cs-4641-project

Crypto Pairs Trading

Introduction/Background

Financial markets present an evolving challenge for traders trying to identify and capitalize on inefficiencies. Our project will focus on pairs trading. We will identify pairs of crypto assets whose price tends to move together and predict the price of one using the other [3]. Our dataset will be a time series dataset of price and volume. We may also incorporate other features, like implied vol, but those require processing.

```
             Adj Close       Volume
Date
2015-01-01   314.248993      8036550
2015-01-02   315.032013      7860650
2015-01-03   281.082001     33054400
2015-01-04   264.195007     55629100
2015-01-05   274.473999     43962800
...                 ...          ...
2024-09-26 65181.019531  36873129847
2024-09-27 65790.664062  32058813449
2024-09-28 65887.648438  15243637984
...
2024-09-30 63329.500000  37112957475
```

Problem/Motivation

Crypto is an emerging asset class, where we can apply proven ideas from traditional finance to capture inefficiencies [2]. There are fundamental economic reasons, like ecosystem tokens, for why these crypto assets should trade in a correlated manner. There are also frequent price-insensitive flows from unsophisticated traders and liquidations causing deviations from these relationships. Thus, we can make pricing more efficient by taking the other side of these trades.

Methods

To choose particular stock pairs, we used the data preprocessing methods of correlation analysis on daily log returns to identify which assets tend to move together on a short term basis and cointegration analysis to identify which assets have a stationary, mean-reverting spread on a longer time horizon.

We used correlation as one of the data processing methods on close-to-close log returns to determine the top 5 crypto asset pairs out of 19 total crypto symbols to use for the linear regression model (listed below). A high correlation would mean that two assets would be linked in some way, making it suitable to use pairs trading strategies on them.

```
ETHUSDT and BTCUSDT: 0.8423116420659079
DOTUSDT and ADAUSDT: 0.791728177932053
SHIBUSDT and DOGEUSDT: 0.7803690843815932
ETCUSDT and ETHUSDT: 0.7728313755045593
DOTUSDT and LINKUSDT: 0.7468021238329571
```

We also used cointegration as a longer term test on the log prices as opposed to log returns. It helps to determine long-term equilibrium relationships which allows us to determine if the spread between two assets is stationary, reverting to a mean value over time. If the pair is cointegrated, this means a strong signal for pairs trading. The top 5 cointegrated pairs are listed below:

```
AVAXUSDT and ICPUSDT: 0.004368030729914953
XRPUSDT and LTCUSDT: 0.007590820581341352
DOTUSDT and FILUSDT: 0.010731569533025098
LTCUSDT and TRXUSDT: 0.01405180807932554
UNIUSDT and ETCUSDT: 0.022047971836758017
```

Model

In pairs trading, the goal is to find a relationship between two assets that can be exploited for profit.

Linear Regression

Linear regression would fit a linear equation to the data, such as BTC Returns=α+β⋅ETH Returns. Because in pairs trading, we assume that the spread will revert to the mean over time, by fitting a regression model to the spread between two cointegrated assets, the regression line can act as a baseline for how the spread, and thus, signal an opportunity to conduct trading. It can also be used as a part of factor analysis, where we can decompose the returns of one name into the returns of many other names.

Ridge and Lasso Regression

Ridge regression fits a linear equation to the data with an L2 regularization term. This means the loss function is the sum of the mean squared error and the squares of the coefficients. Similarly, Lasso regression fits a linear equation to the data with an L1 regularization term. This means the loss function is the sum of the mean squared and the absolute value of the coefficients. In both cases, by adding a regularization term, we can reduce overfitting by penalizing large coefficients. Compared to the baseline linear regression, we obtained a slightly smaller slope with almost identical intercept. This was as expected, because the data seemed quite clean with clear relationships, so the regularization would only slightly decrease the coefficients.

