

Statistical Arbitrage: Asset clustering, market-exposure minimization, and high-frequency explorations.

Aniket Inamdar(ainamdar@stanford.edu), Blake Jennings(bmj@stanford.edu), Bernardo Ramos(bramos@stanford.edu), Yiwen Chen(yiwen15@stanford.edu), Ben Etringer(etringer@stanford.edu)

Stanford University

MS&E 448, Spring 2016

Abstract

In this paper we describe and implement two statistical arbitrage trading strategies. The first strategy models the mean-reverting residual of a cluster of assets whose weights are selected so as to minimize market exposure. The second strategy maintains a portfolio of pairs, each weighted proportional to a measure of mean-reversion speed. We present performance metrics for each strategy, discuss the difficulties of their application into high-frequency trading, and suggest a direction where future research could focus.

1 Introduction

The technique of statistical arbitrage is the systematic exploitation of perceived mispricings of similar assets. A trading strategy built around statistical arbitrage involves three fundamental pillars: (1) a measure of similarity of assets, (2) a measure of pricing mismatch, and (3) a confidence metric for each mismatch. Traditional statistical arbitrage techniques, like “Pairs Trading”, employ these three pillars, holding long-short positions in a pair of strongly “similar” assets. The covariance (or correlation) between two assets is a widely used metric of their “similarity”. However, recent studies have used additional attributes such as co-integration tests to select asset pairs that are better suited for a statistical arbitrage trading strategy.

We base our signal generation strategy on the seminal paper [3], whose theory is briefly described in Sec. 2. Next, Sec. 3.1 describes a strategy to select assets to perform statistical arbitrage on. Sec. 3.2 introduces a mechanism by which we attempt to achieve market neutrality. Sec. 4.1 shows the efficacy of our asset selection procedure, and Sec. 4.2, 4.3 contain a detailed discussion on the performance of the strategy across a wide range of sectors for up to 9 years of out-of-sample testing. Section 5 presents a multiple pairs trading strategy and the idiosyncrasies involved with mean reversion at different frequencies. Section 6 compares the strategy’s performance in both low and high-frequency. Section 7 presents the conclusion and final remarks.

2 Model Basics - Signal Generation

Consider the returns of two correlated assets. We can express the relationship between the returns as a linear relationship,

$$\frac{dQ}{Q} = \alpha + \beta \frac{dP}{P} + \epsilon \quad (2.1)$$

Since $\frac{dP}{P}$ and $\frac{dQ}{Q}$ are correlated returns for a given lookback period (from the literature, this lookback time is 60 days for daily signal generation, which makes sense intuitively for correlated assets with quarterly earnings reporting [3]), we can express their residuals ϵ as a mean reverting process dX_t . That is, we can express dX_t as the instantaneous mismatch in the returns of the assets we are considering by performing a simple linear regression. In practice, $\alpha \approx 0$, so we will set $\alpha = 0$ here for simplification. *Note, problems can arise when we take on this assumption for assets with drift $\alpha \neq 0$ (explained in depth in Section 5).* We then have the following model,

$$dX_t = \frac{dQ}{Q} - \beta \frac{dP}{P} \quad (2.2)$$

In traditional single-pairs trading, dX_t could be interpreted as a 2-asset portfolio with the weight of one asset proportional to the regression coefficient of its returns with the returns of the other asset, but it could also model the residual of a general

cluster of assets. The principle hypothesis of statistical arbitrage is that the mismatch between any two correlated assets is mean-reverting and may be modeled as a stochastic mean reverting process. In [3], the choice of the mean reverting process - which also seems to be the conventional wisdom - is the *Ornstein-Uhlenbeck* (OU) process,

$$dX_t = \kappa(m - X_t) + \sigma dW_t \quad (2.3)$$

where

$$dX_t = \underbrace{\kappa}_{\text{Mean Reversion Speed}} \underbrace{(\overbrace{m}^{\text{Mean}} - X_t)}_{\text{Volatility}} + \underbrace{\sigma}_{\text{Volatility}} \underbrace{dW_t}_{\text{Brownian Motion}} \quad (2.4)$$

Noting that the Autoregressive (AR-1) model is the discretized version of the OU-process, where $\{a, b, \xi\}$ are as in equation (2.6), we can regress to obtain $\{a, b, \text{Var}\{\xi\}\}$ and back out the parameters $\{\kappa, m, \sigma^2\}$.

$$X_{n+1} = a + bX_n + \xi_{n+1} \quad (2.5)$$

$$b = \exp(-\kappa dt) \quad a = m(1 - \exp(-\kappa dt)) \quad \text{Var}(\xi) = \sigma^2 \frac{1 - \exp(-2\kappa dt)}{2\kappa} \quad (2.6)$$

where the trading signal is defined by

$$s = -\frac{m\sqrt{2\kappa}}{\sigma}, \quad (2.7)$$

where s is a normalized signal that indicates the number of units of mismatch in terms of the volatility of the OU process and dt is the frequency of the signal generation. In subsequent sections, we present results for several dt . Consider a pair position to hold $-\beta$ units of asset P and 1 unit of asset Q . Then a negative value for s suggests the pair is underpriced while a positive value suggests the pair is overpriced. Clearly, a signal within ± 1 standard deviation is expected roughly 66% of the time if we take on the convention that normalized returns are i.i.d. Gaussian, and hence is not reasonable to trade on. Thus, the signal resulting from the mismatch is not a statistical anomaly if it only fluctuates ± 1 standard deviation. A “significantly” negative (corr. positive) value suggests that there is statistically significant underpricing (corr. overpricing) of the pair. Thus, it is reasonable to expect gains if we enter into a long (corr. short) position as the signal is expected to mean-revert. We inherit the signal thresholds from [3], see equation 2.8. We enter a long position when the signal crosses $s_{long,enter}$ and hold the position until the signal crosses $s_{long,exit}$. We follow an identical process for a short position (see figure 1). We establish a stop-loss rule in Section 5.3.1.

$$s_{long,enter} = -1.25 \quad s_{long,exit} = -0.5 \quad s_{short,enter} = 1.25 \quad s_{short,exit} = 0.75 \quad (2.8)$$

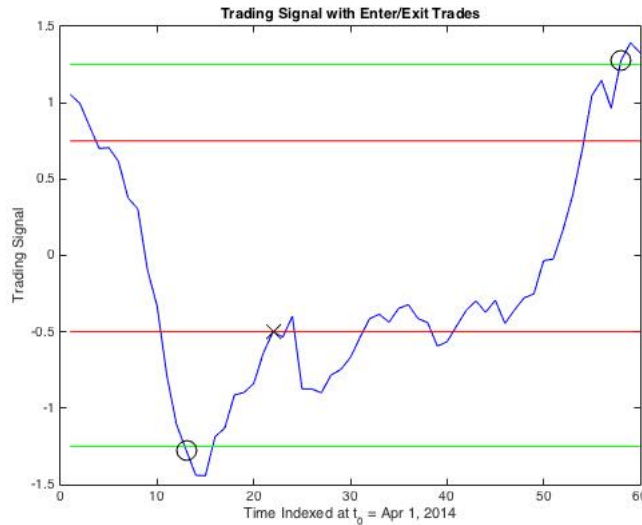


Figure 1: An example of a trading signal generated from daily returns of XLF and JPM in 2014

3 Strategy 1: Cluster Trading

In this section we describe a generalized statistical arbitrage strategy that minimizes exposure to market factors. The asset universe is projected onto its first three orthogonal principle components and clustered based on their mutual proximity in this orthogonal sub-space. Two distinct linear combinations of cluster members (cluster residuals) are obtained that are market neutral under the assumption of being portfolios with either positive or negative expected bank-excess returns. These cluster residuals are amenable to being modeled as mean reverting stochastic processes on which a recent arbitrage strategy is implemented. The feasibility and performance of this strategy are discussed in detail.

3.1 Asset Selection - Subtractive Clustering in Orthogonal Eigenspace

Now that we have a methodology to generate a trading signal, the logical next step is selecting assets from a given universe to trade on. For this we employ a combination of Principle Component Analysis (PCA) and subtractive clustering.

