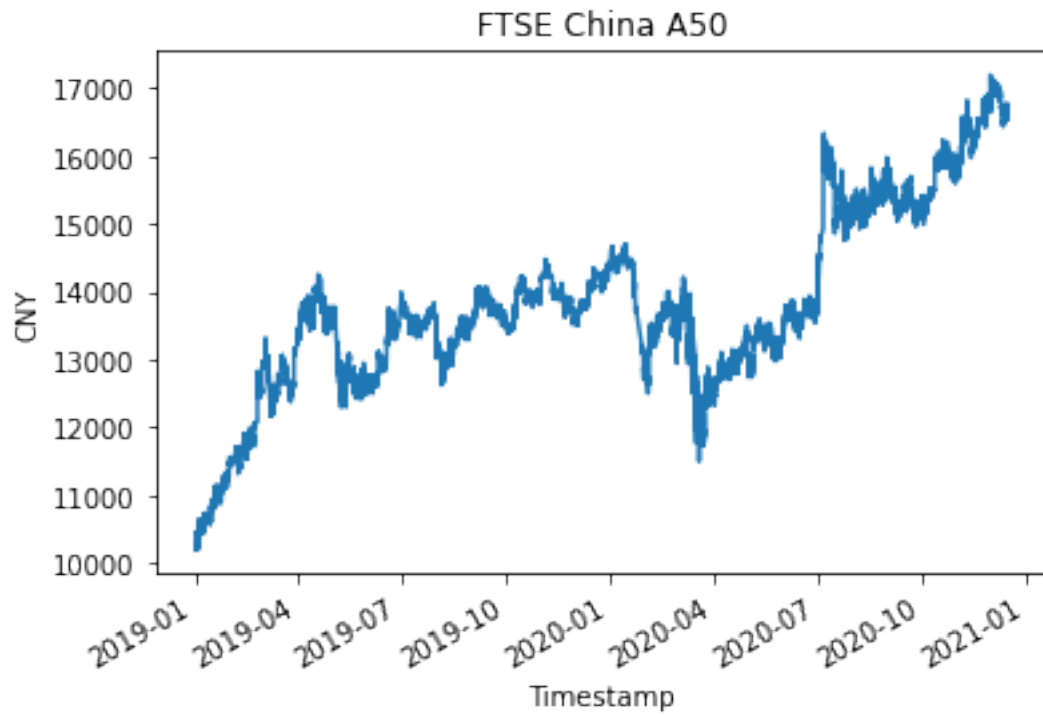
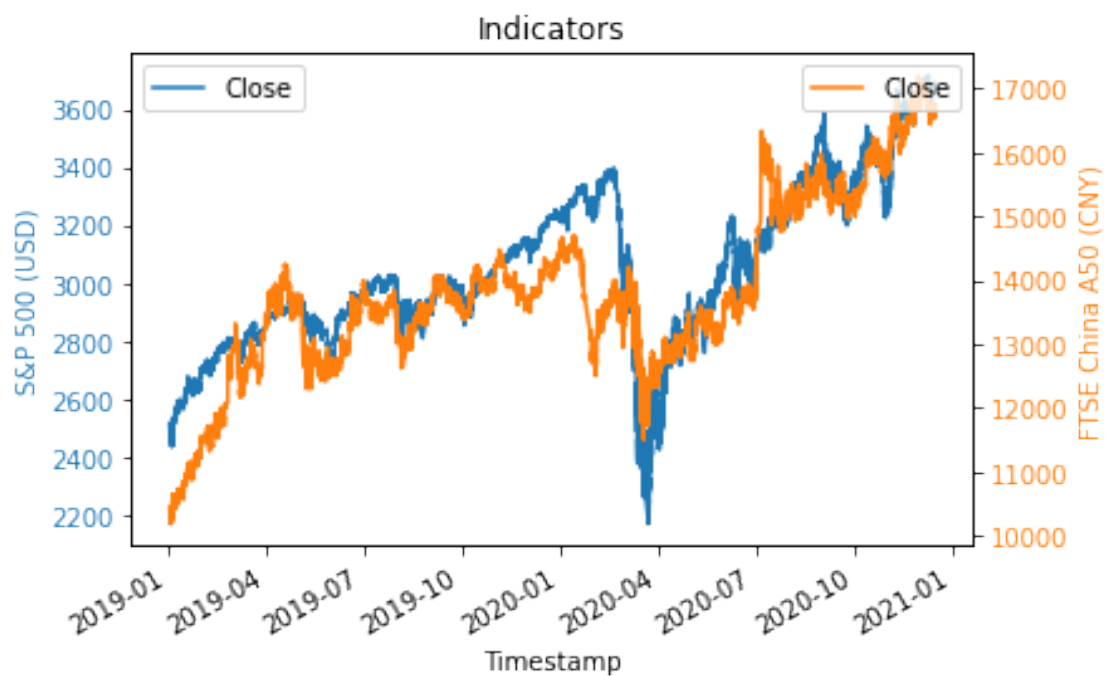**Time Zones on Data Set: EST (Eastern Standard Time)**
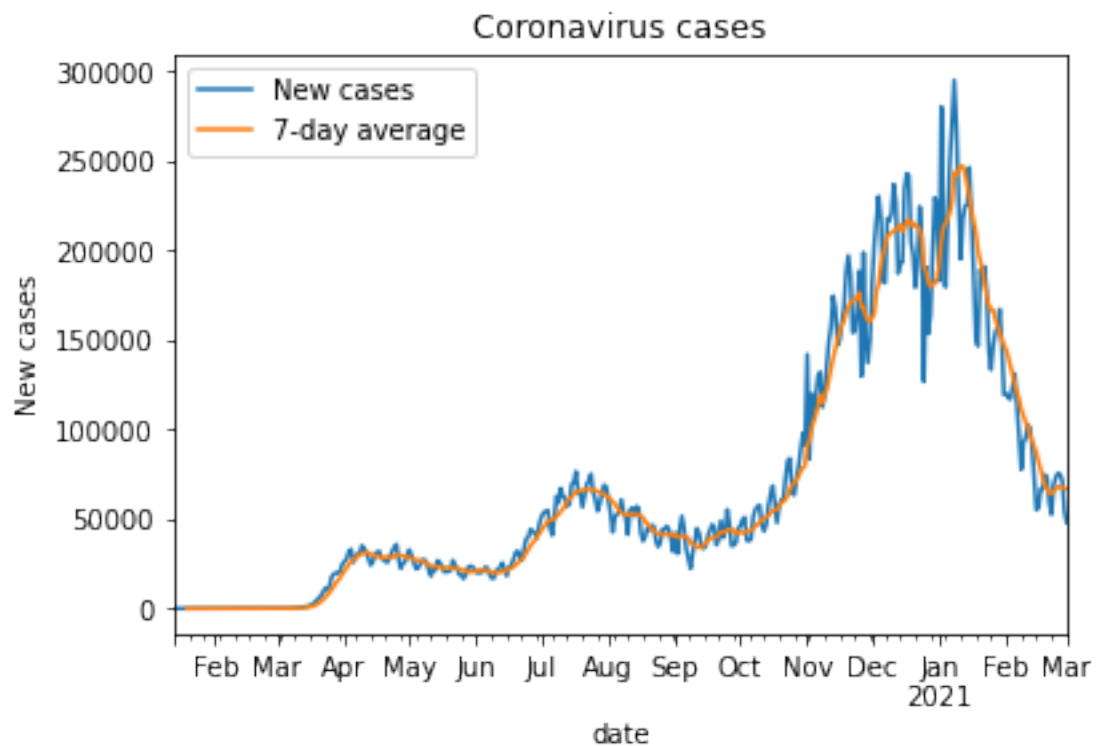
1.2 Data