Elastic Net

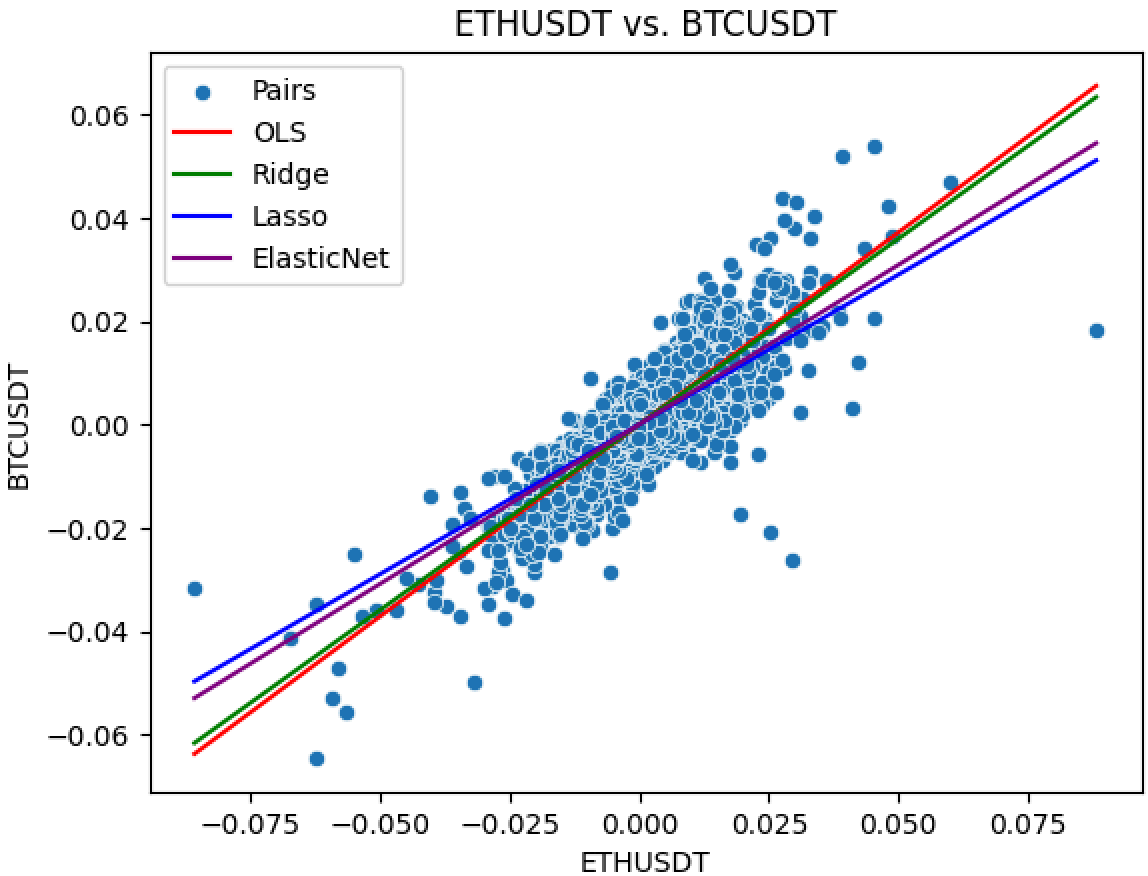
Elastic Net fits a linear equation to the data with a weighted average of L1 and L2 regularization terms. This weighting determines whether it behaves more like Ridge or Lasso. Once again, the regularization term serves to reduce overfitting by penalizing large coefficients that don’t improve mean squared error proportionally. Compared to the baseline linear regression, we obtained a slightly smaller slope with almost identical intercept. Additionally, the line fit by Elastic Net usually rested in between the lines fit by Ridge and Lasso. This was as expected, because we used a 50/50 combination of the 2 regularization terms.

