Back to Community

How to Build a Pairs Trading Strategy on Quantopian?

Delaney Granizo-Mackenzie edited Jan 23, 2016

(pairs trading)   (Show Q)   (stat arb)   (tools and tips)

Pairs trading is a form of mean reversion that has a distinct advantage of always being hedged against market movements. It is generally a high alpha strategy when backed up by some rigorous statistics. This notebook runs through the following concepts

- What is cointegration?
- How to test for cointegration?
- What is pairs trading?
- How to find cointegrated pairs?
- How to generate a tradeable signal?

The notebook is intended to be an introduction to the concept, and whereas this notebook only features one pair, you would probably want your algorithm to consider many pairs at once.

Please find all the lectures here: https://www.quantopian.com/lectures

The notebook was originally created for a presentation at Harvard's Applied CS department and has since been used at Stanford, Cornell, and several other venues. If you're interested in learning more about how Quantopian is being used as a teaching tool at top universities, please contact me at delaney@quantopian.com

Thanks,
Delaney

View Notebook

DISCLAIMER

various reasons, including changes in market conditions or economic circumstances. All investments involve risk, including loss of principal. You should consult with an investment professional before making any investment decisions.

73 responses

Delaney Granizo-Mackenzie May 15, 2015

Here's a very simple algorithm based on the approach presented in the notebook.

Performance   Source Code   Migrated Source Code

```
# Backtest ID: 555610b0b284be0c540e4866
1   import numpy as np
2
3   def initialize(context):
4       context.stock1 = symbol('ABGB')
5       context.stock2 = symbol('FSLR')
6
7       context.threshold = 1
8       context.in_high = False
9       context.in_low = False
10
11  def handle_data(context, data):
12
13      s1 = context.stock1
14      s2 = context.stock2
15
16      p60 = history(bar_count=60, frequency='1d', field='price')
17
18      p5 = p60.iloc[-5:]
19
20      # Get the 60 day mavg
21      m60 = np.mean(p60[s1] - p60[s2])
22      # Get the std of the last 60 days
23      std60 = np.std(p60[s1] - p60[s2])
24
25
26      # Current diff = 5 day mavg
27      m5 = np.mean(p5[s1] - p5[s2])
28
29      # Compute z-score
30      if std60 > 0:
31          zscore = (m5 - m60)/std60
32      else:
33          zscore = 0
34
35      if zscore > context.threshold and not context.in_high:
36          order_target_percent(s1, -0.5) # short top
37          order_target_percent(s2, 0.5) # long bottom
38          context.in_high = True
39          context.in_low = False
40      elif zscore < -context.threshold and not context.in_low:
41          order_target_percent(s1, 0.5) # long top
42          order_target_percent(s2, -0.5) # short bottom
43          context.in_high = False
44          context.in_low = True
45      elif abs(zscore) < 1:
46          order_target_percent(s1, 0)
47          order_target_percent(s2, 0)
48          context.in_high = False
49          context.in_low = False
50
51      record('zscore', zscore)
52
53
```

DISCLAIMER

Delaney Granizo-Mackenzie May 15, 2015

Here's a more sophisticated algorithm written by Ernie Chan. This algorithm computes a hedge ratio rather than just holding equal amounts of each security.

Performance   Source Code   Migrated Source Code

```
# Backtest ID: 555610a7993ef50c61849288
1   import numpy as np
2   import statsmodels.api as sm
3   import pandas as pd
4   from zipline.utils import tradingcalendar
5   import pytz
6
7
8   def initialize(context):
9       # Quantopian backtester specific variables
10      set_slippage(slippage.FixedSlippage(spread=0))
11      set_commission(commission.PerTrade(cost=1))
12      set_symbol_lookup_date('2014-01-01')
13      context.Y = symbol('ABGB')
14      context.X = symbol('FSLR')
15      # set_benchmark(context.y)
16
17
18      # strategy specific variables
19      context.lookback = 20 # used for regression
20      context.z_window = 20 # used for zscore calculation, must be <= lookback
21
22      context.useHRlag = True
23      context.HRlag = 2
24
25      context.spread = np.array([])
26      context.hedgeRatioTS = np.array([])
27      context.inLong = False
28      context.inShort = False
29      context.entryZ = 1.0
30      context.exitZ = 0.0
31
32      if not context.useHRlag:
33          # a lag of 1 means no-lag, this is used for np.array[-1] indexing
34          context.HRlag = 1
35
36  # Will be called on every trade event for the securities you specify.
37  def handle_data(context, data):
38
39      _Y_value = context.portfolio.positions[context.Y].amount * context.portfolio.positions[context.Y].last_sale_price
40      _X_value = context.portfolio.positions[context.X].amount * context.portfolio.positions[context.X].last_sale_price
41      _leverage = (abs(_Y_value) + abs(_X_value)) / context.portfolio.portfolio_value
42      record(
43              X_value = _X_value ,
44              Y_value = _Y_value ,
45              leverage = _leverage
46      )
47
48      if get_open_orders():
49          return
50
51      now = get_datetime()
52      exchange_time = now.astimezone(pytz.timezone('US/Eastern'))
53
54      if not (exchange_time.hour == 15 and exchange_time.minute == 30):
55          return
56
57      prices = history(35, '1d', 'price').iloc[-context.lookback::]
58
59      Y = prices[context.Y]
60      X = prices[context.X]
61
62      try:
63          hedge = hedge_ratio(Y, X, add_const=True)
64      except ValueError as e:
65          log.debug(e)
66          return
67
68      context.hedgeRatioTS = np.append(context.hedgeRatioTS, hedge)
69      # Calculate the current day's spread and add it to the running tally
70      if context.hedgeRatioTS.size < context.HRlag:
71          return
72      # Grab the previous day's hedgeRatio
73      hedge = context.hedgeRatioTS[-context.HRlag]
74      context.spread = np.append(context.spread, Y[-1] - hedge * X[-1])
```

```
 75
 76        if context.spread.size > context.z_window:
 77            # Keep only the z-score lookback period
 78            spreads = context.spread[-context.z_window:]
 79
 80            zscore = (spreads[-1] - spreads.mean()) / spreads.std()
 81
 82            if context.inShort and zscore < 0.0:
 83                order_target(context.Y, 0)
 84                order_target(context.X, 0)
 85                context.inShort = False
 86                context.inLong = False
 87                record(X_pct=0, Y_pct=0)
 88                return
 89
 90            if context.inLong and zscore > 0.0:
 91                order_target(context.Y, 0)
 92                order_target(context.X, 0)
 93                context.inShort = False
 94                context.inLong = False
 95                record(X_pct=0, Y_pct=0)
 96                return
 97
 98            if zscore < -1.0 and (not context.inLong):
 99                # Only trade if NOT already in a trade
100                y_target_shares = 1
101                X_target_shares = -hedge
102                context.inLong = True
103                context.inShort = False
104
105                (y_target_pct, x_target_pct) = computeHoldingsPct( y_target_shares,X_target_shares, Y[-1], X[-1] )
106                order_target_percent(context.Y, y_target_pct)
107                order_target_percent(context.X, x_target_pct)
108                record(Y_pct=y_target_pct, X_pct=x_target_pct)
109                return
110
111            if zscore > 1.0 and (not context.inShort):
112                # Only trade if NOT already in a trade
113                y_target_shares = -1
114                X_target_shares = hedge
115                context.inShort = True
116                context.inLong = False
117
118                (y_target_pct, x_target_pct) = computeHoldingsPct( y_target_shares, X_target_shares, Y[-1], X[-1] )
119                order_target_percent(context.Y, y_target_pct)
120                order_target_percent(context.X, x_target_pct)
121                record(Y_pct=y_target_pct, X_pct=x_target_pct)
122
123
124    def is_market_close(dt):
125        ref = tradingcalendar.canonicalize_datetime(dt)
126        return dt == tradingcalendar.open_and_closes.T[ref]['market_close']
127
128    def hedge_ratio(Y, X, add_const=True):
129        if add_const:
130            X = sm.add_constant(X)
131            model = sm.OLS(Y, X).fit()
132            return model.params[1]
133        model = sm.OLS(Y, X).fit()
134        return model.params.values
135
136    def computeHoldingsPct(yShares, xShares, yPrice, xPrice):
137        yDol = yShares * yPrice
138        xDol = xShares * xPrice
139        notionalDol =  abs(yDol) + abs(xDol)
140        y_target_pct = yDol / notionalDol
141        x_target_pct = xDol / notionalDol
142        return (y_target_pct, x_target_pct)
143
144
145
```

