Exploring Volume as a Predictor for Close Price

Author: Isaac Lee | **Date: 09/03/2022** | Source Code

Theory

> ### Conjecture
> Over some time interval (to be determined), a change in volume (over a certain threshold v) $\implies$ (with high probability) that the close price will either increase or decrease (to be determined) over the next time interval.

To be more precise, let:

$$\delta_V(t) := \text{Percentage change in volume from time } t - \Delta t \text{ to } t \qquad (\text{for some } \Delta t).$$
$$\delta_P(t) := \text{Percentage change in (close) price from time } t - \Delta t \text{ to } t \qquad (\text{for some } \Delta t).$$

We conjecture that:

$$\exists \Delta t, v \in \mathbb{R}_0^+ : P(\delta_P(t + \Delta t) > 0 \mid \delta_V(t) > v) \not\approx 0.5$$

Where v and Δt are constants to be chosen/determined.

Testing the conjecture

Choosing Δt

We will start by choosing Δt to be the length of 6, 4hr candles, i.e $\Delta t = 24hrs$.
We do this because it represents a nice time frame (a day) but also because the number of entries in our dataframe is divisible by 6 so it's a nice number to work with.

Head of the DataFrame

	$\delta_P(t)$	$\delta_V(t)$	t
0	0.040416	6.481959	2017-01-01 23:30:00
1	0.005923	1.284866	2017-01-02 23:30:00
2	0.008942	1.291023	2017-01-03 23:30:00

```
delta_df.describe()
```

	$\delta_P(t)$	$\delta_V(t)$
count	1821.000000	1821.000000
mean	0.003086	0.508092
std	0.038755	1.854694
min	-0.244779	-1.000000
25%	-0.013550	-0.360602
50%	0.002179	0.014613
75%	0.020495	0.721199
max	0.304327	36.795678

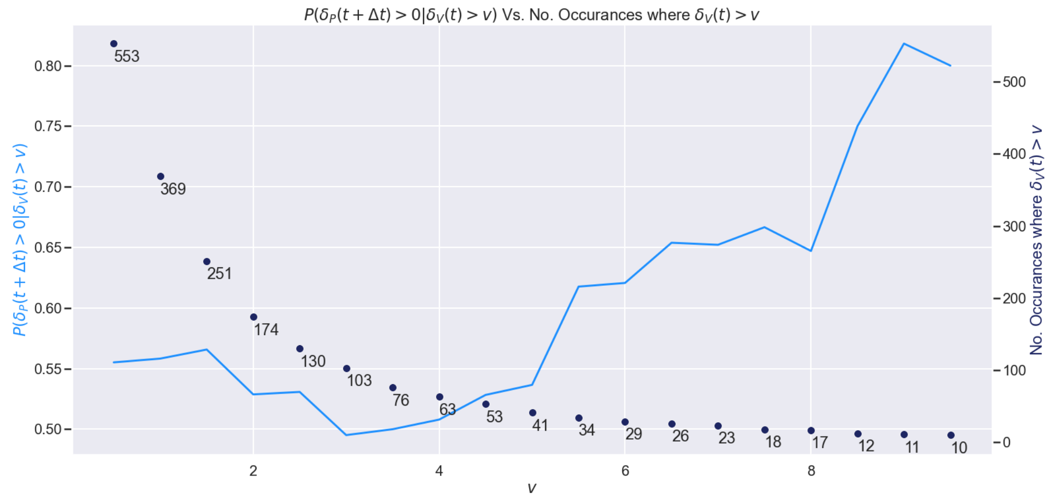
So it appears that on average, the volume increases 50%, which makes sense since more and more bitcoin has been traded over the last 5 years. Also there is a positive average price change, which also makes sense because bitcoin has famously increased in price over the past 5 years.

Next we 'lag' or shift our price delta column, so that we can compare $\delta_V(t)$ against $\delta_P(t+\Delta t)$.

```
delta_df['lag_del_p'] = delta_df['del_p'].shift(1)
```

(Note: we also remove the first row since because lagging one colume means a NaN value).