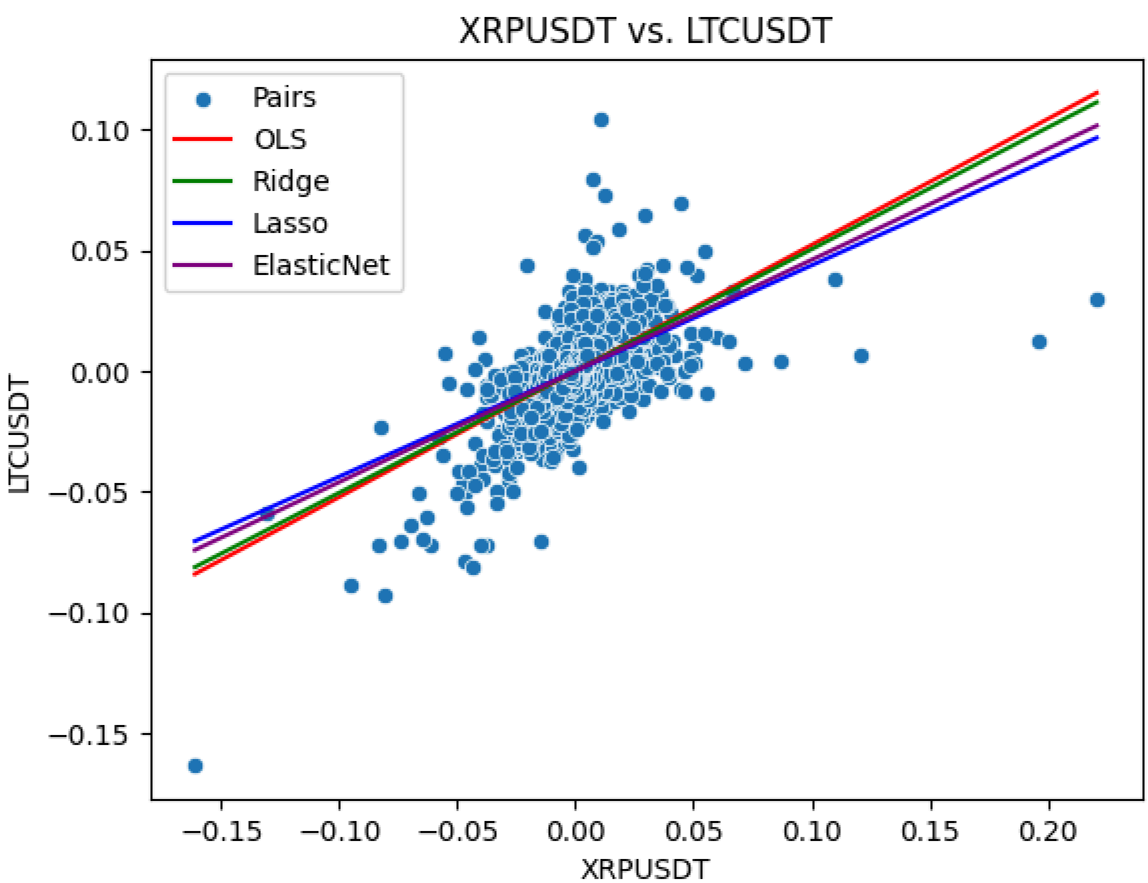
Comparison

All these variants of linear regression performed very similarly. As expected, the regularized variants had a slightly smaller slope than ordinary least squares. The lack of difference is actually a good sign, because it means our data has a high signal-to-noise ratio, which largely negates the effects of regularization. This means that using one asset to predict another is a viable idea and can be used in practice to perform strategies, like simple pairs trading or factor decomposition. This is a very important discovery, because many relationships in finance are noisy and prone to overfitting, especially those obtained from data mining.

Visualizations

Log returns plotted against each other:


ETHUSDT vs. BTCUSDT
Pairs
OLS
Ridge
Lasso
ElasticNet
BTCUSDT
ETHUSDT
0.06
0.04
0.02
0.00
−0.02
−0.04
−0.06
−0.075
−0.050
−0.025
0.000
0.025
0.050
0.075


XRPUSDT vs. LTCUSDT
Pairs
OLS
Ridge
Lasso
ElasticNet
LTCUSDT
XRPUSDT
0.10
0.05
0.00
−0.05
−0.10
−0.15
−0.15
−0.10
−0.05
0.00
0.05
0.10
0.15
0.20

Metrics

For pairs screened via correlation:

