Crypto Trading Strategies

Apr 27, 2021

Amanda Brown MS, MS&E ‘21
Jonathan Ling MS, MS&E ‘21
Arjun Sawhney MS, CS ‘21

Contents

Data Selection

Data universe
Data availability
Handling missing data
Blockchain data

Data Exploration

Asset correlation
Autocorrelation

Methods

Statistical arbitrage
Time series analysis
Deep learning

Evaluation

Toy model
Backtesting
Metrics

Dataset Selection

Data universe

7,800+ cryptocurrencies (as of Jan 2021)[1]

500+ cryptocurrency exchanges[2]

30+ public APIs available[3]; we looked into Kraken and Bitfinex as they had downloadable data without needing an API

BTCUSD is the most traded pair

Data availability: many new currencies have only been in existence for < 3 years

Data is mostly already clean, but missing when exchange is down or trade volume is zero

1 https://e-cryptonews.com/how-many-cryptocurrencies-are-there-in-2021/
2 https://www.cryptimi.com/guides/how-many-cryptocurrency-exchanges-are-there
3 https://github.com/public-apis/public-apis#cryptocurrency

Data source choice: Bitfinex is more liquid and has more complete data than Kraken

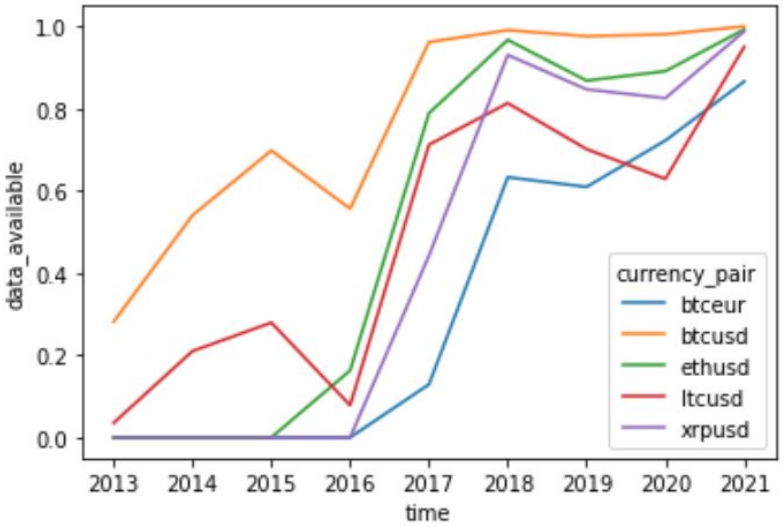
*Data was for BTCUSD from one sampled day (3/19/21)

BTC/ETH/XRP-to-USD are the most data-complete currency pairs across 2019-21

We filtered for all currency pairs whose data availability percentage was above 60% for 2019, 2020 and 2021 (only 5 pairs qualified), then plotted their availability.

From the data, BTC, ETH and XRP to USD are the most data-complete coins. This is confirmed by the fact that they are also the top traded coins on coinmarketcap.com by volume and market capitalization.

Data availability (percentage non-missing) at the minute level for the most data-complete currency pairs in the Bitfinex data