S&P 500 index is measured in US dollars.
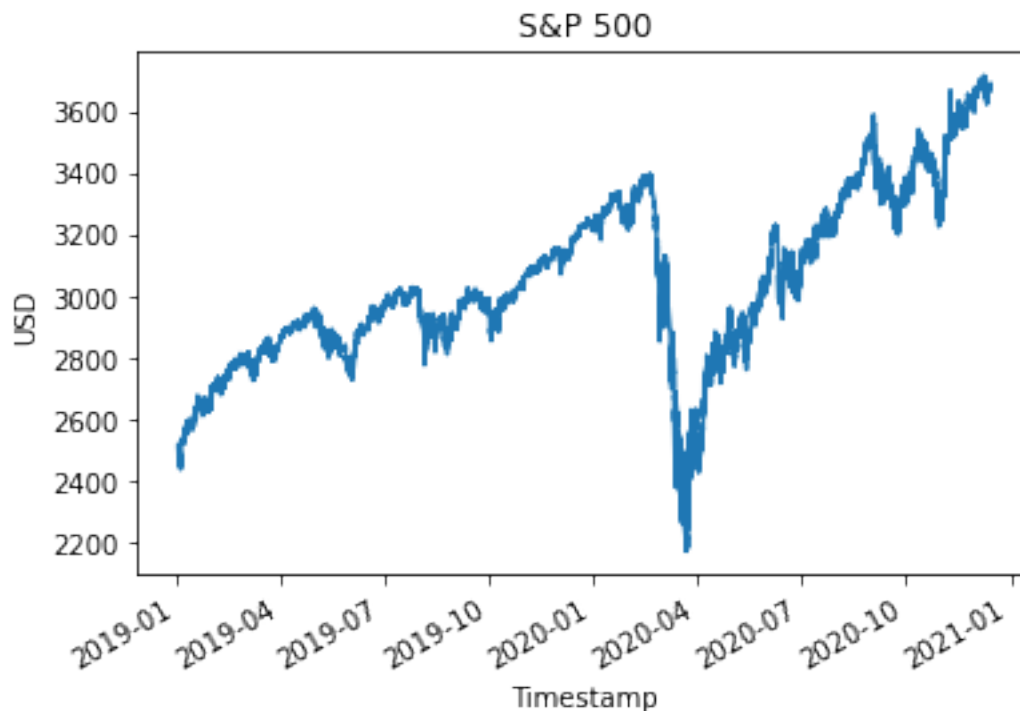
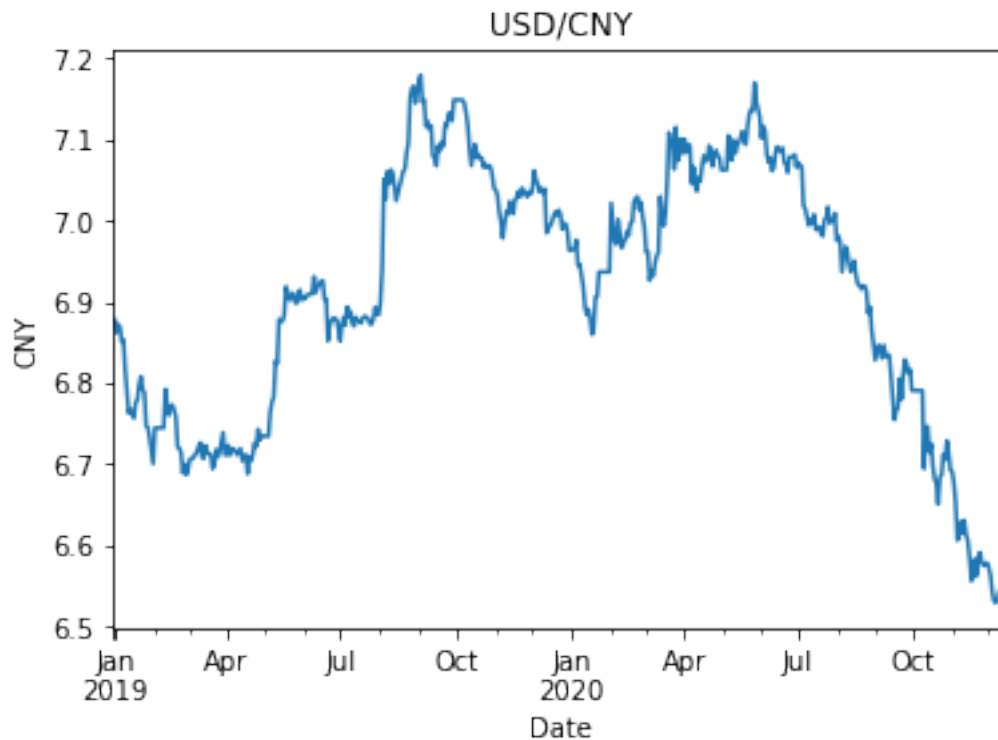While FTSE China A50 is measured in Chinese yuan.

Finally, we will require coronavirus cases data in order to exclude it's influence from the analysis.

Coronavirus cases

Indicators

1.3 Exchange rate

Do we have to make any adjustments to normalize our time series?
In order to answer this question, we should take a look at USD/CNY exchange rate.

Unfortunately, I had no access to minute-by-minute exchange rate and wasn't able to acquire detailed information on which exact rates are used to calculate the indexes. But even with just daily rates, we can still make some heuristics.

USD/CNY Daily Log Returns
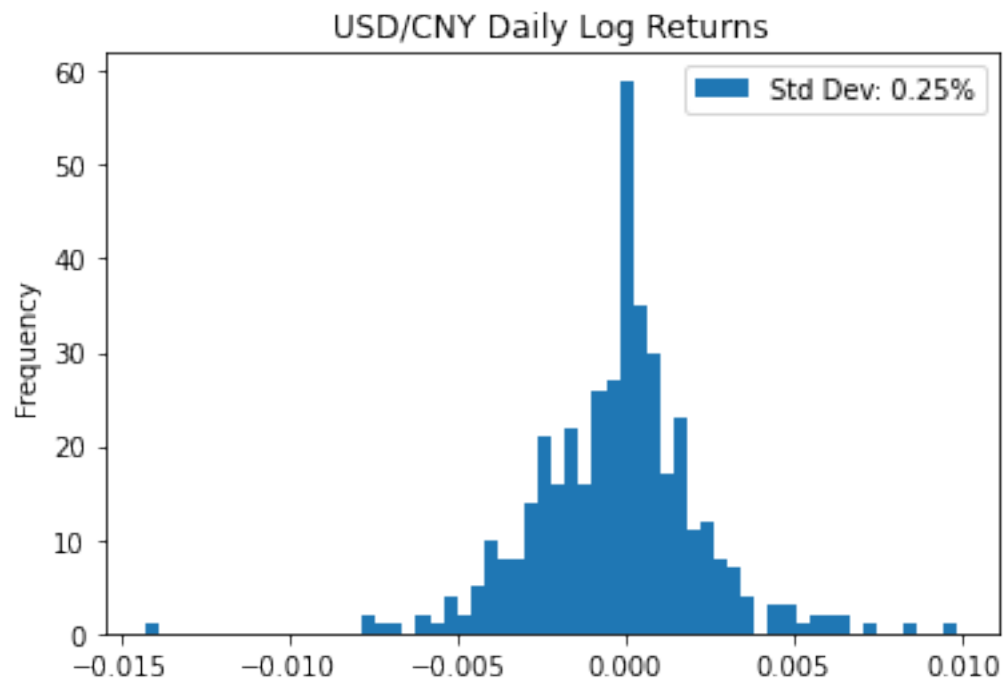
Daily volatility of USD/CNY exchange rate is only 0.25%.

