Expert Systems With Applications 135 (2019) 113–128

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Evolutionary multi-objective optimization for multivariate pairs trading

Jeffrey Goldkamp, Mohammad Dehghanimohammadabadi*

Department of Mechanical and Industrial Engineering, Northeastern University, MA, USA

ARTICLE INFO

Article history:
Received 14 July 2018
Revised 11 April 2019
Accepted 28 May 2019
Available online 29 May 2019

Keywords:
Multi-objective optimization
Genetic algorithm
Mixed integer programming
Pairs trading
Portfolio optimization

ABSTRACT

Forming combinations of comoving assets is a critical step in pairs trading that has only been addressed either manually or through enumerative procedures. Both approaches fail in the multivariate case and do not consider conflicting objectives in the problem structure. This paper is the first attempt to address these novel problems by presenting an intelligent system that recommends profitable pair combinations through a Mixed Integer Programming (MIP) formulation and solving the NP-Hard optimization problem with a multi-objective genetic algorithm (NSGA-II) containing problem specific modifications. Combinations of assets are optimized on two conflicting objectives of risk (mean-reversion) and return (spread variance) to form sets of profitable multivariate pairs trading opportunities. Promising results support the superiority of multi-objective and multivariate pairs trading strategies over their traditional single objective and univariate counterparts. The findings should motivate new directions for pairs trading research and also expand the applications of evolutionary multi-objective optimization for hard problems in finance and other industries.

1. Introduction

Expert and intelligent systems have already made a profound impact on the financial industry. From “robo advisers” making automated asset allocation decisions to high frequency trading algorithms, the market has become ever-more dynamic, efficient, and competitive. Research in the pairs trading space has in some ways kept pace with these advancements, although there still lacks a sophisticated approach to forming profitable pair combinations. Deteriorating profitability of traditional approaches to the strategy have mostly been attributed to increased market efficiency so an advancement of an intelligent selection procedure should be a welcome innovation.

Existing techniques for selecting pairs remain dependent on either expert intuition or computationally intensive enumerative procedures. Not only does this restrict the trading opportunities to univariate pairs, but the usual selection procedures fail to consider conflicting objectives properly. The problem draws similarities from Markowitz portfolio theory (1952) for systematic asset selection that optimizes a trade-off between conflicting risk and return objectives.

Forming pairs of multiple assets under multiple objectives can not be done using existing approaches. The proposed methodology automatically generates a frontier of efficient multivariate pair combinations that satisfy multiple conflicting objectives. The problem is formulated as a mixed integer programming (MIP) model and, due to the non-convex constraints and exponential solution space, a genetic algorithm (GA) is employed to obtain profitable pair combinations. GA is also easily extended to handle multiple objectives, such as the Non-dominating Sorting Genetic Algorithm II (NSGA-II) used in this paper.

We show how multivariate pairs from S&P 500 constituents generate excess returns over their univariate or single-objective benchmarks. The methodology addresses an overlooked aspect in pairs trading through an original application of multi-objective evolutionary optimization, with contributions of:

- Establishing a multi-objective multivariate pair formation model to simultaneously optimize profitability (variance) and risk (mean-reversion)
- Solving the problem through new solution representation and modification procedures in NSGA-II
- Showing the outperformance over traditional univariate and single objective methods with empirical data
- Expanding GA modeling capabilities in financial applications through solution representation and crossover operations

* Corresponding author.
E-mail addresses: goldkamp.j@husky.neu.edu, goldkamp.j16@gmail.com (J. Goldkamp), mdehghani@neu.edu (M. Dehghanimohammadabadi).

https://doi.org/10.1016/j.eswa.2019.05.046

There are numerous decision phases in a complex strategy like pairs trading. This framework only addresses one part of the system, as will be highlighted in following sections. Nevertheless, the practicality of expert and intelligent systems is clearly demonstrated for difficult real-world decision processes. It shares common traits with other realistic problems that have non-convex constraints and multiple conflicting criteria. In addition, the core task of identifying combinations of comoving sequence data is a prevalent problem unique not only to finance.[1] Therefore, implications from this study have potential to impact a wide range of important industries and applications.

Sections 2-3 provide a brief background of pairs trading and formally defines the problem addressed in this paper. Section 4 reviews the most relevant literature to this problem as well as similar metaheuristic applications in finance. Section 5 outlines the framework in which the PF problem is modeled and solved through a multi-objective genetic algorithm called NSGA-II. Empirical experiments in Section 6 are conducted on S&P 500 data to examine how the approach improves pairs trading profitability. Finally, concluding remarks and new research directions are proposed in Section 7.