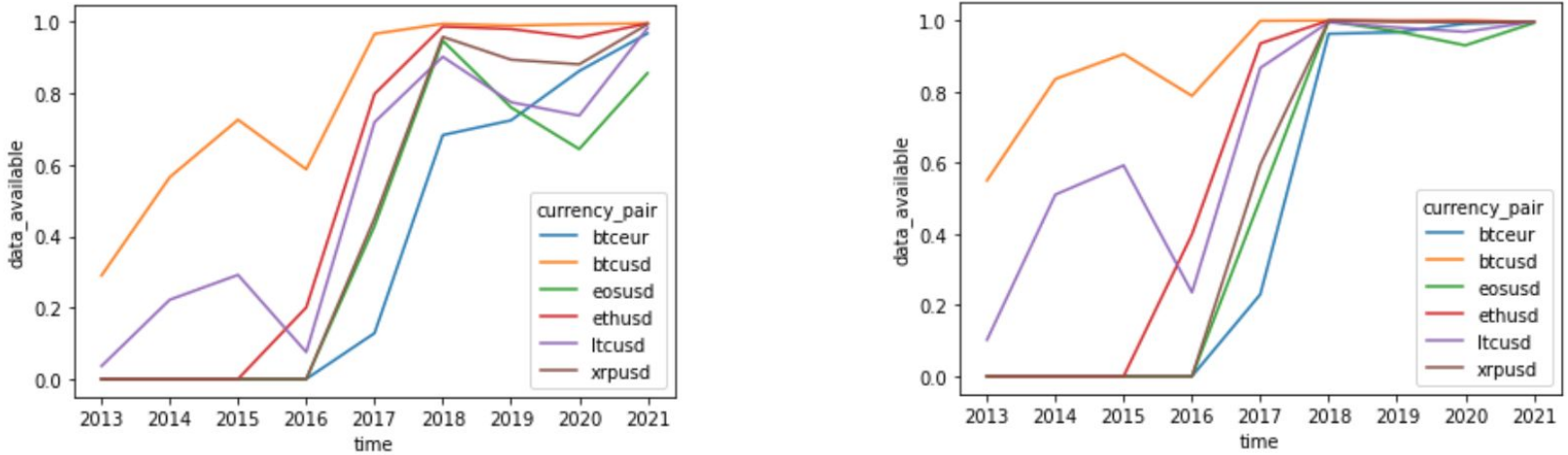


Hourly-level data cleaning and availability was done by syncing ‘close’ timestamps

Resolution technique for syncing “close” time stamps (+/- 2 minutes). This yielded much higher data availability percentage than minute-level data, as expected.

Data availability (percentage non-missing) at the hour level calculated using two methods

Exact filtering on on-the-hour timestamps → +/- 2 minute timestamp leniency



Data Exploration

Time period of observations

Pairs Trading

- Start: 2020-01-01
- End: 2021-04
- Frequency: by hour

Single-Asset Models

- Start: 2018-01-01
- End: 2021-04-01
- Frequency: by minute & hour

Correlation matrix (hourly data), Sept. 2020 - Jan. 2021

Returns

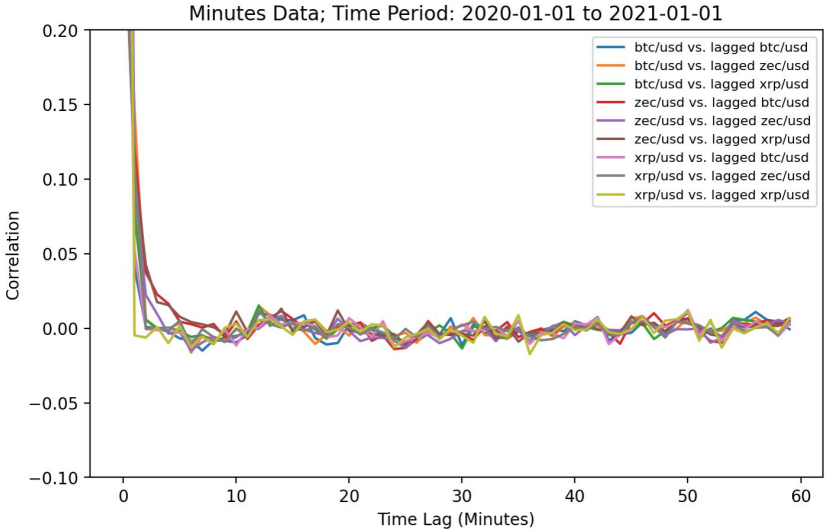
	btc	eth	zec	xrp	dog	dot	ada	ltc	xlm
btc	1.00	0.79	0.59	0.44	0.14	0.56	0.56	0.73	0.50
eth	0.79	1.00	0.64	0.47	0.14	0.63	0.62	0.77	0.55
zec	0.59	0.64	1.00	0.45	0.11	0.55	0.56	0.64	0.51
xrp	0.44	0.47	0.45	1.00	0.10	0.39	0.44	0.52	0.62
dog	0.14	0.14	0.11	0.10	1.00	0.14	0.10	0.15	0.10
dot	0.56	0.63	0.55	0.39	0.14	1.00	0.53	0.58	0.46
ada	0.56	0.62	0.56	0.44	0.10	0.53	1.00	0.60	0.63
ltc	0.73	0.77	0.64	0.52	0.15	0.58	0.60	1.00	0.56
xlm	0.50	0.55	0.51	0.62	0.10	0.46	0.63	0.56	1.00