Let $R_{T \times M}$ be the daily bank excess returns (henceforth the term “return” describes bank-excess return) of M assets for T days, and define \vec{r}_m to be the daily returns of asset m . Define $\vec{y}_m = \frac{\vec{r}_m - \mu_m}{\sigma_m}$ and $Y_{T \times M} = [\vec{y}_1 \vec{y}_2 \dots \vec{y}_M]$, where $\mu_m = \mathbb{E}(\vec{r}_m)$ and $\sigma_m = \sqrt{\text{Var}(\vec{r}_m)}$. Now, using singular value decomposition for Y , we get,

$$Y_{T \times M} = U_{T \times M} \Sigma_{M \times M} V_{M \times M}^T = \tilde{F}_{T \times M} \tilde{\beta}_{M \times M} \approx \underbrace{U_{T \times K} \Sigma_{K \times K} V_{M \times K}^T}_{\text{K-factor approximation}} = F_{T \times K} \beta_{K \times M}$$

Note that \vec{f}_k 's are orthonormal. Thus, the columns of β , \vec{b}_m may be considered the coordinates of the m^{th} asset in the K -dimensional orthogonal factor space. Now that we have a “measure” for each asset, we have access to a plethora of algorithms to group based on proximity. One such algorithm is the subtractive clustering algorithm that is used here to find clusters of assets that have their \vec{b}_m 's within an ellipsoid in the K -dimensional factor space. We define the radii of this ellipsoid as follows:

$$r_i = \frac{\left(\frac{1}{\sigma_i^{svd}}\right)^2}{\sum_{j=1}^K \left(\frac{1}{\sigma_j^{svd}}\right)^2} \quad | \quad i \in [1, 2, \dots, K]$$

This accounts for the relative importance of each principle component as indicated by its corr. singular value. The higher the singular value, the closer the assets must be in the direction of the component, since the contribution of the corr. principle component to systematic part of the variance is greater. Note, this process needs no heuristic thresholds!

3.2 Cluster Signal Generation

In section 3.1 we described how to obtain a set of clusters of assets that are expected to behave similarly. For us to use the signal generation strategy described in Sec. 2, we now have to generate dX_t needed therein. We simultaneously espouse to keep our exposure to the K principle components (and by extension, to the systematic part of market variance) to a minimum. In the case of pairs trading, we know

$$X_t = dr_P - \beta dr_Q = [1 \quad -\beta] \begin{bmatrix} dr_P \\ dr_Q \end{bmatrix}$$

i.e. the dX_t is a linear combination of the returns of the asset pair or equivalently, dX_t are the returns of a portfolio containing assets $\{P, Q\}$.

Let us consider a cluster with P assets and define our $dX_t = Y_{T \times P} \vec{w}$ as a linear combination of the returns of the assets in the cluster (or the returns of a portfolio consisting of the cluster assets). A naive way to minimize market exposure would be to minimize $\vec{w}^T \text{Cov}(Y) \vec{w}$ such that $\vec{w}^T \mathbb{1} = 1$ (normalizing weights). But this puts no restriction on the expected return of the portfolio. If we were to short (corr. long) this portfolio, we would desire for its expected return to be non-positive (corr. non-negative). This is a natural way to minimize stop losses! Hence, we consider two distinct linear combinations (portfolios) of the cluster assets - one with weights \vec{w}_l for when we wish to enter a long position and one with \vec{w}_s when we wish to short the portfolio.

$$\begin{aligned} \min_{\vec{w}_{l/s} \in \mathbb{R}^P} \vec{w}_{l/s}^T \text{Cov}(Y_{T \times P}) \vec{w}_{l/s} &= \|\beta_{T \times P} \vec{w}_{l/s}\|_2^2 \\ \text{s.t.} \quad \vec{w}_{l/s}^T \mathbb{1} &= 1 \\ \text{and} \quad \vec{w}_l^T \vec{\mu} &\geq 0 \quad \vec{w}_s^T \vec{\mu} \leq 0 \end{aligned} \tag{3.1}$$

where $\mu_i = \mathbb{E}(\vec{y}_i) \mid i \in [1, 2, \dots, P]$.

We now have two measures of mismatch in the pricing of the cluster as a whole and use them to find the corresponding trading signals s_l, s_s . We enter a long (corr. short) position in the cluster *iff* $s_l < s_{long,enter}$ (corr. $s_s > s_{short,enter}$).

4 Cluster Trading: Results and Discussion

Before we discuss the performance of the clustering and trading strategies, it is important to note the underlying assumptions that have gone into the strategies:

1. **Look-Back Period** : The clustering and signal generation are both based on a moving window look-back of **60 days**. It is assumed that this period is long enough for any market event at the beginning of the period to have been accounted for.
2. **Systematic Principle Components** : All through our simulations, the PCA was truncated to $\mathbf{K} = \mathbf{3}$. It is assumed that the first 3 principle components are sufficient to describe the systematic part of the market. In other words, they account for $\sim 50\%$ of the observed variance.
3. **Signal Thresholding** : We use the thresholds from [3] given in eqn. 2.8. That is, we consider the signal has deviated “significantly” *iff* its different from zero with 90% confidence ($\Phi(s_{long,enter}) \approx 0.1$, $1 - \Phi(s_{short,enter}) \approx 0.1$).
4. **Risk-Free (Bank) Returns**: We assume that the bank rate is constant and equal to 5% annualized.

Besides the above, the algorithm requires no external inputs. No other heuristic is implemented. The objective is to observe the PDF of trade gains using this strategy and whether the strategy has any inherent merit. If it does (*i.e.* if its PDF is skewed to the right of zero), it can then be improved upon by heuristics designed to sample from the favorable half of the aforementioned PDF. A large number of datasets were used to perform our simulations. These are listed in table 1.

Name	Size (# of assets)	Duration
S&P 500	422	2 yrs.
NYSE	599	9 yrs.
Random(Russel 2000)	1661	9 yrs.
XLB	24	9 yrs.
XLE	33	9 yrs.
XLF	85	9 yrs.
XLFS	57	9 yrs.
XLI	61	9 yrs.
XLK	64	9 yrs.
XLP	32	9 yrs.
XLRE	28	9 yrs.
XLU	28	9 yrs.
XLV	52	9 yrs.
XLY	77	9 yrs.
Concatenated ETFs	456	9 yrs.
All Traded(ongoing)	variable	9 yrs.

Table 1: List of Datasets

4.1 Clustering Efficiency

Here we wish to examine whether the clustering algorithm selects groups of assets that are indeed related. For this purpose, we find clusters in the concatenated ETFs and check which ETF the members of each cluster come from. The clustering algorithm produces 30 clusters from the 456 assets in the concatenated ETF universe. In table 2 the breakdown of each of the 30 clusters is shown. It is clear that the clustering algorithm groups together assets primarily from one major ETF.

It is also able to detect that XLF(Finance) and XLFS(Finance) are similar ETFs. Similarly, the clustering is suggestive of the ETFs XLP(Consumer) and XLV(Health) to have some assets that are strongly similar - *e.g.* Cluster 30 has Coca-Cola, CVS, Mondelez(food & beverage), McCormick(spices & herbs) and Merc. & Co.(pharma). In cluster 11 we have Danaher(design & manufacturing), Minnesota Mining & Manufacturing, Parker Hannifin (hi tech. manufacturing), Rockwell Automation, Stanley Black & Decker (hardware & tools manufacturing) and Walt Disney. While Disney doesn’t quite fit the group, it is a fortune 100 company that helps hedge the rest of the portfolio against the market.

The large cardinality of clusters is a consequence of the underlying universe being well defined (and hopefully related) assets from market ETFs. In fig. 2, the clusters for the “Random” dataset are shown. Clearly, the size of the clusters increase and the number of clusters decrease as the underlying universe is random and weakly correlated. Another observation to be made is the fact that the clustering is the most sensitive to PC1 and almost agnostic to PC2 and PC3. This is to be expected for a random collection of assets as they don’t necessarily belong to a well defined sector/industry.