DISCLAIMER

Bharath Rao May 16, 2015

Very useful stuff.

What makes it lose systematically for nearly 3 months? Does Cointegration fail in that period?

Delaney Granizo-Mackenzie May 16, 2015

Basically yes, they turned out not to be cointegrated in that time frame, but returned to being conitegrated in the long term.

I think the drawdown you point out is a strong case for why you would actually want many pairs trading at the same time. Pairs can be cointegrated over different time scales, and any given one will not always be in a tradable state (big spread, small spread). By increasing your sample size, you can make it far more likely that at least one pair will be strongly tradable state at a given time, and smooth out the weird bumps you see here.

DISCLAIMER

Anthony Ng May 17, 2015

Hi Delaney.
Thanks for this. Very useful indeed. I noticed you used Augmented-Dickey Fuller test for the cointegration test. Do you have similar implementation using Johansen test? I'm not able to find the johansen test with python.

Delaney Granizo-Mackenzie May 18, 2015

Hi Anthony,

It appears that whereas there have been some attempts to add the Johansen test to the statsmodels library, currently there is no built-in implementation. Here, for instance, is a 3rd party implementation. I'm not sure when it will get added to the Python libraries, is there a way you can work around not having it?

DISCLAIMER

Anthony Ng May 19, 2015

Thanks. I did see that link. Pretty complicated to implement and to write it all in the IDE. In fact, Satya B attempted it here
https://www.quantopian.com/posts/trading-baskets-co-integrated-with-spy.
The beauty of Johansen test is that it generates eigenvectors, which I think you can use other methods to calculate though I can't recall at the moment, for up to 12 assets and many other things, which can be used to create a basket. I was looking at one of the index arb strategy of Ernie and attempting to replicate it on Q's platform to assess the performance after fees/comm etc.I noticed fees seemed to chew up a lot of the performance. The ABGB & FSLR pair above has an sharpe ratio of 0.75 but ended with sharpe ratio of -0.29. A lot of seeming profitable pairs turned out to be non profitable after bid/ask spread, fees, commission etc. Hence, I am looking at 3 or more stocks pair trading, and index arb. johansen test will make this easier to implement.
I shall keep trying.

Aaron Brown May 19, 2015

The notebook is an excellent statistical introduction to pairs trading, I recommend anyone interested in the topic also look into some of the financial research. Anatomy of Pairs Trading is a good start, and the references are helpful as well. Two more general papers on risk arbitrage strategies are Characteristics of Risk and Return in Risk Arbitrage and Limited Arbitrage in Equity Markets . There are some expensive lessons people have learned about running these kinds of strategies, and it's worth knowing the lessons in advance. Forewarned is forearmed.

Thomas Wiecki May 19, 2015

Anthony, good to see you here! I have been looking for a good implementation of the Johansen test for a while but couldn't find one. There is a pretty long (but stale) discussion and pull request on github about including it in statsmodels: https://github.com/statsmodels/statsmodels/issues/448 and https://github.com/josef-pkt/statsmodels/commit/bf79e8ecb12d946f1113213692db6dac5df2b6e9 It's really too bad as definitely in quant finance this is pretty widely used.

Anthony Ng May 20, 2015

@Aaron. Thank you for the heads up. Appreciate it coming from your. I shall spend some time with those papers.

@Thomas. Thanks for the link. As you said, it is a bit old. Better than naught I suppose.

Pravin Bezwada May 20, 2015

Here is a python implementation for vector error correction models. You can also use it to find cointegration weights. http://econ.schreiberlin.de/software/vecmclass.py

---

Delaney Granizo-Mackenzie May 28, 2015

Here is a version of Ernie Chan's algorithm modified to trade multiple pairs. This is a good way to obtain multiple uncorrelated return streams and reduce the beta of the overall strategy.
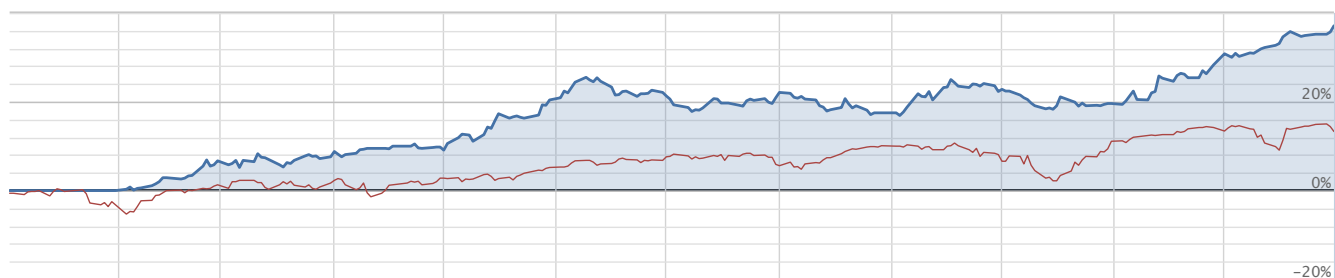
| Performance | Source Code | Migrated Source Code |

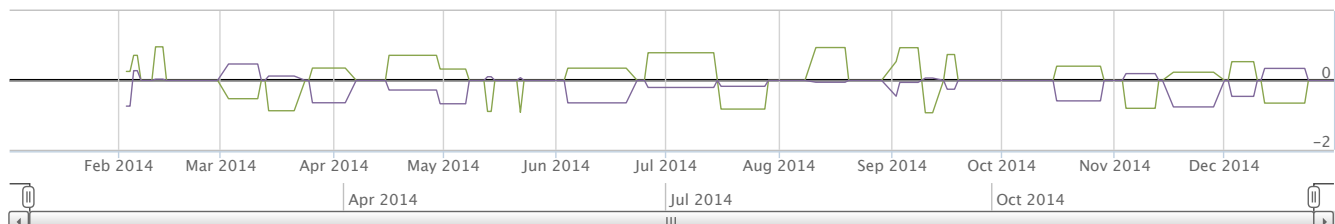Backtest from **2014-01-01** to **2015-01-01** with **$100,000** initial capital

Algorithm   **Benchmark (SPY)**

good trader m.A.A.d Banks May 28, 2015

@Delany, Are there methods available to screen for pairs using stat tests? Or are those usually too computationally expensive?