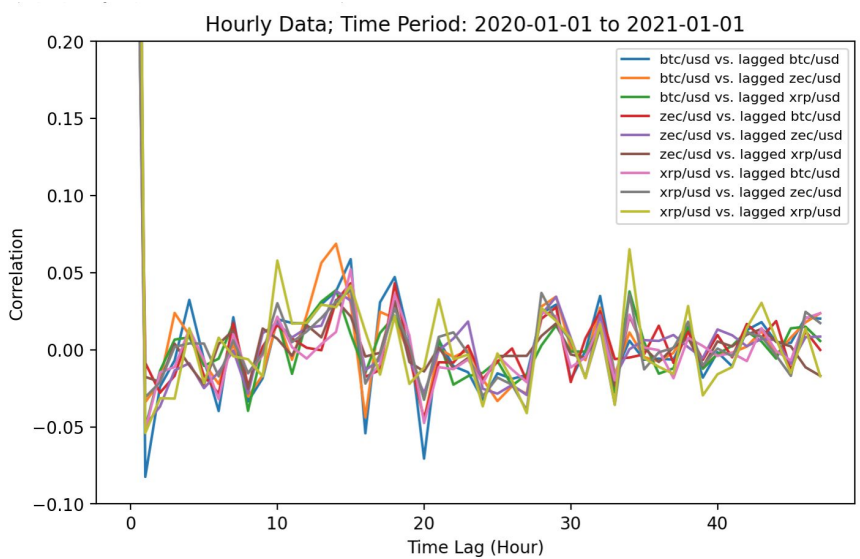
Open Price

	btc	eth	zec	xrp	dog	dot	ada	ltc	xlm
btc	1.00	0.98	0.88	0.61	0.63	0.96	0.95	0.98	0.96
eth	0.98	1.00	0.91	0.65	0.69	0.95	0.93	0.97	0.98
zec	0.88	0.91	1.00	0.82	0.80	0.91	0.90	0.91	0.92
xrp	0.61	0.65	0.82	1.00	0.73	0.62	0.64	0.68	0.71
dog	0.63	0.69	0.80	0.73	1.00	0.65	0.68	0.70	0.68
dot	0.96	0.95	0.91	0.62	0.65	1.00	0.98	0.92	0.93
ada	0.95	0.93	0.90	0.64	0.68	0.98	1.00	0.90	0.92
ltc	0.98	0.97	0.91	0.68	0.70	0.92	0.90	1.00	0.96
xlm	0.96	0.98	0.92	0.71	0.68	0.93	0.92	0.96	1.00