Ordinal Least Squares (OLS)

```
('ETHUSDT', 'BTCUSDT'): Hedge Ratio = 0.7785922771578035
Training: MSE = 6.222184816858084e-06, R^2 = 0.7106221705771293
Validation: MSE = 1.044018722661196e-05, R^2 = 0.6803192024861348
Testing: MSE = 8.06152797560662e-06, R^2 = 0.7272810953642332

('DOTUSDT', 'ADAUSDT'): Hedge Ratio = 0.78468532841248
Training: MSE = 2.5641598591454155e-05, R^2 = 0.5651357656920749
Validation: MSE = 2.4052460136842426e-05, R^2 = 0.7244224207580342
Testing: MSE = 1.9911583375024933e-05, R^2 = 0.6644115312111445

('SHIBUSDT', 'DOGEUSDT'): Hedge Ratio = 0.6966346365978953
Training: MSE = 2.7378305044224046e-05, R^2 = 0.5401522034009614
Validation: MSE = 6.134986814042188e-05, R^2 = 0.6425527230490127
Testing: MSE = 2.4613421221674433e-05, R^2 = 0.6903365360363483

('ETCUSDT', 'ETHUSDT'): Hedge Ratio = 0.48098515382312307
Training: MSE = 1.1407023657697277e-05, R^2 = 0.5474400469437901
Validation: MSE = 1.6701503215741952e-05, R^2 = 0.6564302039575338
Testing: MSE = 1.86262229753769e-05, R^2 = 0.5688074325092807

('DOTUSDT', 'LINKUSDT'): Hedge Ratio = 0.7807668107963401
Training: MSE = 3.308050418015929e-05, R^2 = 0.4993228693826086
Validation: MSE = 3.3466950582913e-05, R^2 = 0.6208457735638564
Testing: MSE = 2.2517962151019723e-05, R^2 = 0.649876002356445
```

Ridge Regression

```
('ETHUSDT', 'BTCUSDT'): Hedge Ratio = 0.7193398262056809
Training: MSE = 6.310677805065621e-06, R^2 = 0.7065065890570723
Validation: MSE = 1.0058980573405015e-05, R^2 = 0.6919918328968382
Testing: MSE = 7.855694953755677e-06, R^2 = 0.7342443604458551

('DOTUSDT', 'ADAUSDT'): Hedge Ratio = 0.7556944398223716
Training: MSE = 2.5687084428292218e-05, R^2 = 0.5643643565485346
Validation: MSE = 2.4298690307291808e-05, R^2 = 0.7216012742340311
Testing: MSE = 2.019837767567353e-05, R^2 = 0.6595779196193735

('SHIBUSDT', 'DOGEUSDT'): Hedge Ratio = 0.6754716437688396
Training: MSE = 2.740798427510396e-05, R^2 = 0.5396537091040048
Validation: MSE = 6.132137590903601e-05, R^2 = 0.6427187294453072
Testing: MSE = 2.5016518447446404e-05, R^2 = 0.6852651368951038

('ETCUSDT', 'ETHUSDT'): Hedge Ratio = 0.46480546306946297
Training: MSE = 1.1422637503130841e-05, R^2 = 0.5468205863931257
Validation: MSE = 1.7061205403749793e-05, R^2 = 0.6490307019023296
Testing: MSE = 1.9137791870385914e-05, R^2 = 0.5569647360281459

('DOTUSDT', 'LINKUSDT'): Hedge Ratio = 0.7519206952809084
Training: MSE = 3.3125536862053855e-05, R^2 = 0.49864129470550633
Validation: MSE = 3.342356081213624e-05, R^2 = 0.6213373455382238
Testing: MSE = 2.2904824399642122e-05, R^2 = 0.6438608151864589
```