Delaney Granizo-Mackenzie May 28, 2015

We are working on a way to make the notebooks clone-able into one's own research environment. In the meantime those interested in playing around with the notebook from the original post can download it here. After downloading upload it into your research account. If you do not yet have a research account, enter an algorithm into the contest to receive access.

@good trader, The method provided in the notebook will screen a given list of securities for cointegration, the underlying condition necessary for pairs trading. The problem is not as much the computational complexity as it is the loss of statistical power. The more comparisons you do, the less weight you must put on significant p-values. This phenomenon is described here. To be statistically rigorous, you must apply a Bonferroni correction to p-values obtained from a pairwise cointegration script. The reason being that the more p-values you generate, the more likely you are to encounter significant p-values which are spurious and do not reflect actual cointegration behavior in the underlying securities. Since the number of comparisons done when looking for pairwise cointegration in n securities grows at a rate of $O(n^2)$, even looking at 20 securities would render most statistical tests useless. A better approach is to come up with a small set of candidate securities using analysis of the underlying economic links. A small number of statistical tests can then be done to determine which, if any, pairs are cointegrated. Let me know if this is what you meant.

DISCLAIMER

Aaron Brown May 29, 2015

I disagree somewhat about the problem with too many comparisons. The Bonferroni correction is appropriate when you are looking for truth. For example, if you have a questionnaire with 1,000 items and you give it to people with and without cancer, you'll find on average 50 items that correlate with cancer at the 5% level of statistical significance, even if nothing on the questionnaire is related to cancer. If you consider combinations of two or more items, you can generate as many correlates are you like.

But when designing automated trading strategies, coincidental relations don't hurt you much. They add random noise and trading costs to your results. Since few results are 100% meaningless, most relations have at least some small degree of persistence, it's not critical to filter your strategy down to rigorously validated ones. Profits matter, not truth. Bonferroni and similar metrics push you to the most statistically reliable relations, which are not generally the most economically useful ones.

If by "analysis of the underlying economic links" you mean starting with natural pairs like two similar companies in the same industry, I have not found that useful. Basically people notice the obvious stuff. If you mean thinking about less obvious relations, especially things that are invisible in the usual data people use, then I agree. Ideally you want a validatable economic story for the pair relation, which explains both why it exists and why it is not arbitraged away. Not only does that guard against data mining, but it means you can measure whether the effect continues to work (without that, the only way you know the strategy isn't working is when you lose money).

Grant Kiehne May 30, 2015

Hi Delaney,

Nice work. I haven't read through your notebook line-by-line, but I can tell that it will be a great addition to the Quantopian example library. And following up with shared algos--good move.

You might have a look at the notebook I posted, https://www.quantopian.com/posts/analysis-of-minute-bar-trading-volumes-of-the-etfs-spy-and-sh. To visualize how a given pair goes in and out of cointegration, you could make a similar plot. Applying the statistical test 390 times per trading day over many years would require some patience, though.

Grant

Delaney Granizo-Mackenzie Jun 2, 2015

@Aaron Am I correct in reading your argument generally as follows?
- In the real world Bonferroni is too restrictive and the number of profitable pairs you lose via the correction outweighs the statistical certainty you gain.
I think we agree as to the final point you make. I think that many of the economic link analysis folks do are simplistic and ignore the potentially interesting relations that are more likely to contain non-arbitraged alpha.

@Grant Thank you. We're actually planning to expand the example library to a full quant finance curriculum taught with notebooks and companion algorithms. We're going to have a series of summer lectures as we develop more topics, so keep an eye out for those. Your notebook is very cool and I do wonder how stable the cointegration scores are even for strongly cointegrated pairs. Unfortunately, I don't think I'll have time to look into that in the near future what with the production of our other curriculum notebooks. We are looking for guest contributors, however. If you have any notebooks you would like to be featured in our curriculum with full credit to the author(s), send them my way and I'll see if they would fit into our current content.

DISCLAIMER

Aaron Brown Jun 2, 2015

> In the real world Bonferroni is too restrictive and the number of profitable pairs you lose via the correction outweighs the statistical certainty you gain.

Not precisely. Yes, Bonferroni is too restrictive in the sense that it gives you too few pairs, but Bonferroni also directs you to the wrong pairs.

In the example of a questionnaire with 1,000 items given to cancer patients and non-cancer patients, it's likely that most of the items have no effect on cancer, or at least such weak and complex effects that it's not worth using them for medical advice. So if you want 5% significance, you test each item at the 0.005% level (that is you want 3.9 standard deviations, not just 1.6). You don't mind that, because any real effect strong enough to matter will likely show up with strong significance. If you didn't do Bonferroni, you'd end up with 50 recommendations even when none of the items mattered, and a lot of useless advice.

Incidentally, Bonferroni is a very conservative correction, and there are more sophisticated ones that allow more items.

But if you have 1,000 pairs to test, it's likely that many of them have some degree of cointegral predictability. Even if there is no predictability, including the extra pair only adds a little noise to your strategy, which is not terrible. Also you don't believe that any of them have predictability so strong that anyone would have noticed it and arbitraged it away. So it's reasonable to consider all the pairs with 5% significance or less, and filter them out using economic or other criteria unrelated to the data. Selecting only the strongest statistical relations is not wise.

You can set this up in a Bayesian framework if you like consistency and precision; or you can just use ad hoc rules of thumb.

Peter Bakker Jun 3, 2015

Just for the il-pair-literated who want to learn.... must there be a story behind the pair? Should there be a logical explanation? I played around with pairs and found for example that MorganStanley and Expedia work.... but why? Or doesn't one want to know why.......

Aaron Brown Jun 3, 2015

> must there be a story behind the pair?

This is actually a semantic question rather than a financial one. If you adopted a pure statistical approach with no consideration of the actual pairs, you would end up with hundreds or thousands of pairs, including some overlapping ones. Then we wouldn't call it a pairs-trading strategy but a long-short equity strategy.

The idea of pairs trading is you can get additional insight by considering specific reasons for the dependence between the stocks; and that insight can result in more accurate positioning, and also avoidance of big losses when the relation breaks.

Obvious relations, like two large-cap stocks in the same industry, tend not to be useful. That's confusing sometimes, because some of the famous early pairs trades involved such pairs, and they're still used for examples in most texts. But too many people are watching those spreads too closely to get the high Sharpe ratios you need for undiversified strategies like pairs trading. Leave those marginal Sharpes to the long-short equity people who have a lot more positions.

Also, when we talk about a reason for the pairs relation, we're talking about both a positive--why is it hard to imagine a world in which the values of these companies diverge from their historical proportions--and a negative--why do these stocks respond to different economic news? So for two near-identical companies the first question is easy, but the second is hard. For two seemingly unrelated companies like MS and EXPE it's the reverse. You might say something like, "In a good economy Morgan Stanley gets a lot of business and people travel a lot," but that's basically true of almost any two companies.

The classic pairs reason was two companies that responded to the same basic economic factors, say oil prices or interest rates or US dollar strength, but at different points in the supply chain, say crude oil prices versus gas station revenues. A single link is not good enough, virtually all companies respond to these factors. But you can find pairs that are matched on narrower factors, say fracking activity in the Northeast US or precipitation in central California, or that match direction on a number of broad factors. Or you can find two companies that are actually in similar businesses today, but that for historical reasons are listed in different sectors. Another common situation is two companies involved at different points of the lifecycle of durable assets; homebuilders and furniture stores with similar geography for example.

