Investigating Pairs Trading Using Deep Q-Networks, Linear Regression, and XGBoost

Advisor: Linda Kreitzman

Rashi Mohta and John "Jack" Passyn

Contents

Summary

We use the results of a 2019 paper on reinforcement learning as a benchmark for the profitability of machine learning (ML) for pairs trading. We aim to investigate and compare the profitability of alternative ML models. We choose a form of XGBoost, as described in our rationale. We run our model on the same stock pairs and with the same train-test time horizons as the benchmark paper and observe poor profitability. This raises the question of whether it was due to the ineffectiveness of our model relative to that in the benchmark. We undertake a close inspection of the results of the benchmark paper and discover that it too, actually posted negative or mediocre performance on most of the pairs it traded, with 50 to 80 percent of its profit coming from the trading of a single pair of two water companies, and 94% of the profit in the best parameterization from just three pairs. We find evidence suggesting that, within these three pairs, the high profit was due to the benchmark algorithm being programmed to unwittingly take trades that would have required leverage ratios in the hundreds or even thousands to execute, far higher than the ~20X leverage offered to the best hedge funds, or the 3X leverage that is more typical (Mandel).

In the context of the low profitability of both our and Brim's models, we take a brief look at the history of pairs trading and conclude that since its discovery over 40 years ago, that the well-worn strategy and its variants have likely been "priced in" by hedge funds and trading firms in the easy-to-trade U.S. markets. We end by encouraging future researchers to look beyond pairs trading or, if they don't, to investigate whether it applies in markets other than the U.S. equity markets, where its alpha seems to have long since decayed due to overuse.

Introduction: Pairs Trading

Pairs trading is a trading strategy that attempts to profit from corrections in the relative prices of two securities. It involves two steps:

1) Pair selection
2) Trading the spread or ratio of the pairs

For example, in classic “statistical arbitrage” (in the oldest use of the term, which has now come to refer to a host of other unrelated strategies), a trader might identify that the prices of two airline companies are closely correlated, with a consistent (though not constant) spread or ratio in the data. If one company’s stock price rises, but the other’s does not, the trader could short the first company and buy the other, thus making a market-neutral and industry-neutral bet that the spread between them will reconverge.

Others have attempted more sophisticated methods, like machine learning, to trade such spreads. One would first use a measure of cointegration or correlation to perform step 1), pair selection, then, one would generate a time series representing the spread or ratio of the pair prices, generate features of that time series, and use machine learning to predict returns or make trades. The idea is that, while applying ML to a large universe of equities directly may be difficult, by selecting a small subset of stocks and “canceling out” market and industry risk by taking offsetting short and long positions within the pairs, one might isolate patterns in the price that would be simpler and therefore more readily identified by a machine learning algorithm. Those patterns could be more complex than the simple reconvergence pattern of classic statistical arbitrage.

Introduction: Objective of the Research

This research is based on a 2019 paper by Andrew Brim that purportedly identifies a successful ML strategy for pairs trading. He uses data from 38 pairs of stocks he selected using a cointegration measure and volatility, and applies a reinforcement-learning method called "Q-learning" to trade them. A "Q-learner" in this case is a neural network that is trained directly to make long and short trades, without the intermediate step of making a numeric prediction of the future return over some horizon and then converting that figure into a trade recommendation.

A Q-learner is trained by modeling the problem in a way almost akin to that in a video game: it consists of an "environment" (the features to which the learner has access), an opportunity to take an "action" (in this case, buy, sell, or take no position), and receive a "reward" which the researcher hardcodes into the problem. In Brim's case, a positive "reward" is offered to the learner for profitable trades, and a negative one for losing trades. Brim applies various levels of a "negative return multiplier" to the negative rewards, increasing levels of which cause the Q-learner to trade more conservatively.

The aim of our research is to investigate other means of algorithmic pairs trading. Using Brim's Q-learning results as a baseline, we seek to evaluate whether similar or greater profit is achievable through the use of other machine learning methods and other parameterizations of the problem (such as the use of future-returns prediction as an intermediate between the machine learning model's outputs and the trades into which it enters).

Our Approach: Rationale For Model Selection

We elected to use an XGBoost model for several reasons: first, XGBoost has developed a reputation for producing the winning model in machine learning competitions. Since one of our research questions related to the comparative performance of multiple machine learning models, choosing a "winning" model seemed fruitful. Secondly, the XGBoost model is capable of sufficient complexity to identify highly nonlinear patterns. Thirdly, a regression-based variant of the model that makes numeric predictions is available, and this would enable us to answer another of our research questions, namely how adding numeric returns prediction as a step between features and trade decisions would affect the trading success.