4.2 Trading Signal Statistics

An ex-post sanity check of our look-back period would be to check if the mean reversion times of the OU processes observed by the traded clusters are less than $\sim 30(= \frac{60}{2})$ days. In the absence of dealiasing, only half the frequencies contained in a dataset with equispaced sampling are uncontaminated (Nyquist sampling theorem). In fig. 3, histogram plots for mean reversion times for all datasets is shown. Clearly, an overwhelming majority of the trades are seen to have a mean reversion time of about 8-10 days.

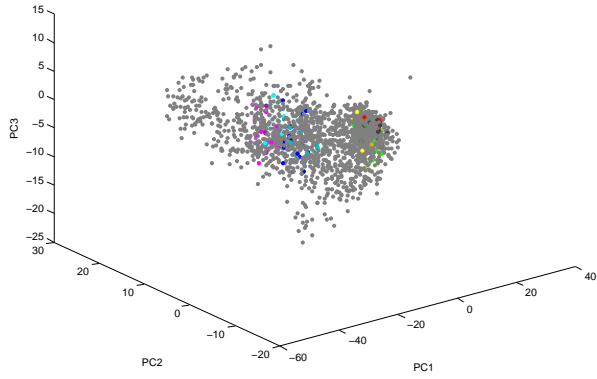
Cluster Number	Cluster Breakdown	Total
1	3(3), 2(4), 7(5), 1(7), 2(8), 2(11)	14
2	1(1), 2(3), 3(4), 3(5), 2(8), 2(11)	11
3	1(4), 5(5), 1(10), 1(11)	8
4	1(1), 1(3), 4(5), 1(8)	6
5	4(3), 2(4), 4(8)	6
6	6(10)	6
7	1(3), 4(5), 1(7), 1(10)	6
8	3(4), 2(5)	5
9	1(3), 1(4), 4(5), 1(6), 1(8), 1(10)	8
10	1(6), 2(11)	3
11	5(4), 1(11)	6
12	1(3), 1(5), 1(7), 1(10)	3
13	8(2)	8
14	1(3), 2(4), 1(8), 1(11)	4
15	3(3), 2(7), 1(8)	3
16	2(6), 2(10)	4
17	4(6)	4
18	1(1), 1(3), 2(4), 1(8)	4
19	2(5), 1(11)	3
20	3(3), 3(8)	3
21	1(6), 3(10)	4
22	3(9)	3
23	2(9), 1(10)	3
24	2(9)	2
25	1(6), 2(10)	3
26	3(10)	3
27	1(5), 1(11)	2
28	1(1), 2(4)	3
29	3(3), 1(7), 2(8)	3
30	4(6), 1(10)	5

ETF	Key
XLB	1
XLE	2
XLF	3
XLI	4
XLK	5
XLP	6
XLRE	7
XLFS	8
XLU	9
XLV	10
XLX	11

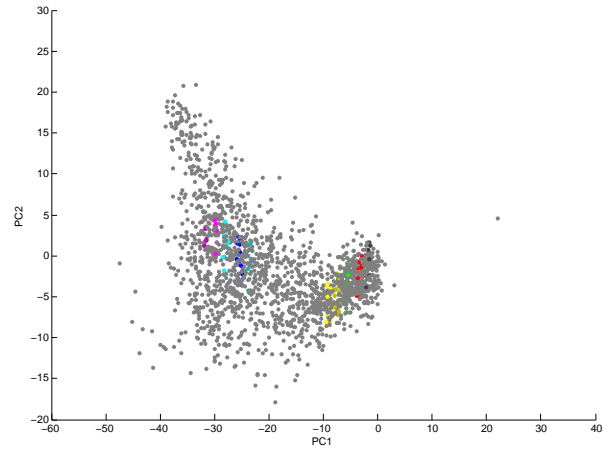
Table 2: Performance of the Clustering Algorithm. Value in brackets is the key of the ETF

4.3 Trade Statistics

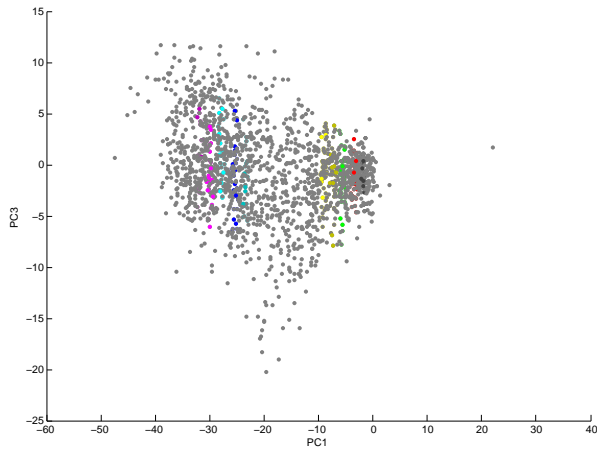
In this section, we discuss the performance of the raw (no heuristics) trading strategy. We desire a PDF of bank excess returns that is positively skewed. In fig. 6 the PDFs for all the datasets are shown. All PDFs, regardless of the dataset, show a characteristic distribution - peak around zero with a narrow negative tail and thicker positive tail - resulting in a hit rate of $\sim 57\%$ across all datasets. This behavior is clearly visible in the PDF of the concatenated ETFs (fig. 5).



