Methodology

1. I want to know which model predicts the next step better.
   1. This could be evaluated as (a)symmetrical percentage deviation from the actual next value
   2. In certain cases (return), this could be evaluated in terms of how often was the correct sign predicted
2. I want to know when the model does well, when it does bad.
3. I want to figure out why it does well sometimes and why it does bad sometimes.

General notes

1. Neither Lag Llama nor Arima are good for predicting raw price. For this taskm Lag-llama is multiple magnitudes WORSE.
2. Neither are good for predicting returns, however, they are better at predicting returns than they are at predicting raw price
3. Even though it is bad at predicting the actual returns, lag llama seems to be quite good at predicting the sign of the return