Technical Details: Input Data and Feature Generation

To maximize the comparability of our results to our baseline in Brim (2019), we elected to use the same data. We trained and tested our model on the same 38 stock pairs (minus three which had issues) as Brim. We used the same train and test periods: four years of training data from 2014 through 2017, and one year of test data from 2018.

We generated a time series representing the arithmetic spread between the two stocks in each pair, and used that series for trading. We treated that spread as if it were a stock that could be bought and sold directly. Later in this report, the reader will observe that we take issue both with the use of an arithmetic spread (as opposed to the ratio of the stock prices) and with the treatment of the spread as if it were a security that could be traded directly (as doing so ignores very real constraints on the simultaneous long-short positions that must be adopted to capture that spread), but we elected to parameterize our trading in this format anyway, again to promote comparability to the baseline.

For that same reason, we even elected to employ the same spread-series derived features as inputs to the model. They are indicated in the table below.

Current Spread	The current value of the spread, which is the arithmetic difference between the two stock prices in the pair.
Spread Returns	The percentage change in the spread.
Sp Mean 15days	The 15-day moving average of the spread.
Sp/SpMean 15days	The ratio of the current spread to the 15-day moving average of the spread.
Sp Mean 10days	The 10-day moving average of the spread.
Sp/SpMean 10days	The ratio of the current spread to the 10-day moving average of the spread.
Sp Mean 7days	The 7-day moving average of the spread.
Sp/SpMean 7days	The ratio of the current spread to the 7-day moving average of the spread.
Sp Mean 5days	The 5-day moving average of the spread.
Sp/SpMean 5days	The ratio of the current spread to the 5-day moving average of the spread.

Figure A: Features used in the Machine Learning Model

These features are used in an XGBoost Regressor model that predicts the next day's return. This short time horizon for the target was chosen to reflect Brim's own "zig-zag" trading where the position is often reversed many times in quick succession, as his reinforcement learning model seems to make new decisions on a daily-frequency basis. If our model predicted a positive return greater than some small tolerance, it would take a long position in the spread; if it predicted a negative return less than the negative of that tolerance, it would go short.

Our Model: A Negative Result

Pairs		*Cumulative Return for 2018 Test Data (as a percent)*	
Stock 1	**Stock 2**	**XGBoost**	**Linear Regression**
BEN	COG	-69.6	24.37
DISCA	RIG	-26.44	-60.01
DISCK	RIG	-14.97	-63.05
ADBE	CRM	0	-1.75
CF	HBI	-21.58	-42.97
ESV	GNW	-17.32	198.19
CNX	HBI	-100	4922263.36
AMZN	CRM	0	-38.16
FCX	GNW	-22.04	96.1
CRM	NVDA	0	8.8
CF	FOSL	-62.16	59.36
FCX	HBI	-99.69	20441670125
DISCK	ESV	-67.25	-6.74
DISCK	NE	0	-33.08
DISCA	NE	0	-33.64
DISCA	ESV	-69.89	-11.03
ESV	RRC	-39.78	182.55
NBL	RIG	0	48.3
CNX	GNW	0	27.86
COG	DO	-99.24	270.42
HBI	NBL	-92.25	-11.23
HBI	MRO	-99.99	1.22E+20
GNW	NBL	-5.93	49.73
DISCA	MA	0	-27.92
DISCK	MA	0	-27.88
RIG	RRC	-100.04	1215.6
CF	CNX	-11.12	-27.18
CF	GNW	0	-2.6
ESV	HBI	-38.92	239.07
NE	RRC	0	26.34
ADBE	RHT	-91.5	242.71
MA	RIG	0	7.11
NBL	SWN	0	26.82
CTWS	WTR	0	-33.18
AWR	WTR	0	-53.44
SLB	PFE	-100.66	16272107.21
	Sum:	-1250.37	1.22106E+20
	Sum/100	-12.5037	1.22106E+18

Figure X (previous page): Results from our XGBoost-based trading simulation, available for 35 out of 38 of Brim's pairs, alongside results from replacing the XGBoost model with a simple linear regression model, as a second baseline. The two models (XGBoost and linear regression) used the same guidelines to convert predictions to trades.

The first item one notices in these charts is the extreme returns seen for some stock pairs. In the succeeding sections, where we evaluate Brim's baseline results, we thoroughly discuss this issue of why returns can register as extreme values when one simulates trading the spread as if it were a single security that could be traded directly. To avoid repetitiousness, we won't repeat that discussion here, but suffice it to say that Brim's and our extreme return values occur for the same reason.