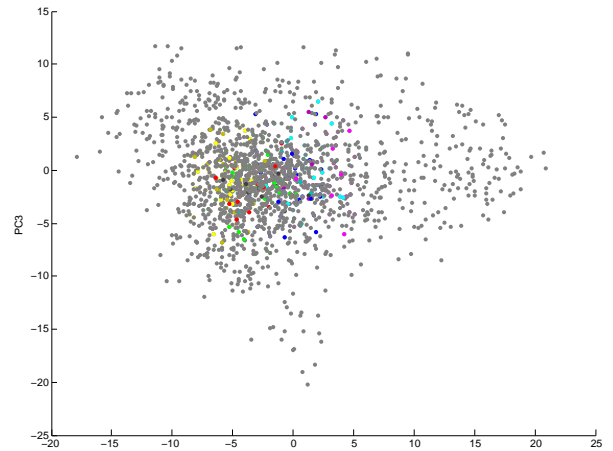
(a) Isometric View



(b) PC1 - PC2



(c) PC1-PC3



(b) PC2 - PC3

Figure 2: Clusters for the “Random” dataset

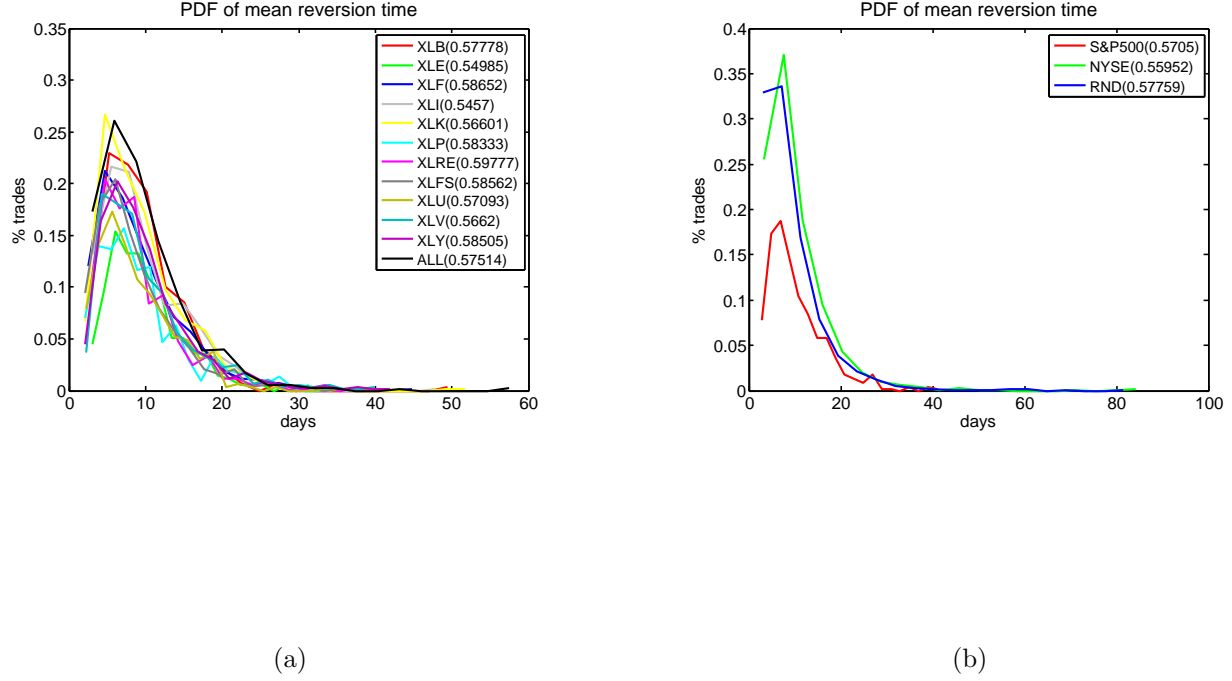


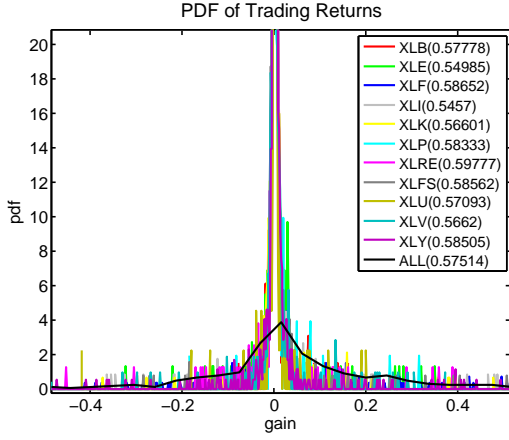
Figure 3: Mean Reversion Time. Value in the legend is the hit rate of the trading strategy for the respective dataset

The interquartile range(IQR) *i.e.* the range of the second and third quartile is a measure of the risk involved in a strategy. An important objective of the current strategy is to obtain better risk performance than a naive risk minimizing Markowitz portfolio. To test the performance of the strategy w.r.t Markowitz, we calculate the IQR for each dataset for a Markowitz strategy using each completed trade within every dataset. That is, given a dataset, for every trade completed, we calculate the performance of a Markowitz portfolio containing the same assets as involved in the trade for the same holding period as the trade such that the expected return of the Markowitz portfolio is the attained return of the stat. arb. trade. The Markowitz problem is defined in eqn. 4.1. In table 3 the IQRs for all datasets are shown.

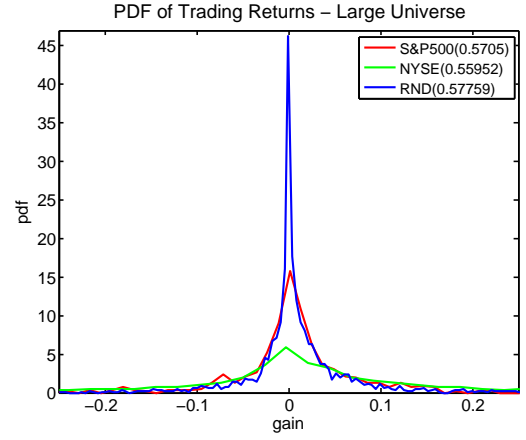
$$\begin{aligned}
 & \min_{\vec{w}_m \in \mathbb{R}^P} \vec{w}_m^T \text{Cov}(Y_{T \times P}) \vec{w}_m \\
 & s.t. \quad \vec{w}_m^T \mathbb{1} = 1 \quad \text{and} \quad \vec{w}_m^T \vec{\mu} = \mu_{stat.arb}.
 \end{aligned} \tag{4.1}$$

The IQR of the strategy is expected to be sensitive to the signal thresholding as the thresholds are a measure of our confidence in the mean reversion of the signal. To validate this hypothesis, in table 4 we have shown the hit rate for XLB, XLU and XLRE with the thresholds given below. These values correspond to roughly 94%(high) and 84%(low) confidence in the mean reversion of the signal. As expected, the hit rate reduces with the lowering of the thresholds and increases marginally with the increasing of the thresholds.

$$\begin{aligned}
 \text{High : } & s_{long,enter} = -1.5 \quad s_{long,exit} = -0.75 \quad s_{short,enter} = 1.5 \quad s_{short,exit} = 1.0 \\
 \text{Low : } & s_{long,enter} = -1 \quad s_{long,exit} = -0.25 \quad s_{short,enter} = 1 \quad s_{short,exit} = .5
 \end{aligned}$$



(a)



(b)

Figure 4: PDF of trade returns. Value in the legend is the hit rate of the trading strategy for the respective dataset

Name	Stat. Arb. IQR	Markowitz IQR	Stat. Arb. Sharpe	Markowitz Sharpe
S&P 500	0.0466	0.0217	1.89	1.60
NYSE	0.1555	0.045	1.88	1.3
Random	0.0299	0.1033	2.29	1.95
XLB	0.008	0.0093	3.16	4.03
XLE	0.0075	0.0114	1.92	2.43
XLF	0.013	0.012	2.18	1.83
XLFS	0.0088	0.0119	2.65	2.59
XLI	0.0068	0.0098	1.58	2.21
XLK	0.0084	0.0101	2.21	2.27
XLP	0.0056	0.0062	3.02	2.68
XLRE	0.0075	0.0114	1.92	2.43
XLU	0.005	0.0065	2.81	2.88
XLV	0.0075	0.0075	2.12	3.29
XLY	0.0091	0.0096	2.70	3.08
Concatenated ETFs	0.2199	0.0367	2.02	1.62

Table 3: Volatility of Trading Strategy and corresponding Markowitz portfolio as measured by the IQR(daily)and Sharpe(annualized)

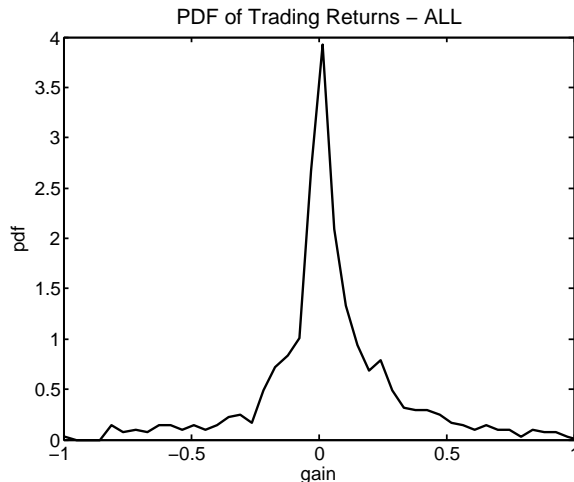


Figure 5: Nature of PDF of trade returns.

Thresholds	XLB	XLU	XLRE
Lower	56.9%	55.6%	58.7%
Baseline	58%	57%	60%
Higher	59%	58%	60.5%

Table 4: Sensitivity of Hit Rate to Signal Thresholding

5 Strategy 2: Combinatorial Pairs Trading at Different Frequencies

In this section we present a strategy built on a generalization of the single-pair framework. Unlike the eigenportfolio contributions from Avellaneda and Lee or the clustering approach presented above, similarity between assets here is not measured via a multi-factor regression model [3]. Instead, we consider all $\binom{N}{2}$ pairs a stock universe of N assets - perhaps the holdings of ETF or a set generated by the subtractive clustering algorithm. At each time-step, single-factor regression and AR(1) model parameters are fit for each pair. We then rank the pairs according to the F -score¹ and correlation coefficient from the OU model, as well as the drift of the residual process.

The F -score is set to 0 if correlation and F -score thresholds are not met, or the drift of the residual process is too large. Since the F -score is related to the amount of times the signal is expected to revert to its mean, it is also set to zero if the F -score is not higher than a certain amount. Before checking if a pair's signal crossed a threshold in the previous time-step, capital is allocated in proportion to signal strength of the remaining pairs with non-zero F -scores. This step is essential to keep cash on reserve for the other pairs with qualifying F -scores, whose signals are imminently expected to cross. Finally, trades are placed only on the pairs that crossed a signal threshold, with the traditional $(1, -\beta)$ weights, but now with scaled with respect to the pair's allocation.

