A statistical study on the soybean and bitcoin future markets

Kaveen Herath Bandara
(kh3001)
Tianmai Jiao
(tj2466)
Chong Hou Lao
(cl4022)
Anne Venezia
(av2744)
Cheng Zhong
(cz2632)
Yidi Zhao
(yz3210)

Department of Mathematics at Columbia University

Math Methods in Financial Analysis
Spring 2021

Introduction
- Backgrounds
- Soybean oil
- Historical prices background - soybean oil
- Bitcoins
- Historical prices background - Bitcoins

Statistical test on the markets
- Push-response test - Soybean oil
- Push-response test - Bitcoins
- Variance-ratio test

Backtesting
- Methodology
- Trade results - Soybean oil
- Trade results - Bitcoin
- The End

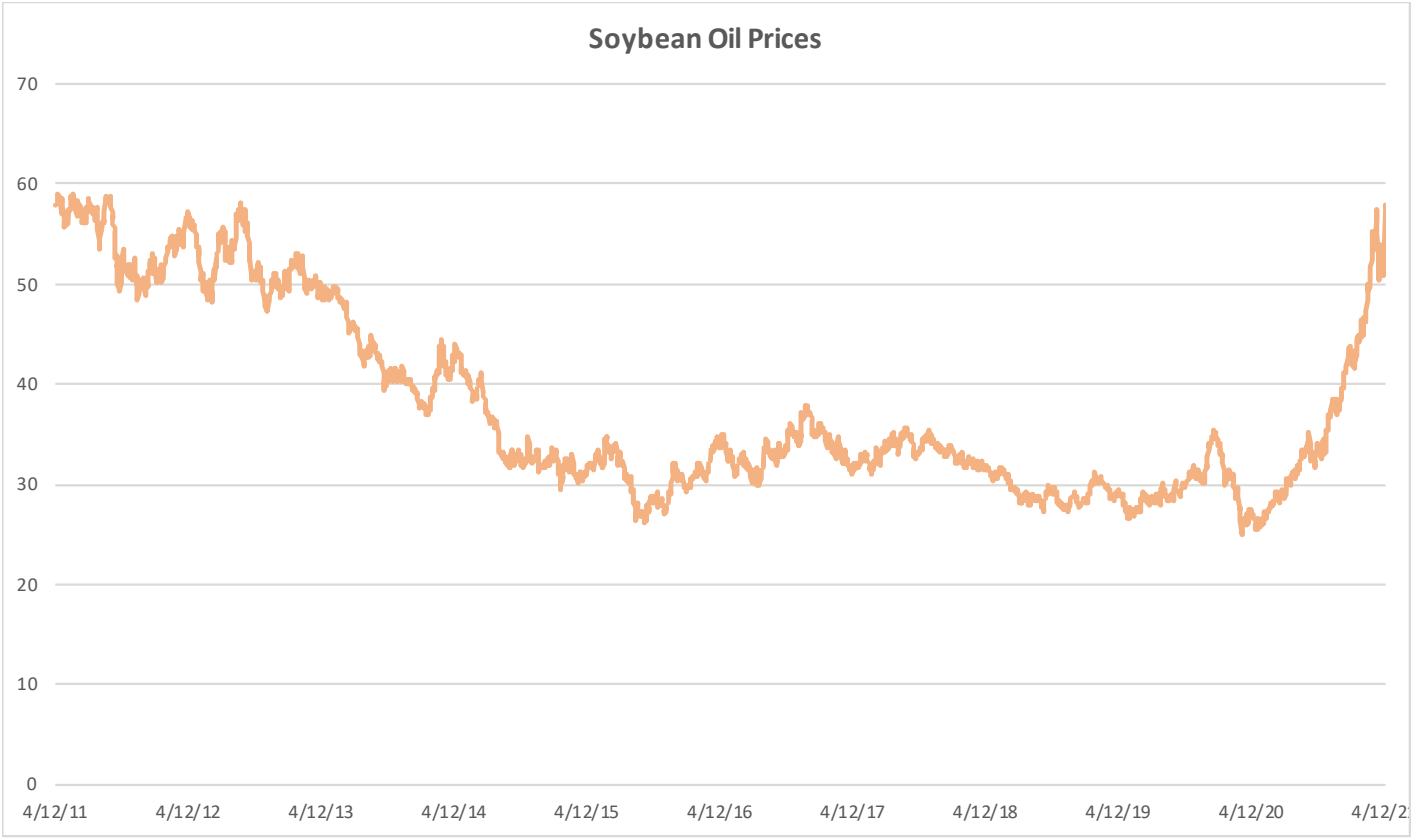
What are futures?

- A futures contract is a binding agreement between a seller and a buyer to make (seller) and to take (buyer) delivery of the underlying commodity (or financial instrument) at a specified future date with agreed upon payment terms.
- Futures contracts protect suppliers and producers from price changes.
  - For example, a corn farmer and a cereal company may enter into a futures agreement to lock in a price for a future delivery of corn in a particular month.

What do we know about soybean oil?

- Chemicals from soybean oil are used for high cholesterol. Soybean oil is also used as a mosquito repellant and as a nutritional supplement in intravenous feedings.
- Due to rising biodiesel purchases, the soybean oil demand as a feedstock has also seen a pickup
- The oil is produced from the seeds of the soybean plant

Historical prices - soybean oil



Summary statistics - soybean oil

Soybean Oil Historical Data	
Average Daily Price Change	$0.01
Standard Deviation of Daily Price	$0.32
Average Price of Soybean Oil	$37.76
Max Price (31 August 2011)	$58.92
Min Price (16 March 2020)	$24.99
Max Daily Price Increase	$2.50
Max Daily Price Decrease	($2.65)

What happened?

- Global Soybean Oil Market Outlook:The industry is currently being driven by increasing demand of soybean meal for livestock products which has, thus, contributed considerably to the rise in soy oil production and rising demand of soybean oil. Other advantages such as easy availability, competitive prices, environment-friendly and various nutritional benefits of soy oil are the factors that are stimulating its consumption.
- The impact of COVID-19 on global soybean oil market: consumer focus has shifted toward the consumption of nutrient-rich food items for boosting immunity, such as soybean oil, which is projected to act favorably for market growth. However, the lockdowns imposed by governments to curb the spread of the pandemic have disrupted the global soybean trade, in turn, influencing the import and export scenario for the soybean oil market.

What do we know about Bitcoins?

- Each bitcoin is a computer file stored in a digital wallet on a computer or smartphone.
- Blockchain: Bitcoin is powered by open-source code, creates a shared public ledger. Each transaction is a "block" that is "chained" to the code, creating a permanent record of each transaction.
- Private and public keys: A bitcoin wallet contains a public key and a private key, which work together to allow the owner to initiate and digitally sign transactions, providing proof of authorization.
- Bitcoin miners: Miners — or members of the peer-to-peer platform — then independently confirm the transaction using high-speed computers, typically within 10 to 20 minutes. Miners are paid in bitcoin for their efforts

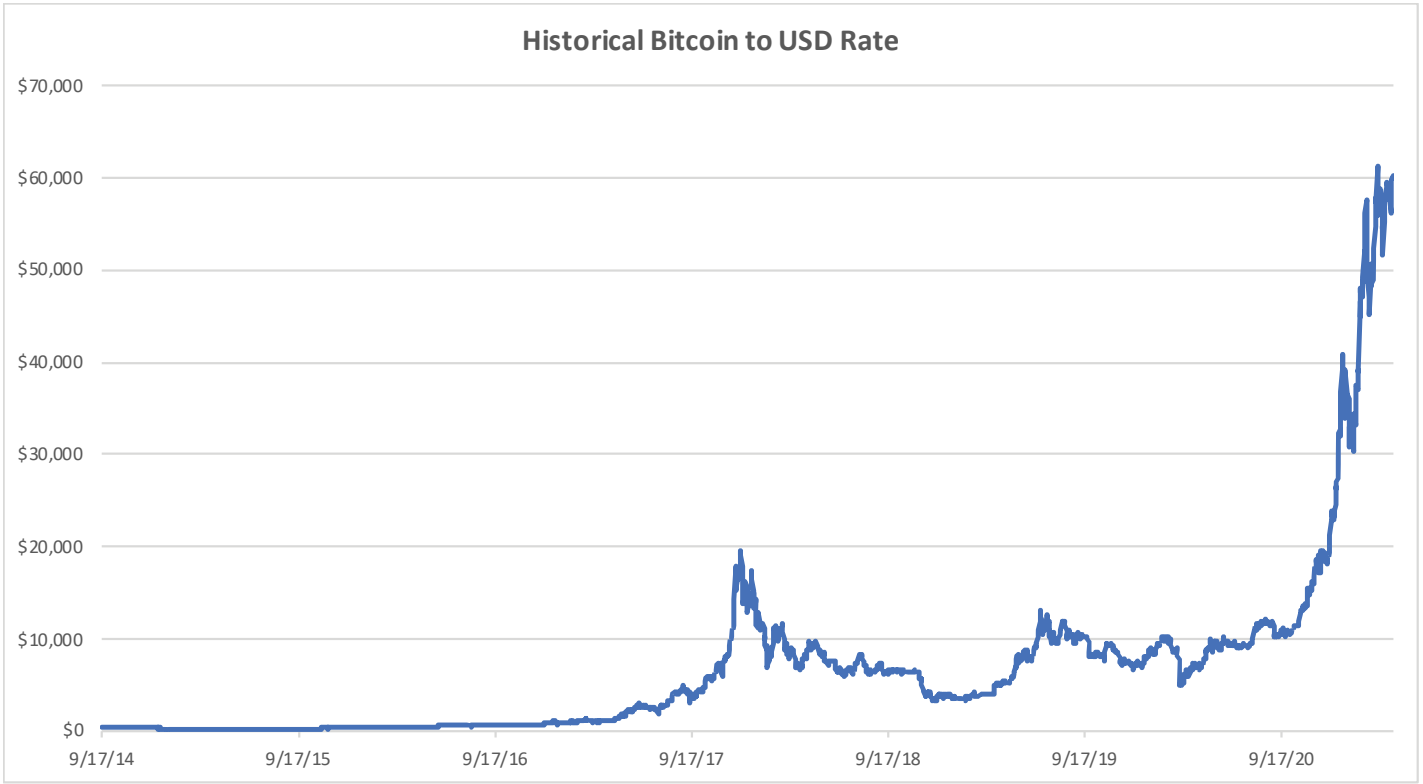
Why is Bitcoin popular?

- Private, secure transactions anytime — with fewer potential fees.
  - you can transfer them anytime, anywhere, reducing the time and potential expense of any transaction. don't contain personal information like a name or credit card number, which eliminates the risk of consumer information being stolen for fraudulent purchases or identity theft.
- The potential for big growth.
  - Investors bet that once Bitcoin matures, greater trust and more widespread use will follow, and therefore Bitcoin's value will grow.
- The ability to avoid traditional banks or government intermediaries. It is outside the control of regular banks, governing authorities or other third parties

How is it different to a traditional currency?

- Bitcoin created a new category and has the network effect as a result.
- The main difference is the traditional currency is a centralized system and bitcoins are decentralized one and peer-peer systems. Both the bitcoins and fiat currency have values which can be used for buying and selling of goods in the market.
- How to get a Bitcoin?
  - Cryptocurrency exchanges: Coinbase
  - Investment brokerages: Robinhood
  - Bitcoin ATMs.
  - Peer-to-peer purchases: from other bitcoin owner
  - Bitcoin mining: technical expertise required and computer cost puts this option out of reach for most

Historical prices - Bitcoins

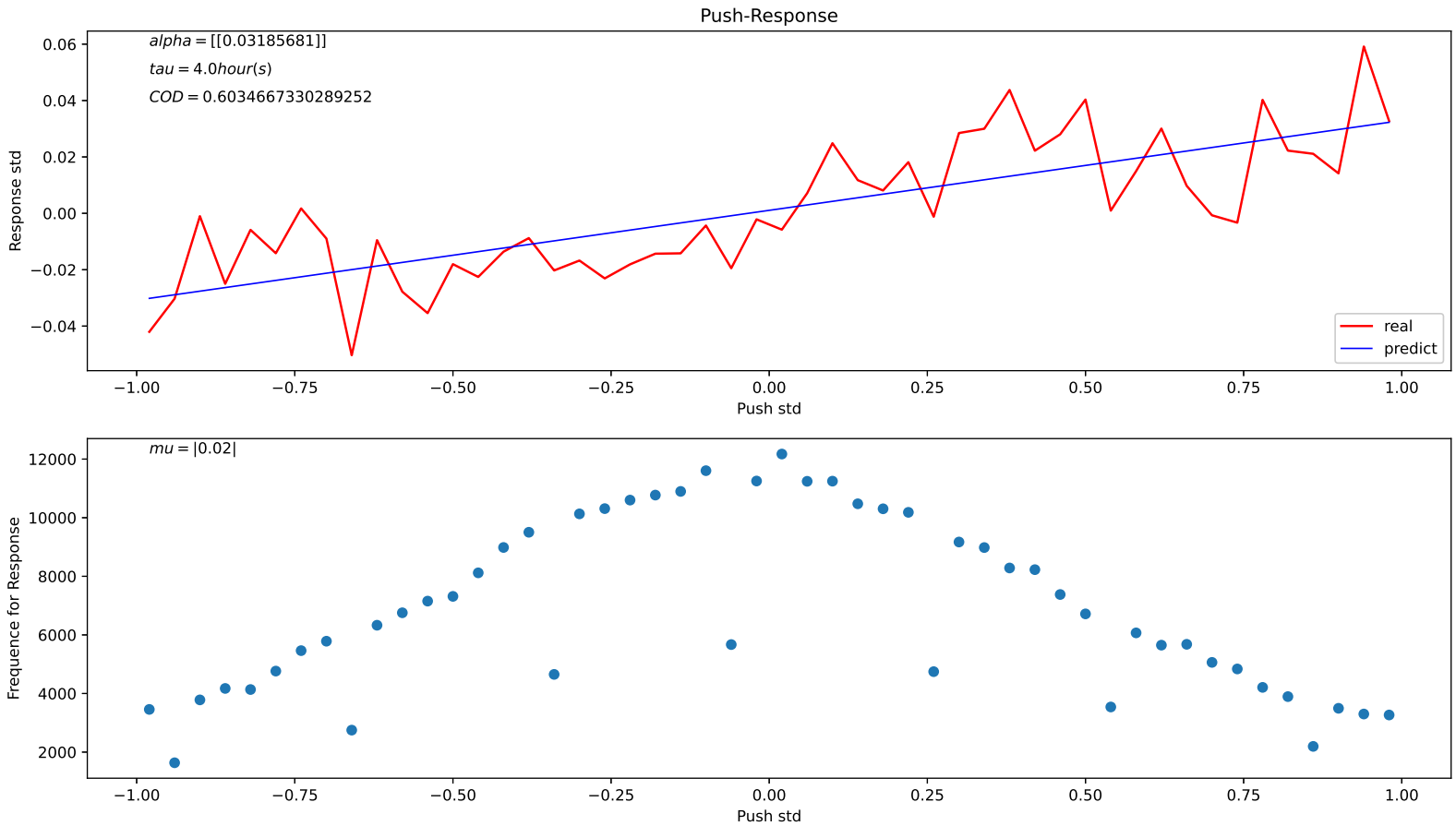


Summary statistics - Bitcoins

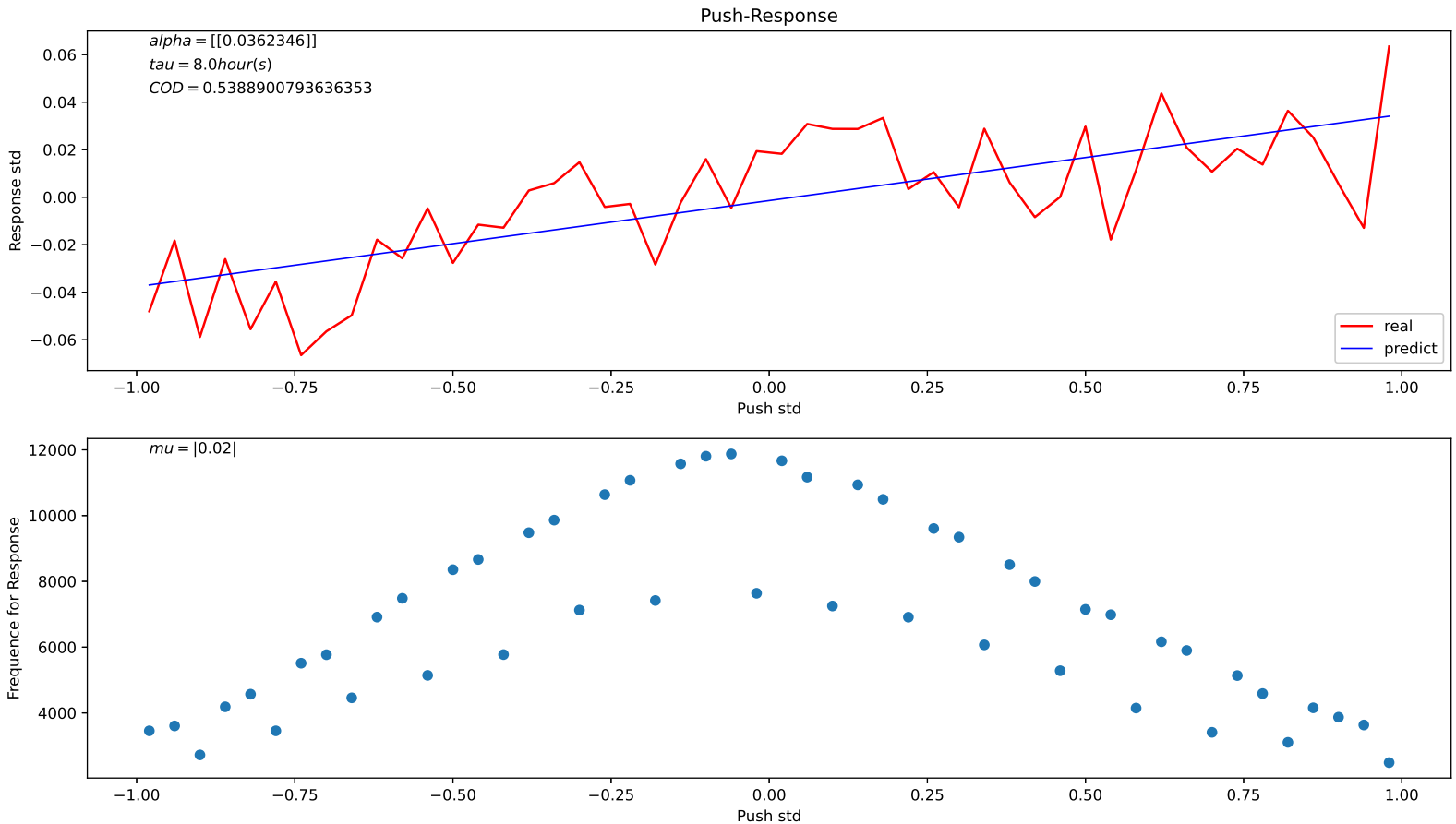
Bitcoin to USD Historical Data	
Average Daily Price Change	($23.82)
Standard Deviation of Daily Price Change	$502.52
Average Price	$6,711.56
Max Price (13 March 2021)	$61,243.09
Min Price (14 January 2015)	$178.10
Max Price Increase	$5,380.50
Max Price Decrease	($7,309.64)

Bitcoin is a highly volatile instrument that has experienced tremendous price movements over the years, sometime gaining hundreds of percentage points or crashing significantly over a relatively short period of time. While it is less affected by happenings in mainstream markets, it can be affected by a variety of factors relating to crypto space, the blockchain industry and by regulatory issues.