The second item that one likely notices is the proliferation of zero values in the table, a feature our results share with Brim's. In our case, the model sometimes failed to make a return prediction of sufficiently large magnitude to surpass the threshold for trading, and hence didn't trade some pairs in the relatively short test set of 2018. In Brim's case, the reinforcement learning model was, for different reasons, also unable to make a strong prediction directionally for some stocks during the training window, and also failed to trade.

The third, and probably most important is the fact that the XGBoost model posted negative returns for nearly all stocks for which it traded, but the linear regression model posted far stronger returns (ultimately the linear model posts better results than those in the baseline paper).

The poor performance of the XGBoost model seems to be a simple case of overfitting. The model fit the in-sample data from 2014 to 2017, but didn't predict well for 2018. This makes sense when one considers that the volume of data is actually relatively small, given that it was collected on a daily frequency for only 38 unique stock pairs (collecting only one data point each day results in less data volume than when doing so, say, every hour). Practitioners have trouble with overfitting even when they train ML models on the entire market---how much moreso might we expect this when we do so on a small subset?

The simple linear model fared far better. The obvious conclusion, that a simpler model is better in this context (with these features and this dataset) is likely correct. However, we don't think that even the linear model's seemingly spectacular profit suggests that pairs trading is profitable, as the issue causing the extreme returns (on order of 10 to the power of 20), which is aforementioned and will be discussed in detail in subsequent sections criticizing Brim' s similar issue, creates concerns about the trade execution feasibility and leverage ratio that are too great. We overall consider this to be a negative result that fails to demonstrate successful pairs trading.

Note About Interpreting Brim (2019)

A comment about interpreting the results in the Brim (2019) paper is in order before the comments we make about those results. Brim reports the results of his pairs trading in the table that follows:

38 Stock Pairs Total Spread Cumulative Returns, by Negative Returns Multiplier during Training											
Negative Returns Multiplier	**1**	**2.5**	**5**	**10**	**20**	**50**	**100**	**200**	**500**	**700**	**1000**
Stock Pair											
BEN_COG	1.23	-0.27	0	1.13	0.27	0.18	0	0	-0.16	0	0.48
DISCA_RIG	-0.22	-0.54	-0.53	0	0	0	0.01	-0.14	0	0	0
DISCK_RIG	1.25	0	1.11	0.45	0.01	-0.14	0	0	-0.14	0	0
ADBE_CRM	1.07	0.09	-0.02	0	0	0	0	0	0	0	0
CF_HBI	0.25	-0.04	-0.14	0.31	0	0	0	0	0	0	0
ESV_GNW	-9.64	-11.54	-11.62	-2.3	-0.32	-11.5	0.89	-9.95	-11.5	11.67	-0.13
CNX_HBI	7.52	8.46	5.78	8.5	11.9	1.42	4.61	3.42	0.44	0	0
AMZN_CRM	0.56	-0.03	-0.01	-0.01	0	-0.07	0	0	0	0	0
MA_VFC	0.01	-0.18	-0.22	-0.26	-0.27	-0.06	0	0	0	0	0
FCX_GNW	-0.19	0	0	0	0	0	0	0	-0.05	0	0
CRM_NVDA	-0.54	-1.22	-3.29	2.94	-0.43	-0.96	-3.56	-3.15	-3.15	0	0
CF_FOSL	-1	-0.04	0.12	0	-0.01	0	0	0	0.02	0.12	0
FCX_HBI	25.67	26.55	17.34	16.12	20.87	22.52	22.27	21.28	20.92	-1.62	0.54
DISCK_ESV	0.08	0	0	-0.18	0	0	0	0	0	0	0
DISCK_NE	-0.08	0	0	0	0	0	0.09	0	0	0	0
DISCA_NE	-0.25	-0.17	0	0	0.12	0	0	0	0	0.03	0
DISCA_ESV	-0.56	-0.16	-0.21	0	0	0	0	0	0	0	0
ESV_RRC	-0.78	0.29	0.15	0.05	-0.28	-0.03	0.2	0	0.11	0.17	0
NBL_RIG	1.14	0.24	-0.04	-0.02	-0.02	0	0.04	0	0	0	0
CNX_GNW	-0.04	0.19	-0.06	0.14	-0.02	0	0.01	0.29	0.03	0	0
COG_DO	2.32	-0.71	0.51	0.54	0.14	-0.97	-0.22	-0.63	-0.41	0	1.06
HBI_NBL	0.24	0.01	0	0.07	0.03	0	0	0	0	0	0
HBI_MRO	27.41	8.6	16.15	29.83	10.68	3.3	13.16	0.6	-0.5	-6.23	0
GNW_NBL	0.09	-0.05	0	-0.02	0.01	0	0	0	0	0	-0.07
DISCA_MA	0.49	0.03	0.01	0.25	-0.04	-0.03	-0.03	0	0	-0.01	0.02
DISCK_MA	-0.21	-0.01	0	0	0	0	0	0	0	0.02	0.03
RIG_RRC	1.32	0.48	0.28	0.31	0	0.44	-0.08	-0.07	0.77	0.58	0.5
CF_CNX	0.09	-0.12	0	0.07	0	0	0	0	0	0	0
CF_GNW	0.12	0.06	0.41	0	0	0.06	0	0	0	0	0
ESV_HBI	0.03	0.17	0	0.1	0	0	0	0	0	-0.02	0
NE_RRC	-0.08	-0.05	0.07	-0.02	0	0	0	0	0	0.01	0
ADBE_RHT	1.58	1	0.55	0	0.07	-0.24	0.05	0	0	0	0
MA_RIG	0.45	0.23	-0.03	0	-0.06	0.08	0	-0.05	0	0	0
NBL_SWN	0.02	0.02	-0.03	-0.32	-0.02	0	0	0	0	0	-0.05
CTWS_AWR	71.28	80.79	87.1	53.6	50.2	44.45	55.92	0.51	1.65	1.39	0.41
CTWS_WTR	-0.48	-0.06	0.19	-0.04	-0.26	0	0.03	0	0	-0.03	0
AWR_WTR	0.62	0.17	0.08	0	0	0	0	0	0	0.15	0
SLB_PFE	0.56	2.73	-2.06	5.75	-0.52	-4.83	-0.72	-6.23	-4.38	-5.08	-5.22
Total Cumulative Returns	**131.33**	**114.92**	**111.59**	**116.99**	**92.05**	**53.62**	**92.67**	**5.88**	**3.65**	**1.15**	**-2.43**