¹For a pair p , lookback period ℓ and frequency window Δ_t , we define the F -score as

$$F_{score}(p) = \kappa_p \ell \Delta_t = \text{Expected \#reversions in } [t, t + \ell].$$

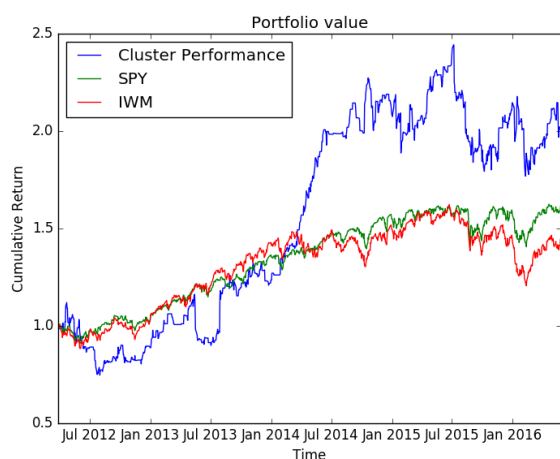
5.1 Simulation

We utilize two distinct backtesting frameworks to simulate our strategy, the first of which we designed to represent a simplified market environment without limit order book considerations. Additionally, only daily trades are supported at the Adj. Close price. All feeds are obtained from Yahoo Finance.

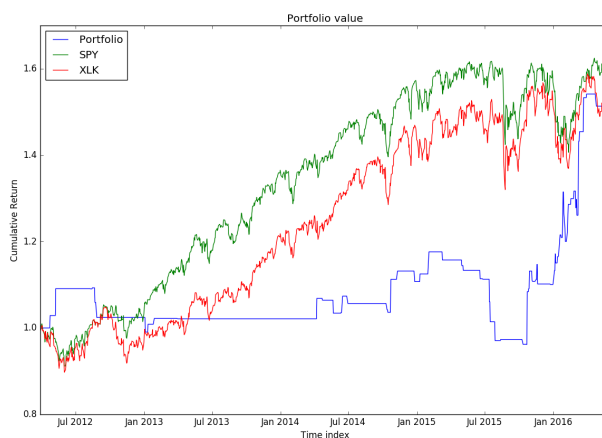
The second simulator is much more robust and is provided by Thesys Technologies. In addition to clean timeseries, it features limit order book data from multiple exchanges at a frequency on the order of microseconds. Additionally, market liquidity is accurately modeled. Accordingly, our strategy is equipped with an order management system for rebalancing and tracking unfilled and partially filled orders. These considerations do not typically arise when trading at a daily frequency, and are addressed in the upcoming high-frequency section.

5.2 Trading at a Daily Frequency

We now present performance metrics of our strategy using the simplified backtesting framework. The graphs below show the portfolio performance of two different asset clusters over the period beginning July 2012 and ending May 2016. The graph on the left is produced when the strategy considers a stock universe identified by the subtractive clustering algorithm, while the graph on the right is the performance of a cluster consisting of large cap popular tech stocks. The overlaid graphs in red and green represent relevant indexes to each cluster.



Cluster 1: ADI, ADSK, AIC, AYI, CCI, FIS, GLW, IPG, KLAC, LLTC, NKE, PBCT, TXN, TXT



Cluster 2: AAPL, GOOG, T, FB, MSFT, INTL

Figure 6: Cumulative return cluster comparison.

	Cluster 1	Cluster 2
Cumulative Return	100.584 %	49.822 %
Annual Mean Return	20.730 %	10.602 %
Annual Volatility	28.355 %	13.756 %
Daily VaR	-2.693 %	0.000 %
Annualized Sharpe Ratio	0.731	0.771

Table 5: Statistics for different clusters

While the first strategy achieves a superior return, it is marked by many periods of high volatility. However, it is clear that it is identifying mispricings: the large positive spike begins in January 2014, precisely when a global selloff began, the DOW dropped 5.9% that month - the worst opening since 2009. Further, the cluster of assets is characterized by large market caps: each roughly between \$5 and \$10 billion, suggesting that the variance in their returns is explained largely by the market factor. The second group of assets are not quite as correlated and are presented for comparison. The low volatility step-like jumps are characteristic of statistical arbitrage. This strategy does not trade very often and achieves a 95% daily VaR of 0.0%. The spike at the end is interesting as it too occurs at the beginning of market turmoil: January 2016 was marked by fear over a slowdown in China along with the price decline in oil, suffering a 10.05% decline in the opening 12 days.

5.3 High Frequency

Limit Order Book Dynamics (potential limits)

We now explore statistical arbitrage in the high frequency domain. If we can make consistent small profits on a small time interval, we can theoretically generate large profits from the accrued small returns. We obtain strong signals (see figure 7), however we find that we cannot easily profit them. The predominant issue is that the expected returns of each trade in the high frequency domain is on the same order as the spreads of the returns (more on this in section 6). In addition, there is a large trade-off between quick order executions and spread-associated costs. Thus, without the guarantee that we trade at the desired price given by the signal, we will lose money on the trade. Furthermore, there is a \$0.01 minimum tick size, which is too large for our strategy to make a profit on. Even if we could overcome these issues, we would need to generate substantial profits from each trade due to transaction costs. In essence, even though we generate strong signals in the high frequency domain, we are unable to make a profit from them.

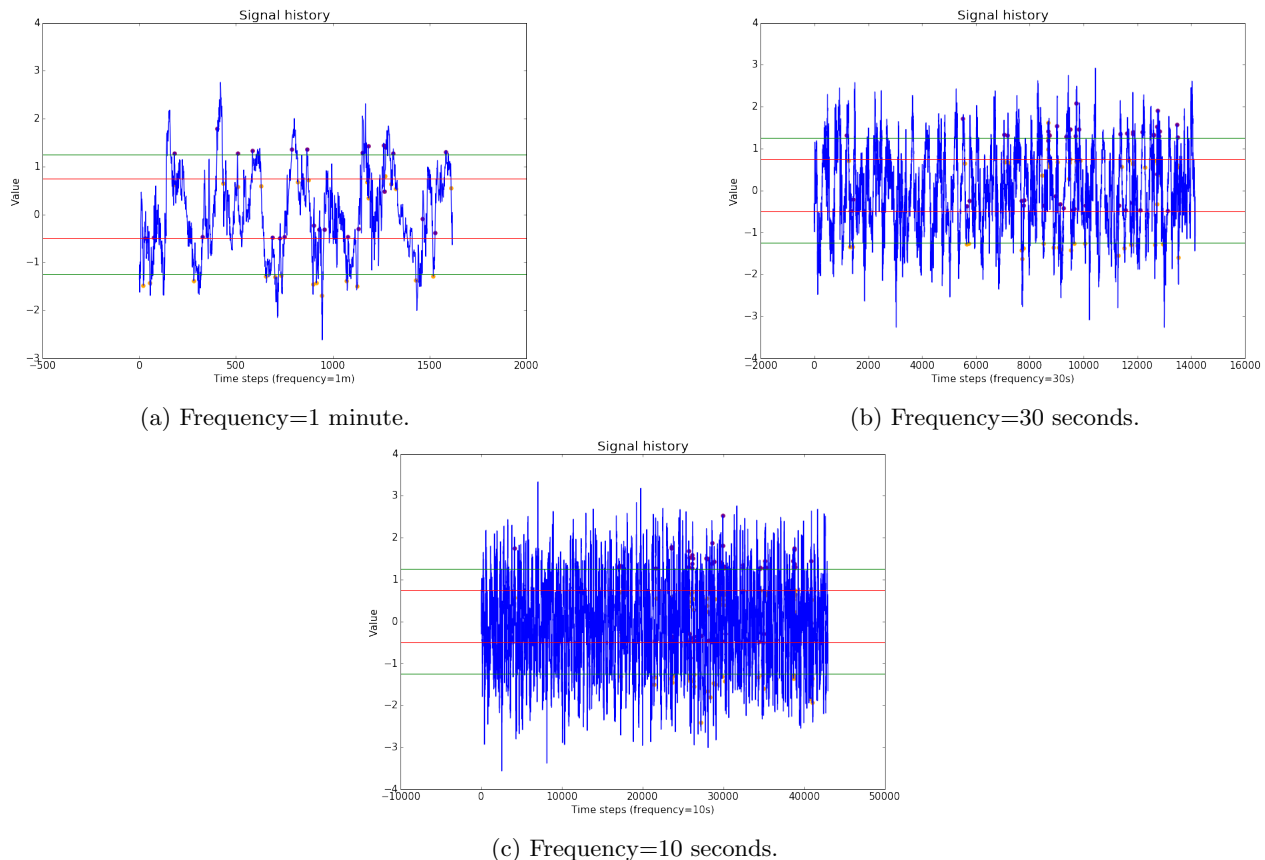


Figure 7: Example of signal behavior for different trading frequencies.