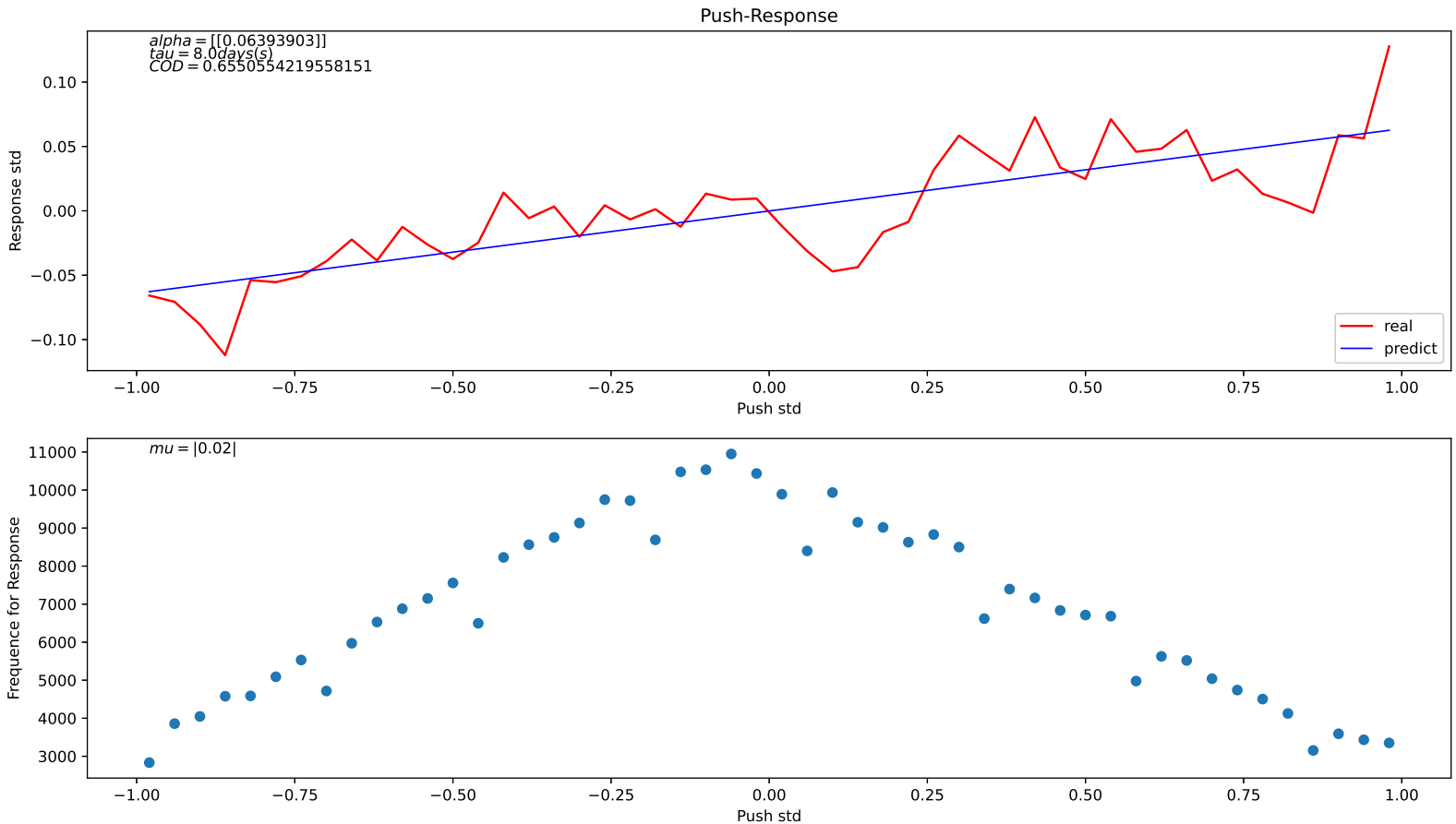
Soybean push response test - 4 hours lag



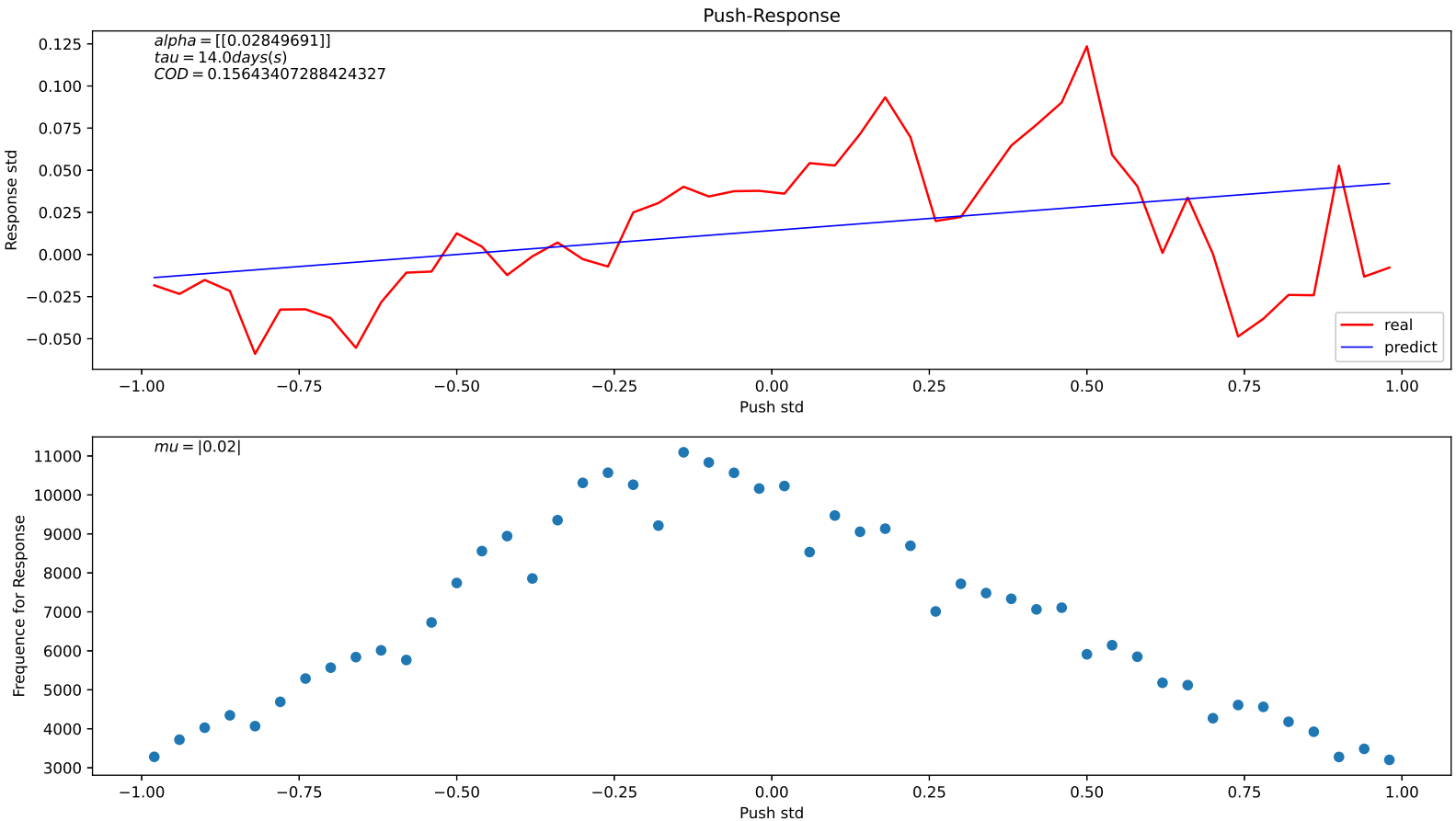
Soybean push response test - 8 hours lag



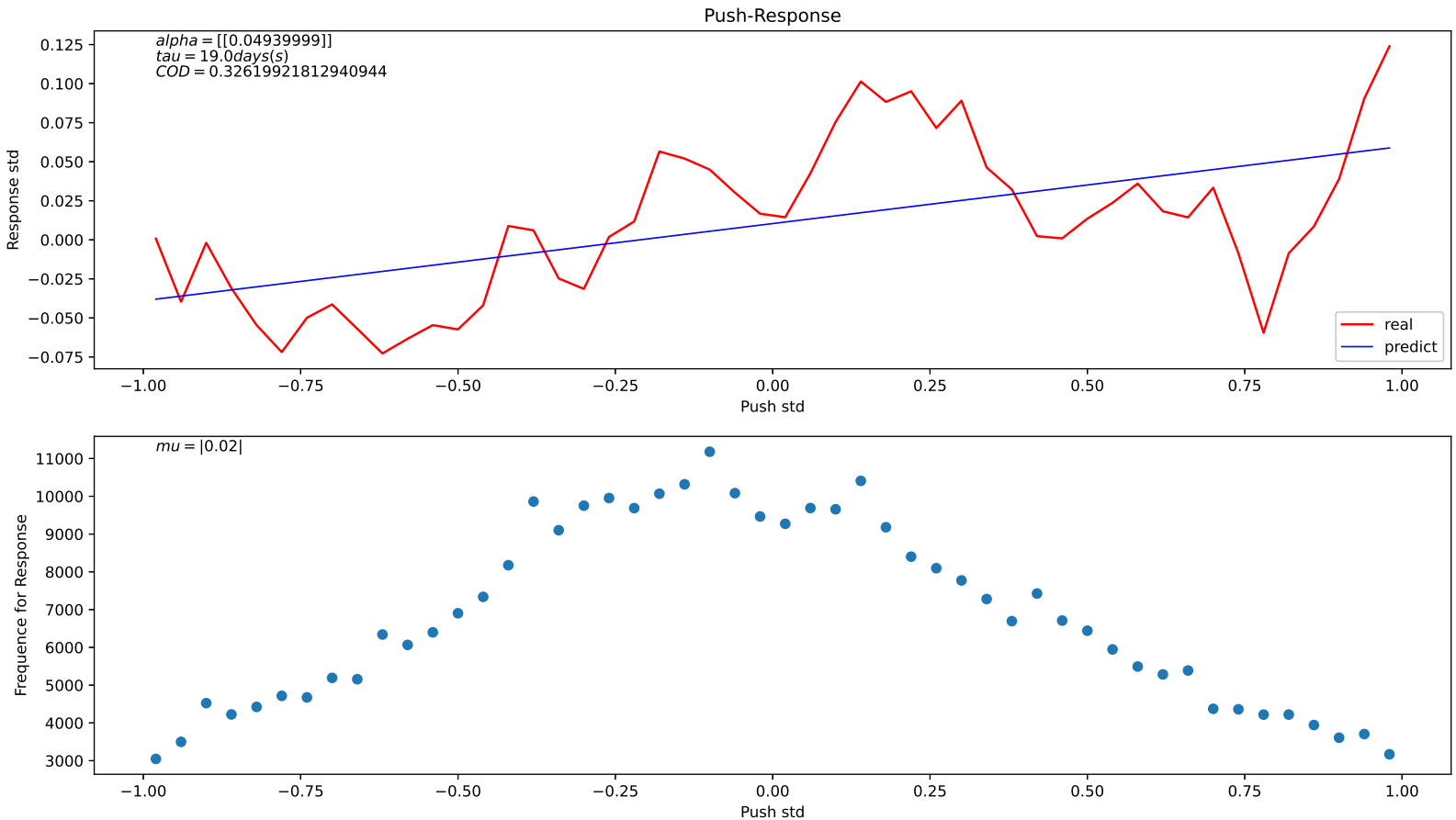
Soybean push response test - 8 days lag



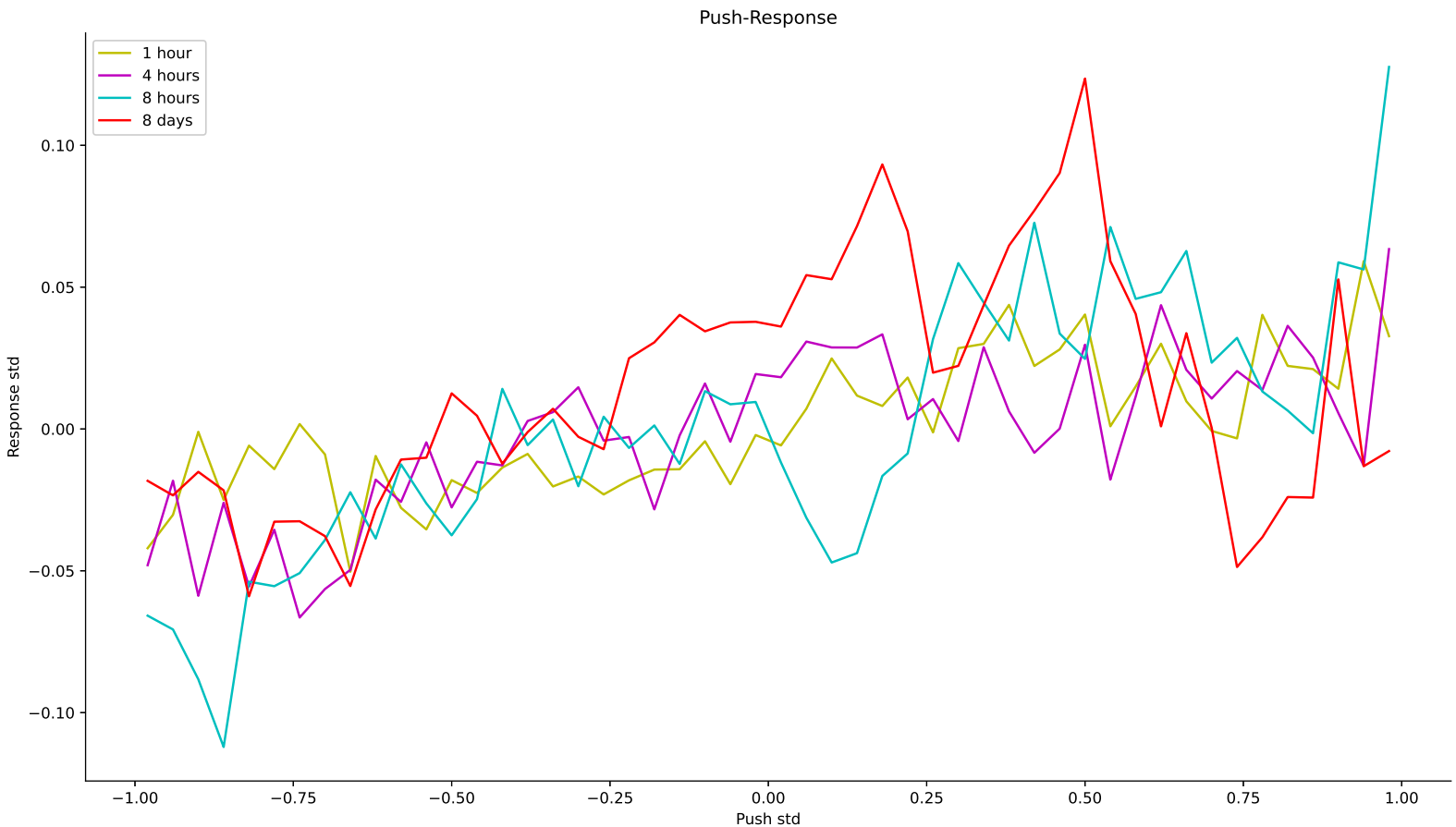
Soybean push response test - 14 days lag



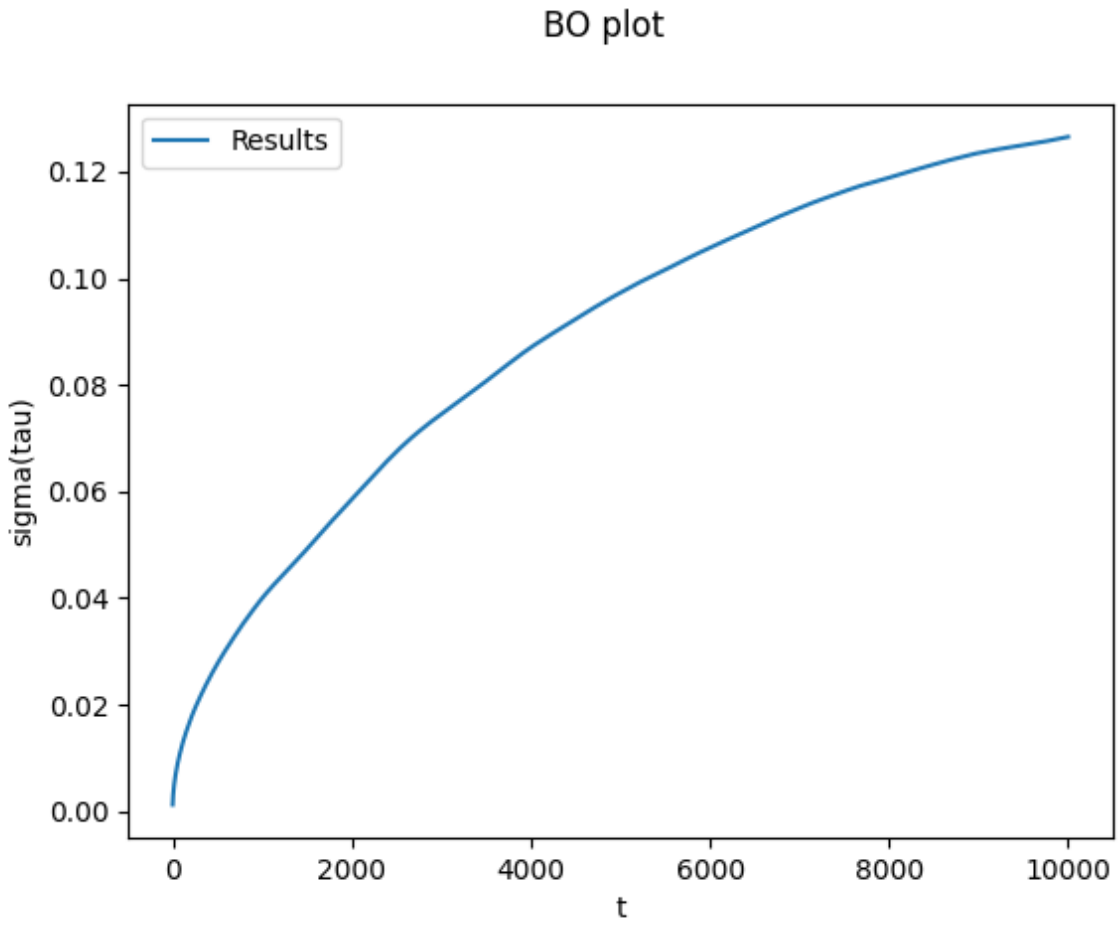
Soybean push response test - 19 days lag