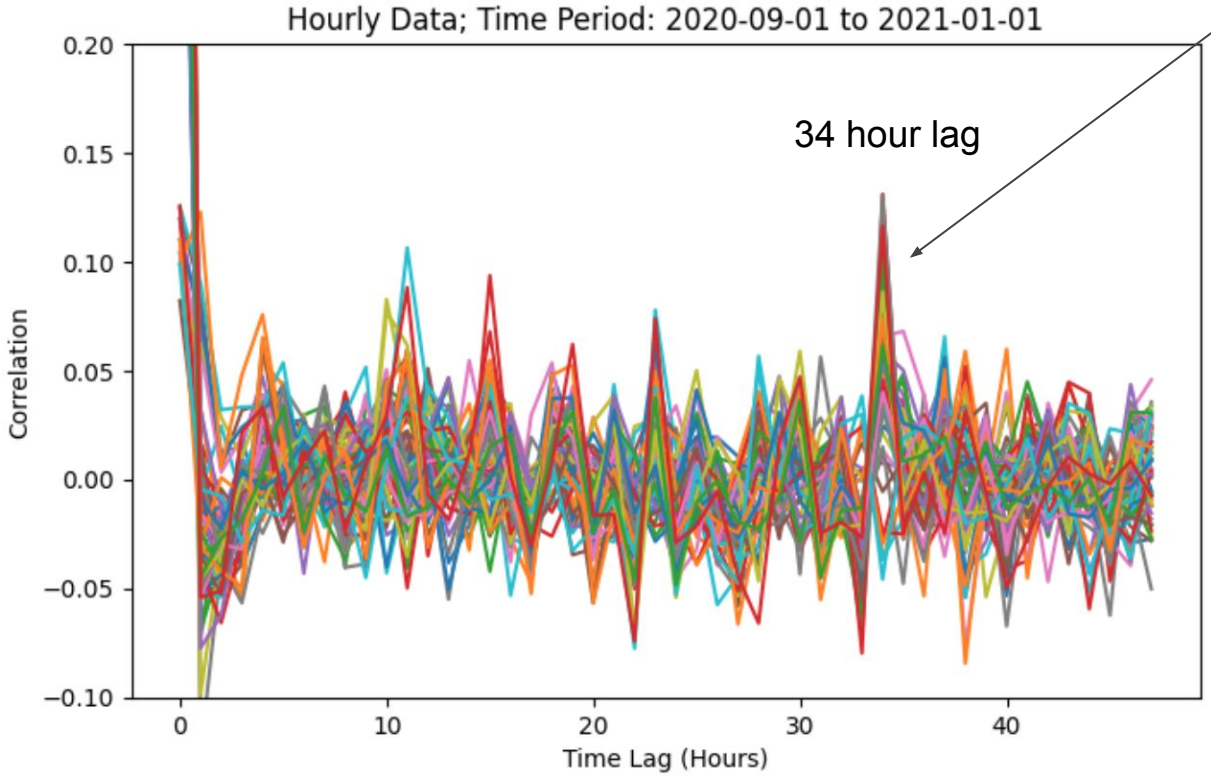
Time series correlations (BTC, ETH, XRP)

We observe extremely weak correlations at the minute level





Hourly returns (8 tokens)