Figure R: Brim's pairs trading results

Notice that Brim neglects to indicate the units of his returns, and they aren't clear from the context provided by the rest of his paper. Should the total of 131.33 in the first column, representing the returns on the test set of year 2018 data with a "negative return multiplier" (a parameter for the deep-Q-network loss function) be interpreted as a 131.33% gain, or as the algorithm making 131.33 times the original capital: 13133%? Which interpretation applies heavily impacts one's analysis of the deep-Q-network results. We favor the returns-multiple interpretation.

Firstly, the majority of returns are between -2 and 2. The percentage interpretation would have it that pairs trading over 1 year on a single, undiversified stock pair that is volatile (Brim filtered out pairs with too little volatility) tends to earn between -2% and 2%. This is unreasonable---the volatility of the trading strategy would be far more than that. The returns-multiple interpretation, under which a return of, say, 0.12 is 12%, is far more reasonable.

Now, how does one make sense of large losses like -9.64 that exceed the initial invested capital? Surely one couldn't lose 964%? Brim's algorithm, however, did, through a spectacularly failed short trade. See the figure below from Brim's paper, representing the performance of his algorithm's trading on the ESV-GNW (Ensco-Genworth Financial) pair.

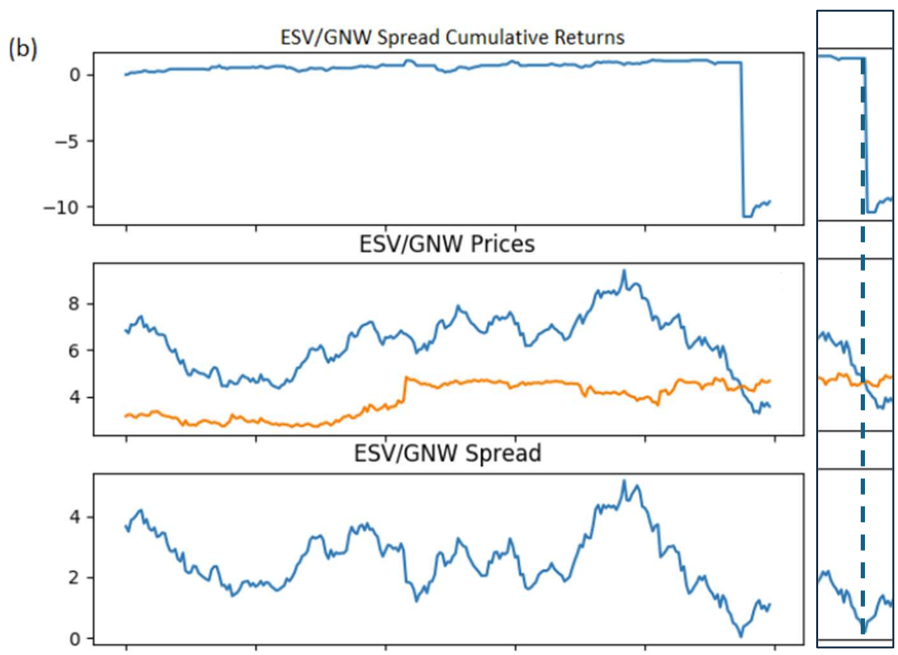


Figure L: Brim's results on the ESV-GNW Pair, with losses highlighted in the inset on the right (dashed line added)