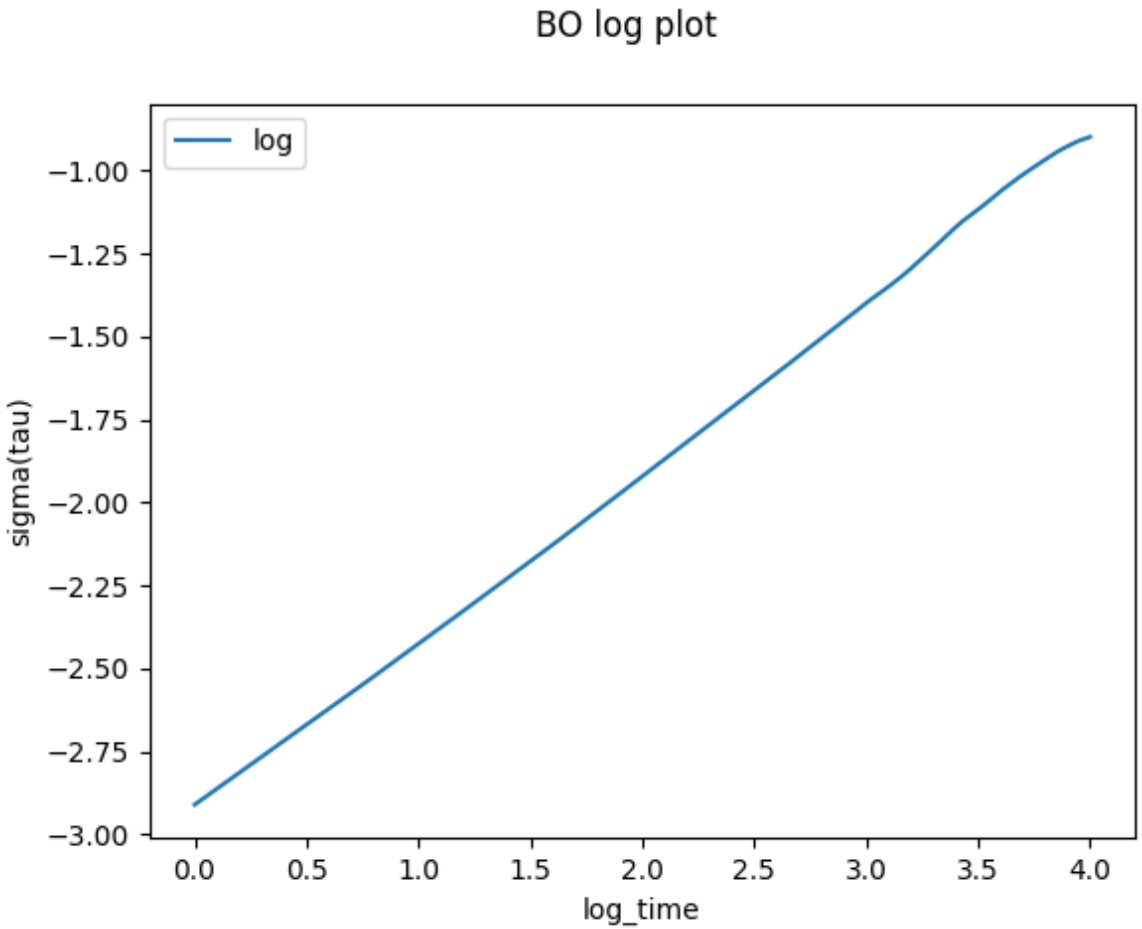
Soybean push response test



Soybean push response test



Soybean push response test



Soybean push response test

OLS Regression Results

Dep. Variable:	log	R-squared:	0.997
Model:	OLS	Adj. R-squared:	0.997
Method:	Least Squares	F-statistic:	3.361e+06
Date:	Sat, 17 Apr 2021	Prob (F-statistic):	0.00
Time:	14:45:16	Log-Likelihood:	29764.
No. Observations:	10000	AIC:	-5.952e+04
Df Residuals:	9998	BIC:	-5.951e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-2.9560	0.001	-2889.911	0.000	-2.958	-2.954
log_time	0.5220	0.000	1833.298	0.000	0.521	0.523

Omnibus:	1228.396	Durbin-Watson:	0.000
Prob(Omnibus):	0.000	Jarque-Bera (JB):	652.323
Skew:	-0.477	Prob(JB):	2.24e-142
Kurtosis:	2.191	Cond. No.	32.1

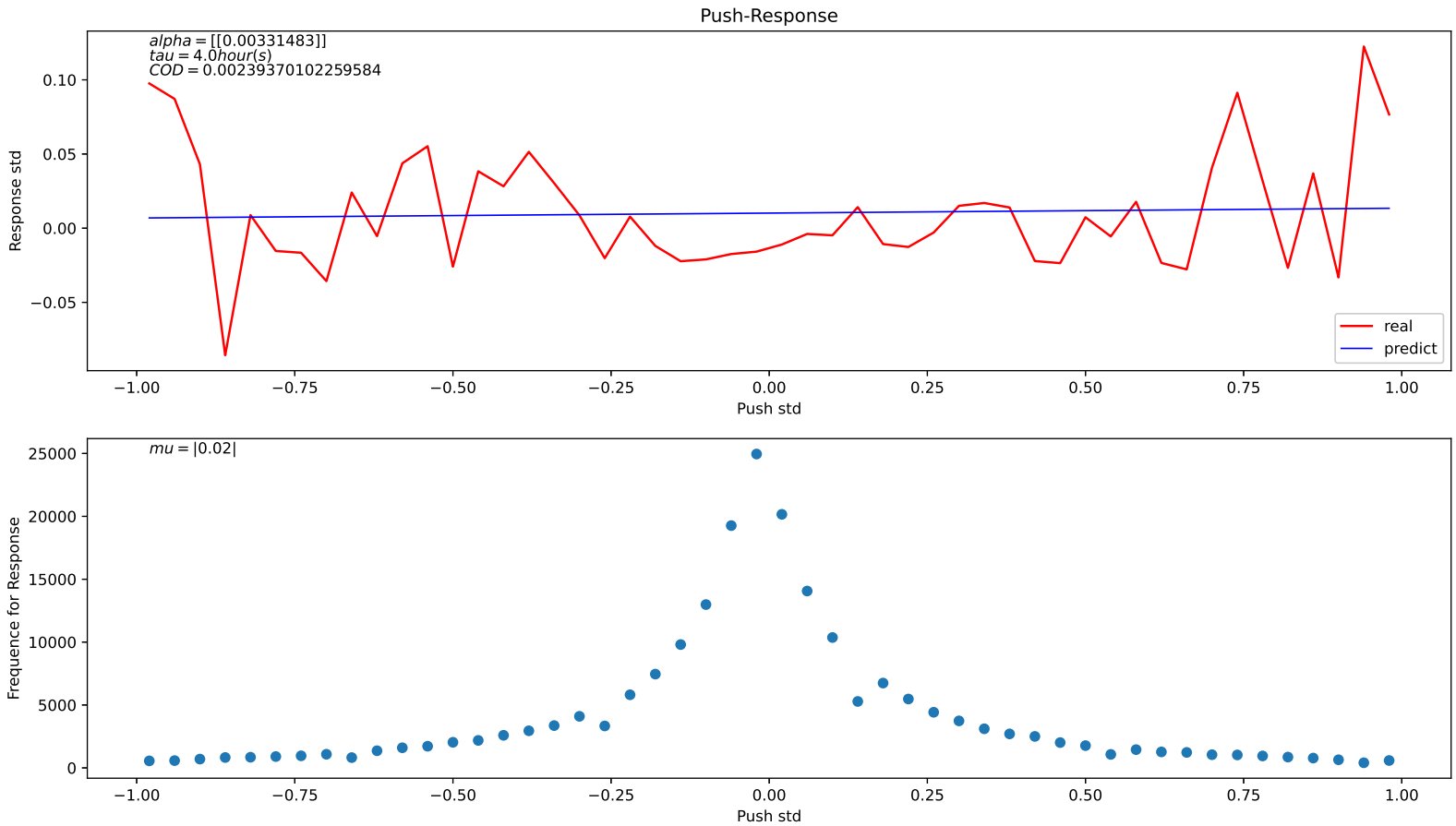
Interpretation

- Recall that for a Levy's distribution:

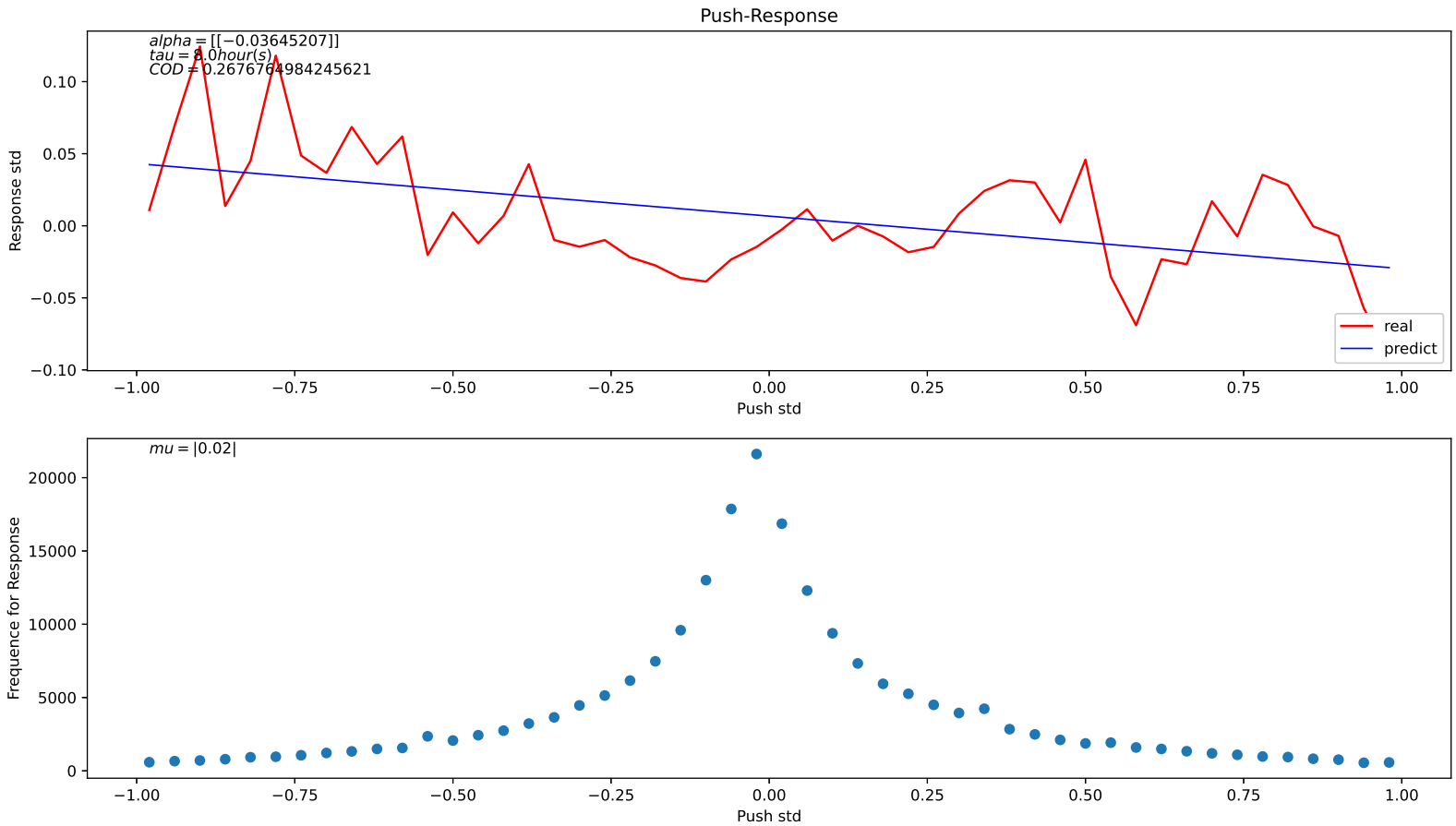
$$\log_{10}(\sigma) = \alpha + \beta \log_{10}(\tau) \implies \sigma = 10^{\alpha} \times \tau^{\beta}$$

- Here we have computed the coefficients that $\alpha = -2.956$ and $\beta = 0.5220$
- This returns the variance of the time series as $10^{-2.956}$.
- More importantly, 0.5220 is just above 0.5 which would correspond to a Brownian motion $\sigma \propto 0.5$, so that we know the auto-covariance for higher lags, is above the 'base level' we would expect on the null hypothesis of a Brownian Motion.
- Hence we conclude that there is some momentum effect in the market.