Daily log return distributions comparison

S&P 500 Daily Log Returns

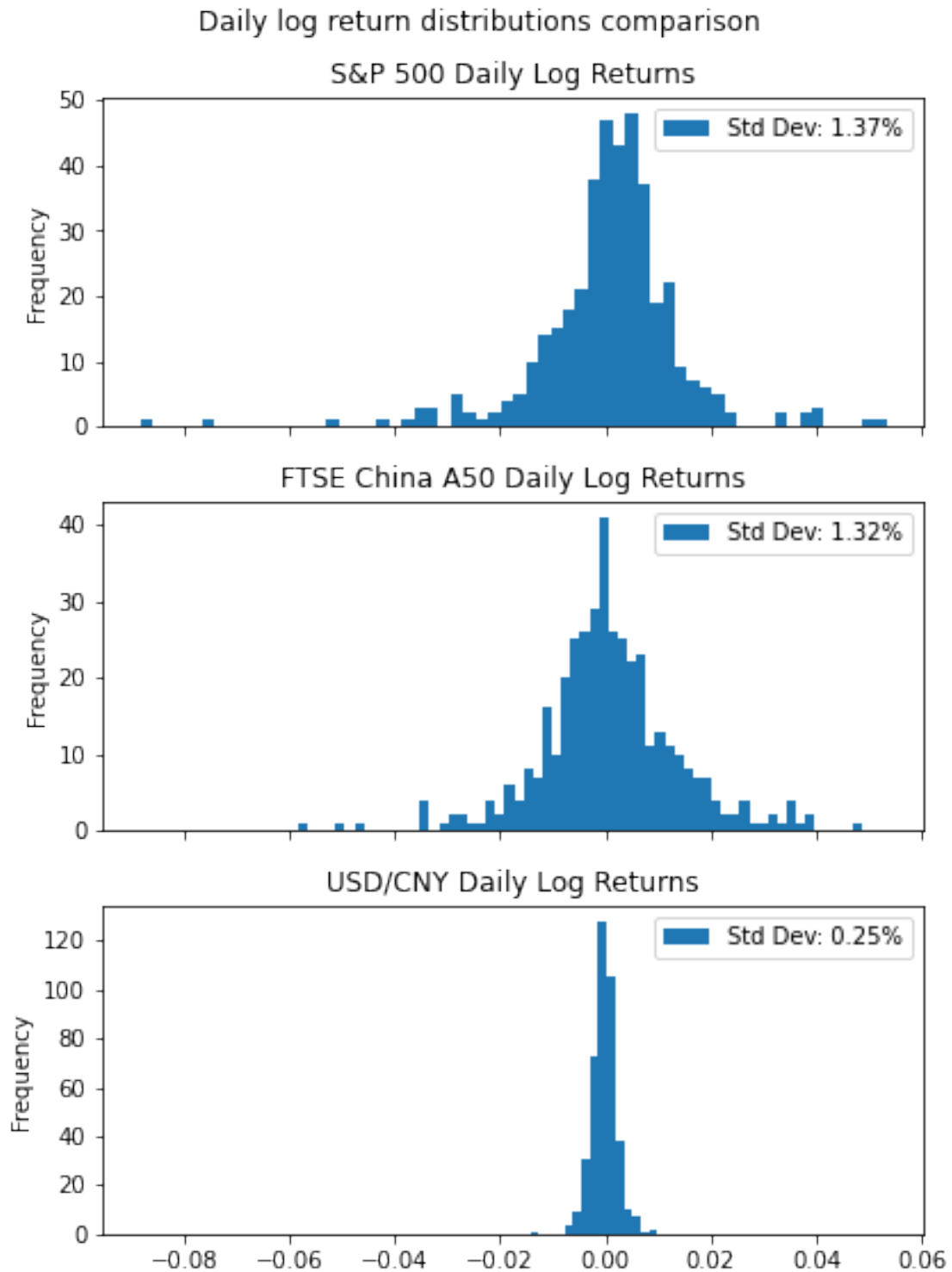Std Dev: 1.37%

FTSE China A50 Daily Log Returns

Std Dev: 1.32%

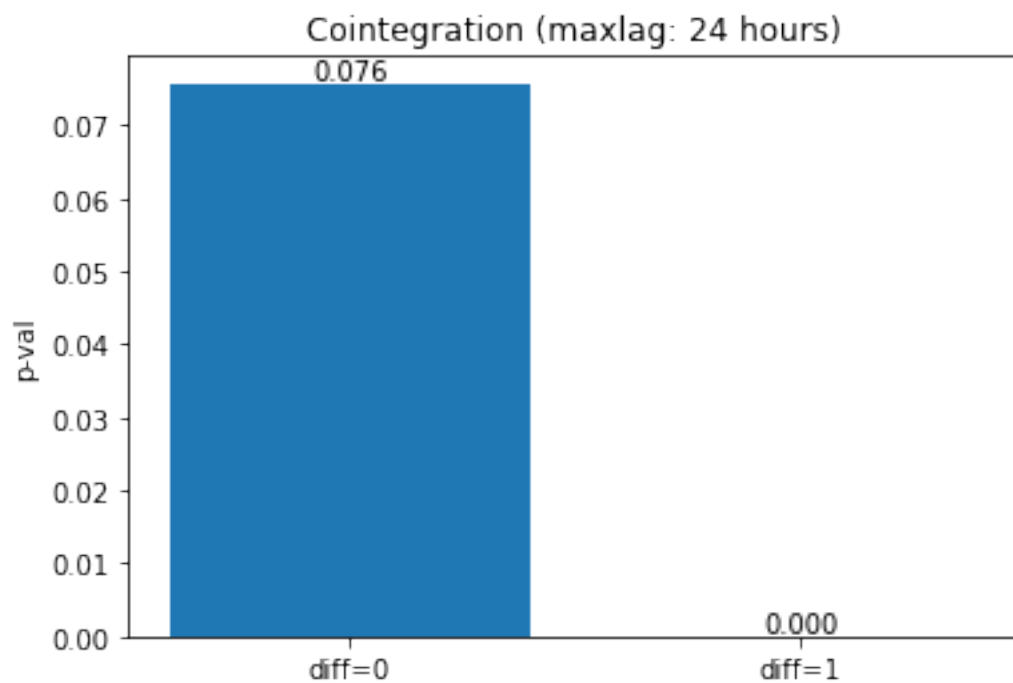### USD/CNY Daily Log Returns

Std Dev: 0.25%

However, volatilities of US and China markets are considerably higher, making 1.37 and 1.32 percent accordingly. Thus making the exchange rate factor negligible for the analysis.

1.4 Statistical tests on hourly data

1.4.1 Stationarity
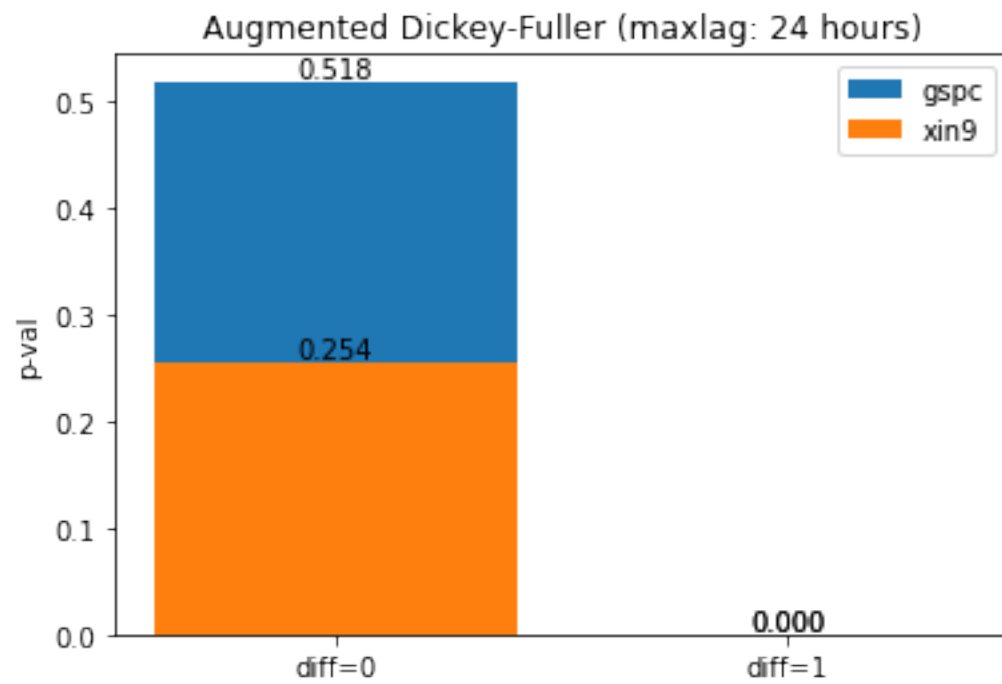
1.4.2 Cointegration

1.4.3 Granger causality
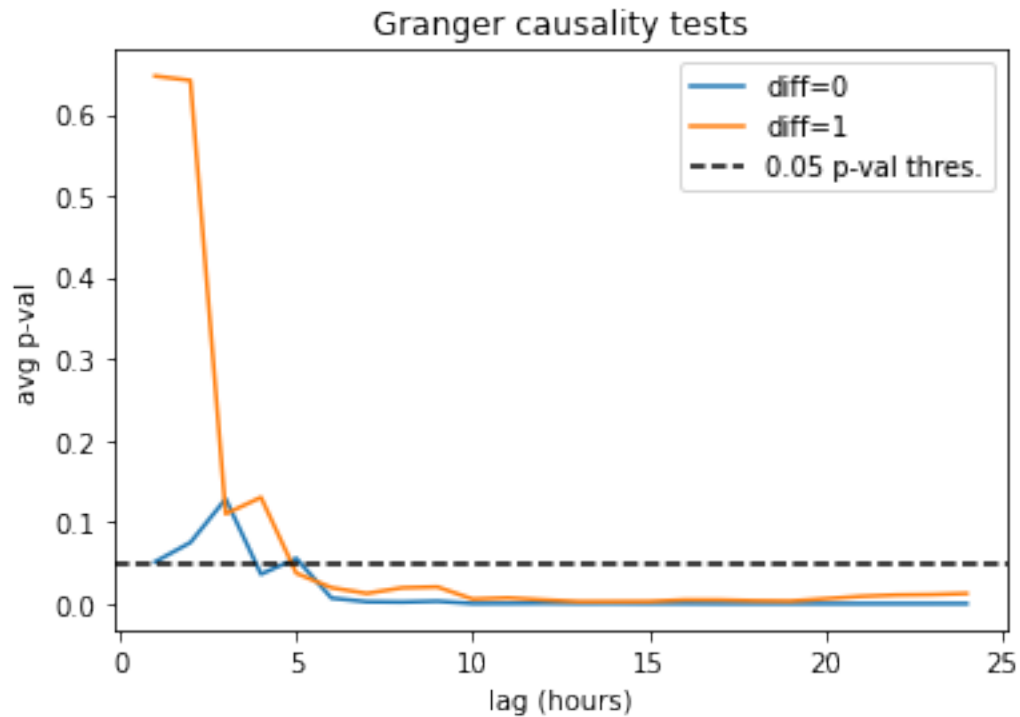
1.5 Time scale

Which time points should we compare?

If we build a plot of average trading volume per hour we will see, that the time ranges when stocks in China and America are traded do not intersect. Taking this into account, the most reasonable assumption would be to believe, that the closing late night price in the Chinese market would most strongly affect the opening price in the American, and then it's influence would weaken during the course of the day.