In order to profit from the signal there must be sufficient time in between peaks for the price of the underlying assets in the pair to move. The signal in the first plot is precisely what we want to see. Our strategy can seemingly detect reversion at a smaller precision than \$0.01, but is unable to trade on this. In order to get filled when the spread is at \$0.01, we must bid at ask, and so we lose any possible profit.

5.3.1 Trading rules

In this section we explain some trading rules we considered in our strategy. They are designed so that we take high-frequency idiosyncrasies into account.

1. **Stop-loss criterion:** In order to minimize risk from a single pair, we block both assets in that pair from being traded if we currently hold a position on the asset, unless the order corresponds to an exit in the position (e.g. if the signal is not strong anymore). We do this both for minimizing transaction costs and setting the maximum loss of a single pair to be the amount of cash we allotted based on the F -score.
2. **Liquidity constraint:** As we discuss in the following sections, high-frequency orders must be performed in a timely manner in order to ensure the expected gains from signal movements. Since orders are not necessarily filled at mid-price,

we propose to introduce a parameter $\alpha \in [-1, 1]$ from order prices are determined as follows:

$$\begin{aligned} \text{OrderPrice} &= \alpha (\text{Ask} - \text{Mid}) + \text{Mid} && \text{(buy side),} \\ \text{OrderPrice} &= \text{Mid} - \alpha (\text{Mid} - \text{Bid}) && \text{(sell side),} \end{aligned} \tag{5.1}$$

so for $\alpha = 0$ we would trade at the mid-price, and at $\alpha = 1$ we would pay the whole spread. This trick will let us have a balance between not paying the full spread and getting orders quickly filled. In our tests we used alpha ranging from 0.2 to 0.5; that is, we always trade higher than the mid-price, but not as much so to pay for the whole spread.

3. Unfilled order management: Our strategy is very highly dependent on trading a particular pair at the right time with the correct weights. Having an exposure from unfilled orders is unwanted since portfolio values can vary wildly in the high-frequency setting. Therefore, we set out the following rule for dealing with unfilled pairs: if we submit an order at time $i\Delta_t$, then at time $(i + 1)\Delta_t$:

- If the signal strength is still within the desired bounds (i.e. it has not crossed any exit bar), then we recalibrate the model, compute the desired proportion (1 to $-\beta$), and submit an order so to rebalance the pair with the current position.
- If the signal is not strong anymore, submit orders so to cancel the current position. Since there is no expected gain from it, we assume the loss arising from having held an unbalanced pair.

When using the previous rules, we see that the strategy performs reasonably well for low and mid-frequencies. As an illustration, figure 8 shows the behavior of holding returns² at a trading frequency of 3 hours. We see that the returns are skewed to the right, which is indicative of a hit rate greater than 50%. In the following section we will evaluate how returns and other statistics behave when we trade at higher frequencies.

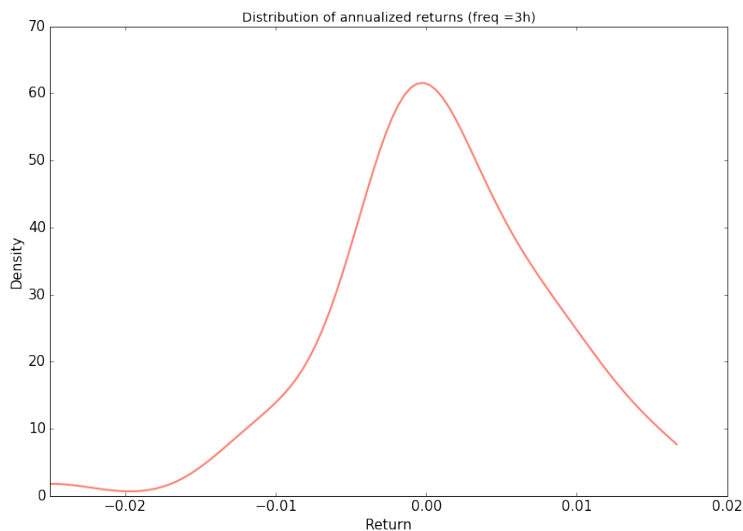


Figure 8: Distribution of annualized holding returns from 2015/06/15 to 2015/12/15 on a 3-hour frequency setting. Assets traded are (JPM, XLF).

6 Results

In this section we assess the performance of pairs strategy in high frequency. To draw sound conclusions about the differences between high and low frequencies, we take the single pair (JPM, XLF)³ and run the strategy over different trading windows. The following visualizations will then be based on this pair. While the results look similar for the multiple-pair case, the interaction between some measures –e.g. portfolio value progression vs. trading signal tracking– is explained in a more straightforward manner for the single-pair case.

²Throughout, by holding return we mean the instantaneous returns of when holding a non-zero position on the pair. In the case where we uninterruptedly hold a position over $[0, T]$, it is computed as the annualized version of

$$\text{Ret}(0, T) = \frac{1}{T} \sum_{t=1}^T \frac{p_t - p_{t-1}}{p_{t-1}},$$

where p_t is the value of the portfolio at time t and T is the total number of steps. In pairs trading, we expect this holding return to be skewed to the right, since the residual signal is expected to return to its mean, thus having instantaneous pair profits until exiting the position.

³We took this pair as the ‘canonical’ example, where we can safely assume the cointegration relationship holds.

6.1 Transaction costs versus gains

First, let us look at how the portfolio values progresses. Figure 9 shows an example of the portfolio value dynamics for both low and high frequencies. Whenever the portfolio value stays flat, the strategy does not hold a pair position. The first difference to spot is how many non-flat periods Figure 9b (high-frequency) shows. The result is as expected: the high-frequency trades a larger number of times, even if the test period is smaller. However, the fact that positions are being evaluated at higher frequencies, the overall portfolio value will be highly volatile relative to the gains. See, for example, how *continuous* changes scale to the gains of exiting the position in the 9a versus 9b.

As we will later see, this effect is a source of hardship for the strategy to make profits. On the one hand, more trading costs have to be paid the more frequently we trade. Second, the timing of exiting the position is fairly influential on the outcome of the trade. Indeed, changing the bid (ask) price even a few cents can drive the outcome from profits to losses had the trade been performed a few seconds (or minutes) earlier. Since profits drawn for the high-frequency setting are of the order of the spread (a few –or even one– cents per trade, versus whole dollars in low frequencies), the timing for entering or clearing positions must be quick. This effect is pervasive for any trading frequencies on the sub-hour, and is our main motivation to believe that the model can become robust if it included predictions about ominous changes in the order book.