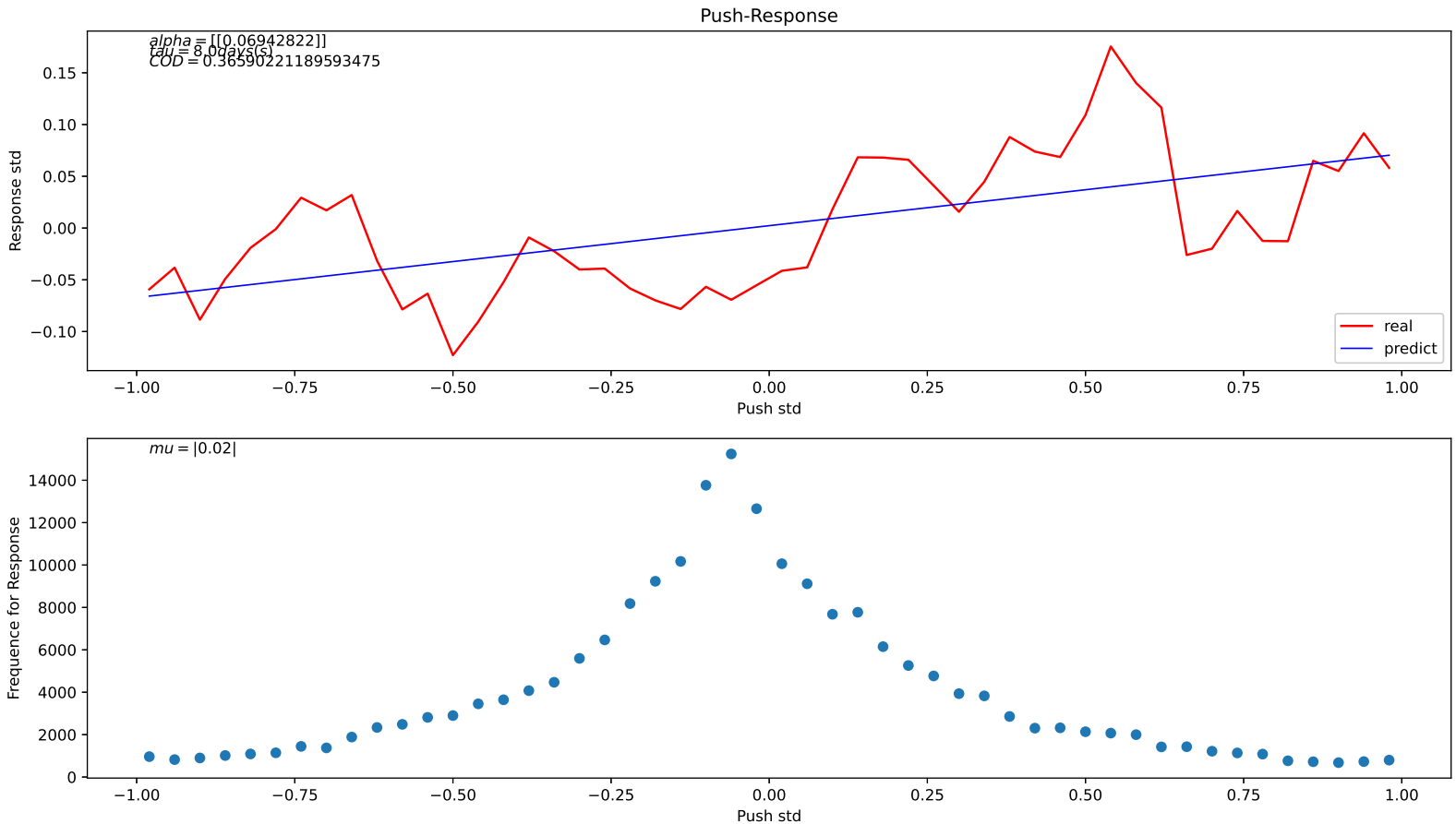
Bitcoin response test - 1 hour lag



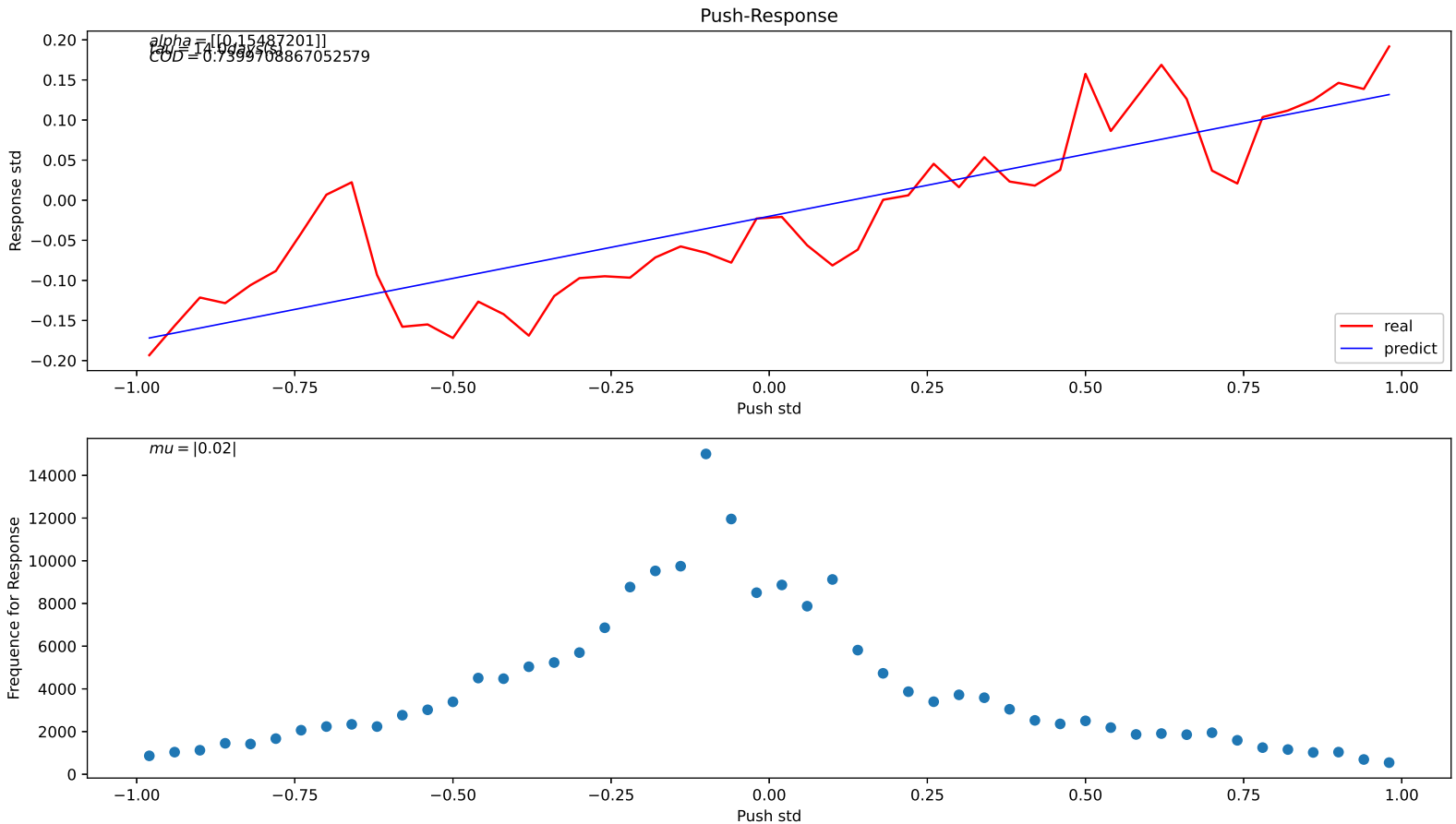
Bitcoin response test - 4 hours lag



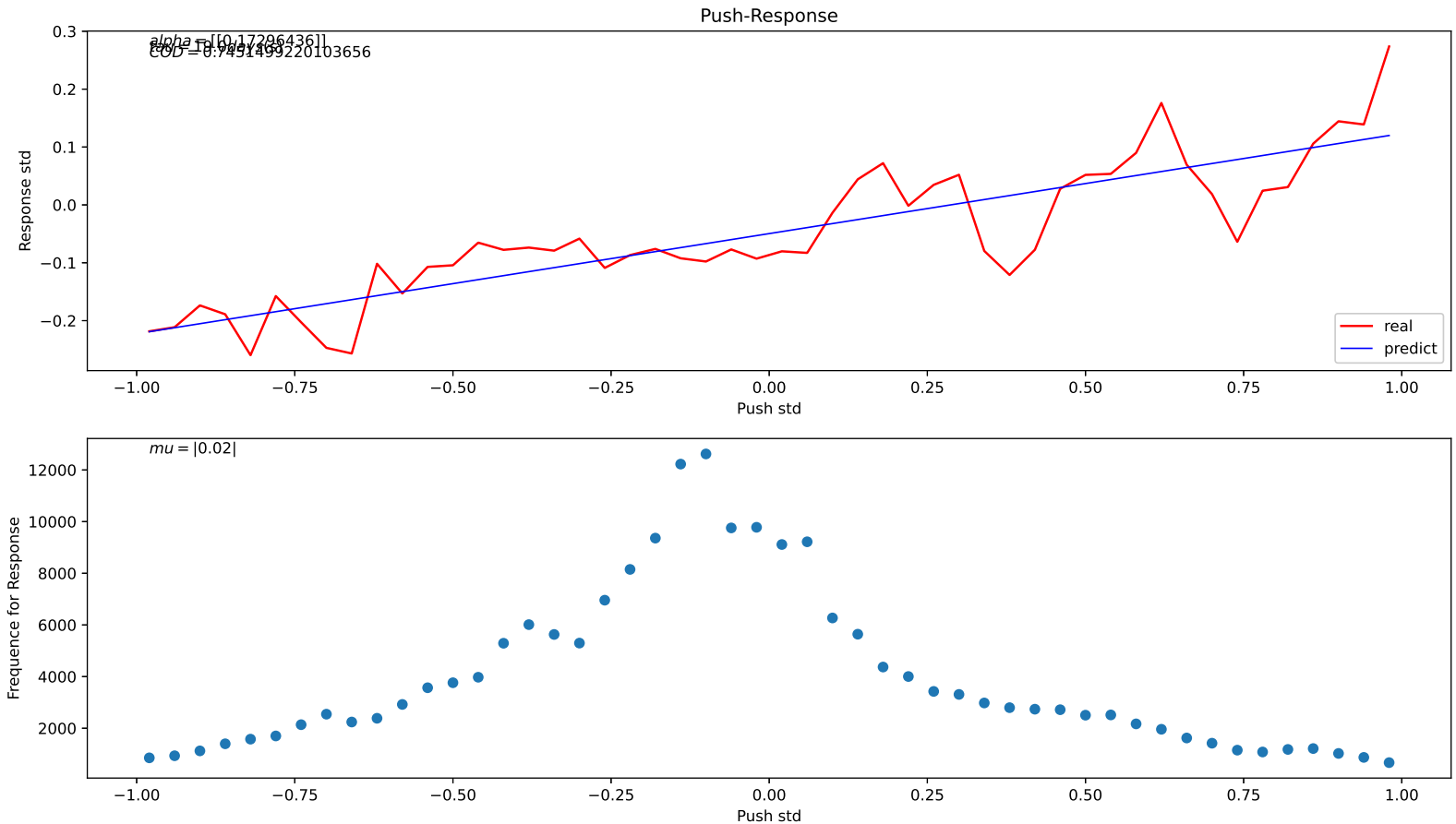
Bitcoin response test - 8 hours lag



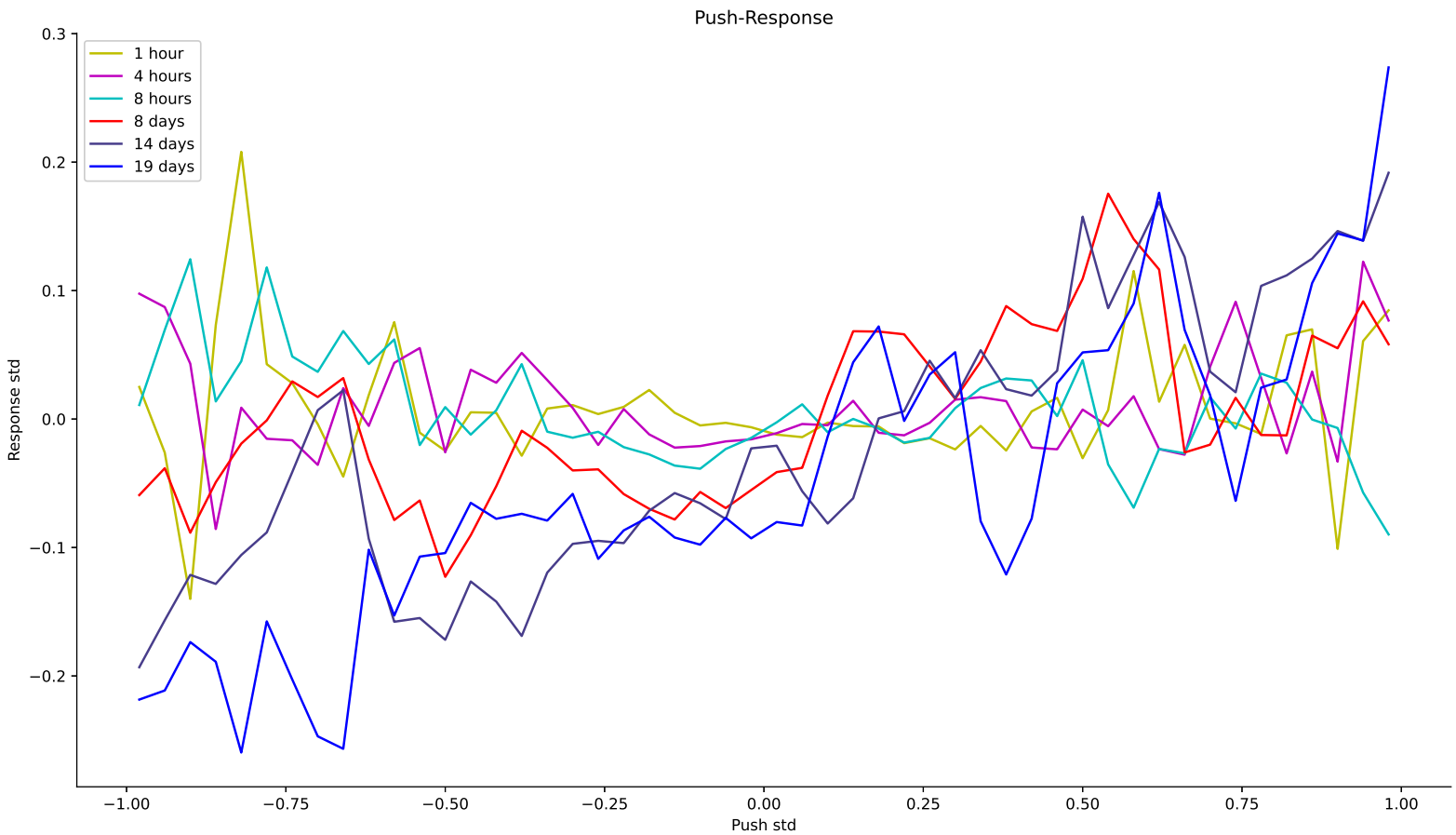
Bitcoin response test - 14 days lag



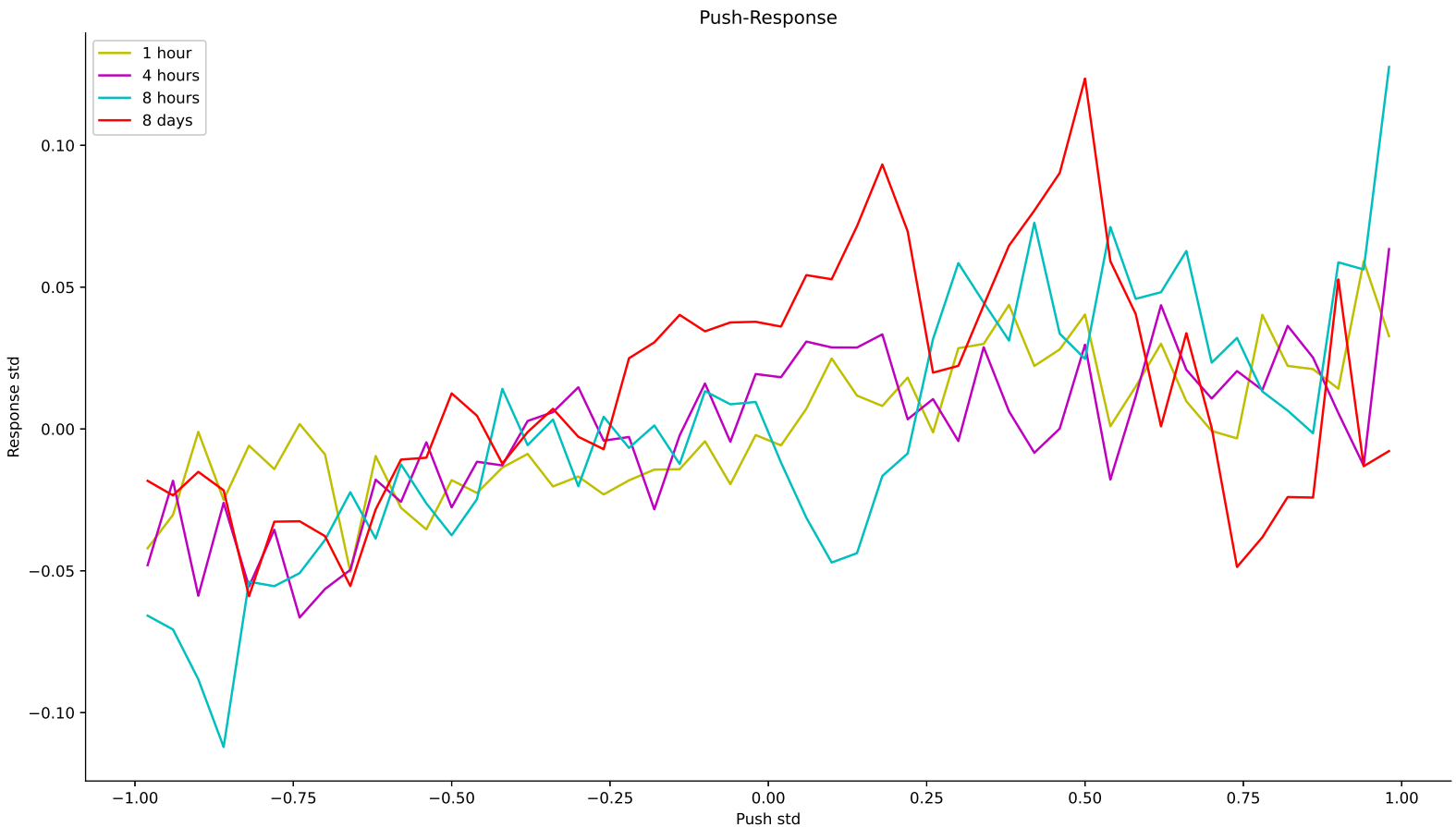
Bitcoin response test - 19 days lag



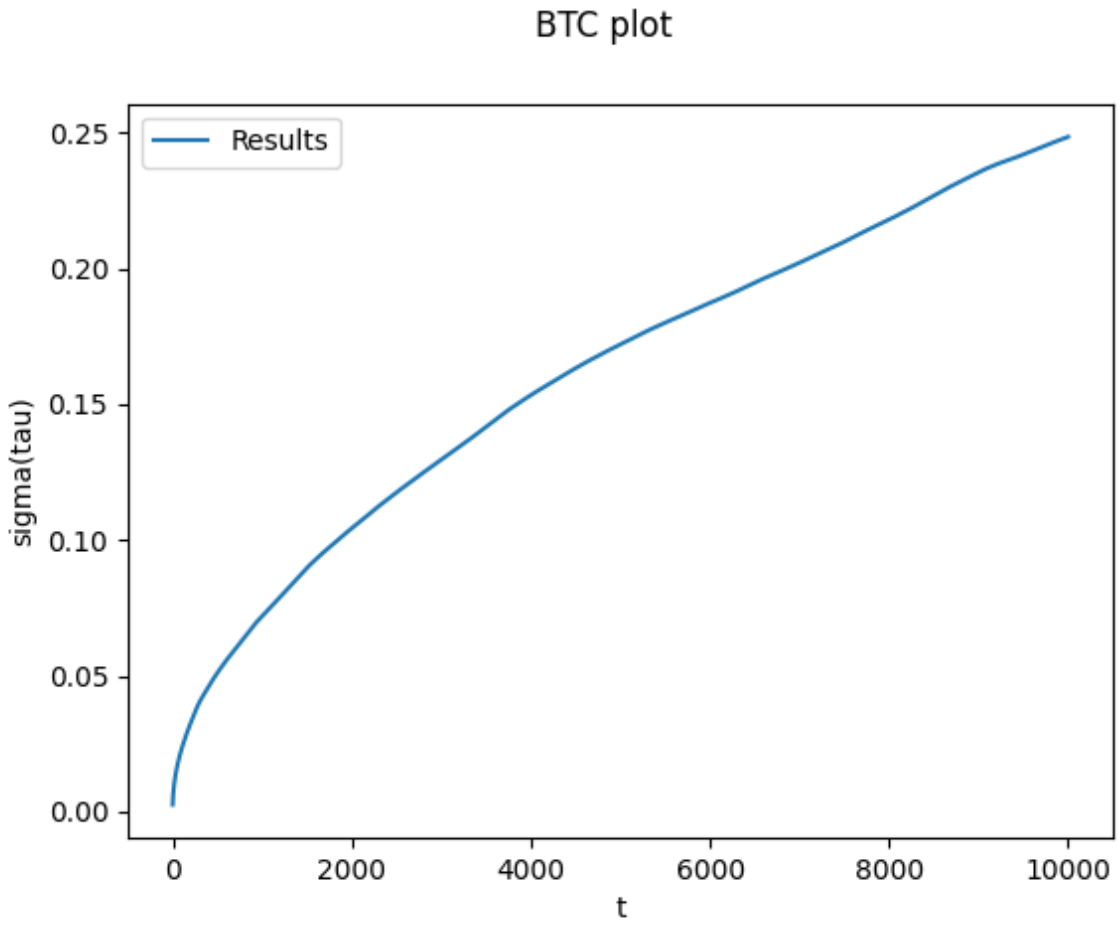
Bitcoin response test



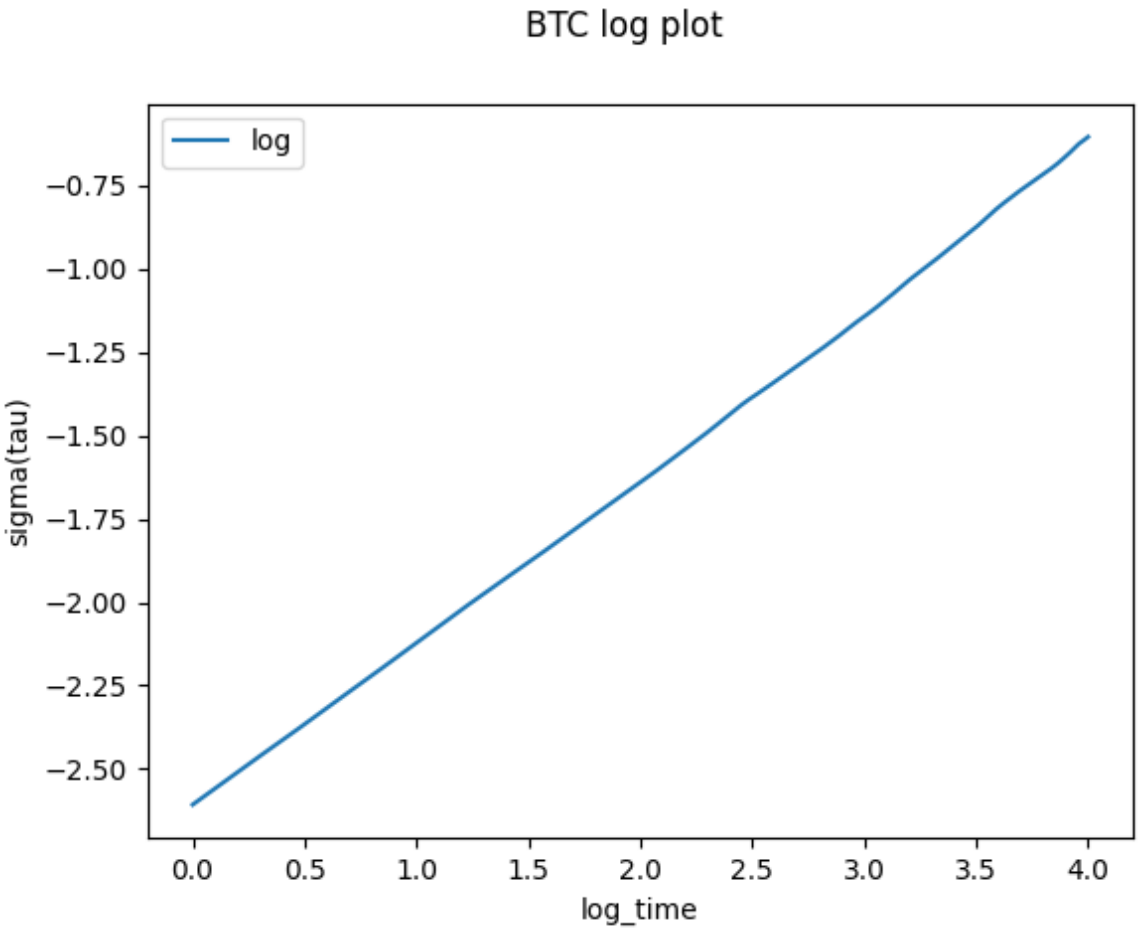
Bitcoin response test



Bitcoin response test



Bitcoin response test



Bitcoin response test

OLS Regression Results

Dep. Variable:	log	R-squared:	0.999
Model:	OLS	Adj. R-squared:	0.999
Method:	Least Squares	F-statistic:	1.358e+07
Date:	Sat, 17 Apr 2021	Prob (F-statistic):	0.00
Time:	14:50:40	Log-Likelihood:	36743.
No. Observations:	10000	AIC:	-7.348e+04
Df Residuals:	9998	BIC:	-7.347e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-2.6994	0.001	-5303.037	0.000	-2.700	-2.698
log_time	0.5222	0.000	3685.301	0.000	0.522	0.523