Average volume per hour of day
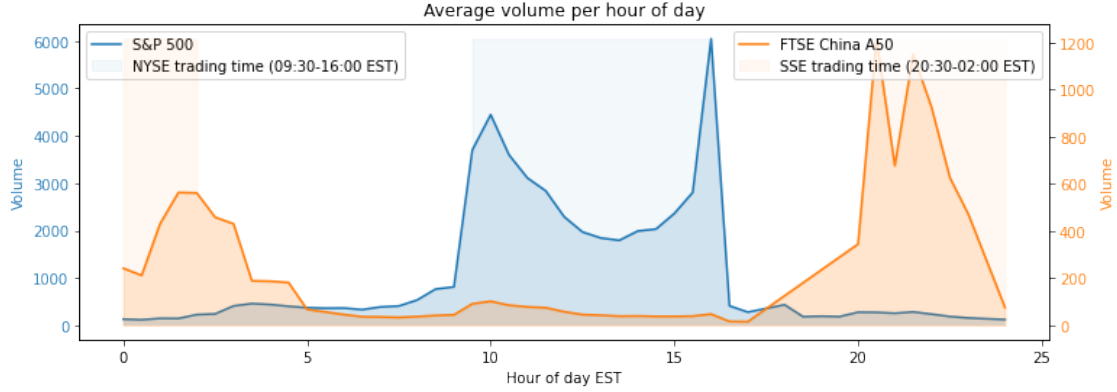
Most trading activity picks up at 9:30 AM EST in US, which corresponds with the opening times of US major stock exchanges such as New York Stock Exchange and Nasdaq. The Chinese market is seen active till 5-6 AM EST. Therefore, it's safe to take 9-10 AM EST American index as daily opening and examine it's depence on Chinese index after 6 AM EST, taken as daily closing.

Take note, that XIN9 is not traded during 5:30-8 PM EST.

1.6 Correlation

The hypothesis is, that relative differnce of US market index prices during the exchanges' opening time today and their closing price the previous day, should correlate with the daily difference in Chinese prices during times of inactivity of Chinese exchanges.

$$US\ diff = intercept + \beta \cdot China\ diff + \varepsilon$$

- The change in US market is measuread in logarithmic difference between the average closing prices at 9:00-9:50 AM today and 6:00-6:50 PM the previous day.

$$US\ diff = log(avg(9AM\ close\ prices\ today)) - log(avg(4PM\ close\ prices\ yesterday))$$

- The change in Chinese market is measuread in logarithmic difference between the average closing prices at 8:00-8:50 AM today and 5:00-5:50 PM the previous day: we're taking the "closing" prices one hour before US opening in order to come up with a trading strategy on time.

$$China\ diff = log(avg(8AM\ close\ prices\ today)) - log(avg(5PM\ close\ prices\ yesterday))$$

We resort to logarithmic differences in order to acquire a relative scale and address some of the mathematical downfalls of using other popular measures, such as percentage change.

Current choice of hours to take was made by finding the right combination of opening/closing hours for US and China so that Pearson's correlation between the two resulting series is the highest.

Since: 1. The resulting function is not easily differentiable, 2. Has several logical constraints, 3. The absense of local extrema is not easily verifiable,
and 4. The variables to optimise are discrete,

I resorted to simply brute-forcing the solution as opposite to using faster but less reliable optimisation methods adaptable for the task, such as Bayesian optimisation.

	gspc_open	gspc_close	xin9_open	xin9_close	pearson_coef	p-value
12784	9	16	17	8	0.701	0.0
12937	9	17	17	8	0.696	0.0
13090	9	18	17	8	0.669	0.0
13243	9	19	17	8	0.620	0.0
13396	9	20	17	8	0.619	0.0

Hours Optimisation result
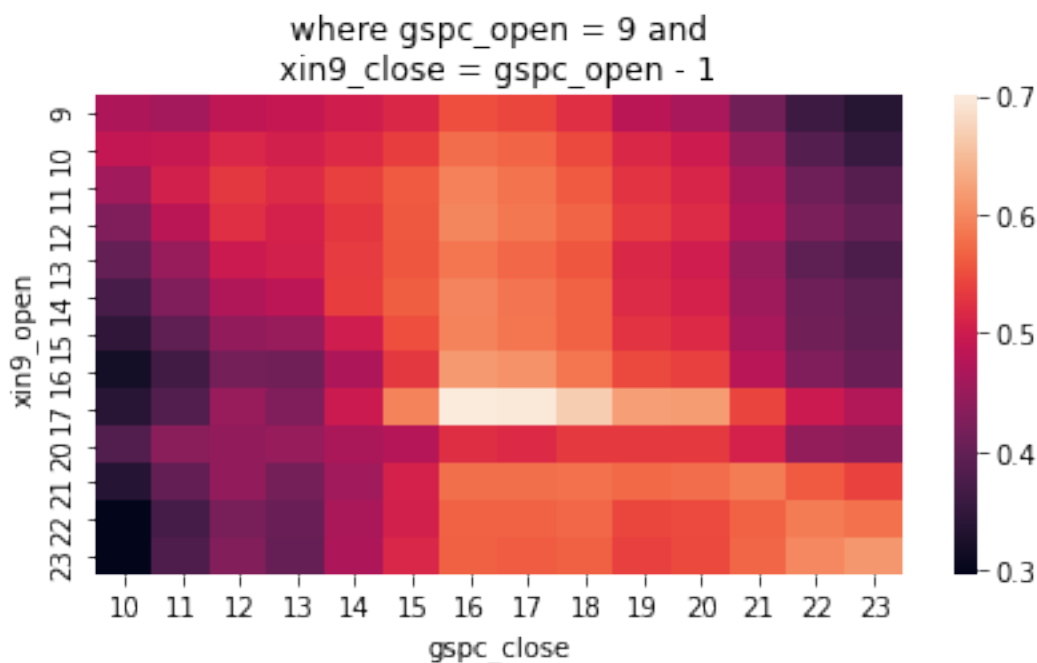
Why is xin9_close = gspc_open - 1?
The search made clear, that this combination leads to the best correlation between the indicators: by the opening time in US we should orient on the latest and most corrected Chinese prices.

Why is gspc_open = 9?
This value was found, while trying to use the Chinese indicator to predict the US: the further away we go from the opening time of US exchanges, the less the predictive power of the Chinese indicator appears to become. Since China is mostly inactive during the day in America, it's influence gradually lowers, and American prices start evolving independently. Therefore, it's best to try predicting the first opening hour's dynamics.
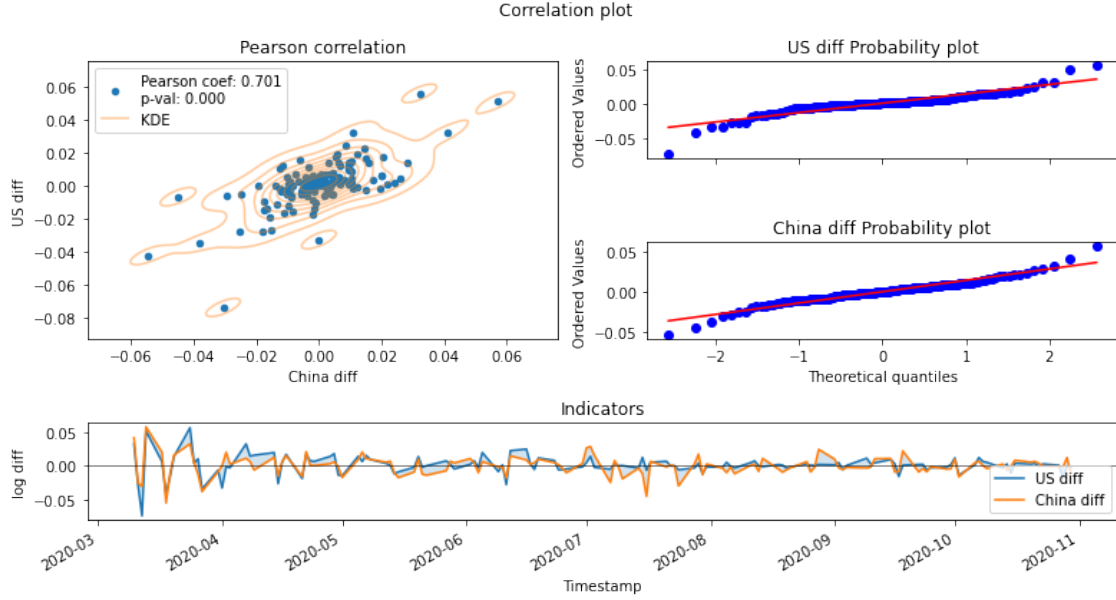
Correlation plot

As we can see, there is a strong linear correlation between the estimates: Pearson's coefficient is 0.7 and the probability of correlation's existence is high (p-value < 0.01).

1.7 Rolling correlation and regression

The presented time series are clearly not dependent on just each other: both American and Chinese markets are subject to a whole array of various factors. Examining each of them would be a life time of work, and is beyond the scope of this paper. What one *can* do, however, is to build a rolling regression model and try playing out the most important driving forces, if possible.

