FOREWORD

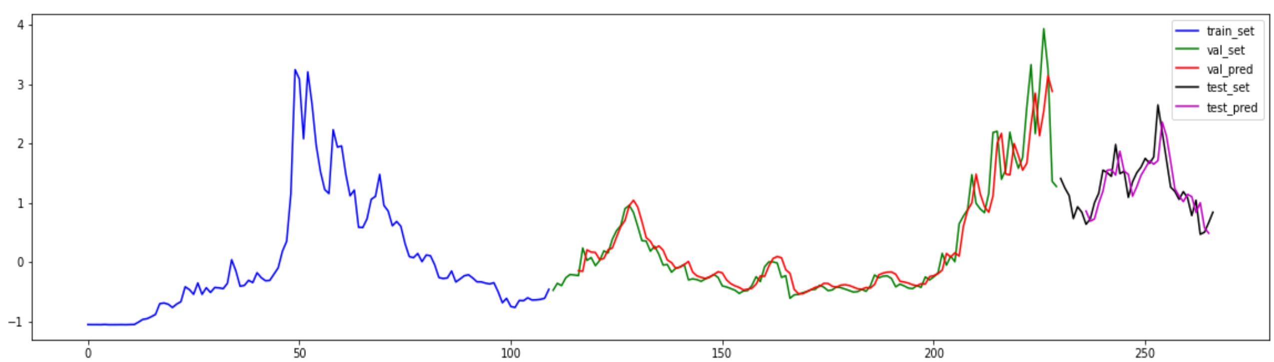
- I am an individual investor but in this article, I will be a trader.
- Please be very careful in case you want to use these findings to make a decision on your investment because I have not tested them in practice.
- Trading fee is not mentioned in this topic.

About Litecoin

As a trader, I'm interested in profit in the short run, specifically weekly and monthly. I'll be happy and lucky if most of the time, I can correctly predict fluctuations in price. I chose Litecoin because there have been many ups and downs which provides enough data to analyse and at least, LTC is not a scam.

1/ Weekly Analysis

I use prices from the previous 6 weeks to predict the price of next week. Here's the result:



In the above graph, we used mean squared error as the metric and it looks great, doesn't it? It could be problematic in real-life decisions because as a trader, I also want to know how it will be next week, an up, or a down.

Let me explain further why mean squared error is not sufficient: let say today's price is 100 and tomorrow's price is 110. What are some possible scenarios? The table below are 2 examples, you can see in scenario 1, the error is just 5% which is much lower than the error in scenario 2. Also, the model predicted it would be a down the next day so you decided to sell it out and you missed the chance of increasing your investment by 10%.

It's the other way round in scenario 2 where the error is higher but its prediction "up or down" was correct.

	Today's price	Predicted price	Error	Tomorrow price (actual)	Decision	Decision's Result
Scenario 1	100	95	5%	110	Sell	Wrong
Scenario 2	100	115	15%	110	Hold	Correct

This means we need to check for the model's accuracy:

Train_accuracy: 0.50 Val_accuracy: 0.56 Test_accuracy: 0.60

It's a bit disappointing :(We just have more or less than 50% of predicting precisely "up or down". More importantly, I only focus on strong fluctuations, let say 20%.

	last_value	predicted_value	actual_value	predicted_change_rate	actual_change_rate	prediction_result
7	175.454086	175.663253	212.358459	0.001192	0.210336	1.0
17	197.866364	192.658190	258.093201	-0.026322	0.304381	0.0
27	147.937408	148.717105	108.428497	0.005270	-0.267065	0.0

During the last 37 weeks, there were 3 times that the price changed by more than 20% but the model could not predict this. In this case, with the expected change of 20%, the accuracy is 1/3.

Lower the expected change rate to 3%

	last_value	predicted_value	actual_value	predicted_change_rate	actual_change_rate	prediction_result
0	133.987289	139.532493	120.027077	0.041386	-0.104191	0.0
1	120.027077	123.892875	126.188675	0.032208	0.051335	1.0
3	144.933212	149.840483	155.864777	0.033859	0.075425	1.0
4	155.864777	164.789921	182.711273	0.057262	0.172242	1.0
18	258.093201	240.992902	227.079605	-0.066256	-0.120164	1.0
19	227.079605	217.999194	195.174423	-0.039988	-0.140502	1.0
28	108.428497	112.856361	110.823677	0.040837	0.022090	1.0

Now we got 6/7 correct predictions but the predicted change rates are not very close to the actual change rates.

2/ Forecast several values ahead.