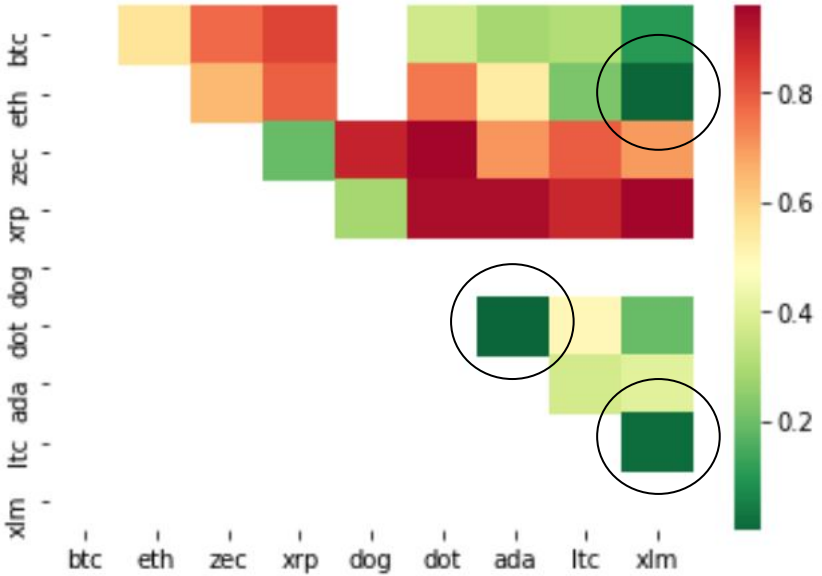


Top 4 Lagged Corrs:

```
(btc, xrp), corr = 0.11
(xrp, xlm), corr = 0.13
(ada, xlm), corr = 0.13
(xlm, xlm), corr = 0.12
```

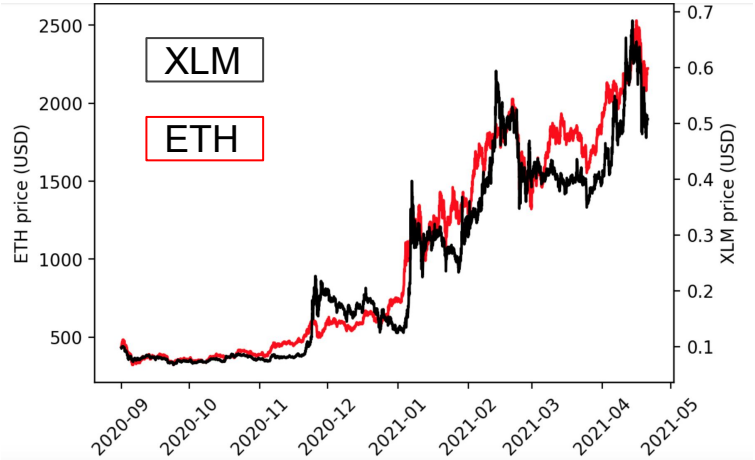
Co-integrated pairs

Co-integration p-values (plotting p < 0.98)



Pairs where p-value is < 0.05:

(ETH, XLM)
(DOT, ADA)
(LTC, XLM)



Methods

Stat arb: pairs trading strategy

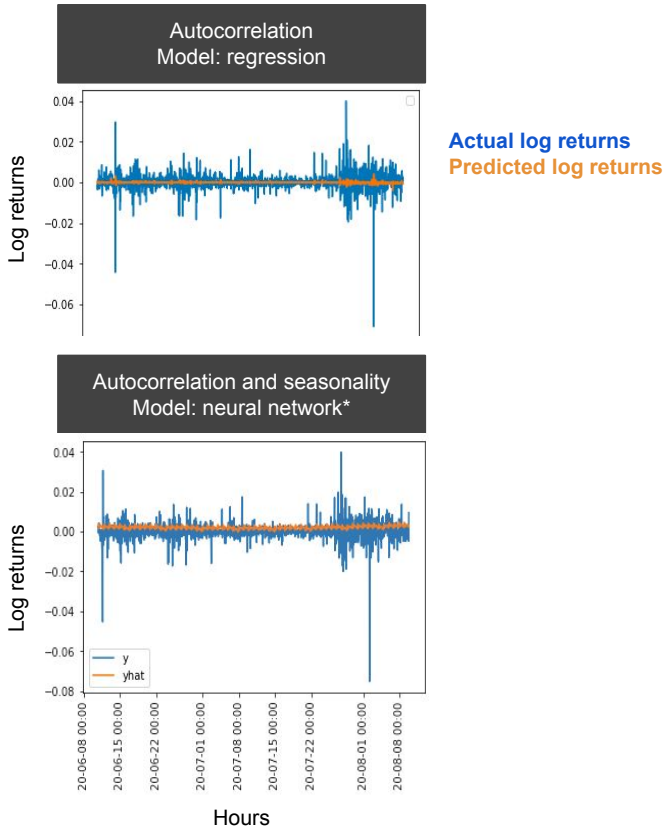
1. Calculate price ratios of cointegrated pairs for all time points in training set (e.g. ‘eth’ divided by ‘xlm’)
2. Get 6 hour moving average of ratios
3. Get 72 hour moving average and standard deviation of ratios
4. Calculate z-score
5. If | z-score | > 2, sell the overperforming coin, buy the other

e.g. If ratio = eth price / xlm price, then
if ratio is low (z < -2), buy ‘eth’ and sell ‘xlm’