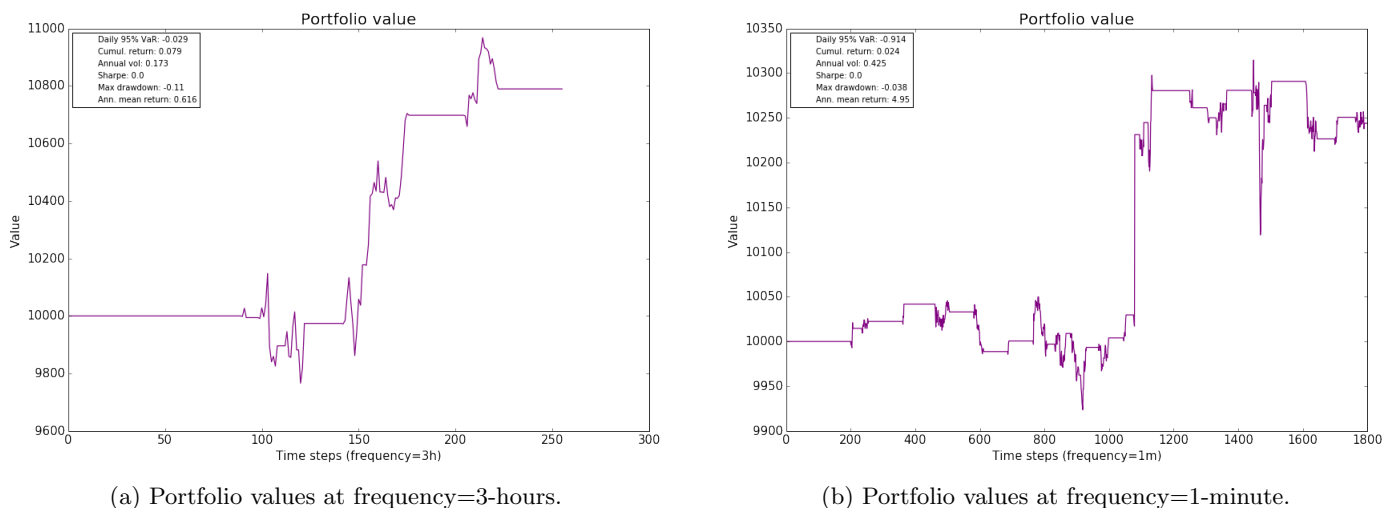


Figure 9: Examples of portfolio value progressions for different frequencies. The dates selected were 2015/06/15 to 2015/12/15 for low frequency and 2015/06/15 to 2015/07/15 for high-frequency.

Figure 10 shows another example of the progression of cumulative clearance (i.e. non-spread related) costs in a high-frequency setting. The number of steps reconfirm the observation that we trade a fairly high number of times in a single trading day. For the particular period shown, cumulative trading costs amount to 10% of the total gains, and that particular day experienced an outstanding return. While bearing in mind that no conclusion is valid for all possible trading seasons and regimes, this illustrates how trading costs can scale to undermine a potentially alarming proportion of gains.

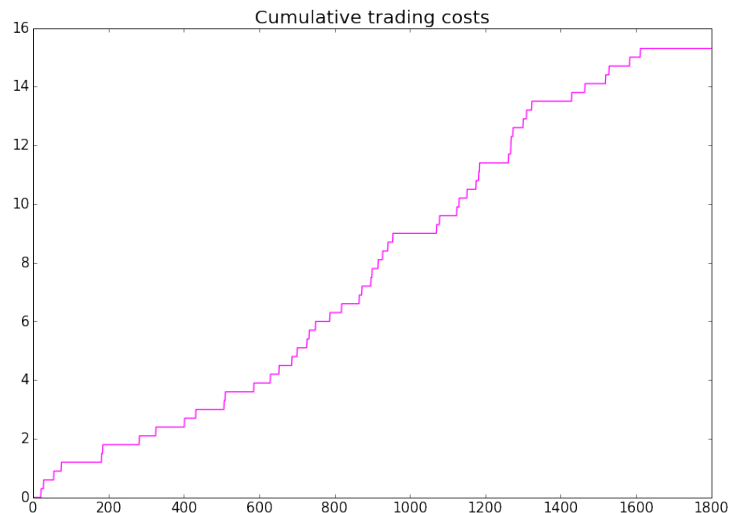


Figure 10: Example of cumulative trading costs in high-frequency.

6.2 Performance and risk measures

While per-trade costs do not increase with frequency (they scale linearly to the number of trades performed), the proportion of trading costs (of a few cents) relative to per-trade gains can increase up to the point where gains are overwhelmed. An example of this effect is shown in Table 6, where we show performance measure as well as risk indicators of the strategy over different frequencies. For this example –all trades start on the same day, but frequencies change–, we can see the annualized cumulative and holding returns being consistently drawn down the higher the frequency. While this is not the case for *all* trading periods (some test periods do have a stable return behavior), this illustrates how either (a) signals generated at higher frequencies are spurious (cf. figure 7), which generates negative returns, or (b) transaction costs overwhelm expected gain on each trade.

Frequency	Return	Vol.	Cum. return	Daily 95% VaR	Max Drawdown	ℓ	Horizon
3 hours	61.55%	17.25%	7.89%	-2.85%	60	6 months	
2 hours	41.26%	17.93%	5.15%	-4.8%		60	6 months
1 hour	89.27%	19.98%	7.71%	-6.03%		60	4 months
30 minutes	-16.7%	12.28%	-0.7%	-6.8%	120	2 months	
1 minute	-8.92%	9.21%	-0.19%	-0.32%	240	1 month	
30 seconds	-1.54%	14.31%	-3.19%	-54.43%	240	1 month	
10 seconds	-113.9%	8.53%	-2.35%	-55.66%	240	1 month	

Table 6: Statistics for different frequencies. The start date was randomly picked to be 06/15/2015. The strategy trades for a period determined by the horizon, with a lookback period determined by ℓ .

Annualized volatility, on the other hand, shows a fairly stable behavior throughout frequencies, so that Sharpe ratios are completely determined by expected returns. This observation shows surprising to us given the belief that high-frequency series contain strong non-continuous behavior, microstructure noise and jumps that undermine the usual estimator of the volatility. The (normalized) maximum drawdown is another measure whose magnitude does not vary enough across frequencies. The decreasing behavior stems for a higher volatility –hence dividing the maximum value drop by a larger amount. However, having a 2.88% 10-second drawdown in a single is indicative of a very highly risky position. In the same line, the sharp increase in Daily 95% Value at Risk indicates that strategies become riskier the more frequently we trade.

Table 6 showed an illustration over one particular period in H2-2015, which one can very well judge as unrepresentative. To show more robust results on the strategy’s performance in high-frequency, we performed thirty independent simulations of the strategy at a 1-minute frequency, where each of the dates was picked at random between 2011 and 2015 and the trading window was four days. The cumulative returns—that is, the return over the 4-day period— density estimation is shown in Figure 11. Note the difference of Fig. 11 with the shape when compared to Fig. 8, where a smooth right-skewed graph is presented.⁴ While the high-frequency strategy can achieve significant returns (seemingly with a high probability), there is a non-negligible possibility that returns are catastrophic (cf. the left clump in figure Fig. 11).

The non-smoothness of the return curve about a mean can also indicative of a high sensitivity of the strategy to model parameters. As we pointed earlier, the presence of microstructure noise and jumps can undermine the confidence on the mean-reverting model. For illustration, Figure 12 shows the behavior of the estimator of β in equation (2.2). We can clearly see that jumps in the data introduce strong discontinuities in one of the model’s most crucial parameters. We would be interested in performing research on this respect –that is, getting jump-robust estimators or modeling the residual as a semi-martingale with a jump component.–

⁴The scale, however, is not comparable.

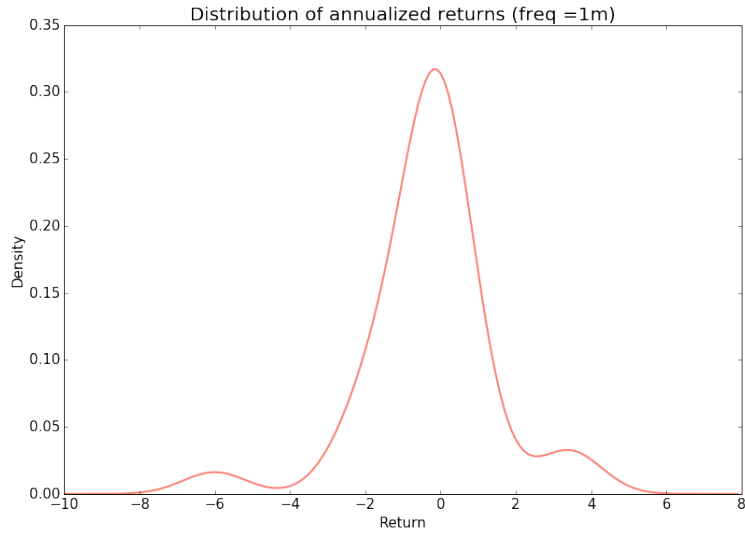


Figure 11: Distribution of returns based on 30 independent simulations of the strategy.