Omnibus:	6484.916	Durbin-Watson:	0.001
Prob(Omnibus):	0.000	Jarque-Bera (JB):	249895.339
Skew:	2.564	Prob(JB):	0.00
Kurtosis:	26.947	Cond. No.	32.1

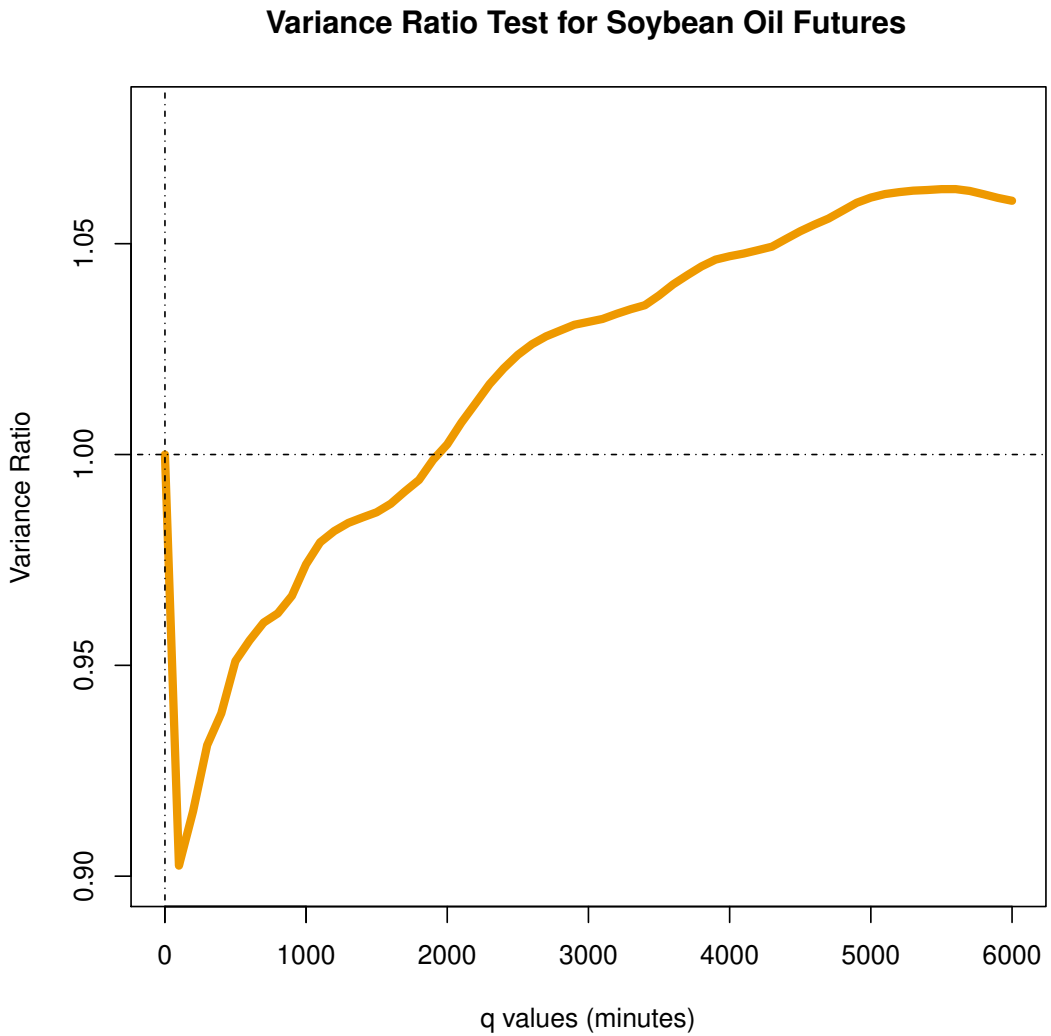
Interpretation

- This time we have $\alpha = -2.6994$ and $\beta = 0.5220$
- So again, we conclude that there is some momentum effect in the market.

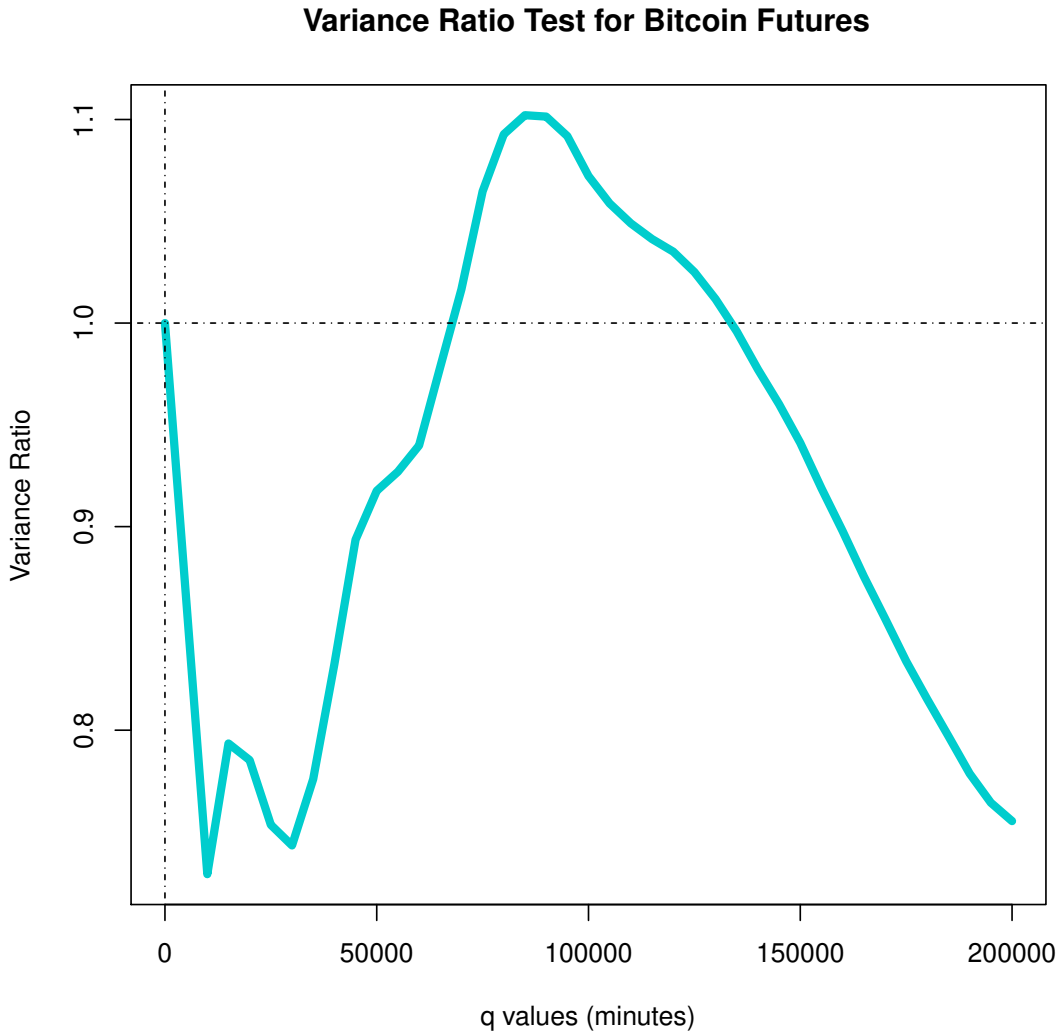
Highlight summary

- The variance ratio tests that were conducted on the two futures markets; the value of q that was used ranged from 1 minute to 200,000 minutes. The tests produced the following results:
- Bitcoin initially produces a pattern of inefficiency of mean reversion. This pattern of inefficiency implies that agents over-react to new information by overselling if negative information is received or over-buying if good information is received. Both trends (overselling or over-purchasing) are later corrected by market agents by reversing the over-purchase or over-selling of BTC. A similar pattern is also seen in a previous experiment (Chekhlov, 2010) with other currency variance ratios, such as the Canadian and Australian dollar. However, it then appears to follow a Trend-following pattern.
- Soybean oil produces a trend-following pattern. By trend following, market agents under-react to new information, and establish a partial position, whilst awaiting confirmations from other market agents. When confirmation is received, these agents continue to increase their position in the same direction (buy or sell). This produces a delayed, chain reaction, wherein the new market information is slowly priced into the market.

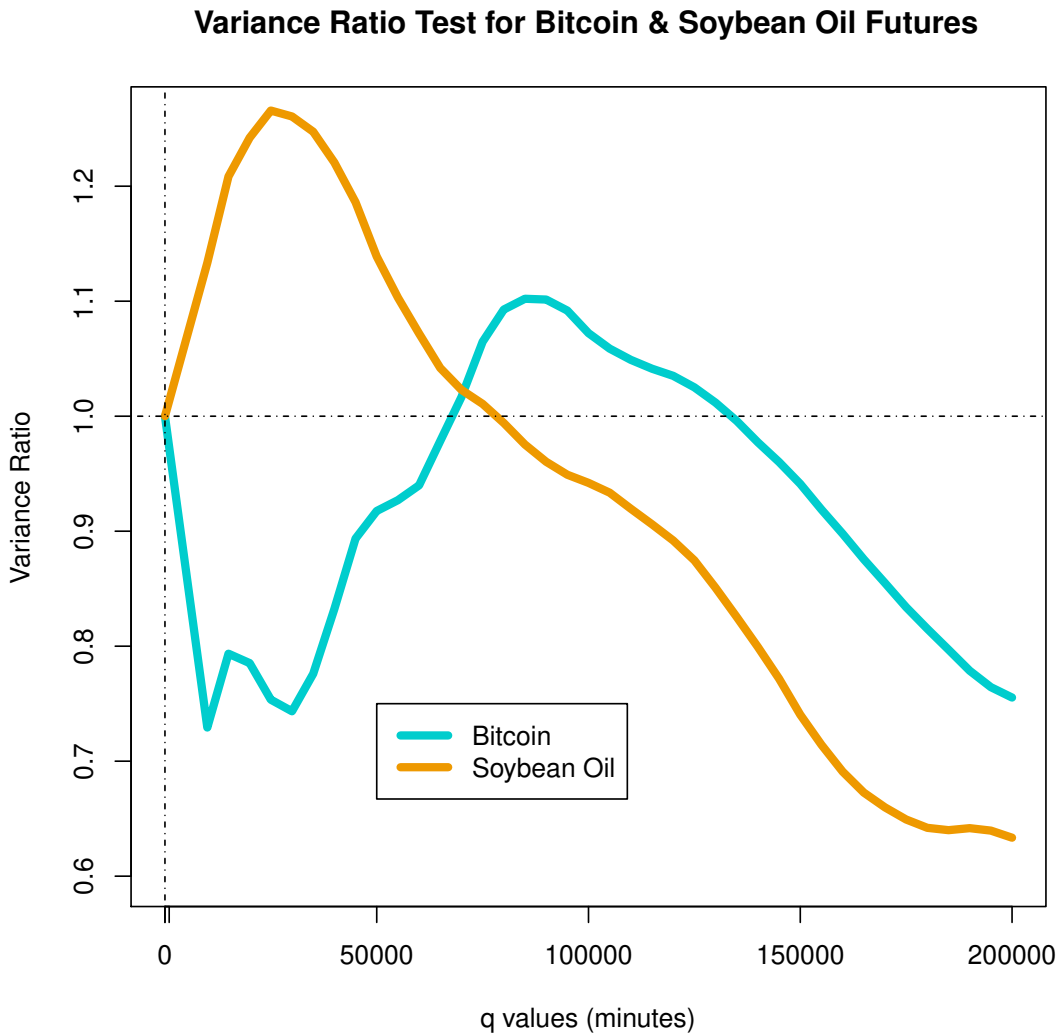
Statistical performance soybean oil $T = 3, \tau = 3$



Statistical performance soybean oil $T = 3, \tau = 3$



Statistical performance soybean oil $T = 3, \tau = 3$



How we did it

- We had originally written the trade code in Numpy but that's not fast enough. We ended up writing the trade code in C++ and compile it into a Python module using Pybind11. By that we achieved a 50-fold time reduction in the trade code itself, or 10-time reduction if we consider all the processing done in Python and C++.
- The trade parameters over 'Channel length' and 'Stop Percent' are optimised using an automatic hyperparameter (parameters in a learning algorithm) framework in Python. More will be explained later.
- To better use the limited computing power, multi-processing was also used to run the code in parallel.

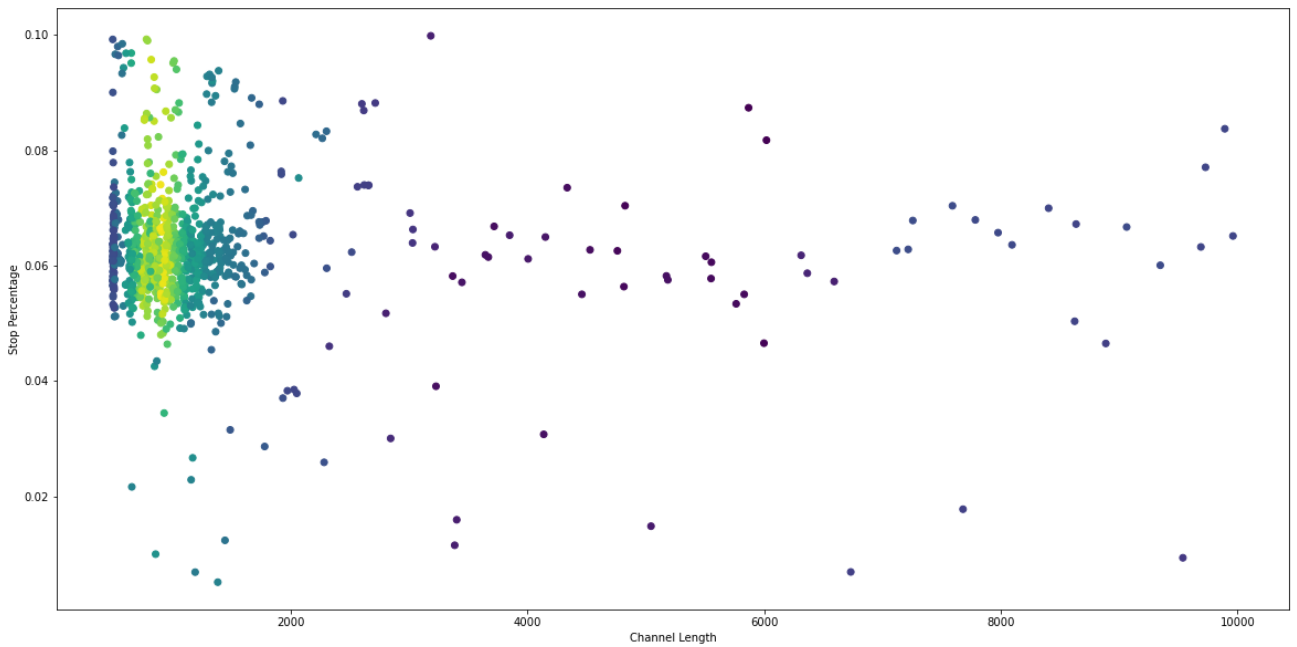
Pybind

- We wrote the trade code/function in C++ and compiled the code into a python package.
- The trade function is wrapped as below:

```
1  #include <iostream>
2  #include <vector>
3  #include <tuple>
4  #include <pybind11/pybind11.h>
5  #include <pybind11/numpy.h>
6  #include <pybind11/stl.h>
7
8  namespace py = pybind11;
9
10 // py::tuple<double, double, double> trade_cpp(py::array_t<double> df_array, double stop_per
11 std::tuple<py::array_t<double>, int, std::tuple< std::vector<int>, std::vector<double>, std:
12 trade_cpp(py::array_t<double> df_array, double stop_per, double slpg, double PV, double E0,
   double init benchmarkShort*/) {
```

Optuna

- We use Bayesian optimization implemented in Python package Optuna for hyperparameter selection (channel length and stop percentage), which sequentially tries out the most promising hyperparameters and is more efficient than grid search.
- Based on our experiments, using 1000 hyperparameter trials in Bayesian optimization can reach a similar performance level as compared to 90,000 hyperparameter trials in grid search.