Anyway, when you have a reason, you have things to monitor to fine-tune your position; and to alert you if a big dislocation is a great trading opportunity or a sign than the historical relation has broken. If you don't have a reason, you'd better have a lot of diversification, meaning you can't afford the specific analysis work for each pair.

Market Tech Jun 3, 2015

Wouldn't you admit though that if a pair has a story then that story is known and therefore unprofitable by the likes of slow to trade retail traders? And if one could mine the data and discover, through the data, stories that were unexpected that one could at least compete in the pairs trading space? I see your point on maintaining a large pool of pairs if the stories that connect the participants are weak or unexplored, but still, if we underlings wish to participate why wouldn't we use such a technique? Or do you maintain that retail traders can capture and profit from anomalous pair spreads of well known couples?

Aaron Brown Jun 3, 2015

> Wouldn't you admit though that if a pair has a story then that story is known and therefore unprofitable by the likes of slow to trade retail traders?

No, I wouldn't agree with that view. Pairs trading tends to be low capacity, especially in lower-cap stocks, and takes a lot of work. It's not attractive for asset managers because the investment amounts and risk characteristics are erratic. It's mostly pursued by individual full-time professional traders, who might follow a dozen pairs in addition to a few dozen other strategies, and semi-pro traders who are willing to take what the market gives them and stay in cash when none of their strategies are attractive. There are more good pairs than there are competent traders chasing them.

In principle, you could find good pairs using a clever automated filter, or by reading and thinking. My general feeling is the first is harder, and if you're going to do it, you'll want to do it to identify large numbers of pretty good pairs rather than two or three great pairs. In that case, I'd say just switch to long-short equity and forget pairs. The good thing about reading and thinking is most good quants are lazy, and would rather let the computer do the work. So you're competing with non-quants, some of whom are pretty good at reading and thinking, but are at a huge disadvantage to someone with a computer who knows a little math.

I don't want to come across as dogmatic, anyone who does what other people tell them is not likely to find great success in any sort of trading. If you think you can design an algorithm to identify good pairs, there's no harm in trying. It just doesn't strike me as the most promising approach.

Market Tech Jun 3, 2015

> … takes a lot of work.

Yeah. The easy pairs trade money was made long ago. Lucrative stories in lower-cap stocks though exposes a pair to the aberrations of smaller company volatility no? "Whoops, that solar stock just lost its major contract. Or, wow, that driller just got a windfall state contract." And then the story gets rewritten, or thee or four pages get torn out. One might catch such preludes to story changes if one only watches a dozen or so stories. But here, where we're looking to avoid story watching -- going fully automated, we would get nailed by such narrative breakdowns in just a few pair relationships.

When you say switch to long/short equities you would seem to advocate abandoning the statistical search for obscure (perhaps whimsical) stories in lieu of broader mean reversion -- is this true? But, if one has the tools, why not create dozens and dozens of strange storied pair trades. Sure the stories may not actually exist. But then again, maybe you discover 10 or 20 that are unique. And through a process of eliminating the poorly paired partners, you end up with a manageable set that are capable of dancing with the stars? This site is nothing if not a massive experiment in data mining no?

Aaron Brown Jun 3, 2015

Again, I'm not trying to law down laws here, but the two straightforward approaches are (a) try to find a few pairs you can understand or (b) forget about pairs and just try to build a large portfolio of longs and shorts without worrying about pairing up stocks or doing unautomated research. In other words (a) niche clever research or (b) massive data mining.

Trying to split the difference by finding dozens of pairs but not doing the tailored research necessary to understand each one seems suboptimal.

Grant Kiehne Jun 4, 2015

@ Aaron,

> try to find a few pairs you can understand

If I'm reading things correctly, by "understand" you mean that there should be some underlying intuitive story behind the relationship, I suppose so that there is less risk that the relationship will suddenly disappear? Are you talking about a kind of narrative, "The reason we think this is happening, but can't really explain with a model, is..." or an explanatory quantitative model that provides the story behind the relationship? Say I find a pairs trade based on the idea that when consumers buy lots of eggs, bacon sales drop off, and vice versa. I could make up a story that people can only eat so much for breakfast, and leave it at that. I have a warm, fuzzy feeling, and if I'm a professional trader, hopefully my management will feel warm and fuzzy, too. But is the risk really any different without the story? Unless I actually find a relevant study on breakfast eating, or conduct one myself, then I could just be deluded. And if the underlying cause can't be coded into a set of rules, then it is not really automated quantitative trading, right? As a Quantopian user who doesn't do this sort of thing for a living, I need to get an algo in the Quantopian hedge fund, let it run, and collect a check. No time for doing lots of offline analyses.

Also,

> There are more good pairs than there are competent traders chasing them

sounds like the land of milk and honey for us inhabitants of Quantopia. This would say that the Quantopian team should think about churning out candidate pairs for their 35,000+ users to examine like a bunch of ants, trying to come up with stories for a subset of them ("I'll take XYZ & PDQ, do some research, and see if I can find a 'story' to support the relationship.").

I'm just trying to sort out if any of this can be reduced to practice for Joe Schmo Quantopian user, or if it is a hopeless endeavor. Is there a path for Quantopian to get hundreds of lucrative, scalable pairs trading algos for their $10B hedge fund (keep in mind that by my estimation, they need several thousand distinct algos in the fund)? Or is this all a bunch of blah, blah, blah?

Simon Thornington Jun 4, 2015

I've tried the automated searching of pairs/baskets, using the public knowledge techniques, and though I haven't gone through them all with my tick-level back-tester, the few that I did examine personally were largely worthless; the supposed spread mean-reversion that my grid search turned up was just spurious or due to bid-ask bounce.

However, I do know for a fact that people run decently profitable automated pairs trading portfolios. I take that to mean that it is possible, but the way that I approached it was naive. Perhaps the legwork method is the way to go, coming up with theses about drivers and *then* looking for portfolios that would express the theses, with the actual hedge ratio construction done "rigorously" using Kalman filters or whatever.

Grant Kiehne Jun 5, 2015

My take is that chatting about pairs trading is wonderful, but there should be a focus on reducing it to practice, with some sort of approachable workflow, so that a Quantopian user can sit down in his pajamas with a cup of coffee on a rainy day and actually come up with a halfway decent algo that would have a shot at getting into the crowd-sourced Q fund. For example, we have:

> ...try to find a few pairs you can understand...
> Perhaps the legwork method is the way to go, coming up with theses about drivers...

O.K. So what's the workflow for your typical Q user? Keep in mind, this needs to be scalable...it won't do Q any good if only users with an advanced degree and 20 years of industry experience can be successful. If the answer is, "Well, there is no workflow...you just need to know" then pairs trading won't be approachable on Q. We have Aaron's "reading and thinking" recommendation above, but read what?

Also, I'd seen somewhere that there are techniques for synthesizing trading pairs, from baskets of securities. Does this work? Or does one effectively end up with the long-short equity portfolio referred to by Aaron Brown above?

Aaron Brown Jun 5, 2015

The kind of warm-and-fuzzy story you mention is worthless for investing, although as you say it can reassure investors and regulators. What you're looking for is covariates to refine your strategy and, most important, warn you when it's not going to work. The quant trap is that when your relation breaks it simply looks more attractive to your model, and you spiral to doom.