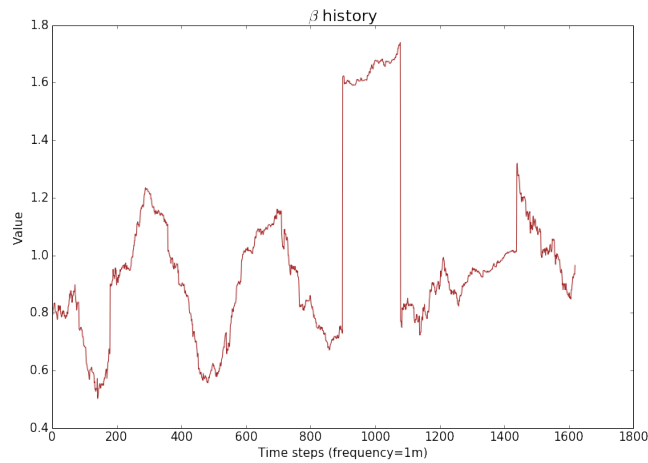


Figure 12: Distribution of returns based on 30 independent simulations of the strategy.

For a final assessment on the shape of the density of returns, Figure 13 shows the distribution of the annualized holding returns at different frequencies. The fact that higher frequencies have a lower holding returns is no surprise –most of the times the portfolio value will not vary too much for very small time windows. However, we wish to highlight how the smoothness of the curve changes as the frequency becomes lower. In particular, low-frequency lines are clearly skewed to the right, thus we are more confident about the model for generating profits. This is not to mean that pairs trading at high frequencies is unreliable; rather, we should be careful in implementing low-frequency strategies to any time window without adjusting for important idiosyncrasies of high-frequency.

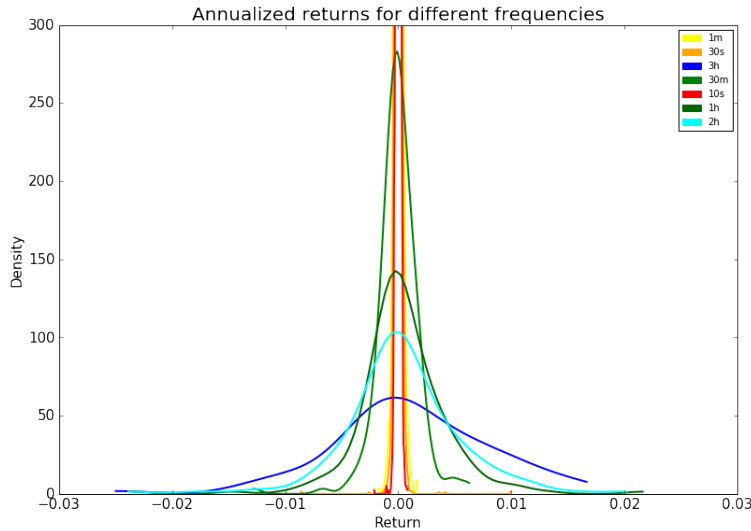


Figure 13: Distribution of holding returns for different frequencies. Cold colors correspond to lower frequencies; hot colors to high ones.

7 Conclusion

Summary - Cluster Trading

We have shown that the asset selection strategy based on subtractive clustering in an orthogonal factor subspace performs reasonably well. Additionally, a market exposure minimization strategy for statistical arbitrage was shown to outperform a naive Markowitz portfolio for large and/or weakly correlated asset clusters.

In the absence of any heuristic aides, the strategy was shown to have favorable returns while being agnostic to the asset universe. The few external inputs to the strategy are seen to be reasonable based on ex-post analysis. We examined the sensitivity to signal thresholding to be minimal. Moreover, we observed the lookback time to be reasonable over all datasets. Finally, across all datasets, singular values of the first three principle components consistently account for more than 50% of the market (systematic) variance.

Summary - High-Frequency Adaptation

We have illustrated how, in the high-frequency domain, trading on signals is not easily profitable. The lack of ability to profit from the signals is due to several factors. First, the expected returns are on the same order as the spreads; that is, tick sizes are large compared to theoretical returns we may get from trading on the signals. This effect is aggravated by the fact that transaction costs become high relative to the expected profits of instantaneous mispricings. In the high-frequency, there are also significant liquidity concerns. When bidding at the ask price, we can easily lose money, but when bidding at the bid price, orders are not instantaneously filled. This makes the whole position to be exposed to movements the model does not account for. Also, high-frequency data can possess such strong noise (generated from microstructure noise or jumps) that model parameters can be easily affected.

At lower frequencies, we can avoid many of these issues. The signals may not cross the trading thresholds as often, but profits we obtain from each trade can offset the transaction costs. Furthermore, we do not run into any liquidity concerns, and the expected returns are significantly greater than the spreads. The higher we go in the time frequency domain, the more we have to consider the limit order book and other players, whence low-frequency signal trading is easier –at the expense, of course, of having to wait longer.

Combination of Strategies

We run into issues with combinatorial pairs strategy partly due to the fact that the strategy does not have an asset class identification mechanism. Daily backtesting suggests that decomposing the stock universe generated by the subtractive clustering algorithm is a reasonable thing to do, as it forms asset cluster based on a sectors. In fact, if we know an asset class is similar and is over/under priced then there must exist at least one mispricing of an individual asset within that cluster (since the cluster asset is defined as a weighted sum of the individual assets in the cluster). Furthermore, we have shown that it is more difficult to find liquidity at high frequency, and it is known that statistical arbitrage is “dry” in well-known correlated pairs (such as XLF and JPM). In principle, we have shown that even though daily backtest results hold true for certain asset pairs, they are unrealistic as they do not represent backmarket conditions.

8 Future work

There are many dimensions to consider in order to overcome the difficulties discovered in our work. As we pointed in section 6.1, the incorporation of the volume in the order book could be highly useful for predicting when is the best time to trade on a pair. Incorporating knowledge about mispricing signal behavior paired with fundamental analysis could be promising for taking advantage on the spreads. Additionally, having a dynamic liquidity parameter α (cf. equation 5.1) based on fundamental data could also reduce transaction costs.

Another direction we could take is to let signal thresholds to vary. In theory we would expect to have potentially higher mispricings—thus returns—in periods of high volatility. We could then set thresholds based on measures of future volatility.

We observed in section 6.2 that high-frequency estimators can alarmingly undermine the variance of estimated parameters. In particular, jumps (Figure 12) can be an important source of estimation error. We contemplate the incorporation of jump behavior into the residual dX_t as a possibility for making the model more robust. As an alternative, we could imagine estimators of the OU parameters to include thresholding to neglect jumps, similar to what recent developments in high-frequency econometrics have developed. For a survey on this literature, see [1].

Acknowledgments

The authors would like to thank Dr. Lisa Borland and Enzo Bussetti for providing their insight into the world of mathematical finance in the age of large data and guiding us through the hurdles we faced during this project.

References

- [1] Yacine Aït-Sahalia and Jean Jacod. *High-frequency financial econometrics*. Princeton University Press, 2014.
- [2] Ghazi Al-Naymat. Mining pairs-trading patterns: A framework. *International Journal of Database Theory and Application*, 6(6):19–28, 2013.
- [3] Marco Avellaneda and Jeong-Hyun Lee. Statistical arbitrage in the us equities market. *Quantitative Finance*, 10(7):761–782, 2010.
- [4] Evan Gatev, William N Goetzmann, and K Geert Rouwenhorst. Pairs trading: Performance of a relative-value arbitrage rule. *Review of Financial Studies*, 19(3):797–827, 2006.
- [5] George J Miao. High frequency and dynamic pairs trading based on statistical arbitrage using a two-stage correlation and cointegration approach. *International Journal of Economics and Finance*, 6(3):96, 2014.
- [6] Merton H Miller, Jayaram Muthuswamy, and Robert E Whaley. Mean reversion of standard & poor’s 500 index basis changes: Arbitrage-induced or statistical illusion? *The Journal of Finance*, 49(2):479–513, 1994.
- [7] Jesus Otero and Jeremy Smith. Testing for cointegration: power versus frequency of observation—further monte carlo results. *Economics Letters*, 67(1):5–9, 2000.