Observe that the bulk of the loss occurs through a single trade that is placed right as the spread between the two companies reaches zero, then sharply increases. Brim's algorithm seems to treat the spread itself as if it were a security, so it short-sells the spread at a low price, say 10 cents, and when the spread rises to $1.00, the percent gain registers on order of 1000% and the short trade loses several times the invested capital.

Brim's largest gains, such as that of 71.28 in the CTWS_AWR spread, are achieved in similar manner: the algorithm treats the spread as a security, and correctly buys or sells when the spread is near-zero, and registers large gains if its prediction is correct. While the stocks themselves don't have insanely high volatility, when the spread between two stocks is near zero, a quotidian 1% move in one stock could mean a 1000% or larger change in the spread between the two stocks. The volatility (in percentage terms) of the spread approaches infinity as the spread approaches zero, allowing Brim's massive gain-or-loss results, which might have otherwise seemed so large that they could have resulted only from the reader misinterpreting the units! Hence, our discussion of the units themselves.

Deep Q-Learning and Pairs Trading: Evaluating Brim' Results

The premise with which we entered into this project is thus 1) Brim's results show that pairs trading with a deep-Q-network is possible 2) we should evaluate whether better or equal profits are attainable through another machine learning method or methods. However, upon further inspection, it seems that Brim's results prove little. With the disclaimer that "absence of evidence is not evidence of absence," and that Brim's results seem to merely depict an absence of evidence, it seems likely that pairs trading was never possible in the 2018 and subsequent periods, possibly due to years of alpha decay as hedge funds and trading firms exploited this decades-old trading strategy of pair-spread reversion until all or nearly all the alpha had been priced out. Allow us to justify these claims about Brim's lack of evidence in the sections that follow.

Brim's Security Selection

In Brim's paper, he acknowledges 78,000 pairs from the S&P 500 Stock Index, which he evaluates. He selects those with an augmented Dickey-Fuller p-value, less than the 5% level, suggesting the pair is co-integrated (since a low p-value for the ADF test normally signifies a lack of integration, one might infer that Brim is using a "reversed" variant of the ADF test where a low p-value signifies integration of the pair-spread time series). From amongst 145 stocks meeting that criterion, he selects 38 that have sufficient volatility to justify trading. Having selected only 38 from amongst 78,000 pairs, less than one in 2,000, one would expect this narrow universe to share common patterns, particularly the one that is typical of co-integrated pairs, namely that the spread between them, whenever it widens, should eventually converge, forming the basis for the most common pairs trading strategies.

It would be expected that strategies that might not work in the general case, with equities broadly speaking, would work within this carefully-selected universe. While it may be difficult to train a machine learning model that makes on-average-profitable trading decisions on, say all S&P 500 stocks, one would expect it to be far more feasible to train one to do so for a narrowly-selected subset of pairs that purportedly have shared patterns of mean-reversion.

Furthermore, this larger stock universe of 78,000 pairs provides important context for the results that follow. We will highlight how a very small number of trades in just one pair is responsible for most of Brim's simulated profit---when those trades are seen in the context of a larger 78,000-pair universe, the perceived likelihood of multiple-testing bias or p-hacking is much greater.

Brim’s Trading Strategy

We would like to note two aspects of how Brim structures the spread that his algorithm trades. Firstly, in his pairs trading, the algorithm trades the arithmetic spread between the two stock prices rather than their ratio. Secondly, Brim treats the spread itself as if it were a security that could bought and sold like a stock. This allows him to make trades that would have otherwise been infeasible.

Brim trades the spread and not the ratio. One would think that if two companies' fates were closely linked, say, because they worked in the same narrow sub-industry, had the same customers, the same suppliers or otherwise the same driving factors behind their market capitalization, that if one stock was at $10 and the other at $1, that if the price of one doubled, say, the first to $20, the second would double to $2. With such a move, the market capitalizations are thus “properly” linked. However, under Brim's model, this change would be considered aberrant as the spread increased from $9 to $18. Under the traditional view of pairs trading, whereby spreads tend to mean revert, one would short sell this spread, even though the relative move in the two firms’ market capitalizations’ was not abnormal in the least.