The eggs-and-bacon story is actually the reverse of what you want. That says there is a fixed total consumption, so the total amount consumed of both products is fixed, meaning they are negatively cointegrated. If they were positively correlated, say because investors bid up or down all breakfast foods as a group, you would do anti-pairs trading. You're looking for things that have to be in some kind of long-term balance, but move is opposite directions in the short-term. A warm-and-fuzzy story might be residential construction and furniture sales, in the short run if people are saving for down payments they're not buying furniture, and newly house poor families are making due with old furniture and underfurnishing. But in the long run, houses will get furnished. This would never be a pairs trading story because it's relating entire sectors. To exploit this, you'd build a model tracing the full life cycle, and likely involving other factors like interest rates and family demographics and migration patterns, and trade large numbers of stocks.

To keep this practical, here is a Pairs Trading for Dummies recipe (I mean that respectfully, I'm a big fan for For Dummies books).

1. Run some kind of statistical screen to identify promising pairs trading targets. Don't look for extreme statistical significance, just some moderate level to screen out the noise like 5% or 1%. It can help to limit one member of each pair to companies or regions you know something about.
2. Look at the pairs, concentrating on the ones that seem somewhat related but not completely obvious. Don't just ask why they appear cointegrated, also ask why they deviate in the short term. Ultimately you need both forces to be strong for a robust pairs trade. Also, don't just look at times the relation worked (deviation/correction) but at times when it didn't. Most of the time you'll conclude that either the apparent cointegration or apparent deviations were random noise, discrete events not likely to be repeated or unexplainable.
3. Sometimes you'll find a good story. Say both companies manufacture parts that are used in similar products, and the manufacturers of these products like to keep multiple suppliers healthy to have a robust supply chain. So both companies go up and down with the health of the manufacturers they serve, and neither can succeed too much as the expense of the other. But due to location of their facilities, one has a shipping cost advantage during the Great Lakes shipping season, and the other is has the advantage in the winter. A cold winter will result in lost business and inflated inventory for the first company, but it will be made up later; a warm winter will do the reverse.
4. If you do this pairs trade, you'll want to monitor the overall industry for this type of company, plus Great Lakes shipping. As long as the sum of the two companies is moving up and down with the industry, and the deviations are occurring in the expected direction when there are changes in Great Lakes shipping costs or volume, you like the trade. But if the two begin to diverge from the industry, they could both be winning or losing due to some unrelated reason, and the shipping relation may no longer hold. Also if you see deviations increasing without any shipping news, it could be that some other factor is at play, say quality problems at one company or the emergence of a new competitor.
5. Based on your research, you may decide to adjust the standard pairs trading algorithm, perhaps only entering into new trades from November to April, or only at certain levels of Great Lakes shipping costs. These kinds of refinements can make major improvements to pairs trading. You'll also construct an alert that says the deviation is too large relative to your assumed explanation, and you should get out of the strategy until you figure things out. All of this, except the figure things out, can be automated. If you want complete automation, the strategy will have to kill itself whenever unusual things occur (for human pairs traders, these signal times of opportunity for qualitative trading).

Clearly this is for someone who has quant skills, but also general research skills and business judgment.

Grant Kiehne Jun 6, 2015

Thanks Aaron,

Based on:

> Run some kind of statistical screen to identify promising pairs trading targets. Don't look for extreme statistical significance, just some moderate level to screen out the noise like 5% or 1%. It can help to limit one member of each pair to companies or regions you know something about.

it sounds like it could be productive for Quantopian to open-source some efficient tools for the screening (and maybe up their game in terms of computing resources). Let's say I'm an expert on company XYZ and maybe I could narrow down my field of candidate securities for comparison to NASDAQ-listed stocks, of which there are about 3,000. So, it is an O(N) computing problem, not O(N^2) as Delaney mentions above for the general screening problem. But, I'd like to compute the statistics on a rolling basis, every trading minute over 2 years. I'd have:

> (3000 comparisons/minute)(390 minutes/day)(252 days/year)(2 years) = 589,680,000 comparisons

Is something like this at all feasible on the Quantopian research platform? If not, how would I scale it back to something that would actually run in a reasonable amount of time (a few days at most) but still provide useful results?

Grant

---

roman brodetski Nov 3, 2015

Hi Delaney,

I'm playing around with the algorithm by Ernie Chan that you posted.
Surprisingly, it fails entirely when I swap the pair, see the attached backtest (I've only changed the order).

Also, how to treat the negative hedge (beta from OLS)?.. With the current implementation we go long (short) on both positions when the sign of the hedge is the same as the sign of the z-score, which you don't expect from pair trading. What economic reason can lead to such cointegrations?
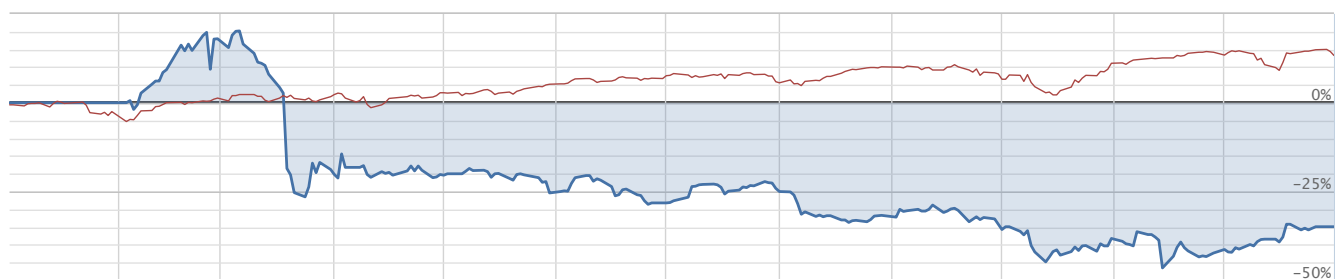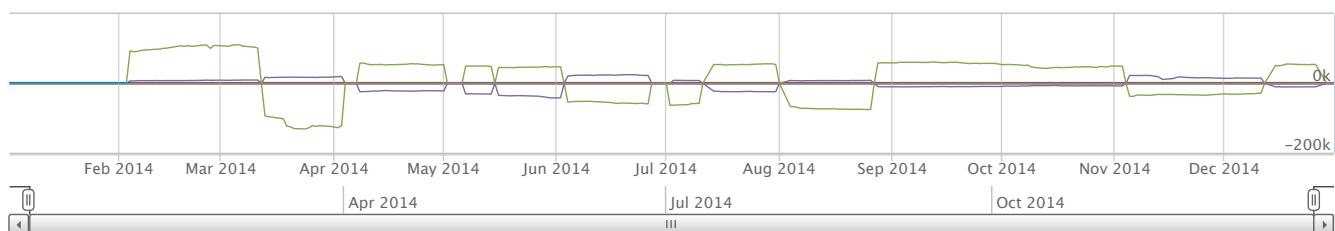
Thanks,
Roman

Performance   Source Code

Backtest from **2014-01-01** to **2015-01-01** with **$100,000** initial capital

---

Delaney Granizo-Mackenzie Nov 3, 2015

Hey Roman,

Not sure exactly why it's failing when you swap the order. Seems like the math may not be robust to an 'upside-down' pair. The hedge ratio comes from the formal definition of cointegration, which is that for some b and $u_t = y_t - b * x_t$, u_t is stationary (the mean stays the same). Therefore we try to estimate the b parameter in each trade so that we can correctly produce a stationary drift between the two securities. It can be the case that the two are negatively cointegrated, whether there's a strong economic reason for this I'm not sure. You might try putting in place restrictions to not trade when you have double long or double short positions, or employing a better estimation method for b (more data points for example).