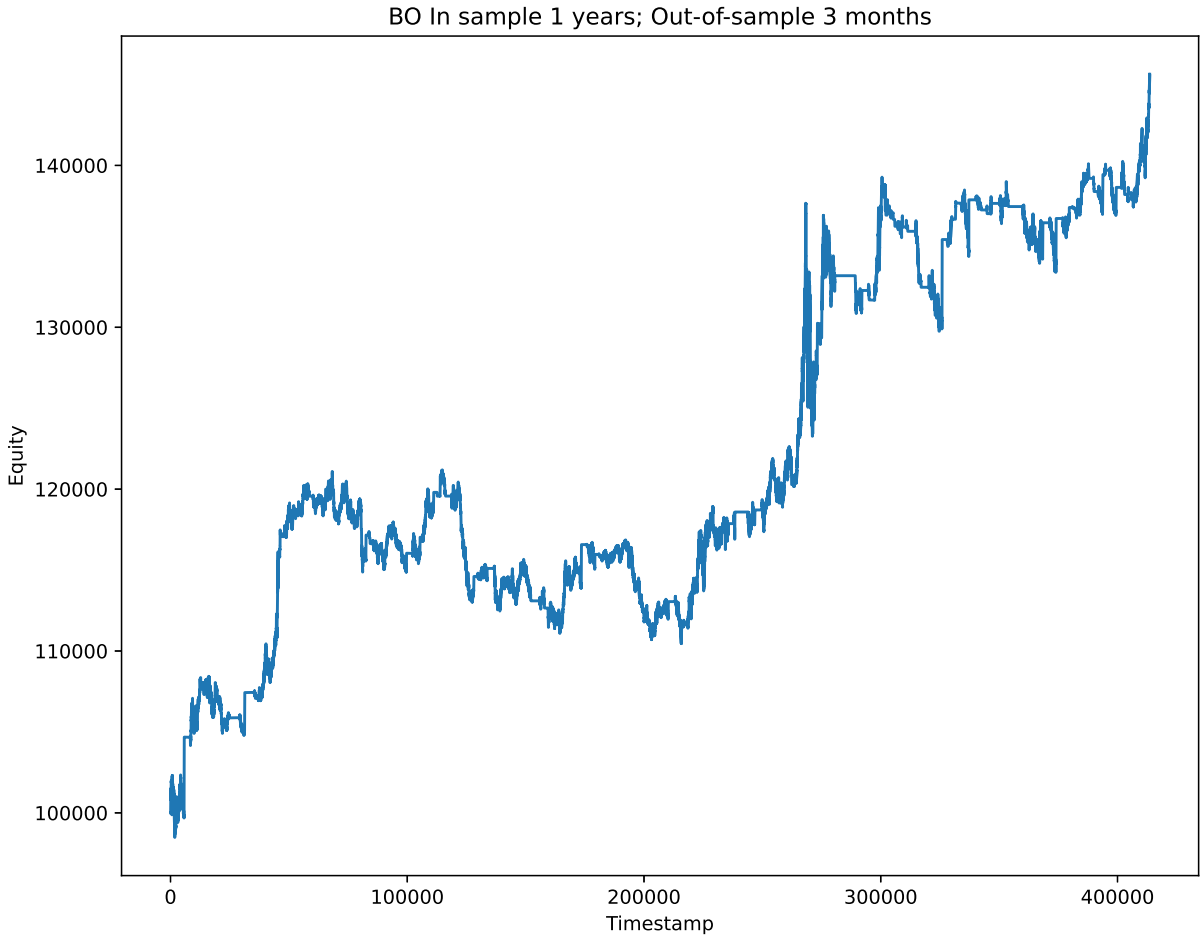
Multi-processing

- We use multiple processing parallelization to leverage multiple CPU cores in laptop to accelerate the computation

High level result - soybean oil

	Soybean Oil (BO) ROA Results	
	τ=3	τ=6
T=1	20.3777	10.8610
T=2	18.6381	12.1540
T=3	9.3845	10.1919
T=4	16.5989	9.7485

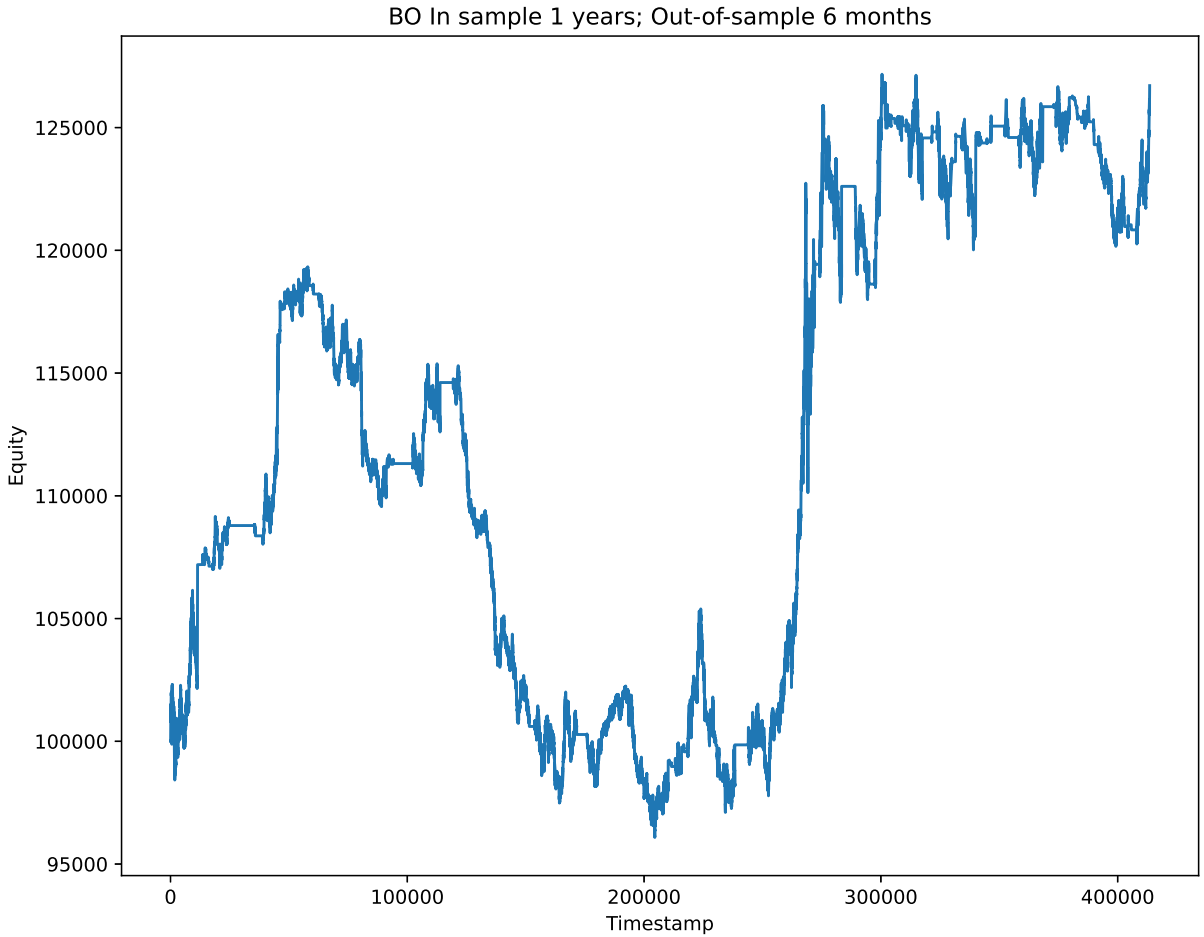
Equity curve $T = 1, \tau = 3$



Statistical performance soybean oil $T = 1, \tau = 3$

Soybean Oil (T=1, τ=3)	
Total Number of Trades	324
Average Return	0.41
Net Profit	134.09
Average Rate of Return	0.7247%
Standard Deviation of Returns	0.0541
Total Number of Winners	137
Percentage of Winners	42.28%
Total Value of Winners	272.40
Average Winner	1.99
Total Value of Losers	(138.32)
Average Loser	(0.74)
Worst Drawdown	(6.58)
Sharpe Ratio	0.1339
Profit Factor	0.9694
Net Profit to Worst Drawdown (ROA)	20.3777

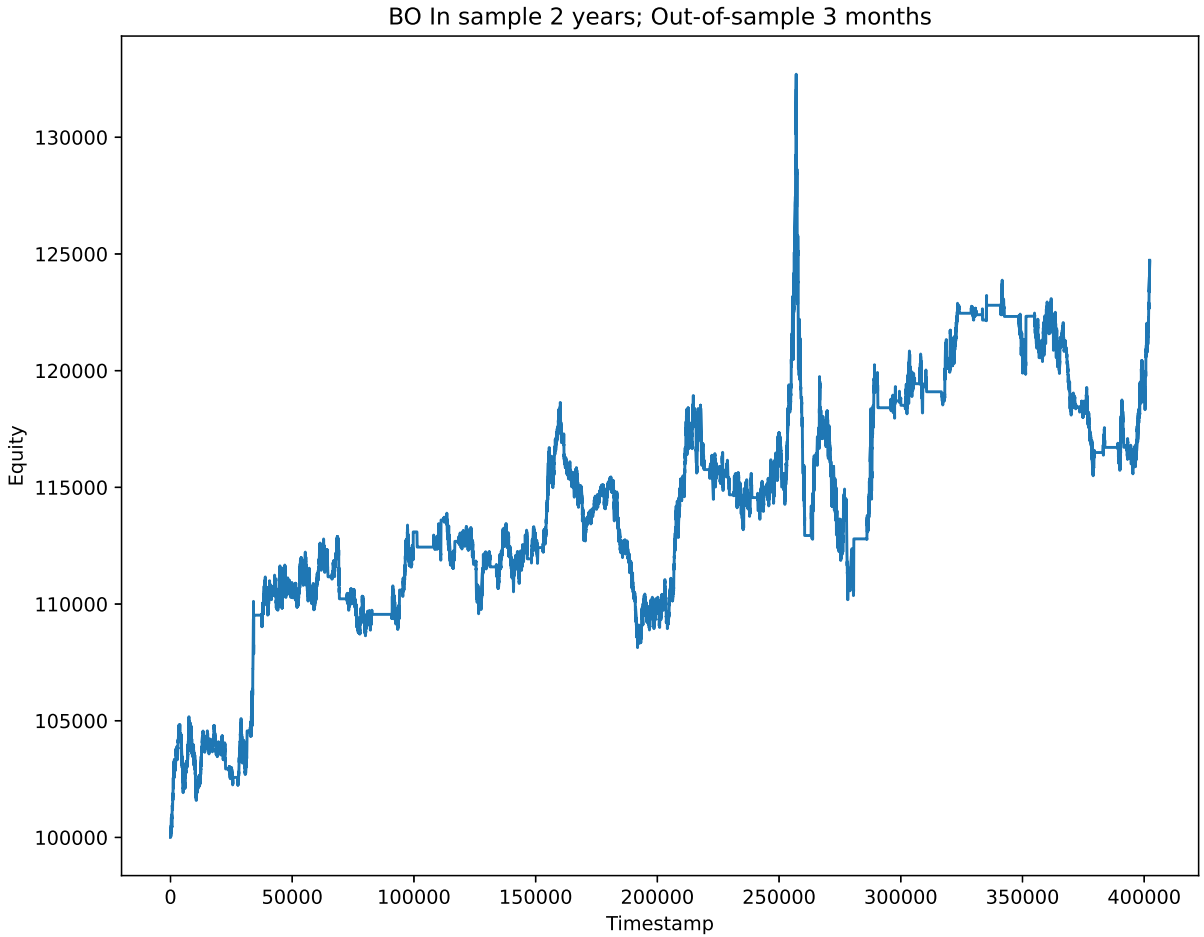
Equity curve $T = 1, \tau = 6$



Statistical performance soybean oil $T = 1, \tau = 6$

Soybean Oil (T=1, τ=6)	
Total Number of Trades	336
Average Return	0.19
Net Profit	64.84
Average Rate of Return	0.3149%
Standard Deviation of Returns	0.0415
Total Number of Winners	132
Percentage of Winners	39.29%
Total Value of Winners	202.10
Average Winner	1.53
Total Value of Losers	(137.26)
Average Loser	(0.67)
Worst Drawdown	(5.97)
Sharpe Ratio	0.0759
Profit Factor	0.4724
Net Profit to Worst Drawdown (ROA)	10.8610

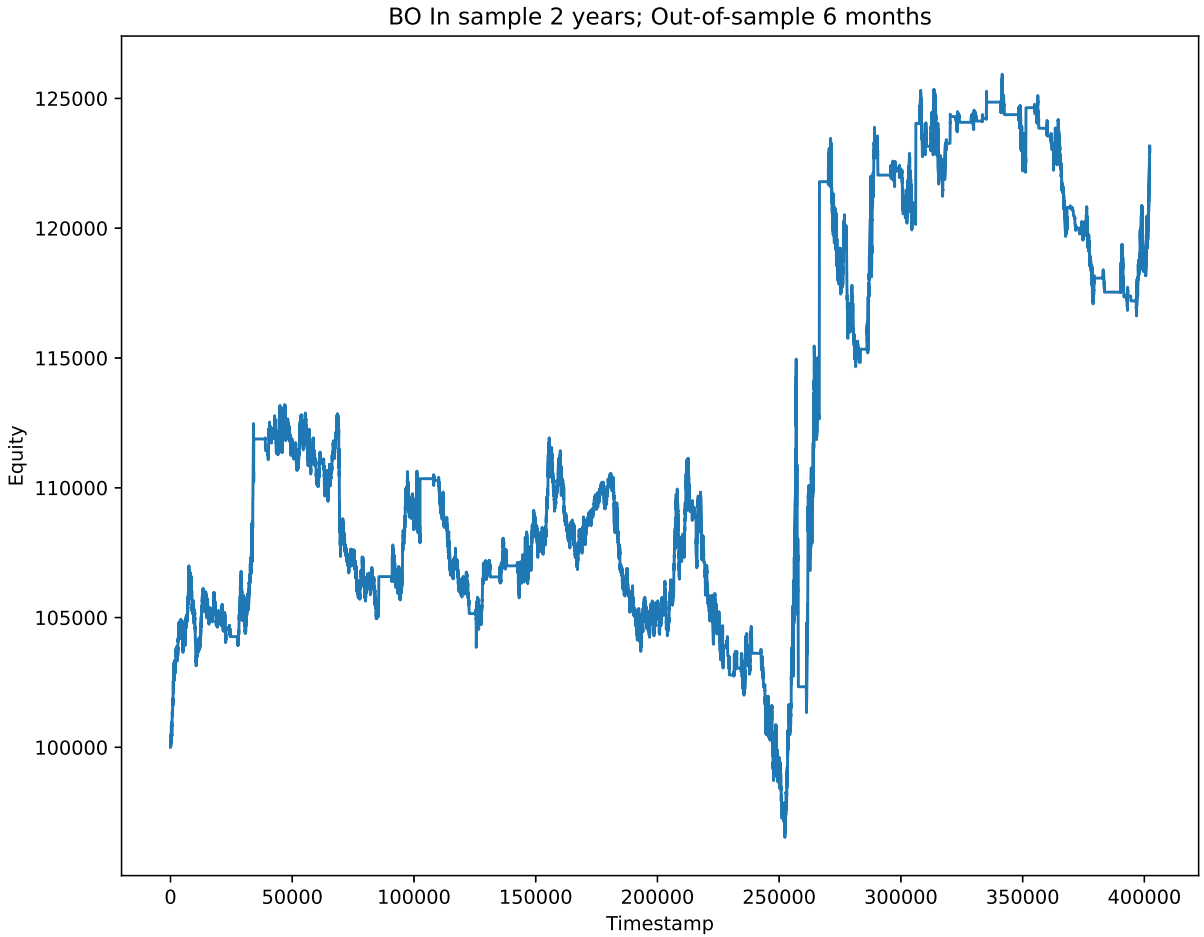
Equity curve $T = 2, \tau = 3$



Statistical performance soybean oil $T = 2, \tau = 3$

Soybean Oil (T=2, τ=3)	
Total Number of Trades	286
Average Return	0.22
Net Profit	63.12
Average Rate of Return	0.3879%
Standard Deviation of Returns	0.0315
Total Number of Winners	131
Percentage of Winners	45.80%
Total Value of Winners	177.83
Average Winner	1.36
Total Value of Losers	(114.71)
Average Loser	(0.74)
Worst Drawdown	(3.39)
Sharpe Ratio	0.1230
Profit Factor	0.5503
Net Profit to Worst Drawdown (ROA)	18.6381

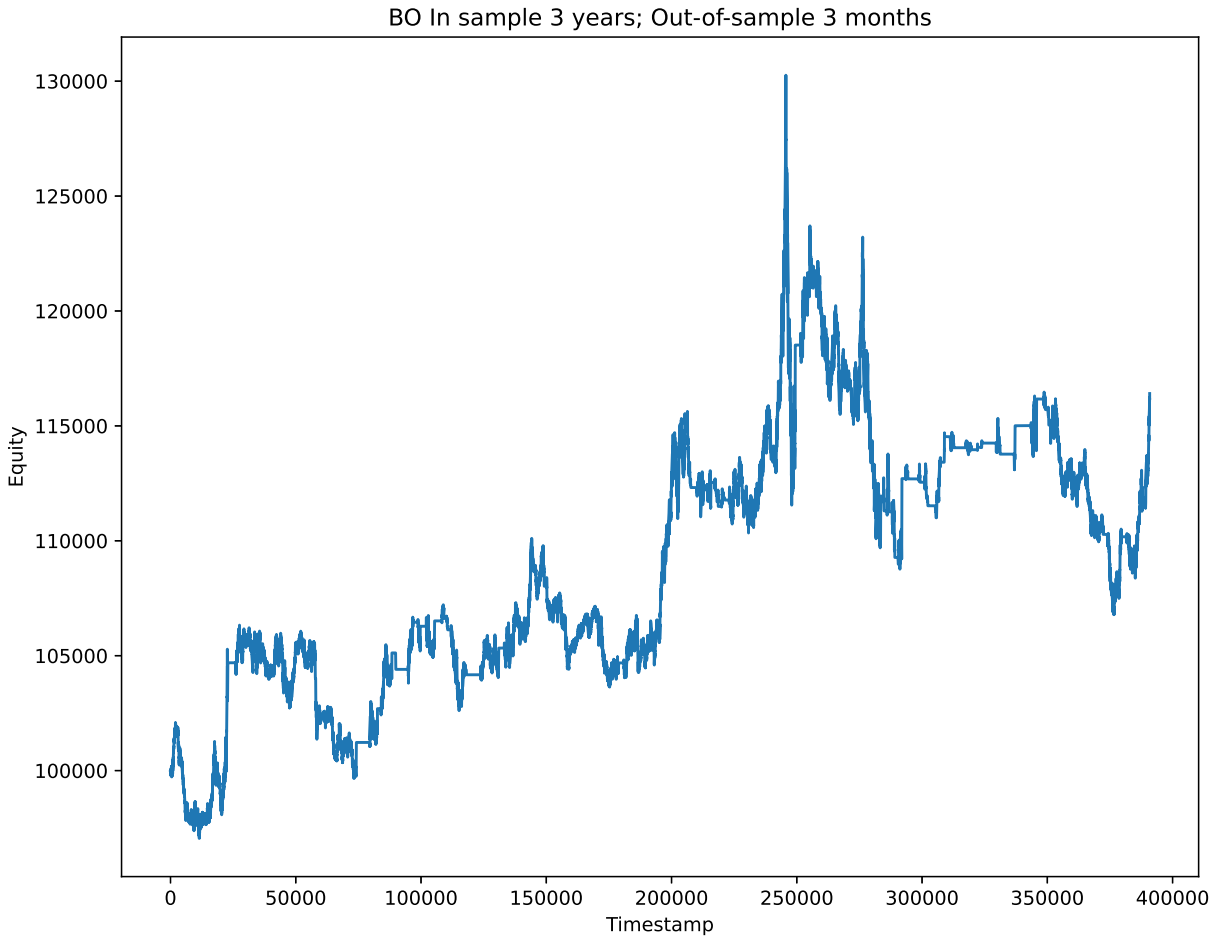
Equity curve $T = 2, \tau = 6$



Statistical performance soybean oil $T = 2, \tau = 6$

Soybean Oil (T=2, τ=6)	
Total Number of Trades	292
Average Return	0.19
Net Profit	56.52
Average Rate of Return	0.2896%
Standard Deviation of Returns	0.0352
Total Number of Winners	120
Percentage of Winners	41.10%
Total Value of Winners	184.25
Average Winner	1.54
Total Value of Losers	(127.74)
Average Loser	(0.74)
Worst Drawdown	(4.65)
Sharpe Ratio	0.0822
Profit Factor	0.4424
Net Profit to Worst Drawdown (ROA)	12.1540