2. Problem background

2.1. Pairs trading

Pairs trading places directional bets on the gap (spread) between pairs of asset prices. Stock prices have been widely accepted to follow a random walk process[2] and therefore hinders most prediction efforts. In order to achieve a higher degree of predictability, pairs trading creates a mean-reverting spread between coupled assets. Subsequent spread deviations and reversions are traded through buying (*long*) one component and selling (*short*) the other component of the pair. We first clarify two principal terms before further discussion.

Definition 1. Pair: The combination of two components, each exhibiting their own individual time-series. In the univariate case, each component consists of one stock. In the multivariate case, each component has more than one stock. Quasi-multivariate pairs have one univariate component against another multivariate one.

Definition 2. Spread: The price difference between the long and short components of a pair. The component prices are simply the sum of their constituent normalized price vectors.[3] The spread (d_t) at time t is the difference of the wieghted ($\mathbf{w}$) normalized prices ($\mathbf{A}$) between the long (L) and short (S) components.

$$d_t = \left(\sum_{i \in L} w_i A_{i,t} - \sum_{j \in S} w_j A_{j,t} \right) \quad t = 1..m \tag{1}$$

While the profit in pure arbitrage is deterministic, statistical arbitrage is stochastic and only requires the expected value over a sufficient number of trades to be profitable. There should be some advantage where the profit likelihood must only exceed the loss likelihood by some margin to accumulates wealth over the long run. The statistical property *cointegration* arises when a stationary linear combination exists from multiple non-stationary time series. For a pair in which both time-series have historically moved together, we expect them to continue moving together until there is a fundamental change in the relationship. Given such a pair, one would aim to open a position once the spread between the two series has widened to some upper threshold and then close the position upon the spread's convergence to an equilibrium level.

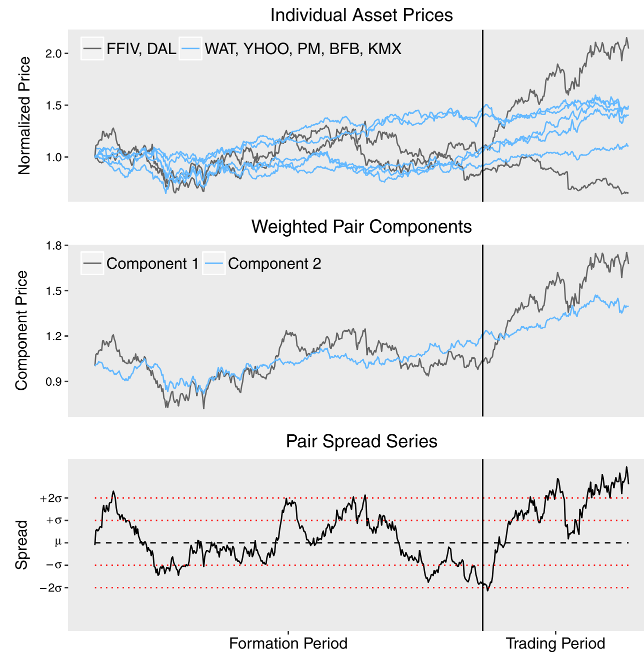


Fig. 1. The top plot displays individual asset prices which are all normalized to 1 at the start of the formation period. The middle figure shows a weighted combination of those same assets to form a multivariate pair with a mean-reverting spread as shown in the bottom figure. The $\mu_{\mathbf{d}}$ and $\sigma_{\mathbf{d}}$ refer to the spread's mean and standard deviation during the formation period only. These values are frequently used for the subsequent trading thresholds.

Fig. 1 demonstrates an example of a multivariate pair with (FFIV, DAL) belonging to one component and (WAT, YHOO, PM, BFB, KMX) belonging to another component. The covered GICS (Global Industry Classification Standard) codes are quite diverse including Communications Equipment (FFIV), Airlines (DAL), Health Care Distributors (WAT), Internet Software & Services (YHOO), Tobacco (PM), Distillers & Vintners (BFB), and Specialty Stores (KMX).[4] There is no clear relation between these seven companies, so this could very well be spurious or due to some complex latent factors. Nevertheless, this example formed with historical price data yields profitable opportunities for the out-of-sample trading period denoted after the vertical line.

Although one could certainly reduce the pair's cardinality in Fig. 1, it presents the potential benefit of including multivariate pairs in an overall pairs trading strategy.