All of the issues you bring up are very sophisticated improvements, and making these improvements to the algorithm could result in something very good. I don't have cut and dried solutions for you, as you are now dancing around the edge of what is known about algorithmic trading. A lot of it comes down to rigorously testing different signal processing methods to see which yield the best out of sample performance. Also, like you said it's important to let the economic reasoning drive the creation of your model.

Thanks,
Delaney

DISCLAIMER

---

roman brodetski Nov 3, 2015

Thank you for your quick reply.
This is actually a very valuable response, as I was afraid I might have missed something obvious.

Roman

Delaney Granizo-Mackenzie Nov 3, 2015

Happy to help.

DISCLAIMER

---

daniel Snyder Jan 2, 2016

Simon,
Here is a temp website which has similarity of movement information, which is about the same idea as pairs. StockA is the stock you are comparing to, row is how this pair ranks to all pairs, (its row count). It only contains information for the top 5000 or so pairs.
The data is pulled from the period of Aug 2014 to Feb 2015 and is an average of each day.

http://tandemwalk2016.azurewebsites.net/Home/Inverse/IYR
(Change IYR to symbol wanted)

The idea behind the algorithm is not actually for pairs trading, but is for similarity of how a pair moves. I will leave this test site up for a few weeks.

Adam Smith Jan 16, 2016

Thanks Delaney. It's a great starting step for pair trading technique.

I am working on the missing piece of this strategy which is how to use Quantopian Research environment to find statistical cointegration stock/ETF pairs from entire universe or from the same sectors. After I construct good pairs, then I can use the Notebook you provided for further analysis and backtest.

Does anyone have any suggestion for me?
Thanks.

Ciaran Youd Jan 17, 2016

I have a question for those trading pairs

How do you deal with the large processing requirements?

I coded some tests for co-integration and results per combination take roughly 1 second.
I can get this down with parallel processing and by storing data locally but a universe of 2000 stocks will still have 4000000 potential combinations.

Tim Vidmar Jan 17, 2016

Perhaps pointing out the obvious, but ...

A pre-screening tool, or pre-screening done for you for a fee ...

https://www.pairtradinglab.com/

Simon Thornington Jan 17, 2016

When I was researching this sort of thing a couple of years ago, the baskets of 3 and 4 of only a few hundred ETFs took months on my MacBook. And they were all mostly garbage, though I never actually went through them all. I probably should.

Simon Thornington Jan 17, 2016

If I remember correctly, that was 1.6T combinations, or something like that.

---

daniel Snyder Jan 17, 2016

Simon,
The formula is R to the Sterling S, divided by S!
so, for 4000 stocks, it would be
(4000 x3999)/2! or, about 8 million pairs made from the 4000 typical stocks. for 3 stocks considered together, there would be 4000 x 3999 x 3998 /3!
etc
You can prune the possible tree pretty easily though. I believe most stocks behave as if they really were ETFs (at the market neutral way of looking at it only) and can be represented by a group of other stocks, that move with their same fundamentals. You only have to know what sectors they move with, and then check for pairs against this.
So, for example, with HLF, it moves with consumer, several currencies, emerging markets, and a few others. It is hard to separate out exactly as emerging markets also move with currency, so which is which becomes the question.
For two typical tech stocks that appear to be very similar, it may well be the case that their main difference is which currencies they move with. So, for most of the time, they may appear co-integrated, but then, when there is a difference in currencies that affects one a lot, and not so much the other, they then move apart.
I was working on an algorithm to determine the underlying components, (so to speak) that collectively make each stock behave with the same logic as if it was a multi-sector ETF. (where the underlying stocks are a mystery to be solved) I have most of it done, and I believe I have enough done to prove it does work this way, but I lost my real time quote stream a few months ago, and so stopped working on it.
since my algorithm would need to consider up to 15 underlying components to solve this problem, it would be 4000 x 3999 x3998 ... 3985/15! So, I have to trim it. The link I posted a few messages above shows some of the results of this work, where I first determine the possible stocks to consider, for each symbol.
It is my belief that the market is essentially swamped out with pairs trading, and this is why it works so mathematically perfect for each stock to behave as if it is an ETF.

Delaney Granizo-Mackenzie Jan 17, 2016

There is certainly a high computational cost to looking at all possible pairs. However, there is a tradeoff to this approach, as you put yourself at a high risk for multiple comparisons bias. Please see earlier in this thread for a fairly complete discussion of this issue. Regardless of which method you use to select pairs, you'll want to do some additional validation using the notebook and then use the algorithms in this thread to try backtesting a strategy.

DISCLAIMER

Simon Thornington Jan 18, 2016

Indeed, Aaron Brown's advice is gold.

Grant Kiehne Jan 22, 2016

What is "multiple comparisons bias"? I'm lazy and don't feel like sifting through this rather extensive discussion thread.

I find it hard to believe that pairs trading would work as a scalable hedge fund strategy (be able to pour $10's of millions into a single pair). Is there any evidence? In other words, why is Quantopian promoting this?

Simon Thornington Jan 22, 2016

This is one of the best threads on the site.

It scales; you can trade hundreds of pairs.

Delaney Granizo-Mackenzie Jan 22, 2016

Multiple comparisons is a core problem in all of statistics, right up there with overfitting. The general idea is that if you run 100 statistical tests on random data, you should still expect to get 5 below a 5% cutoff and 1 below a 1% cutoff based on random chance. This is true when testing various iterations of a model, or many pairs. Because the number of pairs is $O(n^2)$ you should expect to get a lot of spurious p-values when looking for pairs. A naive strategy of just looping through pairs won't work, you need to be a bit more sophisticated.

https://en.wikipedia.org/wiki/Multiple_comparisons_problem

And yes you trade many pairs with low exposure to each. That said, I think that long-short equity strategies may be a better first bet to get into the fund at this point, just based on robustness and capacity.

DISCLAIMER

---

daniel Snyder Jan 22, 2016

Grant,
There is more electricity used in the state of New Jersey doing calculations on the market than there is electricity used in that state for manufacturing. Pairs strategy likely accounts for at least 50% of this usage as even HFT likely often uses some version of deviation from the mean. It is my opinion that the market is so saturated with pairs trading that given the price of any ten tickers that had no big news, one could deduce the price of the rest of the market and be within 0.7% of the actual price, 90% of the time for the top traded 4000 stocks. (and it could probably be done with less than ten tickers.) So, for a 30 dollar stock, the margin of error would be about a quarter. This is how precisely, compared to each other, I think they move. Until there is news.

Market Tech Jan 22, 2016

It sounds like a corollary to the reciprocal of the law of large numbers; given enough samples you will always find something to fit.

I would reintroduce the concept I proposed in an article in S&C last spring ; the directed acyclic graph or DAG. Using thousands of correlated or cointegrated pairs I built groups from them. Those groups were essentially social graphs of securities. You can search here for DAG, but briefly, you can use the concept of pair trading, that is, fade and favor the divergences, but with a correlated group. And such a group is assembled, dynamically, from a list of pairs that are "friends of friends". It's a pairs strategy, essentially, but with lower risk and less work managing hundreds of separate strategies.