Notwithstanding that this practice of using the spread instead of the ratio seems not to make sense, We would like to note a second aspect: Brim's treating the spread like a security. Normally, when one trades a spread between two stocks, one would short one stock and take a long position in the other stock. One’s P/L exposure would be purely to the arithmetic spread between them. However, one’s position size remains limited by the nominal prices of the two stocks and the amount of leverage offered by one's broker, typically about two times for a retail trader, and up to 20 times leverage for excellent and low-risk hedge funds like Renaissance. If the spread is only $1, but stock A is $400 and stock B is $399, then $400 of capital and 2X leverage would allow one to trade only 1 unit of the spread, with the value of the short and long legs ($400 plus $399 nominal value) adding to double one's capital, at about $800.

Brim, however, calculates the arithmetic spread as a time series. His simulation algorithm buys and sells the spread itself as if it were a stock. This means that, in the aforementioned example with stocks at $400 and $399, the spread is only $1, and Brim's algorithm could, with $400, buy 400 units of a security that represents the spread! He thus takes an implied position in (400)(399+400)=$319,600 worth of long and short positions in the two stocks, for a leverage of ($319,600/400)=799 times, far higher than even the roughly 20X leverage afforded to the very best, most leveraged hedge funds. For a typical hedge fund in 2024, "for every $100 of their own capital, the hedge funds had $300 in long and short positions" for a leverage ratio of only 3X (Mandel). As subsequent sections will discuss, the unrealistically extreme implied leverage ratio Brim uses is a contributor to his seemingly high cumulative returns (and also to ours).

One might imagine Brim countering that such extreme effective leverage ratios are possible using options portfolios, where the aggregate value of the portfolio, or even of a single option, can be extremely sensitive to small changes in the price of the underlying securities. However, constructing such a portfolio would require accounting for things like the time decay of the options, the theta, their gamma, and a host of other additional complexities that accompany options trading (relative to that of the underlying), which Brim does not.

Even if one could achieve such extreme leverage, a tiny move against one's position would put one out of business, making subsequent profitable trades impossible, and this “survivorship problem” is something for which Brim also fails to account in his research.

A Few Trades Amongst 78,000 Pairs Fail to Substantiate the Use of Q-Learning for Pairs Trading

Return to Figure R, where Brim presents the results of his pairs trading for various levels of the so-called negative returns multiplier, a parameter of the loss function for the deep Q network. We'll be charitable to Brim and evaluate primarily his most profitable results for negative results multiplier 1.

His total cumulative returns in the far left, most-profitable column are 131.33, suggesting (under the unit interpretation that I earlier proposed) that he multiplied his initial capital by 131.33 across all of these different pairs. I think one, however, should take the initial capital to be implied at 38, since that's the number of pairs over which he sums, and he normalizes the initial capital to 1. This adjustment leads to a more conservative evaluation of his model’s profit, making roughly 4 times his initial capital rather than 131 times.

Now, upon further inspection, one will see that the majority of this profit is from a single pair, the CTWS-AWR pair (Connecticut Water Service and American States Water Co.) with a return of 71.28. 54% of profits are made trading this pair. If one looks at the page of Brim's report (see figure F) where he provides a more detailed plot, one sees that, within this pair, a small number of CTWS/AWR trades are responsible for the majority of its profit. In fact, it's a single trade towards the beginning of the period where one sees most of the pair’s return. If one zooms in on this CTWS-AWR trade, one sees that it occurs at the point where the two prices cross and the spread between them is near zero.