2.2. Multi-objective optimization

Most real-world problems have conflicting objectives that require trade-off decisions. For example, portfolio optimization is inherently a multi-objective problem with a trade-off between risk and return. Traditional Markowitz portfolio optimization, however, was formulated as a single objective problem by modeling either the expected return or variance as a target constraint. Solutions to the problem are frequently visualized in the multi-objective space referred to as the efficient frontier. This frontier is obtained by iter-

[1] For example, forming clusters of products that have comoving inventory levels. A deviation of some magnitude might signify an issue and trigger actions to be taken

[2] Also, frequently referred to as Brownian motion in the continuous time space. Some may dispute that assets to not follow a random walk precisely, but the lack of individual predictability holds true.

[3] For this paper, all prices are normalized to 1 at the start of the formation period.

[4] The full names are F5 Networks (FFIV), Delta Airlines (DAL), Waters Corporation (WAT), Yahoo! Inc.(YHOO), Philip Morris International (PM), Brown-Forman Corp. (BFB), and Carmax Inc (KMX).

atively changing a target constraint and resolving for a new point on the frontier.

Multi-objective optimization treats each function independently by forming a frontier of non-dominated solutions. A solution dominates another if all objective values are no worse and at least one is distinctly better. As a combination of objective functions, Pareto fronts can be linear, convex, non-convex, and discontinuous. NSGA-II is an algorithm proposed by Deb, Pratap, Agarwal, and Meyarivan (2002) that sorts candidate solutions in nondominated sets within each iteration of a genetic algorithm to form a Pareto front of non-dominated solutions for multi-objective problems.

Definition 3. Pareto dominance: A solution of i objectives in $\mathbf{f^a}$ dominates another solution $\mathbf{f^b}$ if and only if at least one objective f_i^a is better than f_i^b and all other objectives f_i^a are no worse than f_i^b for $i \in \{1, 2, ..., k\}$. For minimization, we have the following conditions for dominance:

- $\mathbf{f}^a \preceq \mathbf{f}^b$
- $\{f_i \mid f_i^a < f_i^b, f_i \in \mathbf{f}\}$

Definition 4. Pareto optimal: A candidate solution is considered Pareto optimal if it is non-dominated, such that no other candidates show improvements without degrading another objective.

3. Problem definition

An overall pairs trading strategy consists of two main decision problems: *pair formation* (PF) and *pair trading* (PT). We define the PF problem below.

Definition 5. PF Problem: Given a set of n stocks,[5] assign weights to a subset so that the spread between the resulting long and short component, each consisting of up to v stocks, exhibits a high degree of profit potential.

3.1. Decision variables

For a total of n possible stocks to choose from, those belonging to each long/short component of a pair are represented by the Boolean vectors $\mathbf{x^c} \in \mathbb{B}^{n\times 1}$, where $c \in L, S$ indicates the long or short component. The corresponding weights assigned to each chosen stock in the long/short components of a pair are represented by the positive Real vectors $\mathbf{w^c} \in \mathbb{R}_+^{n\times 1}$. Therefore, there are four total vectors of length n as the decision variables.

If, for example, one has 100 stocks to choose from and wants to find an optimal multivariate pair with component sizes of 3 each, they need to find the optimal settings of 12 out of 400 variables. The mixture of cardinality-constrained Real and Boolean decision variables from which only a small subset of assets are selected leads to a non-convex and sparse solution space. Moral-Escudero, Ruiz-Torrubiano, and Suárez (2006) prove how the cardinality constrained portfolio optimization problem in NP-hard due to the asset selection part of it, which they relate to the NP-compete *Subset Sum* problem. The PF problem shares this same asset selection subproblem since one must choose stocks for each long/short components of the pair.

3.2. Constraints

The cardinality for each component of a pair is controlled by constraint (2) where $\mathbf{1}$ is a $n \times 1$ vector of ones. It ensures the number of chosen stocks satisfies a lower bound ℓ and upper bound u.

$$\ell \le \mathbf{x}^{c\prime}\mathbf{1} \le u \tag{2}$$

[5] Pairs can span different types of assets, but only stocks are referenced for brevity.