Lasso Regression

```
('ETHUSDT', 'BTCUSDT'): Hedge Ratio = 0.5802232868599575
Training: MSE = 7.214029768347314e-06, R^2 = 0.6644940101907079
Validation: MSE = 1.0505575203780067e-05, R^2 = 0.678317008441603
Testing: MSE = 8.565209772731275e-06, R^2 = 0.7102417017886656

('DOTUSDT', 'ADAUSDT'): Hedge Ratio = 0.6922970511300134
Training: MSE = 2.6103539977866485e-05, R^2 = 0.5573015510435276
Validation: MSE = 2.5396046507477278e-05, R^2 = 0.7090284744666567
Testing: MSE = 2.1138363509141387e-05, R^2 = 0.6437354624628844

('SHIBUSDT', 'DOGEUSDT'): Hedge Ratio = 0.6211825018551167
Training: MSE = 2.7755565717937 94e-05, R^2 = 0.5338157085277186
Validation: MSE = 6.22166335007329e-05, R^2 = 0.6375026238851569
Testing: MSE = 2.6412785139632295e-05, R^2 = 0.6676985915284399

('ETCUSDT', 'ETHUSDT'): Hedge Ratio = 0.3971549423617758
Training: MSE = 1.1826174715004015e-05, R^2 = 0.5308107325397478
Validation: MSE = 1.91035033951939e-05, R^2 = 0.6070182018707737
Testing: MSE = 2.15149184897538e-05, R^2 = 0.5019348283753287

('DOTUSDT', 'LINKUSDT'): Hedge Ratio = 0.6883785335138738
Training: MSE = 3.354244556657162e-05, R^2 = 0.49233133483395863
Validation: MSE = 3.388815730819729e-05, R^2 = 0.616073832669534
Testing: MSE = 2.407054102653583e-05, R^2 = 0.625735490932412
```

Elastic Net Regression

```
('ETHUSDT', 'BTCUSDT'): Hedge Ratio = 0.6181016790241669
Training: MSE = 6.871410082627159e-06, R^2 = 0.6804283714945736
Validation: MSE = 1.0197549275670714e-05, R^2 = 0.6877488291758085
Testing: MSE = 8.206277913143803e-06, R^2 = 0.7223842514246368

('DOTUSDT', 'ADAUSDT'): Hedge Ratio = 0.7058835090878808
Training: MSE = 2.597766540163404e-05, R^2 = 0.5594362990397153
Validation: MSE = 2.5096299458308758e-05, R^2 = 0.7124627828793856
Testing: MSE = 2.0900774231138947e-05, R^2 = 0.6477397759573851

('SHIBUSDT', 'DOGEUSDT'): Hedge Ratio = 0.6349542426890687
Training: MSE = 2.7630416598503966e-05, R^2 = 0.5359177212975033
Validation: MSE = 6.185762003558231e-05, R^2 = 0.6395943706059988
Testing: MSE = 2.600924827879694e-05, R^2 = 0.6727755217543466

('ETCUSDT', 'ETHUSDT'): Hedge Ratio = 0.4214067400285349
Training: MSE = 1.1618736555016871e-05, R^2 = 0.5390405922089263
Validation: MSE = 1.8271463212512216e-05, R^2 = 0.6241342585616254
Testing: MSE = 2.0618552107511678e-05, R^2 = 0.5226854938367211

('DOTUSDT', 'LINKUSDT'): Hedge Ratio = 0.702138011506305
Training: MSE = 3.3415096850335956e-05, R^2 = 0.4942587718973551
Validation: MSE = 3.372219114149702e-05, R^2 = 0.617954098796378
Testing: MSE = 2.3781531727689418e-05, R^2 = 0.6302291964635622
```

For pairs screened via cointegration:

Ordinal Least Squares (OLS)

```
('AVAXUSDT', 'ICPUSDT'): Hedge Ratio = 0.6361523223890483
Training: MSE = 7.025600759008873e-05, R^2 = 0.332200543102271 54
Validation: MSE = 7.969596843511086e-05, R^2 = 0.44525900912320526
Testing: MSE = 4.3495653848474356e-05, R^2 = 0.5061390120384779

('XRPUSDT', 'LTCUSDT'): Hedge Ratio = 0.5218811474844558
Training: MSE = 3.441821388475662e-05, R^2 = 0.3190928059479259
Validation: MSE = 3.5672086870892905e-05, R^2 = 0.4734586336987744
Testing: MSE = 2.845502685277849e-05, R^2 = 0.3253365985687119

('DOTUSDT', 'FILUSDT'): Hedge Ratio = 0.9248351521289037
Training: MSE = 5.057916245949115e-05, R^2 = 0.47785795744834414
Validation: MSE = 6.132442195218051e-05, R^2 = 0.5703227412389006
Testing: MSE = 2.1766254503304957e-05, R^2 = 0.6754775622535351

('LTCUSDT', 'TRXUSDT'): Hedge Ratio = 0.3171486271893613
Training: MSE = 1.813477130349501e-05, R^2 = 0.21896885502451613
Validation: MSE = 1.0072927983230006e-05, R^2 = 0.11745511732770131
Testing: MSE = 8.622380375083978e-06, R^2 = 0.004046913665603835

('UNIUSDT', 'ETCUSDT'): Hedge Ratio = 0.6750214968586625
Training: MSE = 3.307563090614275e-05, R^2 = 0.4454525733841156
Validation: MSE = 8.098365790843175e-05, R^2 = 0.1468026873114735
Testing: MSE = 1.9267673607617534e-05, R^2 = 0.5410446170980718
```