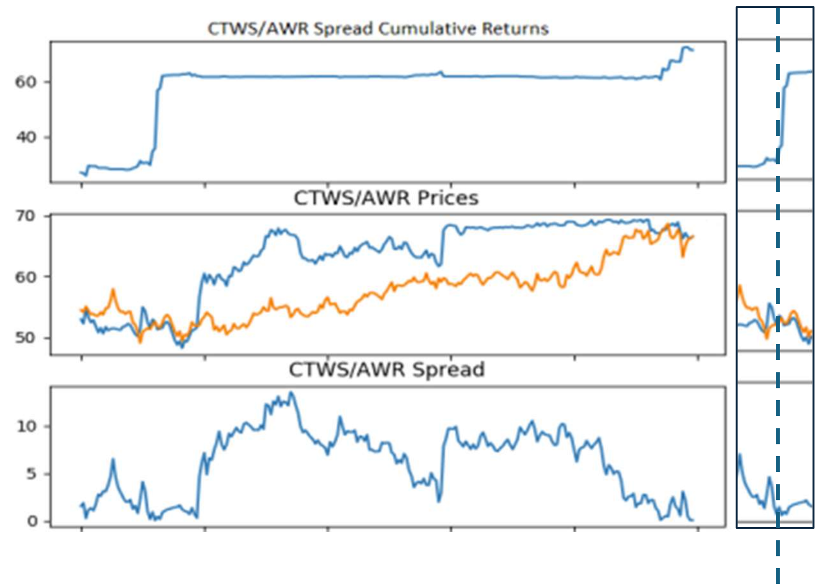


Figure F: Returns from Brim's CTWS/AWR spread with negative returns multiplier 1, with inset at right (dashed line added) showing the key trade being made where the spread approaches zero.

Recall the aforementioned issue where Brim treats the spread as if it were a security that he could buy or sell directly. When the spread approaches zero, Brim's algorithm can buy a large amount of it at pennies and then, when the spread rises in accordance with his algorithm's prediction, the factor by which the trade capital increases can be several fold, thus allowing a single large trade in the CTWS-AWR spread to drive his entire return across all 38 pairs. The percentwise volatility in the spread approaches infinity as the spread goes to zero.

Extend the analysis to the top 3 pairs (for negative returns multiplier 1). In addition to CTWS/AWR (cumulative return of 71.28), FCX/HBI (25.67) and HBI/MRO (27.41) have outsize returns due to the same "near-zero spread hacking." Together, these three comprise 94% of the cumulative profit of 131.33 for Brim's best model.

The evaluation is even more severe if one looks at other columns where Brim posted large profit. For example, for negative returns multiplier 5, the single CTWS-AWR spread has a return of 87, of a total return across all tickers on 111, comprising 78% of his returns (up from 54% in column 1).

Q-Learning: Conclusion

With a good trading model, many profitable trades will be made. And by the law of large numbers, the strategy is likely to be generalizable because it's unlikely that chance was the reason for the strategy's overall returns. It’s hard to “get lucky” for a large number of trades. However, when the strategy’s overall returns across all instruments are so heavily loaded on a single trade, as Brim’s are, the law of large numbers does not apply. It's highly likely that the returns of the strategy overall were merely due to chance.

This raises issues about multiple testing bias, where the researcher can “roll the die until it comes up six” by re-running the Q-learning model, which is initialized each time with small random weights that affect its evolution during training. The slight differences in the weights on which the model converges, due to different initial conditions or different stopping times, etc., can have a profound impact on the profitability of the single trade that makes the results. Results become highly sensitive to the parameters. One can have no confidence from Andrew Brim's results that it is possible to successfully perform pairs trading with deep Q-learning for the 38 stock pairs that he selected, let alone with the broader universe of 78,000 possible stock pairs.

Comparing our Results and Brim's: Conclusion

Pairs trading may have been in use before this time, but it was made famous by Gerry Bamberger and Nunzio Tartaglia, who applied it in Morgan Stanley's quantitative trading group in the 1980s ("A History of Quantitative Trading…"). The practice of taking opposing long and short positions in correlated stocks had become mainstream enough to make its way into academe (and therefore into public access for all trading firms) by the time of Gatev, Goetzmann, and Rouwenhorst's 2006 paper "Pairs Trading: Performance of a Relative-Value Arbitrage Rule."

All told, the alpha decay for pairs trading in U.S. equities has had a 40 year history to run its course, during which sophisticated trading firms have attempted to optimize it using all varieties of predictive algorithms. We find it quite plausible that the strategy's alpha had been priced out by the time of Brim's paper, explaining his and our poor profit. We would encourage future researchers either to investigate strategies beyond pairs trading, or, if pairs trading is an interest of theirs, to seek to apply it in less-liquid, less-developed, or otherwise more difficult-to-trade markets where algorithmic trading firms would have been less able to "price out" the deviations in correlated assets' prices over the past four decades.

Bibliography

"A History of Quantitative Trading: The First Pairs Trade." *Financial Frontier*, www.financialfrontier.com/history-of-quant-trading. Accessed 4 Oct. 2024. [1]

Brim, Andrew, "Deep Reinforcement Learning Pairs Trading" (2019). All Graduate Plan B and other Reports. 1425. https://digitalcommons.usu.edu/gradreports/1425

Gatev, Evan, William N. Goetzmann, and K. Geert Rouwenhorst. "Pairs Trading: Performance of a Relative-Value Arbitrage Rule." *The Review of Financial Studies*, vol. 19, no. 3, 2006, pp. 797-827. Oxford University Press, doi:10.1093/rfs/hhj020. Accessed 4 Oct. 2024.