It's worse in this case. If the first next prediction is an up, the model tends to output all ups for the rest and vice versa.

In conclusion, this is not a very reliable model to predict LTC's price. If we use this model to predict the price and make decisions, we might still make profit from it. Another noticeable point is that the model gives poor predictions on strong fluctuations (it cannot precisely tell you when the price changes by more than 10%).

About VN30

VN30 is a stock index of the 30 largest market cap companies in VietNam (just like S%P500 in the US). The data was collected from 2009 to 2022 (around 10 years, 3284 days).

Research questions: How many days do we have to wait until there is an increase in the VN30 index?

(For example, today's index decreased by 1% compared to yesterday, how is it likely to increase in 1 day, 2 days or 3 days or more? It's just like Geometric distribution where a success is described as an increase and a failure is described as a decrease in this example).

1/ Getting started

Let's have a look at this table

change_rate	up_or_down	waited_days
0.19	1.0	0.0
1.59	1.0	0.0
-1.42	0.0	3.0
0.19	1.0	1.0
-0.61	0.0	1.0
0.29	1.0	1.0
-1.09	0.0	1.0
0.42	1.0	1.0
-0.07	0.0	1.0
0.13	1.0	1.0

Change_rate is the rate of change in the index

Up_or_down: 1 is an up in the index and 0 is a down in the index

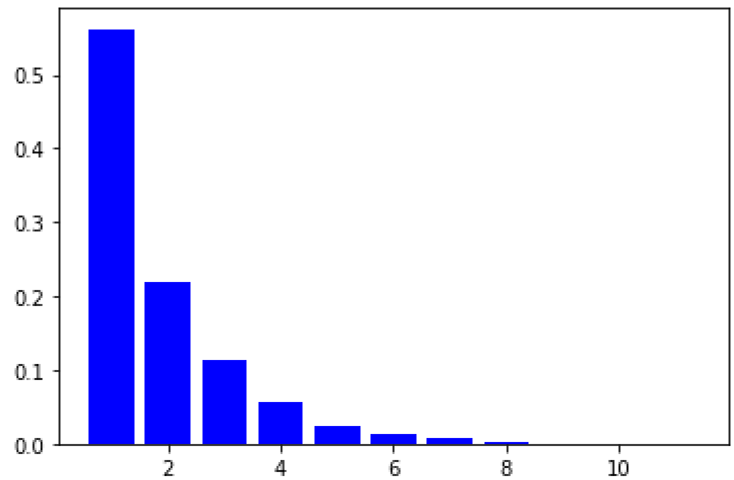
Looking at the first 3 rows, the index had increased for 2 days before it saw a drop (-1.42%) on the third day. This means we had waited for 3 days for a change in the index.

Similarly, the index changed every day for the following rows.

0.6	1.0	0.0
0.59	1.0	0.0
1.03	1.0	0.0
-0.57	0.0	4.0

Sometimes we have to wait for 4 days for a change in the index.

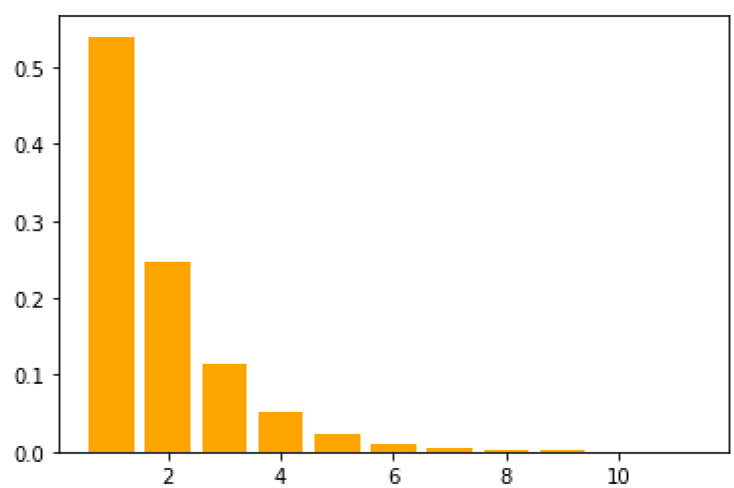
What are the unique waiting days and their frequency?
unique waiting days: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]
their frequency: [734, 380, 220, 100, 60, 30, 17, 5, 4, 2, 1, 2]
And these are their proportions



Now you may notice it looks very similar to the Geometric distribution. We will shortly check that.