To prevent stocks belonging to both long and short components, constraint (3) requires each $x_j^L + x_j^S \le 1 \; \forall j \in 1...n$. A chosen stock can belong to either the long or short component of a pair, but not both.

$$\mathbf{x}^L + \mathbf{x}^S \preceq \mathbf{1} \tag{3}$$

All weights range between $0 \le w_j^c \le 1$ for stocks in each component, while those stocks not chosen are forced to $w_j^c = 0$. This is accomplished by constraint (4) because $x_j^c = 0$ if the asset is not chosen for component c. Weights greater than one would result in a leveraged position which we do not consider in this study.

$$0 \preceq \mathbf{w}^c \preceq \mathbf{x}^c \tag{4}$$

The last constraint (5) ensures that the weights for each component sum to one. By doing so, the constraint forces equal amounts of capital being allocated to each side of the pair, resulting in a dollar-neutral bet. For example, this constraint avoids cases where twice as much capital is allocated to the short side than the long side of a pair.

$$\mathbf{w}^{c\prime}\mathbf{1} = 1 \tag{5}$$

A set of objectives are optimized instead of aggregating them into one value. The vector $\mathbf{f} = [f_1, f_2, ..., f_k]$ represents the set of multiple objectives functions to be minimized over the problem specifications.

3.3. General multi-objective PF model

The following formulation shows the multi-objective MIP representation of a multivariate pair formation model. If only univariate pairs are desired, parameters ℓ and u are simply set to one in constraint (2).

$$\begin{array}{lll}
\text{minimize} & \mathbf{f} = [f_1, f_2, ..., f_k] & \\
\text{subject to} & \ell \le \mathbf{x}^{c\prime}\mathbf{1} \le u & (2) \\
 & \mathbf{x}^L + \mathbf{x}^S \preceq \mathbf{1} & (3) \\
 & 0 \preceq \mathbf{w}^c \preceq \mathbf{x}^c & (4) \\
 & \mathbf{w}^{c\prime}\mathbf{1} = 1 & (5) \\
\text{and} & c \in \{L, S\} & \\
 & \mathbf{x}^c \in \mathbb{B}^{n\times 1} & \\
 & \mathbf{w}^c \in \mathbb{R}_+^{n\times 1} &
\end{array}$$

The spread calculation is vectorized in order to simplify the notation in objective functions. Let $\mathbf{A} \in \mathbb{R}^{n\times m}$ be the normalized price matrix for n stocks across m price observations and $\mathbf{d} \in \mathbb{R}^{1\times m}$ be the resulting spread vector for the pair's formation period of length m. The spread vector $\mathbf{d}$, as shown in Equation (6) below, is the product of $\mathbf{A}$ for the selected assets $\mathbf{x^c}$ along their respective weights $\mathbf{w^c}$. This also simplifies the model because both components are restricted to non-negative weights.

$$\mathbf{A}'(\mathbf{w}^L - \mathbf{w}^S) = \mathbf{d} \tag{6}$$

SSD is commonly used in existing approaches, as in Gatev, Goetzmann, and Rouwenhorst (2006), which is merely a least squares estimation, or the spread vector's ℓ^2-norm. In some techniques, an additional criteria for the number of times the spread crosses zero (NZC) is used, as shown in Equation 7 below.

$$NZC = \textstyle\sum_{t=2}^{m}[sgn(d_t) \neq sgn(d_{t-1})] \tag{7}$$

3.4. Benchmark single-objective model

Although NZC is an important criteria from literature, the combination of both minimizing SSD and maximizing NZC is redundant since a lower SSD often implies higher NZC. The two objectives are highly correlated which are demonstrated later in Figs. 8, 10, and 12. We therefore use this combination of objectives to represent the *Benchmark Single-objective Model* in Equation 8.

$$\mathbf{f}_{BS} = [||\mathbf{d}||_2, -NZC] \tag{8}$$

Table 1
Pair combinations.

		Max component size (c)				
		≤ 1	≤ 2	≤ 3	≤ 4	≤ 5
Total Stocks (n)	10	45	1225	4950	124,750	499,500
	50	1485	10^5	10^7	10^9	10^{11}
	100	15,225	10^8	10^{10}	10^{14}	10^{16}
	500	73,920	10^{10}	10^{12}	10^{18}	10^{20}
	1000	10^5	10^{12}	10^{15}	10^{22}	10^{25}

3.5. Distance multi-objective model

This alone will find multivariate pairs, but against sub-optimal search criteria. The improved model needs one criteria to maximize profitability and another one to minimize risk of a pair's spread permanently diverging. This is accomplished by simply maximizing[6] spread volatility (σ_d) instead of minimizing SSD.[7]. A third objective of minimizing the final spread's magnitude (on day m) is also introduced to prevent spreads that are "running away". The set of three objectives is shown in Equation 9, which is referenced as the *Distance Multi-objective Model*.