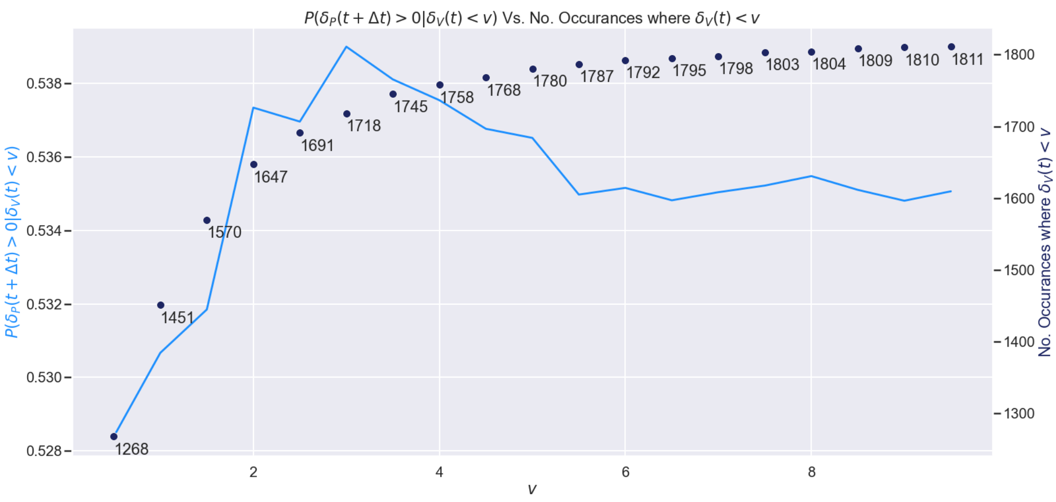
Determining v



This is a plot of the probability that the close price of the candle 6 (4hr) candles from now will increase, given the (percentage) increase in volume is greater than v, against v.
We also plot the sample size or number of occurances where the percentage increase in V is greater than v.

Interpretations

From this plot we see that the probability that the price will increase following an increase in volume over the threshold v, increases as we increase v and is significantly greater than 50% for large enough v, which provides evidence to support our conjecture. However we also see that as we increase v that sample size decreases and this means our probabilities become (perhaps) less representative of the true value, so it's harder to make definite conclusions.

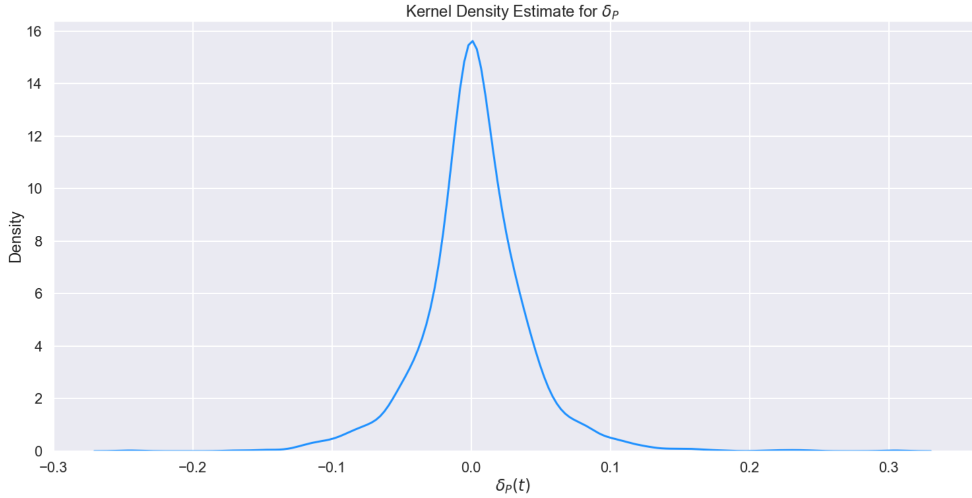


Interpretations

So we see that the probability that the price will increase, following a *non-significant* volume change (approximately) approaches 50%, which is what we would expect.

Distribution of $\delta_P(t)$

Just out of curiosity, we can also use a kde (kernel density estimate) to get a sense of the distribution of the delta p values:



So the evidence suggests that:

$$\delta_P(t) \sim N(0.0031, 0.0015) \quad \forall t$$

(Where $t \in (t_0, t_0 + \Delta t, t_0 + 2\Delta t, \ldots, t_0 + n\Delta t)$ for some starting time t_0.)