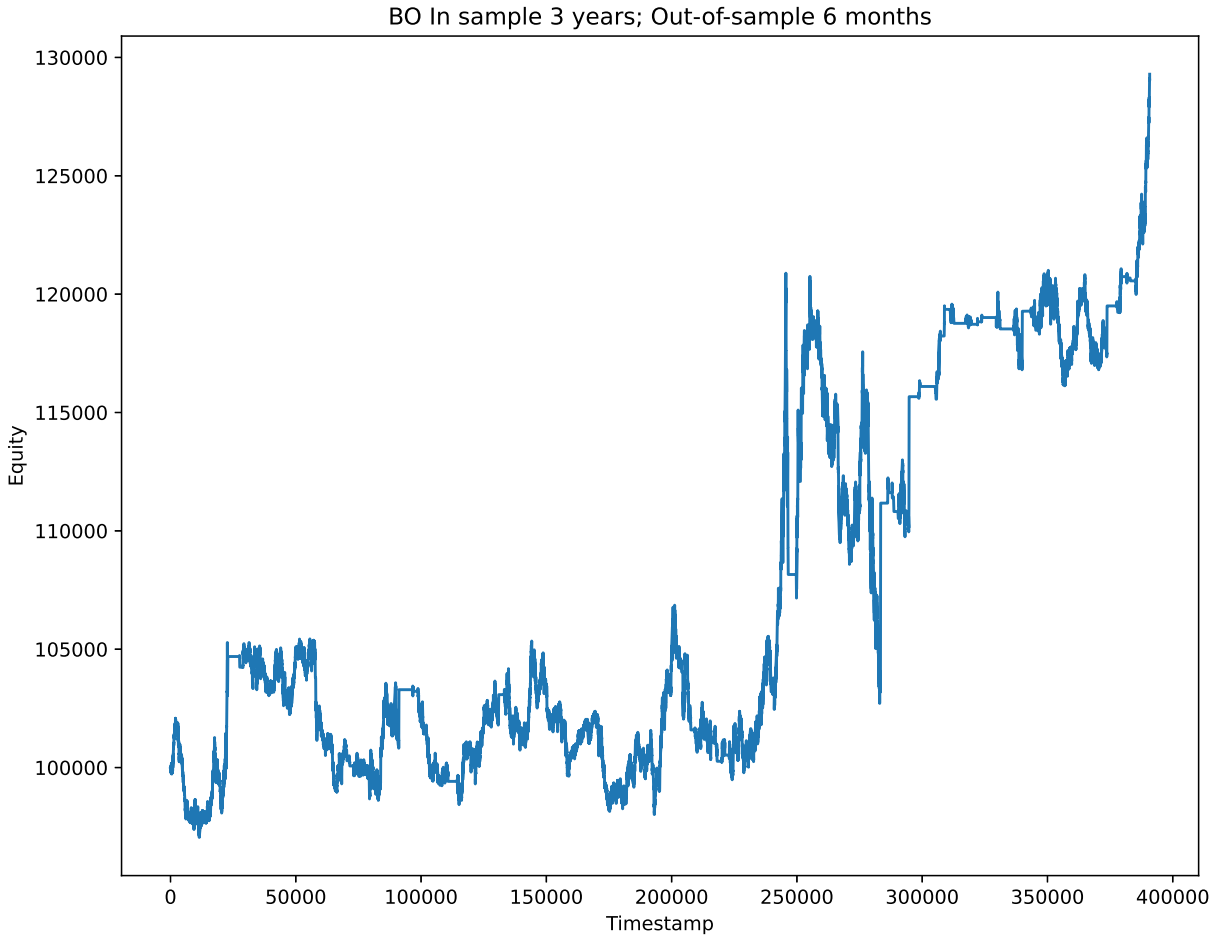
Equity curve $T = 3, \tau = 3$



Statistical performance soybean oil $T = 3, \tau = 3$

Soybean Oil (T=3, τ=3)	
Total Number of Trades	280
Average Return	0.17
Net Profit	46.66
Average Rate of Return	0.3046%
Standard Deviation of Returns	0.0346
Total Number of Winners	124
Percentage of Winners	44.29%
Total Value of Winners	176.45
Average Winner	1.42
Total Value of Losers	(129.79)
Average Loser	(0.83)
Worst Drawdown	(4.97)
Sharpe Ratio	0.0881
Profit Factor	0.3595
Net Profit to Worst Drawdown (ROA)	9.3845

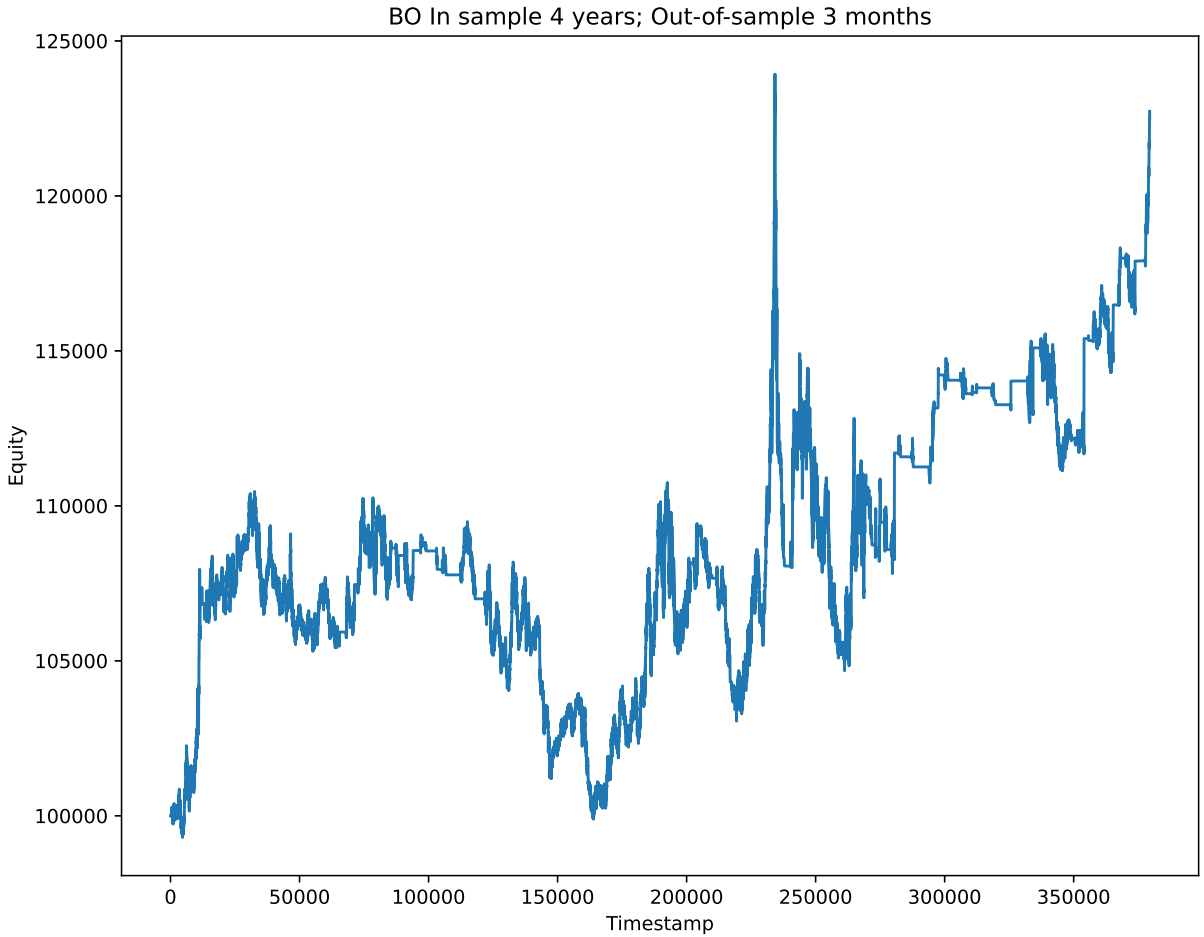
Equity curve $T = 3, \tau = 6$



Statistical performance soybean oil $T = 3, \tau = 6$

Soybean Oil (T=3, τ=6)	
Total Number of Trades	255
Average Return	0.23
Net Profit	59.01
Average Rate of Return	0.4234%
Standard Deviation of Returns	0.0363
Total Number of Winners	110
Percentage of Winners	43.14%
Total Value of Winners	183.46
Average Winner	1.67
Total Value of Losers	(124.45)
Average Loser	(0.86)
Worst Drawdown	(5.79)
Sharpe Ratio	0.1167
Profit Factor	0.4742
Net Profit to Worst Drawdown (ROA)	10.1919

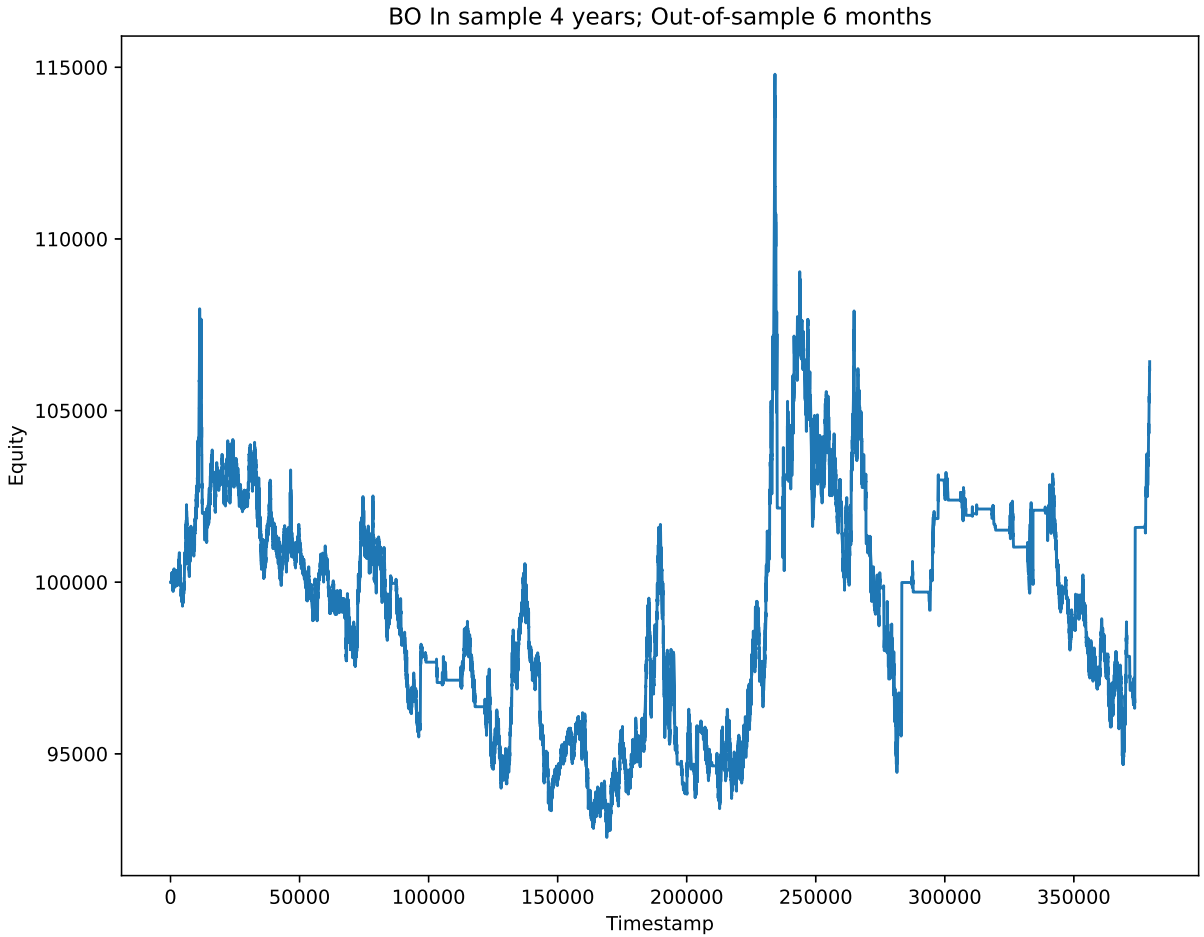
Equity curve $T = 4, \tau = 3$



Statistical performance soybean oil $T = 4, \tau = 3$

Soybean Oil (T=3, τ=6)	
Total Number of Trades	290
Average Return	0.20
Net Profit	58.26
Average Rate of Return	0.4114%
Standard Deviation of Returns	0.0304
Total Number of Winners	124
Percentage of Winners	42.76%
Total Value of Winners	166.55
Average Winner	1.34
Total Value of Losers	(108.29)
Average Loser	(0.65)
Worst Drawdown	(3.51)
Sharpe Ratio	0.1355
Profit Factor	0.5380
Net Profit to Worst Drawdown (ROA)	16.5989

Equity curve $T = 4, \tau = 6$



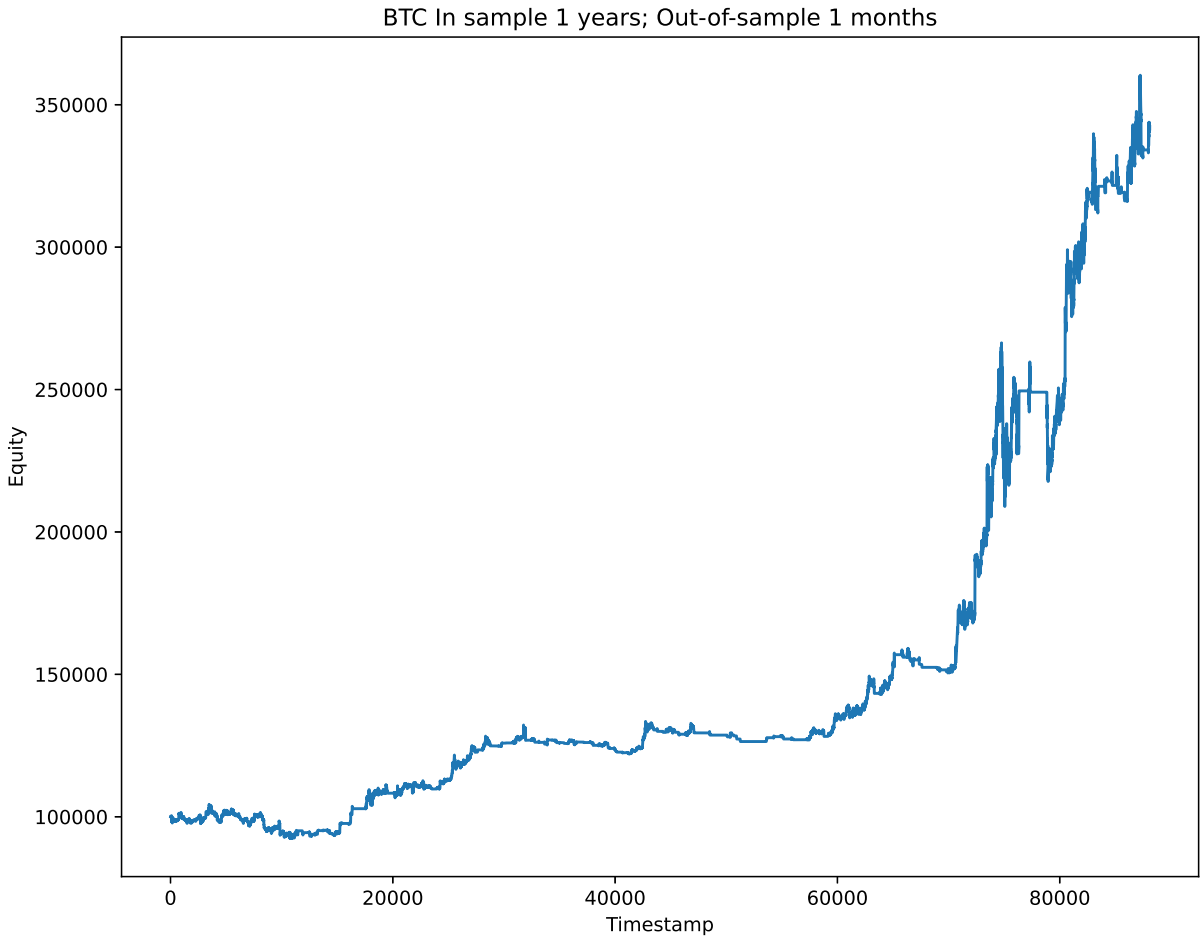
Statistical performance soybean oil $T = 4, \tau = 6$

Soybean Oil (T=3, τ=6)	
Total Number of Trades	314
Average Return	0.11
Net Profit	34.70
Average Rate of Return	0.2171%
Standard Deviation of Returns	0.0331
Total Number of Winners	127
Percentage of Winners	40.45%
Total Value of Winners	173.67
Average Winner	1.37
Total Value of Losers	(138.96)
Average Loser	(0.74)
Worst Drawdown	(3.56)
Sharpe Ratio	0.0655
Profit Factor	0.2497
Net Profit to Worst Drawdown (ROA)	9.7485

High level results - Bitcoins

Bitcoin (BTC) ROA Results			
	τ=1	τ=2	τ=3
T=1	62.7747	43.3588	11.7709
T=2	42.1601	30.2966	75.4938

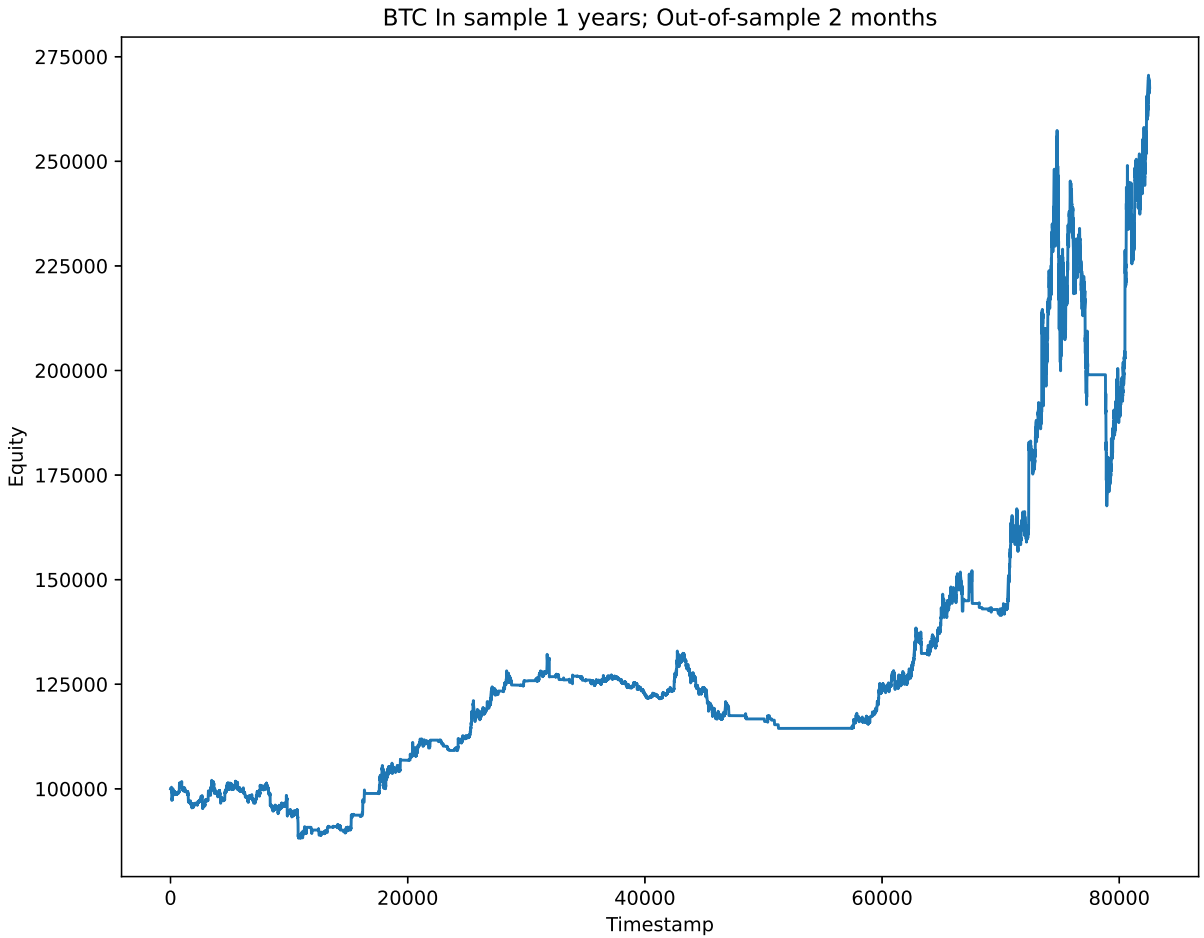
Equity curve $T = 1, \tau = 1$



Statistical performance Bitcoin $T = 1, \tau = 1$

Bitcoin (T=1, τ=1)	
Total Number of Trades	217
Average Return	225.84
Net Profit	49,007.70
Average Rate of Return	1.1708%
Standard Deviation of Returns	0.0769
Total Number of Winners	95
Percentage of Winners	43.78%
Total Value of Winners	64494.68
Average Winner	678.89
Total Value of Losers	(15,486.98)
Average Loser	(126.94)
Worst Drawdown	(780.69)
Sharpe Ratio	0.1523
Profit Factor	3.1644
Net Profit to Worst Drawdown (ROA)	62.7747