Window Optimisation result
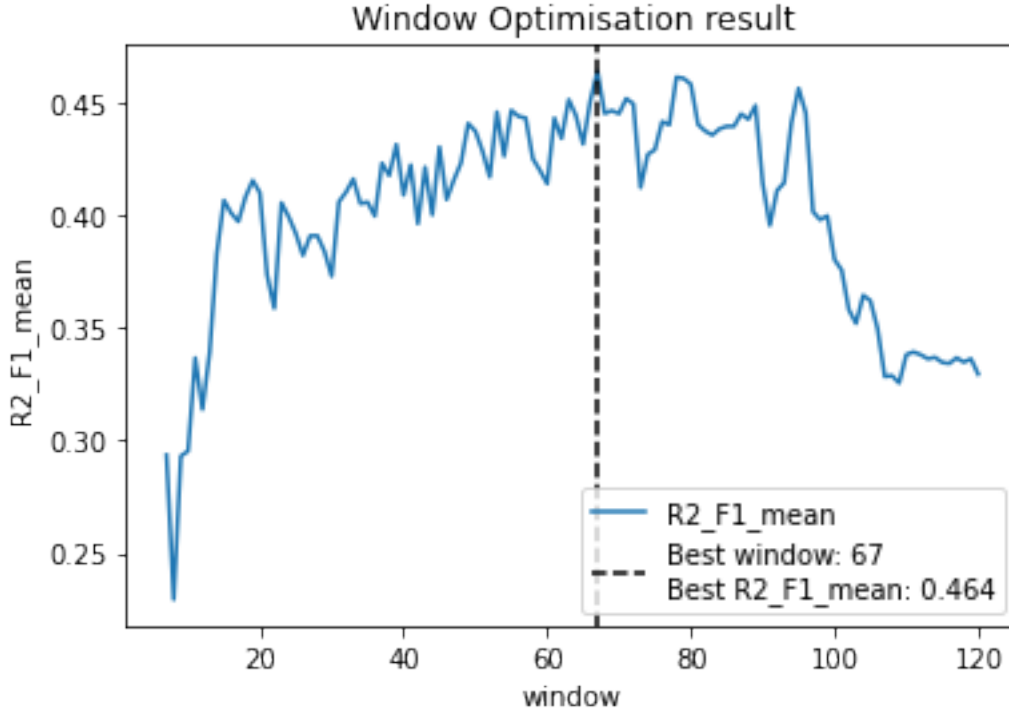
The size of the window for rolling regression was optimised by searching through windows of size 7 to 120 days and measuring mean of R2 and F1 scores, acquired by predicting one following day's US difference, located among the latest 31 observations.

$$US\ diff = intercept + \beta * China\ diff + \varepsilon,$$

where β and intercept are approximated fitting an OLS on previous 67 days of observations

```
R2-score: 0.166
F1-score: 0.762
```

The F1-score measures, how accurately the model predicts the direction in which the price will go - up or down, while also accounting for probable class disbalance, thus making up for the limitations inherent in simpler accuracy metrics. Therefore, the F1-score of 0.762 means, that the model should make the correct guess about the direction of the price change in about 76.2% cases if probability of the market going up is 50 percent.

Rolling OLS Residuals Probability plot
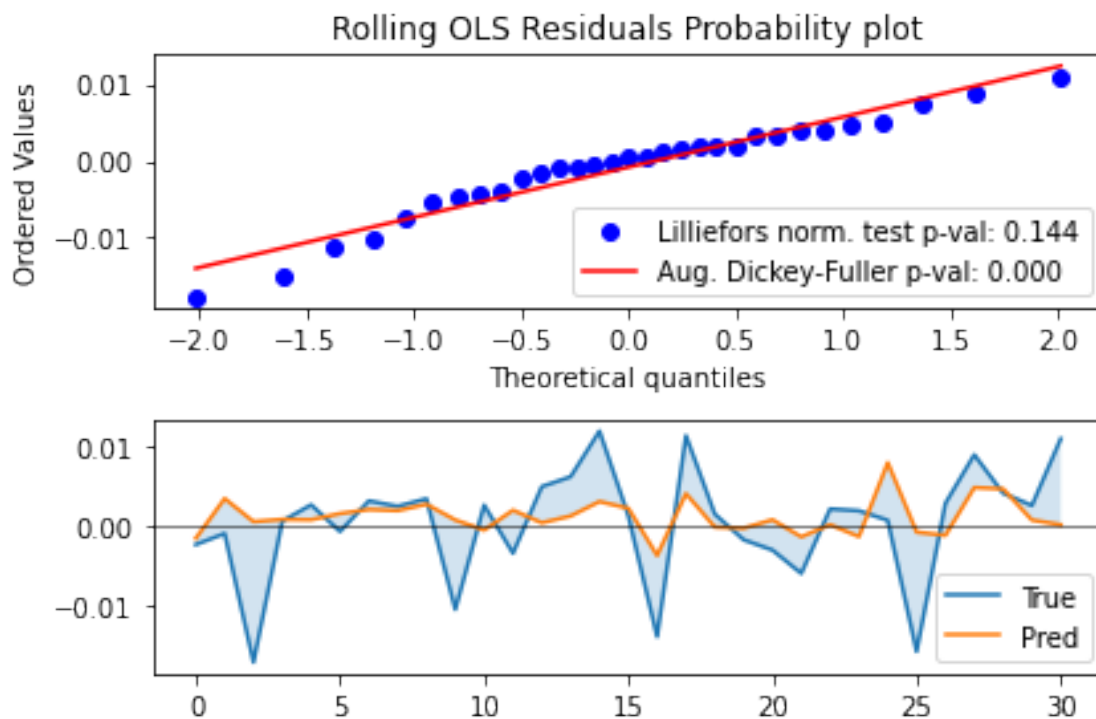
Further improvements could include changing the rolling estimator to a polynomial regression, neural network or a random forest, tuning them, adding regularisation etc. (a polynomial fit of order 7 and 102 days window size, found during optimisation, resulted in 0.276 R2 and 0.75 F1 on the test set).

```
poly 7, 102 days window
R2-score: 0.276
F1-score: 0.750
```
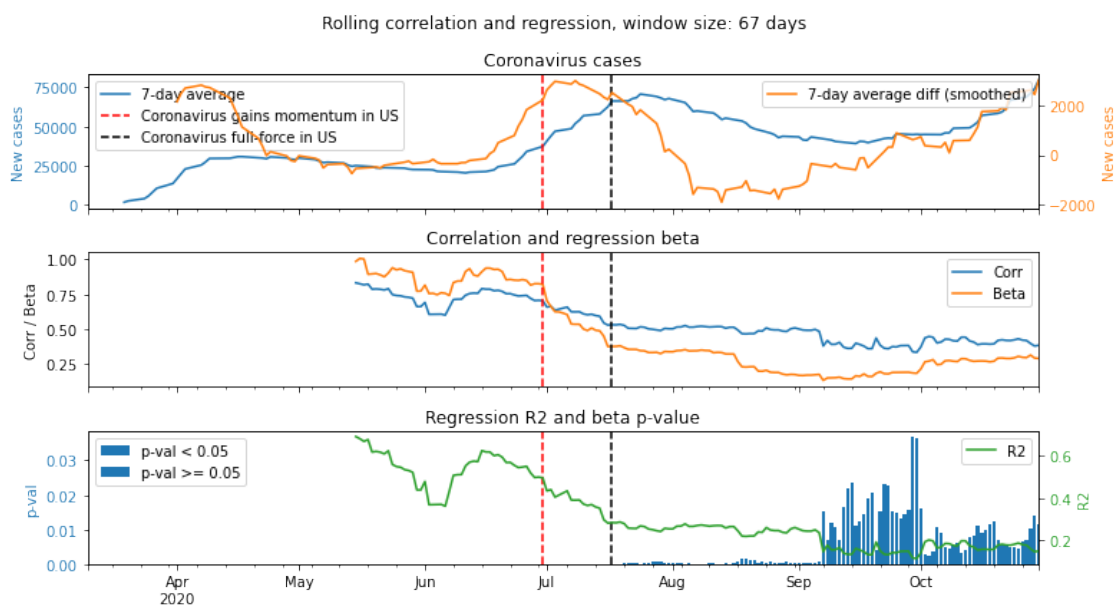
Rolling Polynomial OLS Residuals Probability plot
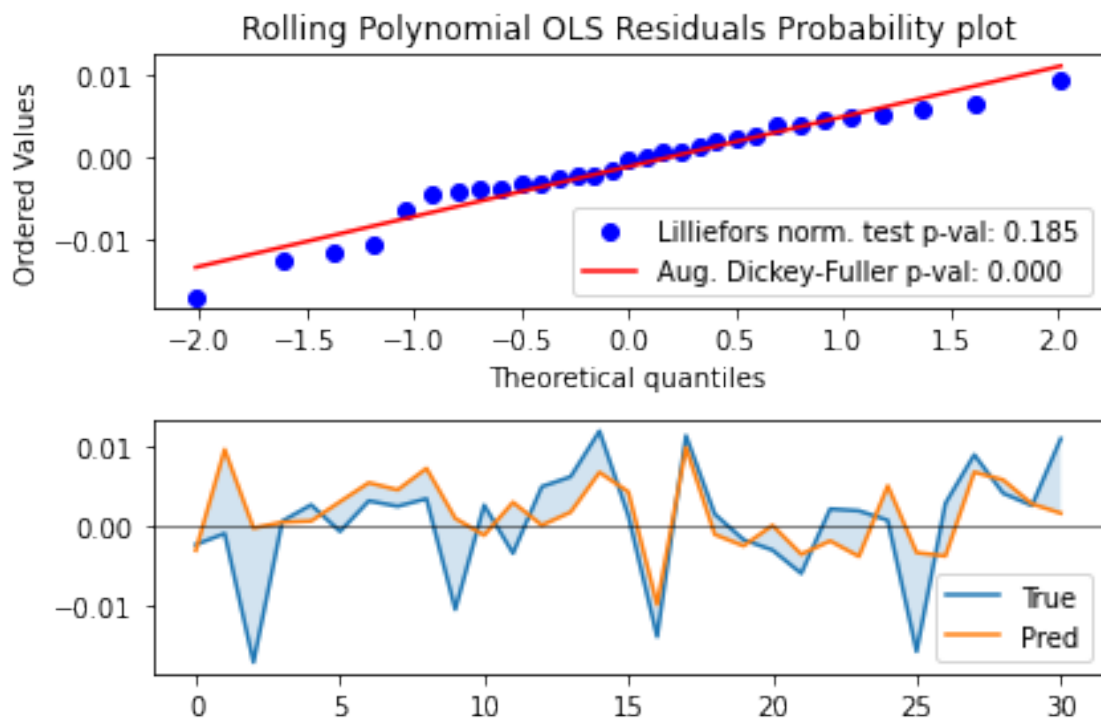
However, let us keep with the simple rolling OLS on the monomial of degree 1, estimated above, to simplify the further analysis.