If ratio is high (z > 2), sell ‘eth’ and buy ‘xlm’

Time series regression: autocorrelation and seasonality

- We use lagged regression features up until some lookback time period to predict the target
- Feature engineering can then be performed (as a function of the lookback) to account for non-linear signals and interactions
- Seasonality settings will require further adjustments to make the model predictions more granular



**NeuralProphet library from Facebook*

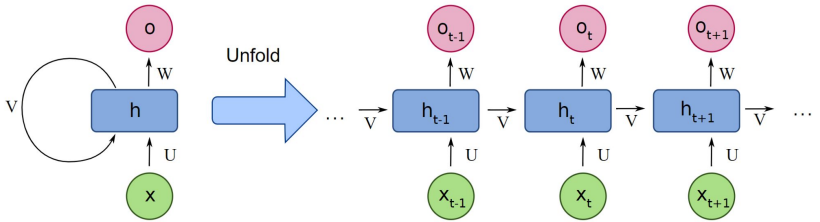
Deep learning approaches

Underway:

- RNNs (recurrent neural networks)
  - LSTMs (long short-term memory)

Approaches to try next:

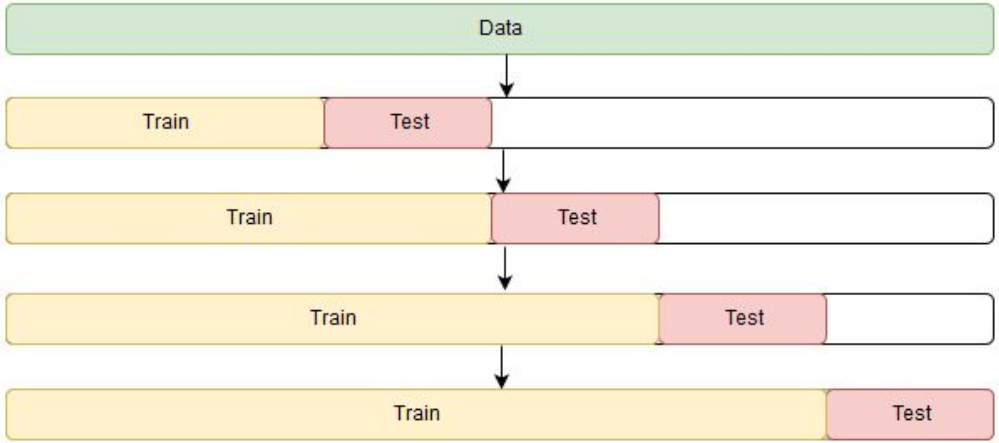
- CNNs (convolutional neural networks)
- Transformers



Evaluation

Evaluating models: train, validate & test

- Rolling-window cross-validation approach
- Accounting for seasonality or trends in model performance over time



Evaluating strategies: backtesting

- Need to account for factors such as transaction costs and market impact
- We aim to use Backtrader as our backtesting framework
- It allows us to define data feeds to feed our models and also account for transaction costs, initial investments and the possibility of going long/short and trading on margin

Metrics for evaluation

We use different sets of metrics to evaluate our models and our strategies

- For the model level, given that we focus on a regression task, we focus on the **validation adjusted R^2**
- For strategies, we consider the risk-adjusted return as our benchmark and so consider the **Sharpe ratio** as our strongest metric
- To get a sense of our downside, we also consider our **max-drawdown** and **win-ratio**