Ridge Regression

```
('AVAXUSDT', 'ICPUSDT'): Hedge Ratio = 0.6212176160091072
Training: MSE = 7.027526992309497e-05, R^2 = 0.3320174502115375
Validation: MSE = 8.002472565201155e-05, R^2 = 0.4429706235518589
Testing: MSE = 4.382770826286256e-05, R^2 = 0.5023687796902522

('XRPUSDT', 'LTCUSDT'): Hedge Ratio = 0.50420445872914
Training: MSE = 3.443671832841072e-05, R^2 = 0.3187267262655753
Validation: MSE = 3.621405808886873e-05, R^2 = 0.4654588139337642
Testing: MSE = 2.8404157078000446e-05, R^2 = 0.3265427114801269

('DOTUSDT', 'FILUSDT'): Hedge Ratio = 0.8906663052182207
Training: MSE = 5.0642347429140274e-05, R^2 = 0.47720568233134153
Validation: MSE = 6.142805931817019e-05, R^2 = 0.5695965930273709
Testing: MSE = 2.1815169107939017e-05, R^2 = 0.6747482734025368

('LTCUSDT', 'TRXUSDT'): Hedge Ratio = 0.30463597929404007
Training: MSE = 1.8142685354663182e-05, R^2 = 0.218628011440562
Validation: MSE = 9.867948025081266e-06, R^2 = 0.13541454416126275
Testing: MSE = 8.463038137692531e-06, R^2 = 0.02245220155710359

('UNIUSDT', 'ETCUSDT'): Hedge Ratio = 0.6518125493890569
Training: MSE = 3.3107039362945e-05, R^2 = 0.4449259778690342
Validation: MSE = 7.842666111851585e-05, R^2 = 0.17374173706613127
Testing: MSE = 1.885095800553304e-05, R^2 = 0.550970769710512
```

Lasso Regression

```
('AVAXUSDT', 'ICPUSDT'): Hedge Ratio = 0.5782555311316591
Training: MSE = 7.054549154637425e-05, R^2 = 0.3294489317394832
Validation: MSE = 8.128525267113882e-05, R^2 = 0.43419645314715416
Testing: MSE = 4.497455838162049e-05, R^2 = 0.4893471445938655

('XRPUSDT', 'LTCUSDT'): Hedge Ratio = 0.4374513393818314
Training: MSE = 3.484036292526974e-05, R^2 = 0.31074128835872106
Validation: MSE = 3.86050415535876e-05, R^2 = 0.4301664660295553
Testing: MSE = 2.8538121039770607e-05, R^2 = 0.32336645082907634

('DOTUSDT', 'FILUSDT'): Hedge Ratio = 0.8324468748464369
Training: MSE = 5.1041103845903476e-05, R^2 = 0.4730892146042204
Validation: MSE = 6.211789561533596e-05, R^2 = 0.564763168435269
Testing: MSE = 2.2185786696479728e-05, R^2 = 0.6692225765819523

('LTCUSDT', 'TRXUSDT'): Hedge Ratio = 0.21823193204475377
Training: MSE = 1.8629354779218042e-05, R^2 = 0.19766805713381053
Validation: MSE = 9.031528061836378e-06, R^2 = 0.20869791911993318
Testing: MSE = 7.72320226574889e-06, R^2 = 0.10790909257671288

('UNIUSDT', 'ETCUSDT'): Hedge Ratio = 0.5892714636688043
Training: MSE = 3.350438107209205e-05, R^2 = 0.4382641299689921
Validation: MSE = 7.261267287517676e-05, R^2 = 0.2349945783594748
Testing: MSE = 1.815276929836038e-05, R^2 = 0.5676016028854909
```