## 1.8 Statistical tests on modified indicators

1.8.1 Structural breaks

Rolling correlation and regression, window size: 67 days

Could we probably use a single OLS model instead?
July 2020 appears to be the month, during which coronavirus hit the US particularly hard. On the graphics above we can see a clear structural break in the dependence during exactly this time: correlation and predictive power of the Chinese indicator drops dramatically, albeit staying staying significant (p-values of the coefficients are consistently below 0.05). We can observe the same picture if we fit and compare two OLS models, one fitted on pre-July and the other on post-July data.
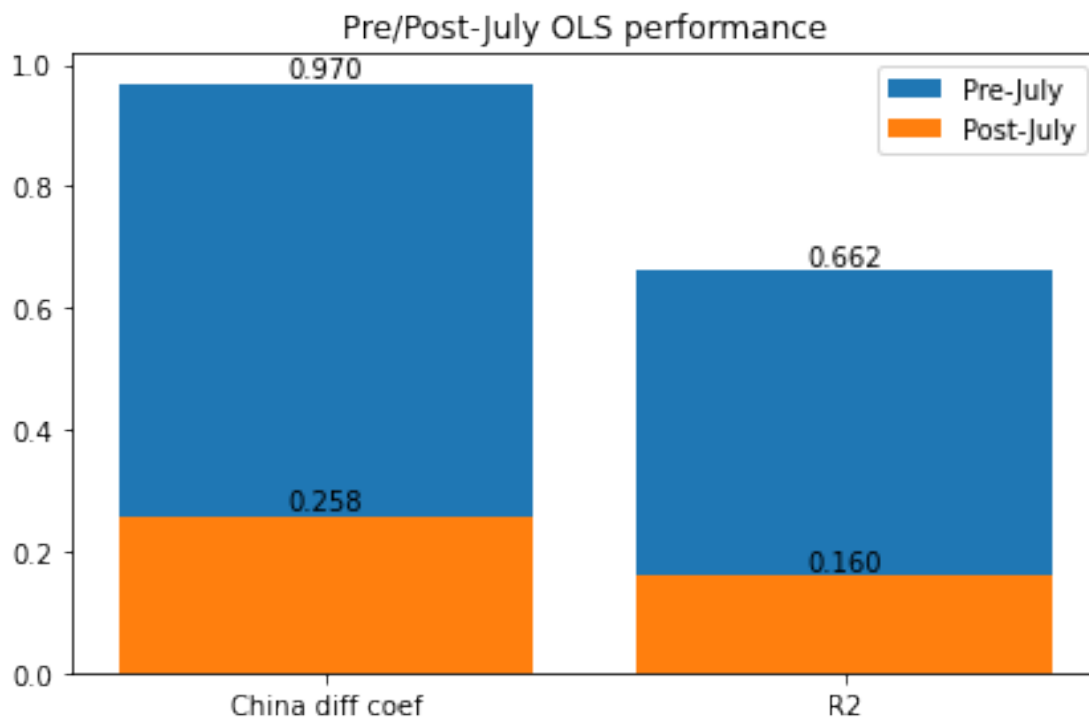
Chow test is used to verify the data set can not be represented with a single regression line (as opposed to several). It can be used to support our hypothesis, that the regression is different for pre- and post-July markets.

$$CHOW = \frac{(RSS_p - (RSS_1 + RSS_2))/k}{(RSS_1 + RSS_2)/(N_1 + N_2 - 2k)}$$

```
F-val: 10.9748
P-val: 0.0000
```

Our degrees of freedom are 60 and 58 accordingly. Critical F-value for this combination is approximately 1.5343, while the test resulted in 10.9748, which is significantly higher. This means the p-value is less than 0.05, and we can reject the null-hypothesis, that there is no structural break.
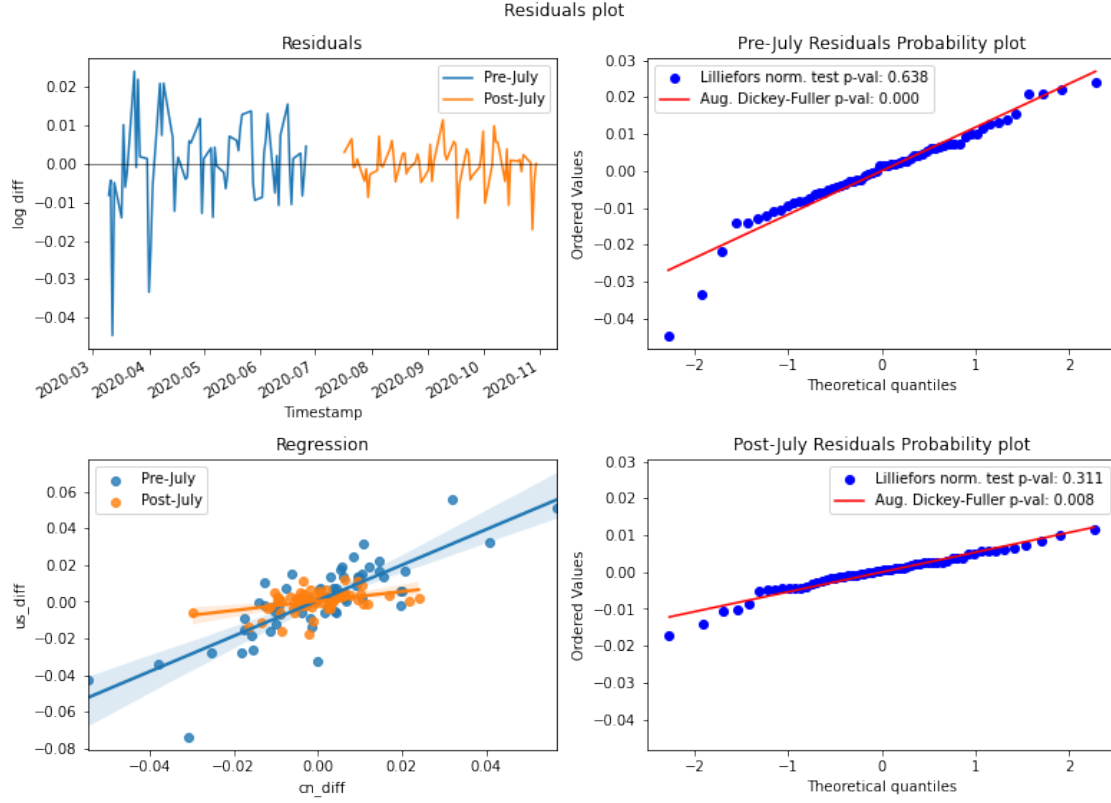
Residuals plot

Augmented Dickey-Fuller unit root test rejects the null-hypothesis of resudals of both models being non-stationary. While Lilliefors test can not reject them coming from a normal distribution. Both support our assumptions and conclusions so far.

In practice, nothing can guarantee us from such structural breaks. If we take a look at the graphics of rolling regression coefficients, we will observe the coefficient constantly evolving in complex patterns, which do not resemble a random process: the interaction of external factors is setting the trend for the strength of the dependence between Chinese and US markets. Therefore, it is more reasonable to use a rolling regression instead of a single one for making a prediction for the next day.