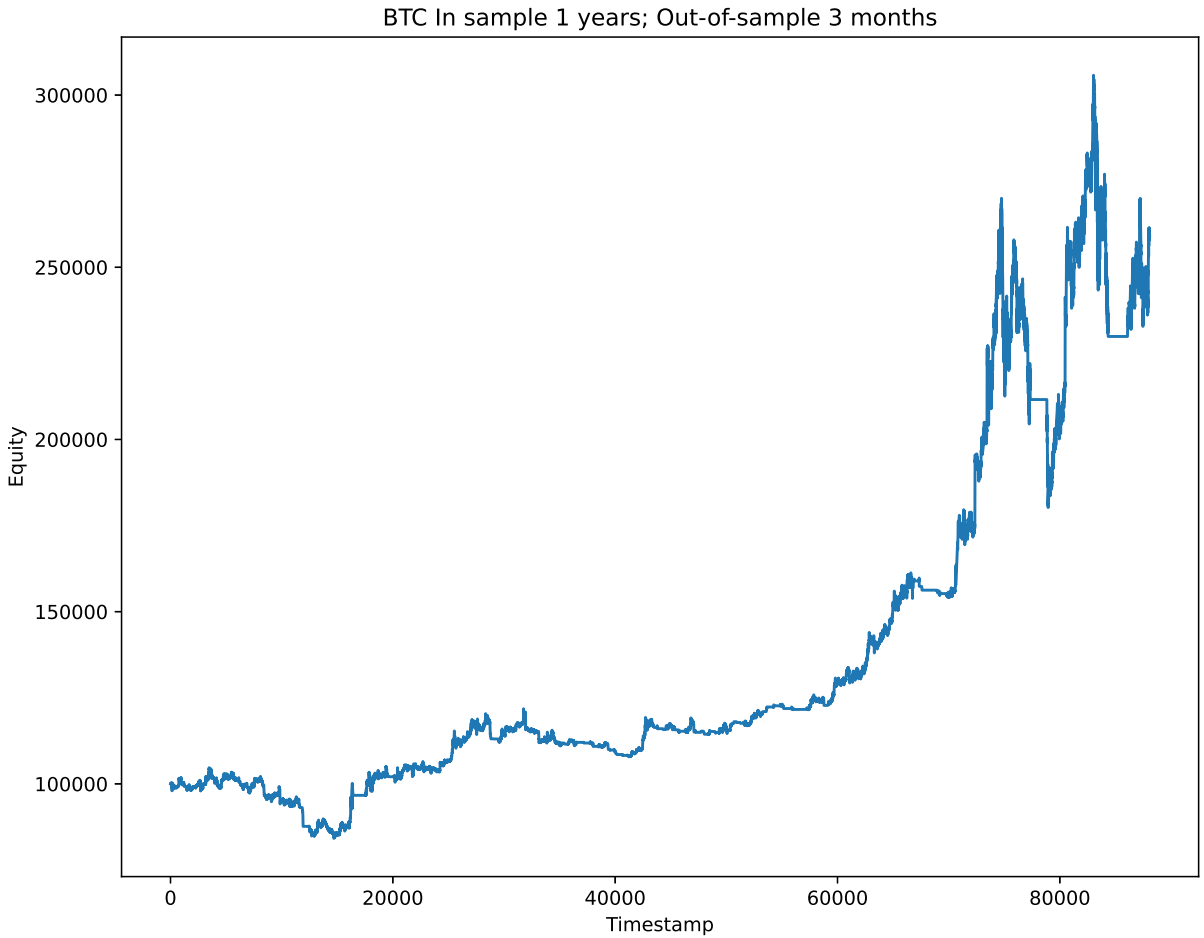
Equity curve $T = 1, \tau = 2$



Statistical performance Bitcoin $T = 1, \tau = 2$

Bitcoin (T=1, τ=2)	
Total Number of Trades	173
Average Return	134.09
Net Profit	23,196.98
Average Rate of Return	0.9537%
Standard Deviation of Returns	0.0587
Total Number of Winners	72
Percentage of Winners	41.62%
Total Value of Winners	36334.46
Average Winner	504.6452773
Total Value of Losers	(13,137.48)
Average Loser	(130.07)
Worst Drawdown	(535.00)
Sharpe Ratio	0.1625
Profit Factor	1.7657
Net Profit to Worst Drawdown (ROA)	43.3588

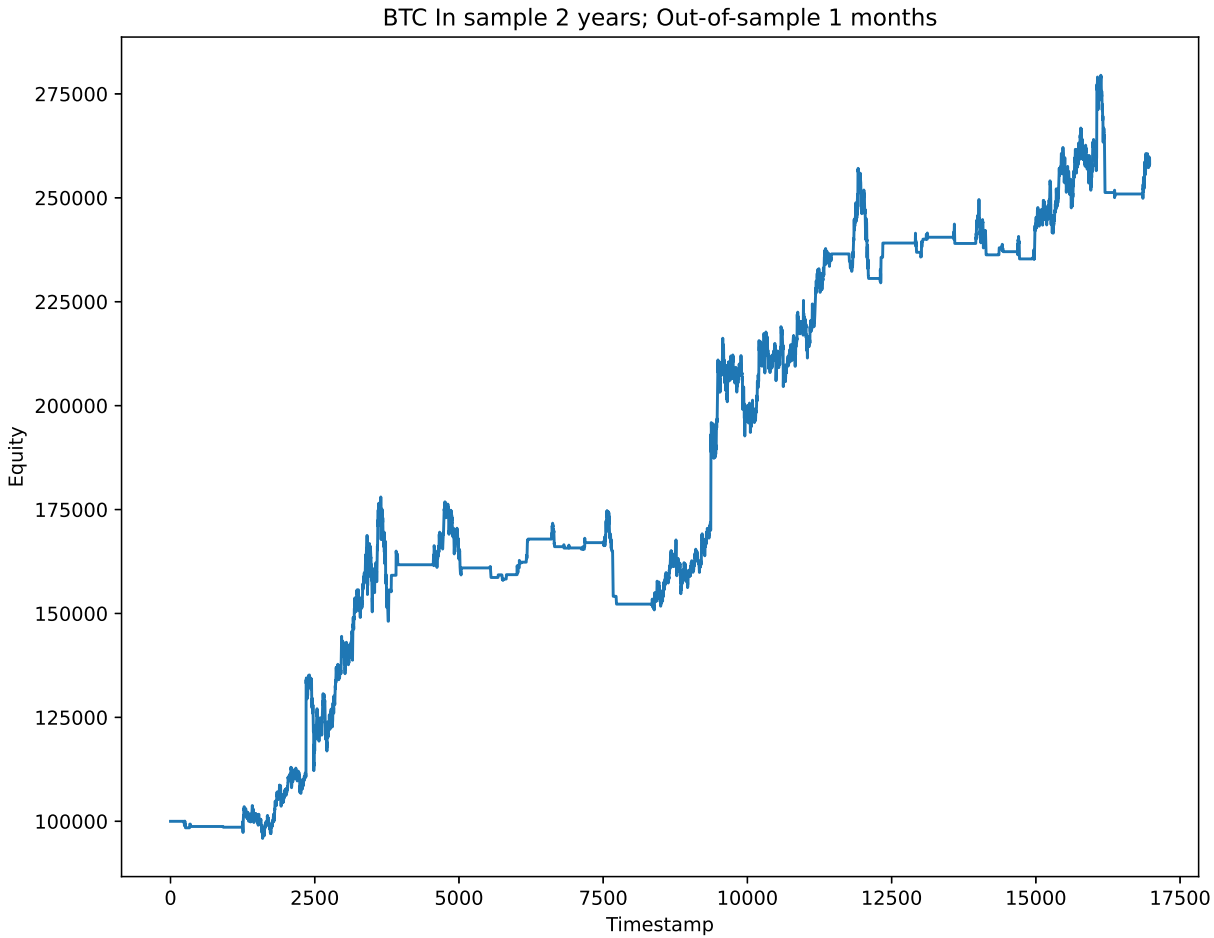
Equity curve $T = 1, \tau = 3$



Statistical performance Bitcoin $T = 1, \tau = 3$

Bitcoin 1.25 Year	
Total Number of Trades	139
Average Return	199.38
Net Profit	27,714.06
Average Rate of Return	1.2582%
Standard Deviation of Returns	0.0730
Total Number of Winners	54
Percentage of Winners	38.85%
Total Value of Winners	42111.89
Average Winner	779.8498187
Total Value of Losers	(14,397.83)
Average Loser	(169.39)
Worst Drawdown	(2,354.46)
Sharpe Ratio	0.1723
Profit Factor	1.9249
Net Profit to Worst Drawdown (ROA)	11.7709

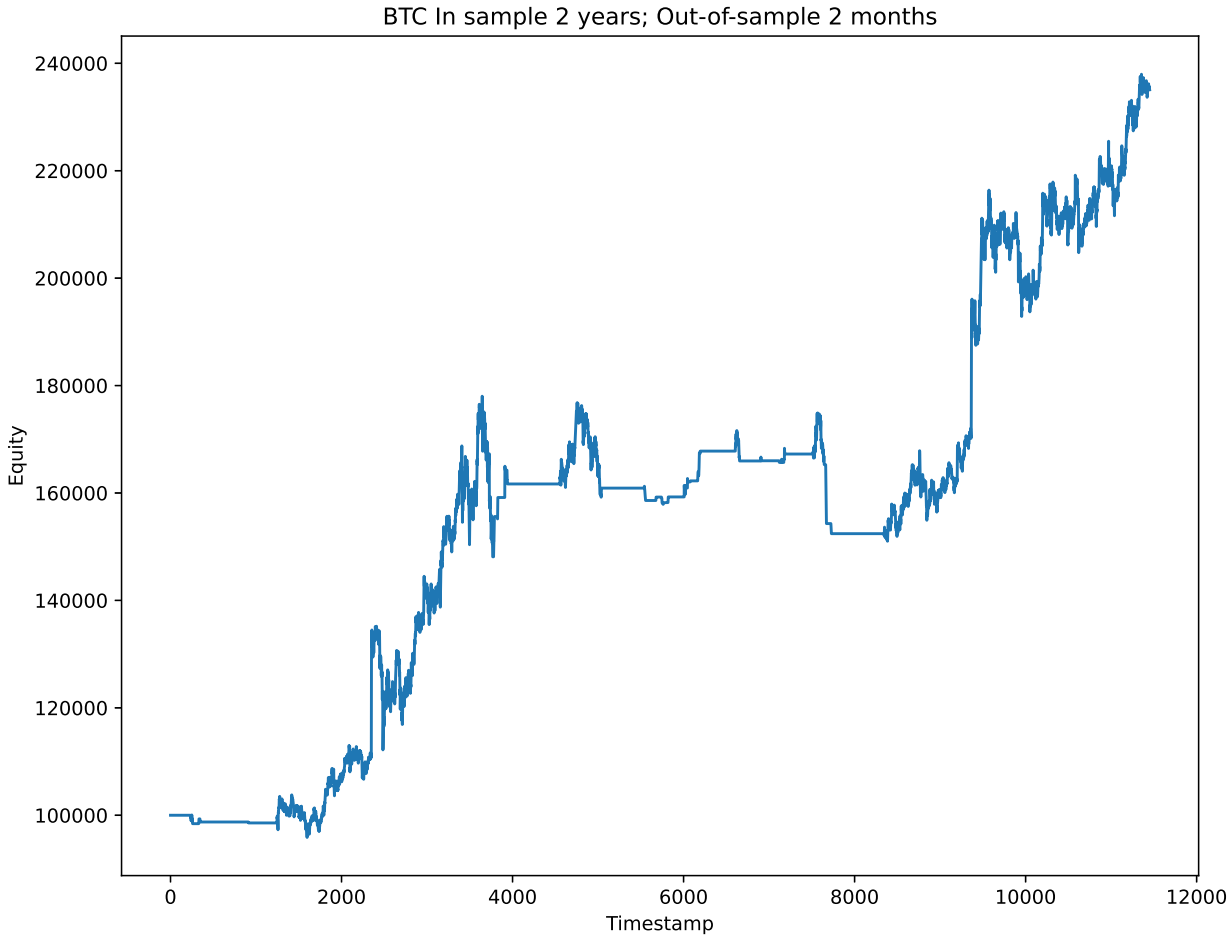
Equity curve $T = 2, \tau = 1$



Statistical performance Bitcoin $T = 2, \tau = 1$

Bitcoin (T=2, τ=1)	
Total Number of Trades	86
Average Return	375.33
Net Profit	32,278.10
Average Rate of Return	1.1699%
Standard Deviation of Returns	0.0716
Total Number of Winners	38
Percentage of Winners	44.19%
Total Value of Winners	42477.93
Average Winner	1117.840174
Total Value of Losers	(10,199.83)
Average Loser	(212.50)
Worst Drawdown	(765.61)
Sharpe Ratio	0.1635
Profit Factor	3.1646
Net Profit to Worst Drawdown (ROA)	42.1601

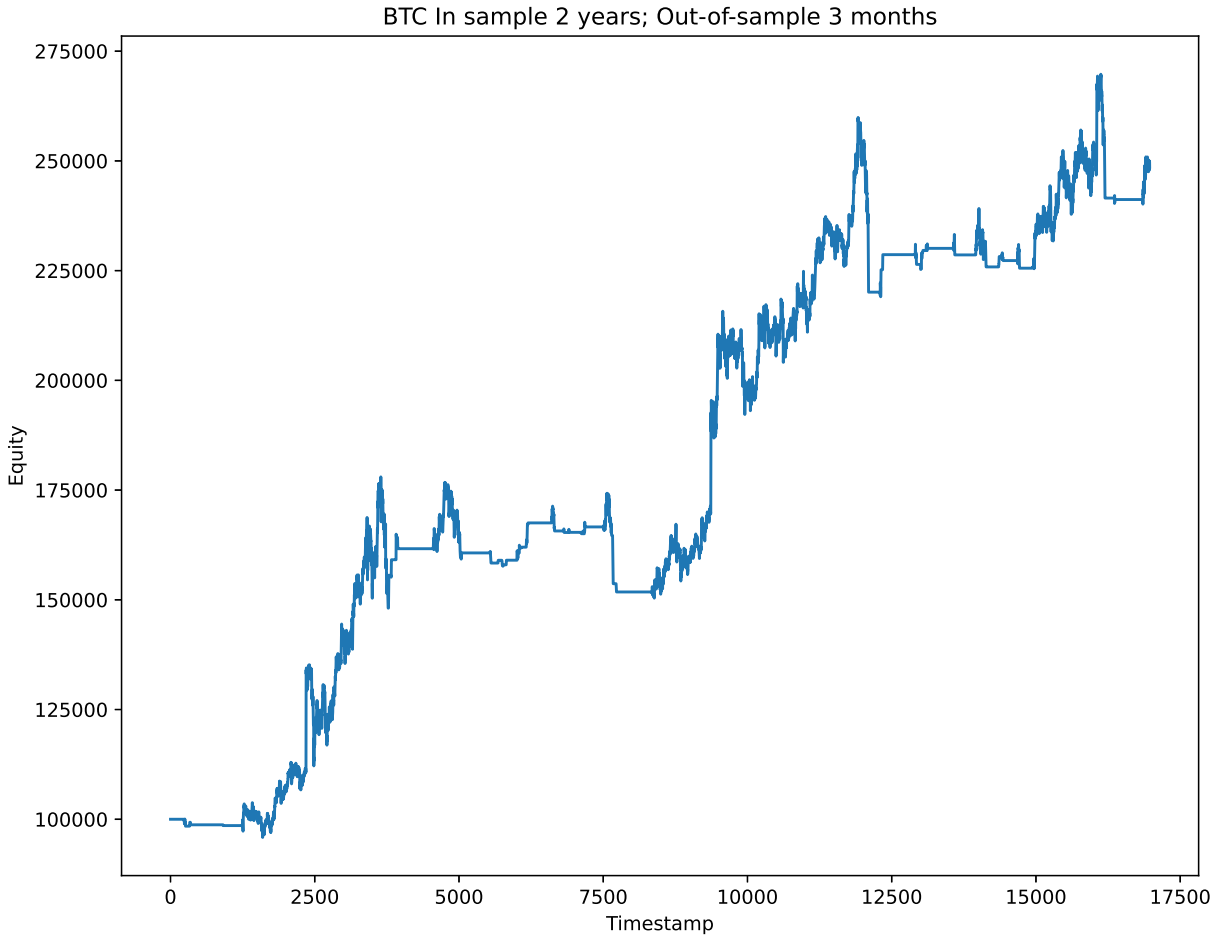
Equity curve $T = 2, \tau = 2$



Statistical performance Bitcoin $T = 2, \tau = 2$

Bitcoin (T=2, τ=2)	
Total Number of Trades	51
Average Return	251.15
Net Profit	12,808.68
Average Rate of Return	0.9667%
Standard Deviation of Returns	0.0667
Total Number of Winners	20
Percentage of Winners	39.22%
Total Value of Winners	18271.41
Average Winner	913.570346
Total Value of Losers	(5,462.72)
Average Loser	(176.22)
Worst Drawdown	(422.78)
Sharpe Ratio	0.1450
Profit Factor	2.3447
Net Profit to Worst Drawdown (ROA)	30.2966

Equity curve $T = 2, \tau = 3$



Statistical performance Bitcoin $T = 2, \tau = 3$

Bitcoin (T=2, τ=3)	
Total Number of Trades	86
Average Return	365.08
Net Profit	31,397.03
Average Rate of Return	1.1337%
Standard Deviation of Returns	0.0671
Total Number of Winners	38
Percentage of Winners	44.19%
Total Value of Winners	40,706.32
Average Winner	1,071.22
Total Value of Losers	(9,309.29)
Average Loser	(193.94)
Worst Drawdown	(415.89)
Sharpe Ratio	0.1689
Profit Factor	3.3727
Net Profit to Worst Drawdown (ROA)	75.4938

Key technologies used

- Git
- R
- Python 3
  - Optuna
  - SKlearn
  - Pandas DataFrame
  - Multi-processing
- C++
  - Through pybind 11
- LaTeX

Thank you for listening!