Simon Thornington Jan 22, 2016

> That said, I think that long-short equity strategies may be a better first bet to get into the fund at this point, just based on robustness and capacity.

Have people been coming up with good ones? If so, what proportion are using the new data sets? If not, why not, do you think that is?

I haven't been focusing on them at all, mostly because there's a problem of opportunity cost; if I spend all my time looking for equity long-short algos, not only is there a chance I don't find anything, but if I do, there's still a chance that Quantopian doesn't select it, and since I cannot trade them myself, that time is wasted (unless I pitch it to other funds I suppose). If I look for algos that I personally can trade, and I find some, then I trade them.

I realize there's an unfortunate schism wherein I am using your platform but not contributing to your business model, so if you have any ideas how I can help without wasting my time writing algos that only work high account levels, please let me know. Pairs trading/statistical arbitrage might be one solution, but I've found them very difficult to implement; anything that looks promising in Quantopian fails the backtest when using dividend-adjusted bid-ask tick data, so I might shift my focus back to building my own lower latency infrastructure for a while.

Simon.

Simon Thornington Jan 22, 2016

> I would reintroduce the concept I proposed in an article in S&C last spring ; the directed acyclic graph or DAG. Using thousands of correlated or cointegrated pairs I built groups from them.

This sort of thing? https://tr8dr.wordpress.com/2009/12/30/equity-clusters/ (also http://www.slicematrix.com/)

Market Tech Jan 22, 2016

Cool. Yeah, pretty similar. The DAG though was used specifically to find the networked graph. Those trees might embody the same thing, not sure. But I'd guess the idea is approximate.

Why would anyone want to pairs trade when trading a Minimum Spanning Tree or correlated network graph of stocks is so much safer and easier? I've built dozens of pairs strategies and the directionality of the pair always broke the model. And all pairs I ever tested all went directional at some point -- beyond the account's ability to Martingale down.

Delaney Granizo-Mackenzie Jan 22, 2016

> Have people been coming up with good ones? If so, what proportion are using the new data sets? If not, why not, do you think that is?

I can't release any specific data on this. I can say that there's a lag between when we update product features/try to educate people about algorithm writing techniques (larger universe size, shorting), and when new strategies start appearing. We'd love more large universe strategies right now and I'm trying to figure out ways to make it easier for folks to develop large universe long-short strategies using pipeline.

I haven't been focusing on them at all, mostly because there's a problem of opportunity cost; if I spend all my time looking for equity long-short algos, not only is there a chance I don't find anything, but if I do, there's still a chance that Quantopian doesn't select it, and since I cannot trade them myself, that time is wasted (unless I pitch it to other funds I suppose). If I look for algos that I personally can trade, and I find some, then I trade them.

I realize there's an unfortunate schism wherein I am using your platform but not contributing to your business model, so if you have any ideas how I can help without wasting my time writing algos that only work high account levels, please let me know. Pairs trading/statistical arbitrage might be one solution, but I've found them very difficult to implement; anything that looks promising in Quantopian fails the backtest when using dividend-adjusted bid-ask tick data, so I might shift my focus back to building my own lower latency infrastructure for a while.

Totally reasonable. We don't release our product with the expectation that everybody will use it to develop strategies for the fund, we also want to support your use case of personal trading. We also understand there's a conflict between pushing people to write high capacity market neutral long-short strategies, when those will never work on their own money. What I'm trying to figure out is ways to make the workflow of producing and evaluating factors easier, because once you have a factor-based ranking system, it's pretty easy to slot that into an existing long-short algorithm using pipeline. I'm working on sharing a pipeline algorithm with the community and attaching it to the lectures page in an effort to get more cloning and tweaking going on.

DISCLAIMER

Grant Kiehne Jan 22, 2016

I share Simon's sentiment. I've continued to participate in the contests, but the idea of spending tens (hundreds?) of hours trying to come up with an uber algo that will compete with the big dogs sounds like a lot of work, with a very uncertain pay-off (it's not even clear that you are still working on the hedge fund...any substantive news?). The pipeline thingy has a bit of a learning curve, so I haven't taken that on yet (the fact that lots of obscure modules need to be imported is a red flag). That said, if there were good working examples that could be tweaked, I might give it a go.

Grant Kiehne Jan 23, 2016

What I'm trying to figure out is ways to make the workflow of producing and evaluating factors easier, because once you have a factor-based ranking system, it's pretty easy to slot that into an existing long-short algorithm using pipeline.

Why don't you get all of the Q eggheads together for 1 week and see if you can come up with a long-short algo that would be Q hedge-fundable, and publish it (and better yet, actually fund it). Not only would this provide an existence proof, but you should also gain some insight into the workflow and the person-hours to accomplish the task.

Delaney Granizo-Mackenzie Jan 23, 2016

Here is a pipeline algorithm that I just published as the goto example of a long-short equity strategy. I'm sure it will go through many improvements as the public eye turns to it, but it should at least be a start. It's tricky because we do want to publish algorithms that are 95% of the way done, so that users can take the last 5% and improve the strategies in many different uncorrelated ways. With long-short equity most of the work is in choosing good factors and factor ranking techniques. Unfortunately those are the type of signals that will disappear when shared publicly, but the actual machinery to trade within the algorithm should stay pretty consistent. If you're maybe looking to learn pipeline a bit, I would recommend going through Lectures 17 and 18, then looking at the algorithm.

https://www.quantopian.com/posts/quantopian-lecture-series-long-short-equity-algorithm

I can say for certain we are working on the hedge fund. Even if you have strategies that aren't consistently winning the contest, we may be interested in an algorithm that can consistently do ok. Ultimately, my job as the one overseeing the lectures is to keep trying to make it easier so people don't have to spend as much time working on algorithms that may never pay off for them, and so we get more algorithms that do pay off in the long run.

DISCLAIMER

Adam Smith Jan 29, 2016

I start to implement pair trading backtesting in research environment instead of IDE. The main reason is to automatic run multiple pairs performance analysis before I jump into IDE for full backtest. Another reason for this work is to do further analysis for returns from many pairs.

I am wondering where I can find the example of backtesting in research environment to start with. Any comment is very appreciated.

Thanks.

Delaney Granizo-Mackenzie Feb 1, 2016

Hi Adam,

In your research environment there should be a 'Tutorials and Documentation' folder. Inside the folder should be a notebook with the title 'Tutorial (Advanced) - Backtesting with Zipline'. Make a copy of that and let me know if that's enough to get you started.

Thanks,
Delaney

DISCLAIMER

Blue Feb 2, 2016

May 28 algo falls below benchmark if extended to date and -43% PvR with default slippage and commissions, tanking thru 2015.
Hope it can be rescued b/c it shows good potential.

Delaney Granizo-Mackenzie Feb 3, 2016

The example strategies cheat and run on the same timeframe over which we did research and found the securities to be cointegrated. In a real strategy you'd want to find pairs that were cointegrated into the future and not just historically cointegrated. The template should stay largely the same, so it's an issue of swapping in new securities that you have statistical evidence will stay cointegrated.

DISCLAIMER

Pravin Bezwada Aug 19, 2016

Hi Delaney,

Could you post a tutorial on calibrating an Ornstein Uhlenbeck process for mean reverting series residuals?