1.8.2 Stationarity
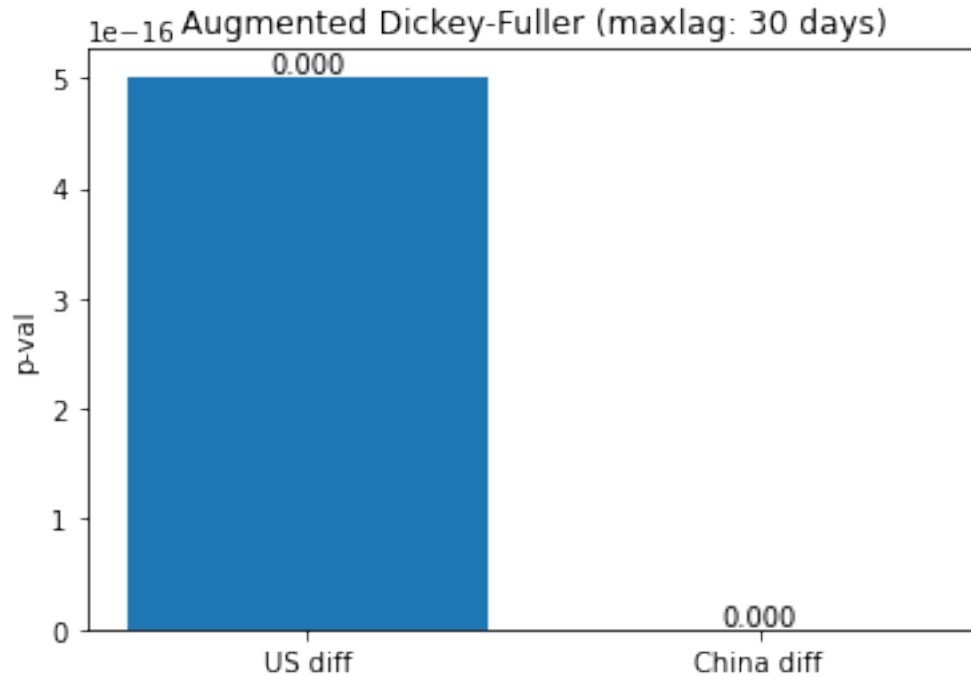
1.8.3 Cointegration

Could it be, that there is no real dependence between the indicators, and they are driven by external factors alone?
This is not a simple question to answer. However, we can try making as many tests hinting otherwise in order to rule out some of the non-causality hypotheses. One such hypothesis, is that the indicators are not cointegrated. Cointegration means there is an amplifying effect in between x and y - at any time point, we can always multiply x with the same parameter to get a value which is close to y.

```
Cointegration p-val: 0.0000
```

P-value is less than 0.05. Therefore we reject the null-hypothesis that the series are not cointegrated.

1.8.4 Granger causality

Another hypothesis we can check is that the Chinese indicator consistently precedes the US. This is called Granger causality.

```
Granger Causality
number of lags (no zero) 1
ssr based F test:         F=117.7788, p=0.0000  , df_denom=128, df_num=1
ssr based chi2 test:   chi2=120.5392, p=0.0000  , df=1
```

```
likelihood ratio test: chi2=85.4646 , p=0.0000  , df=1
parameter F test:        F=117.7788, p=0.0000  , df_denom=128, df_num=1
```

Statsmodels package has 4 various Granger causality tests in its grangercausalitytests method. Each of them results in p-values less than 0.05. This means we can reject the null-hypothesis that the US indicator is not Granger-caused by the Chinese.

1.9   Complex model example

A multilayered perceptron is a good example of an algorithm capable of approximating a smooth function of several variables, which appears to be the most probable case. Rather than reimplementing the rolling window with just one variable (China diff) as a predictor, we can reinforce our prediction by adding several lagged values of the target variable (US diff). The basic intuition behind this decision is that each time series contains information about itself, such as its trend and "elasticity". Therefore, the neural network is capable to deduce the several-days trend from the lagged values of the time series itself. This approach is comparable to classical models, such as SARIMAX (or, to be more specific, ARX - autoregressive model with exogenous regressors). However, instead of extracting individual components with statistical technics, the network will supposedly estimate the correct equation more precisely while also demanding less parameter turning and assumptions met.
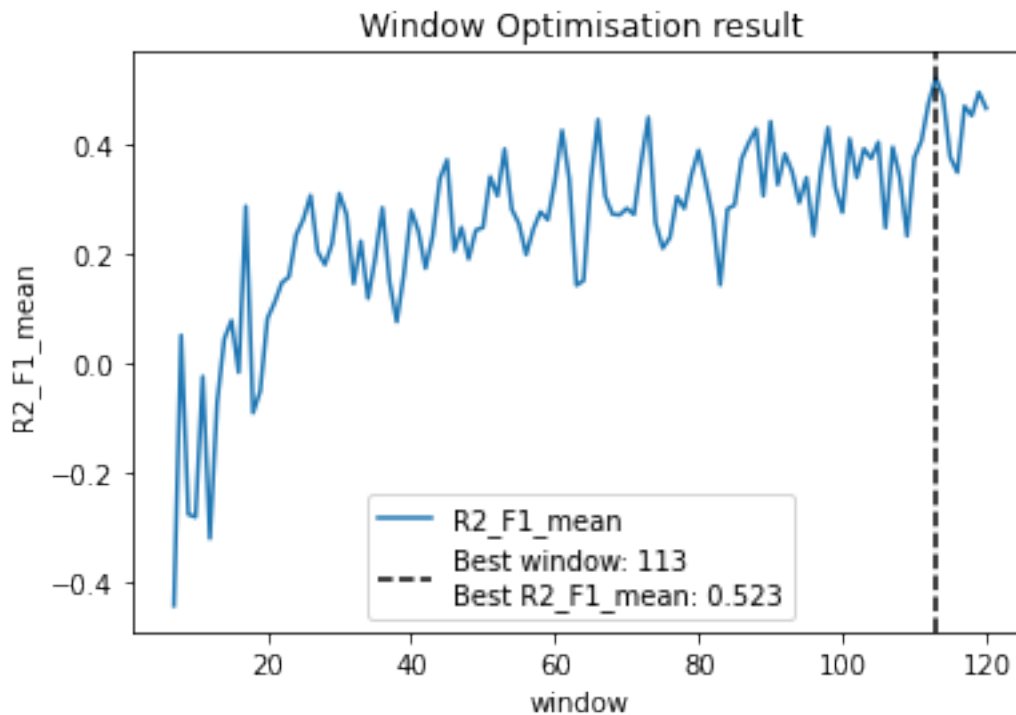

1.9.1   Model specifications and feature interpretation

- The MLP model uses a window of 113 previous days (rows) to train.
- In addition to China diff it also uses 4 lagged values of US diff as predictors (columns).

- The network architecture is 1 hidden layer of 100 perceptrons with ReLU activations and Adam solver.
- The loss function is squared-loss.
- 0.2 validation fraction is split from the window before training to compute loss used for early stopping.

Partial dependence plot   Partial dependence plots are used to examine how black-box model's features impact the prediction when changed. It is a simple method, which fixes all columns except one and then computes mean predictions, replacing the values in the examined column with an array of exactly same constant values and gradually increasing them.

Partial dependence plot
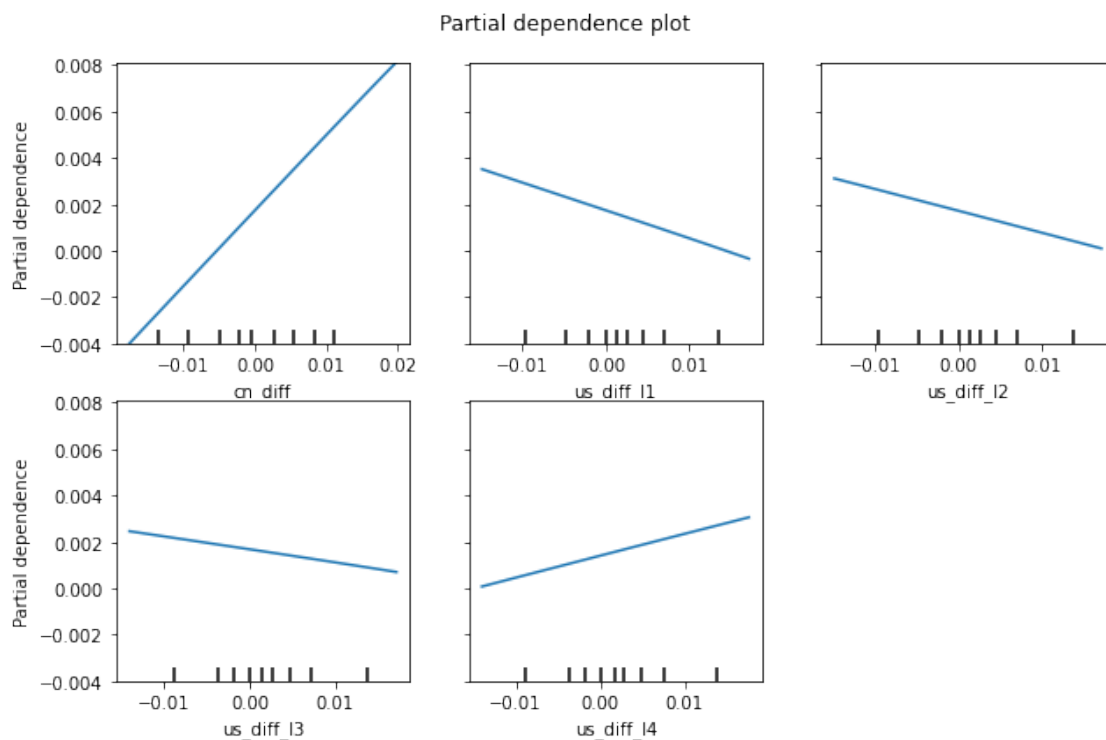
The relationships, approximated by the MLP, appear to be linear.