Elastic Net Regression

```
('AVAXUSDT', 'ICPUSDT'): Hedge Ratio = 0.5901208742245502
Training: MSE = 7.043899650658021e-05, R^2 = 0.3304611915045371
Validation: MSE = 8.089042042162305e-05, R^2 = 0.43694476824548734
Testing: MSE = 4.462939131654052e-05, R^2 = 0.49326626139496066

('XRPUSDT', 'LTCUSDT'): Hedge Ratio = 0.46023736713856067
Training: MSE = 3.464325026353852e-05, R^2 = 0.31464083497264295
Validation: MSE = 3.772767945767405e-05, R^2 = 0.44311685601921913
Testing: MSE = 2.8434432041708127e-05, R^2 = 0.32582489771390166

('DOTUSDT', 'FILUSDT'): Hedge Ratio = 0.8398450928322192
Training: MSE = 5.097008381833246e-05, R^2 = 0.47382237309192177
Validation: MSE = 6.19943577602693e-05, R^2 = 0.5656287519210728
Testing: MSE = 2.2118609700665488e-05, R^2 = 0.6702241472673823

('LTCUSDT', 'TRXUSDT'): Hedge Ratio = 0.25507470614369926
Training: MSE = 1.8329539824033934e-05, R^2 = 0.21058053415430034
Validation: MSE = 9.264470780098946e-06, R^2 = 0.18828852035319232
Testing: MSE = 7.961655447798668e-06, R^2 = 0.0803658652685092

('UNIUSDT', 'ETCUSDT'): Hedge Ratio = 0.6061574687326796
Training: MSE = 3.335214708889667e-05, R^2 = 0.4408164913695666
Validation: MSE = 7.402771407022545e-05, R^2 = 0.22008651695375026
Testing: MSE = 1.828022261283884e-05, R^2 = 0.5645656689196251
```

Analysis

Our preprocessing techniques, coupled with our linear regression model, yielded highly promising results, demonstrating the effectiveness of our model. From our visualization, we can see that we are able to find currencies that are highly correlated which is the basis of our pairs trading approach. In addition, we can see our R^2 values are high which shows that our model really fits the data. On the most predictive pairs, we were able to explain over 70% of the variance using the other asset.

It’s pretty explainable that our model performed well, because there are fundamental economic drivers that impact the price of both assets in a predictable manner. For example, interest rates, economic growth, and crypto policy drive all cryptocurrencies together. [1] Further, for pairs like SHIB and DOGE, they likely share a holder base and are heavily impacted by speculative fervor. It also makes sense that the relationship is linear, because that’s how their fair values should change proportionally in reaction to new information.

Next Steps

One thing we noticed was heteroskedasticity with respect to returns. The variance in returns grew as the absolute level of returns (up or down) increased. This breaks the fundamental assumption in linear regression that variance stays constant. We could experiment with autoregressive conditional heteroskedasticity (ARCH) models that are commonly used in finance to capture nonconstant, autocorrelated variance. This would help our model out a lot in the tails.

Gantt Chart

You can view the detailed project timeline and milestones in the Gantt chart here.

Contribution Table

You can view the proposal contributions here.

References

[1] Polizu, C. (2023). Are crypto markets correlated with macroeconomic factors? https://www.spglobal.com/content/dam/spglobal/corporate/en/images/general/special-editorial/are-crypto-markets-correlated-with-macroeconomic-factors.pdf

[2] H. Yang and A. Malik, "Reinforcement Learning Pair Trading: A Dynamic Scaling approach," arXiv.org, 2024. https://www.arxiv.org/abs/2407.16103

[3] X. Zhu, "Examining Pairs Trading Profitability *," 2024. Accessed: Oct. 03, 2024. [Online]. Available: https://economics.yale.edu/sites/default/files/2024-05/Zhu_Pairs_Trading.pdf