Best regards,

Delaney Granizo-Mackenzie Aug 25, 2016

Hey,

We've added a lecture on this to our queue. No idea when we might currently get to it, but it's on there.

Thanks,
Delaney

DISCLAIMER

Market Tech Aug 25, 2016

Ages ago I posted, perhaps as anonymole, that a "pair" needn't be made of only two securities. In fact, the whole "we only allow low beta strats" mantra is pretty much an argument that all strategies should be a variation of a pairs strat. That is, over all, a market neutral position is best.

Taking this further however, and applying a more formal model to the pairs strategy (that the security set have a "story" attached to it) I wonder if the two halves of the pair would do better as independent baskets of securities. That if one approached a pairs strategy with the mind to match up two behaviorally opposed baskets of securities that instead of trying to search all pair combinations looking for all the super-great-marvelous attributes a pair should have, that instead, one determine the two sides of the pair coin and fill each side with the most appropriately identified securities -- for each side.

A simplistic model might be described thusly:
Equities which cycle up in the spring/summer and down in the fall/winter would be bundled together and set against equities which cycle oppositely (down in the summer, up in the winter).

No doubt there are more interesting or undiscovered cycles that exist. My point is that rather than identify securities that yin and yang, one discover technical, or macro, or fundamental classifications which zig when the other zags. Then find securities which fit each of those baskets of behavior.

MT

Delaney Granizo-Mackenzie Aug 25, 2016

This is a very interesting idea and definitely something that professional quants do. At the core we just want two assets on either side of a pair, and a portfolio of assets will do just as well as a single equity. There are probably pros and cons of each method, but the idea of using a basket of things rather than a single thing can greatly reduce your position concentration risk and lead to a better algorithm. I'd say it's worth research. You'd still likely want a few different pairs of baskets as each would smooth out the return curve of the other and produce a lower volatility algorithm.

DISCLAIMER

daniel Snyder Aug 25, 2016

Delaney Granizo-Mackenzie
I have to run an errand, so I only have five minutes, but hopefully I can be clear in that time.
To demonstrate the chops of an AI system, I created an algorithm that can represent the small changes in stocks price, as the sum of a set of ETFs. For example, with MSFT one might have XLK, XLY, FXE, FXI, and some others.
I can show that the typical price movements during a day can be represented in this way. However, when there is specific news, then it is no longer true, if the news is strong.
What I believe this shows is that instead of things "returning to the mean" they are in fact not moving arbitrarily and so, if they return to the mean, it is because one of the underlying components in fact moved. (Of all the underlying components, usually only one or two have news, and the rest are balancing each other out, once the price has adjusted.)
How might one design a trading platform for this as even if you do know it is the sum of other waveforms that are causing one waveform, one still doesn't know what causes them to move until after the fact.
(the reduction in influence is 1/1.6 when looking at the components, so after a couple of feedback loops, the influence is not measurable. Thanks, and sorry for the hurried note,
Daniel

King Azadbakht Aug 26, 2016

Hi everybody
Have you read Algorithmic Trading written by Ernie Chan? For sure you read it, I have a question: in fact I am not good in programming and working with Matlab, I am really interested in Currency cross rate part of the book and I want to implement the positions in live trading but I don't know how to do that in fact I can't understand what the numbers as positions mean! If somebody can guide me I'm really appreciated.

James Christopher Aug 26, 2016

daniel

Not entirely sure I'm understanding your thesis but it seems that you've created an expression that models the returns of a specific stock from it's sector exposures. This is actually a common risk modeling tactic, check out my notebook here. To build a trading strategy off of this I would take your hypothesis about changing news and use that to alter the coefficients of your model. A cool place to start would be to check out the lectures on factor modeling and then maybe look at some news/sentiment data sets to see if you can find any anomalies.

DISCLAIMER

daniel Snyder Aug 26, 2016

James,
That is close. It models the returns to within a few cents usually, at any moment in time, depending on the stock and its volatility as a sum of its sectors. (except when it has specific news.) What I envision behind it is a large set of funds using NLP to invest by sector based on news. Because they are so large, then they tend to swamp out the market during normal times.
I can also show that stock prices changes are directly proportional to the sum of the underlying sectors information, for most time periods. For example, the price changes for three months show this and also for three weeks, which is a bit chaos like, as it would seem they wouldnt be so perfectly in tune. Anyway, with this I can sort stocks by their overall market efficiency (the more efficient you are, the more you sync with the relationship stated above).
I also believe that there are huge funds that are interested in doing nothing more than treading water (as one possible explanation) and they move their money around the world, just trying to stay even, and so the result is that at any given time, the sum of everything stays near zero. (when one thing goes up somewhere, something else somewhere else goes down.)
These relationships also break down during periods of very high volatility such as fall 2015.

daniel Snyder Aug 26, 2016

There are other things I am able to quantify, but again have no idea how to use. When information about a specific stock or sector hits the market, it is my observation that the more objective the information, the faster the market responds, and the more subjective it is, the slower the market responds.
For example, when Ackman says that HLF is a pyramid scheme, then it can sometimes be hours, and sometimes even days before that news is no longer affecting the price of the stock, but when an analyst upgrades or downgrades a stock, that is more objective and the entire price adjustment is over in fifteen minutes. (If you subtract out market movements then an analysts announcement looks like a log curve, with most of the action in the beginning and a bit of a ringing at the last.)
Again, this all happens too fast to be of use, and it is after the fact that I can say, "That was subjective."
I don't think I am able to alter the coefficients as you suggest. I am using a hard coded take on a system of recursive polynomials for my modeling,so there are billions of coefficients.

Pandasaurus Sep 3, 2016

Hi, I have a quick and possibly dumb question. Why did you use the ratio instead of the difference between S1 and S2 in the Quantopain pairs trading lecture? In the co-integration lecture, you use the difference instead. In other sources, they use the difference as well.

Delaney Granizo-Mackenzie Oct 13, 2016

There's an updated notebook, algorithm, and video available on the lecture series page.

https://www.quantopian.com/lectures#Introduction-to-Pairs-Trading

DISCLAIMER

Delaney Granizo-Mackenzie Oct 13, 2016

And as a response to pandasaurus' question, which I unfortunately just saw, we have removed the ratio as it was a typo in the lecture.

DISCLAIMER

Andrew Hamlet Oct 23, 2016

Greetings Quantopian Community,

I was at the NYC Event on Pairs Trading, and the current example algorithm is deprecated, such that one cannot deploy it in live trading. With this fix, users can now deploy the algorithm in live trading. The fix is hosted as a pull request on github--thanks.

https://github.com/quantopian/zipline/pull/1550

Delaney Granizo-Mackenzie Oct 24, 2016

Hello Andrew,

Thanks very much. Could you please submit your PR to the following repo? It's where we store lectures and examples. Doesn't quite fit in the current form of zipline.

https://github.com/quantopian/research_public/tree/master/lectures/pairs_trading

DISCLAIMER

Andrew Hamlet Oct 24, 2016

Thanks, Delaney. I submitted the PR to the specified branch.

Best wishes,

Andrew

Please sign in or join Quantopian to post a reply.

About Quantopian

Careers

Community

Events

QuantCon

Help

Academia

Lectures

Workshops

Investor Relations

Status

Facebook

Twitter

LinkedIn

Blog

Terms of Use

Privacy Policy