Feature importances   Permutation feature importance is a method to calculate relative feature importances. Here we apply permutations to an examined column to see, how much worse the performance of an estimator becomes according to our metric of choice.

Feature importances
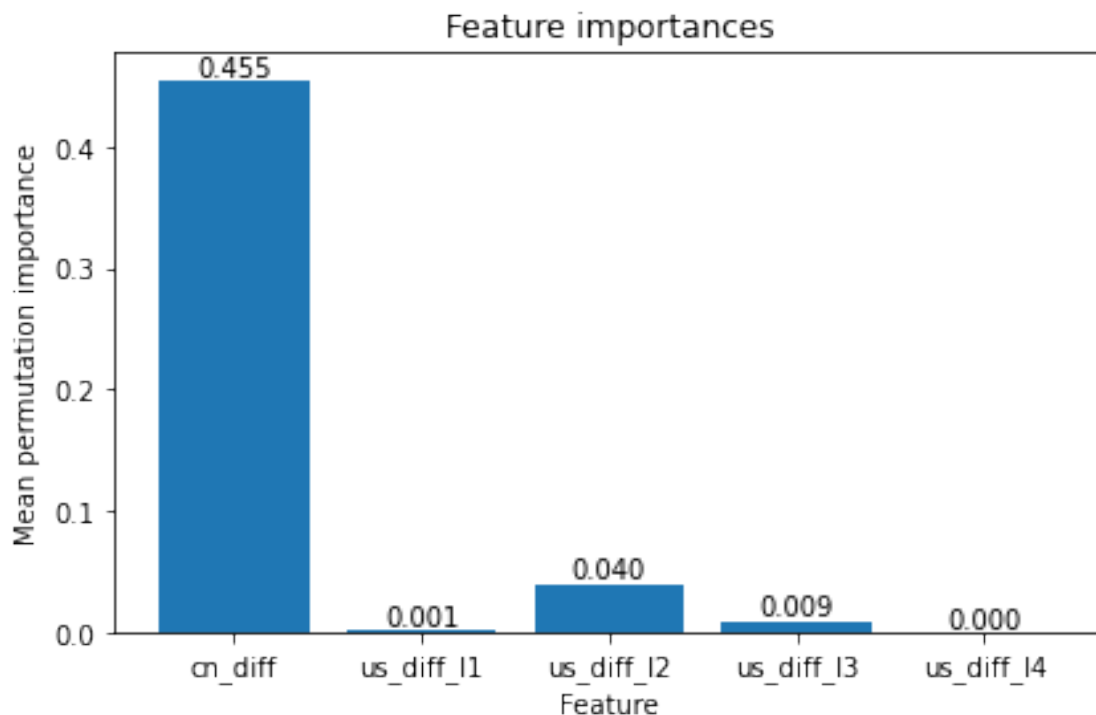
The most significant feature with the strongest impact on prediction is still China diff, while the other factors only play a supporting role.

Why not add lagged values of China diff as predictors?
Experimenting with such approach proved to give worse results, than leaving only the current latest. This can probably be explained by the fact, that Chinese and US markets, even if correlated to some extent, evolve in individual fashion, and the latest several-days trend in US does not necessarily repeat the Chinese.

Why not add more hidden layers?
On the partial dependence plot above, we can see that the approximated function is linear. Experimenting with more layers produced almost linear curves as well, while lowering the scores. The training is early-stopped the moment when the model approximates a near-linear function, hinting at a possibility, that the linear model could probably appear to be the best describing the data.

What other models have been tested?
Random forest and ARMA with fully automatic parameter selection were also tested using the same rolling window framework described above. However, their performance did not match the MLP. Remembering our previous conclusions, a rolling ARX or ARMAX model with fixed number of lagged values and carefully chosen optimiser could have probably gotten on par.

1.9.2   Model comparison

```
                    Model     R2     F1
0        Simple Rolling OLS  0.166  0.762
```

20

```
1                  Rolling Polynomial OLS  0.276  0.750
2  Rolling MLP with lagged predictors  0.202  0.844
```

As we can see, the MLP outperforms the simple rolling OLS in terms of both R2 and F1. It doesn't beat the polynomial OLS in terms of R2, however, it's F1-score is significantly higher, making it arguably a more practically applicable trading model: it doesn't matter so much, if we don't win as much profit as we predict, since we make correct assumptions on the price movement directions, thus minimizing the risk of losing.

1.10   Prediction
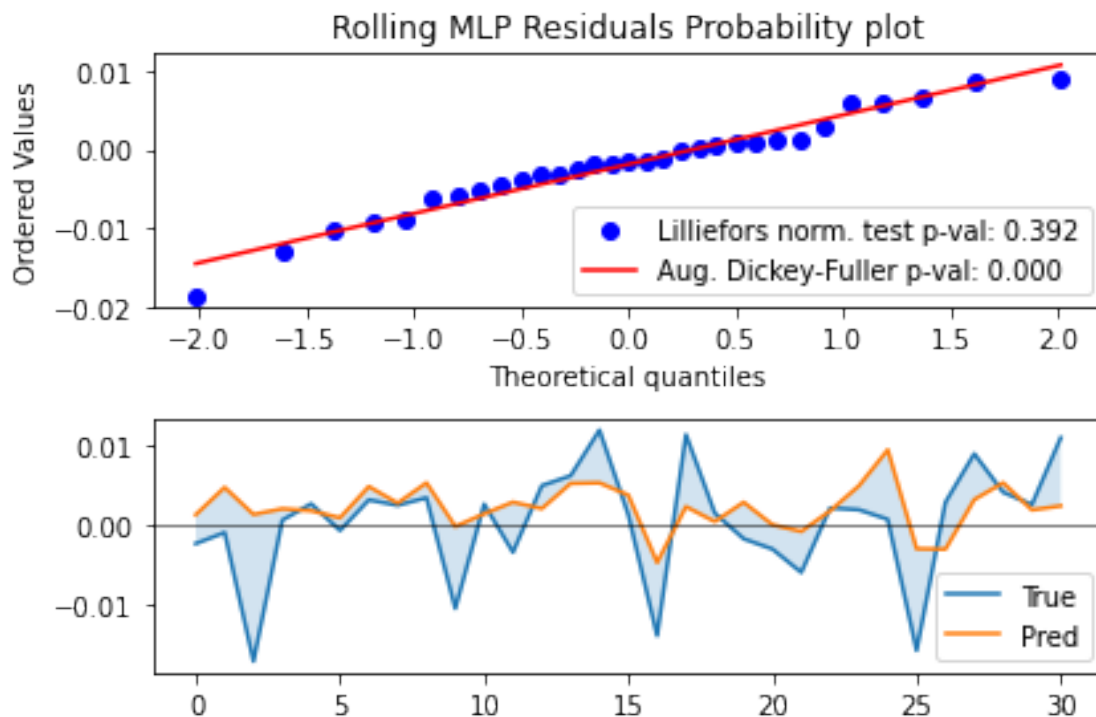
1.11   Conclusion

1. We can not reject that Chinese market's nightly performance influences US.
2. If there really is such influence, its strength changes over time.
3. Opening dynamics of S&P 500 can be predicted by FTSE China A50 nightly performance using an OLS model fitted on previous 67 days results. If given fresh data in time, the model is ready to make prediction about the dynamics of the next hour at 9AM EST.
4. However, predictive power is not guaranteed to be strong: R2 is around 0.17 when predicting 31 latest observations, and the model is able to guess the direction of price movement correctly in around 76% cases. We can expect to observe comparable results if there is no change in external factors.
5. If simplicity and interpretability are not in priority, and our goal is performance, the rolling inner model can be replaced by a more complex one (polynomial fit, ARMA, random forest, neural network), with added predictors (*probably*: volume, rolling volatility, lagged val-

ues/moving averages - ARIMA style etc.) and the window - readjusted. One such experiment brought R2 score up to 0.28 while keeping F1 at approximately same level of 0.75, which is probably still not applicable to use in full automatic mode, but a considerable improvement nevertheless.

6. The best model I was able to compose is an MLP on a rolling window of size 113, using 4 lagged values of US diff in addition to China diff as predictors. The model arguably outperformed both simple and polynomial rolling OLS, coming second in terms of R2 score (0.2) and first in F1 (0.84) by a wide margin.

```
                           Model    R2    F1
0                Simple Rolling OLS  0.17  0.76
1            Rolling Polynomial OLS  0.28  0.75
2  Rolling MLP with lagged predictors  0.20  0.84
```

7. Taking everything above into account, blindly using the model is not advised: it can be used as a supporting tool or may make profit on statistically significant numbers, but in order to make it a fully automated decision-making tool, more factors should be added and complex trading strategies should be applied, especially if it is a big one-time sell/purchase.