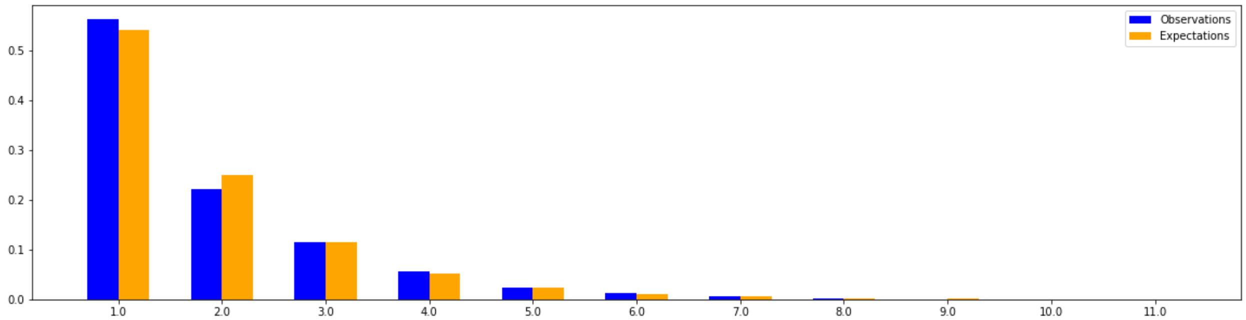
2/ Checking and hypothesis testing

If it does follow a Geom distribution, then the probability of an increase occurring = total number of increases / numbers of days = 54.02



Geometric distribution (p = 54.02%).

Here I plot them side by side



Since we want, we expect it to follow a Geom distribution, I would call the actual data is "observations" and our expected numbers are "expectations"

The next step is to do a hypothesis test.

If it does follow a Geom distribution, the observed waiting days and the expected waiting days are:

	1	2	3	4	5	6	7	8	9	10	11
observations	1004	393	204	100	41	24	12	5	1	1	1
expectations	965	444	204	94	43	20	9	4	2	1	0

I binned the last 4 values to make it at least 5, in order for the chi square test condition to be met)

	1	2	3	4	5	6	7	8
observations	1004	393	204	100	41	24	12	8
expectations	965	444	204	94	43	20	9	7

Now we can start calculating chi-square value and the p value

H0: The indices are independent of each other.
HA: The indices are independent of each other, particularly, they follow geometric distribution.

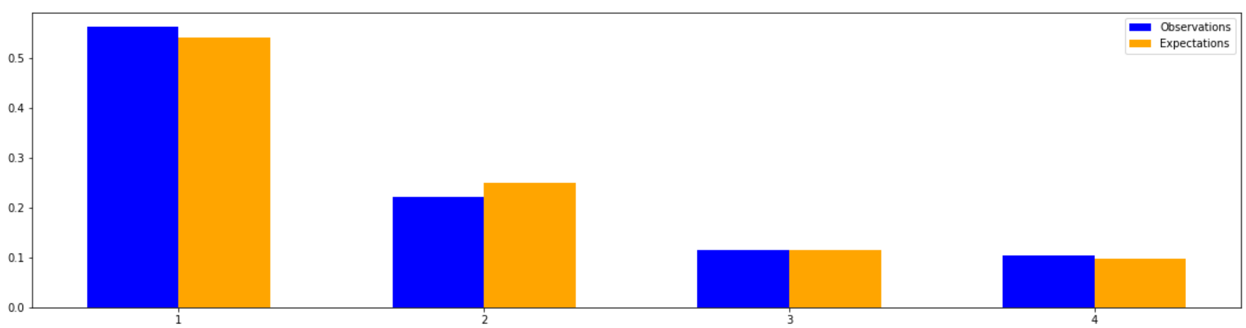
With the above data, we will find that the p-value is 0.197 which evidently supports us to accept the null hypothesis.

3/ What if I narrow down 8 groups to 4?

Mathematically, narrowing down the number of groups means narrowing down the degrees of freedom which might result in a smaller p-value (since chi-square value is almost unchanged). Let's try it out

	1	2	3	4
observations	1004	393	204	185
expectations	965	444	204	173

The frequencies of observed positive days and the frequencies of expected positive days



The proportions of observed positive days and the proportions of expected positive days

Doing the hypothesis test again but with this data, our new p-value is 0.04. In other words, we have sufficient evidence to say that the price being up or down is independent of the previous day, but follows the Geometric distribution.

* Further question: this makes sense mathematically but what does it mean financially?

AFTERWORD

- Saying whether tomorrow's price goes up or down is almost an impossible task.
- The waiting days (1, 2, 3 or more than 4) until there is a positive rate of change (in VN30 index) follows the Geometric distribution.