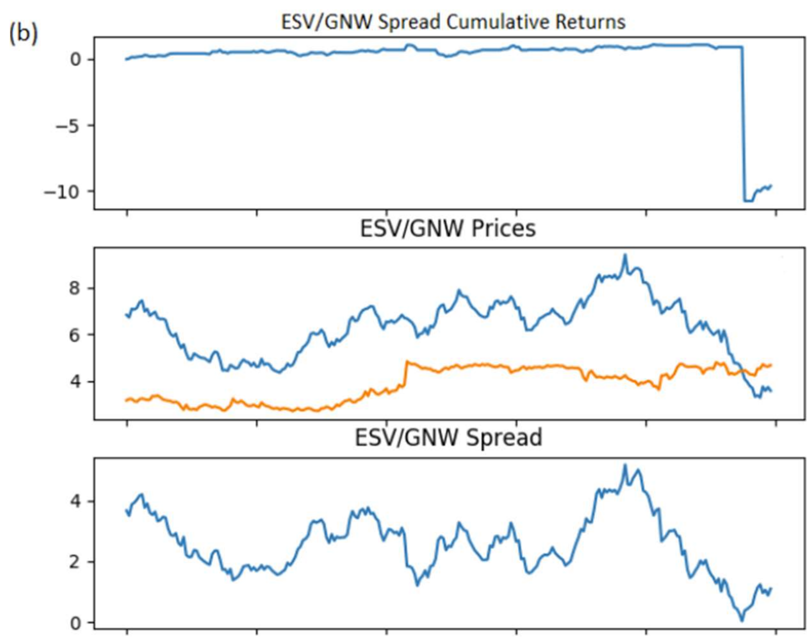
Mandel, Carolina. "US Hedge Flow: Hedge Funds Ramp up Leverage to Near Record Highs to Juice Returns." *Reuters*, 12 Mar. 2024, www.reuters.com/markets/us/hedge-flow-hedge-funds-ramp-up-leverage-near-record-highs-juice-returns-2024-03-12/. Accessed 4 Oct. 2024. [2]

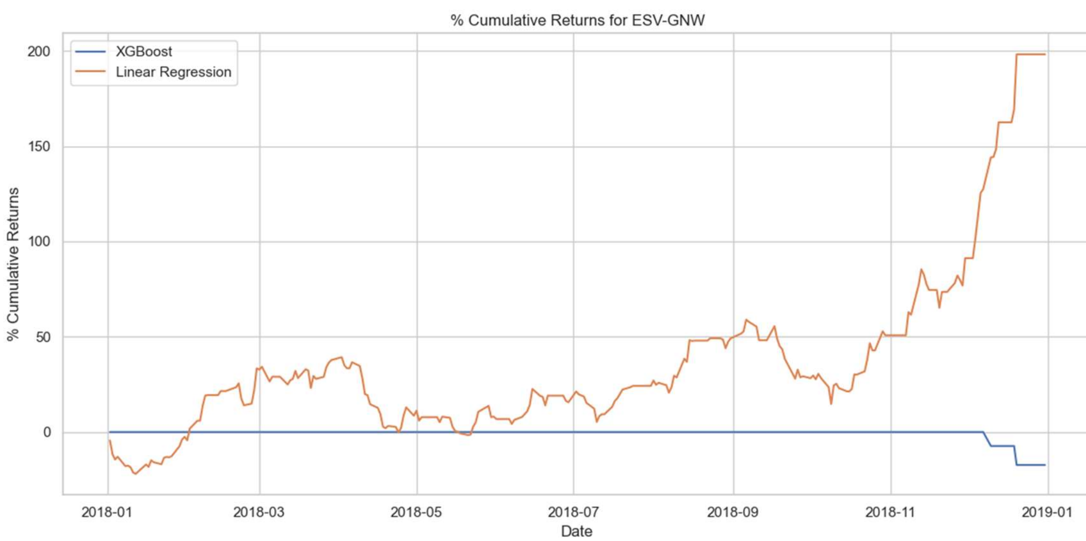
Notes:

1: This source refers to Bamberger and Tartaglia’s use of pairs trading in the 1980s

2: This source provides information about the typical leverage ratios of hedge funds.

Appendix: Example Chart for Our Results





Appendix Images: Brim's results (above) and our cumulative returns chart (below) for the ESV-GNW pair. One can observe that our XGBoost model didn't often make returns predictions of high enough magnitude to trade, and our linear regression model outperformed Brim's for this particular pair.