$$\mathbf{f}_{DM} = [-\sigma_{\mathbf{d}}, -NZC, |d_m|] \tag{9}$$

3.6. Cointegrated multi-objective model

An alternative approach to pairs trading is based on the statistical property of cointegration. As will be discussed in the literature review, two time series are cointegrated if there exists a stationary linear combination between them. This linear combination represents the pair's spread so having a stationary spread fluctuating around some average would naturally present attractive trading opportunities. Genetic algorithms, and gradient-free approaches in general, are capable of minimizing the test-statistic from a cointegration test because they are simply evaluating the test instead of calculating gradients. For this reason, the test-statistic from the Augmented Dickey Fuller test for stationarity is used in the *Cointegrated Multi-objective Model* which maximizes volatility and minimizes the test statistic for stationarity, denoted by *ADF* in Equation 10. The ADF test statistic replaces the NZC measure as a means to finding mean-reverting spreads. Lower t-stats indicated higher likelihoods of stationarity. This model can be viewed as a pseudo-cointegrated framework since we are testing the linear combination of our multivariate pairs for stationarity, but not necessarily placing a confidence threshold level.

$$\mathbf{f}_{CM} = [-\sigma_{\mathbf{d}}, -ADF] \tag{10}$$

3.7. PF problem structure: combinatorial aspect

The constrained search space of existing univariate methods limit the potential trading opportunities, and one must compete with other market participants on many of the same pairs. The combinatorial nature lends itself to a harder problem and nullifies existing enumerative PF techniques.

Given n different stocks, Equation (11) yields the number of pair combinations P for component sizes ranging up to u. Forming a pair of two unique multivariate components can be viewed as a combination of combinations that grows exponentially as shown in Table 1.

$$P = \binom{\sum_{v=1}^{u} \binom{n}{v}}{2} \tag{11}$$

[6] In our algorithm, all objectives are minimized so any functions to be maximized have their values multiplied by −1

[7] Standard deviation and SSD have the same effect since SSD gives rise to variance. Standard deviation is simply a better representation of volatility and therefore profit potential of a pair's spread

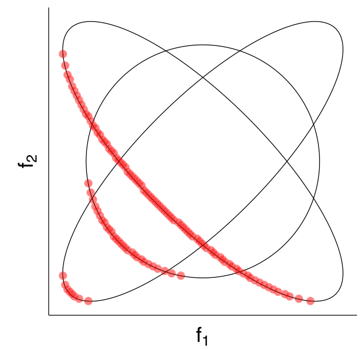


Fig. 2. Pareto fronts for correlated objectives where the feasible space is constrained to the 95% confidence interval of multivariate Gaussian distribution with correlations $\rho = 0.9$, $\rho = 0$, and $\rho = -0.9$, respectively.

Evaluating all 73,920 univariate pair combinations for 500 stocks takes only a few seconds on any modern machine, regardless of evaluating Euclidean distance or for cointegration. With component sizes of up to two, however, the enumeration time jumps to 58.7 hours. For sizes up to 5, just 50 candidate stocks equate to as many combinations as there are stars in our galaxy while an instance with 500 candidates exceeds the grains of sand on earth[8], the latter taking 1.8×10^8 years!

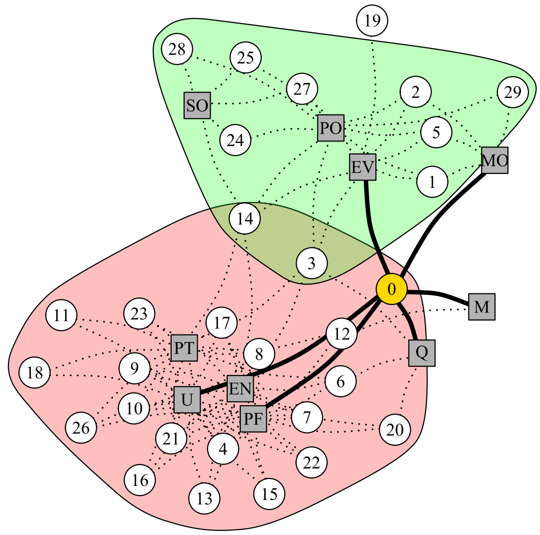
3.8. PF problem structure: multi-objective aspect

Existing PF methods have limited profitability due to the unintended minimization of spread volatility when selecting pairs with minimum *Sum of Squared Differences* (SSD). One can ameliorate this deficiency by introducing terms that encourage some degree of mean-reversion while exhibiting higher volatility. The naïve approach of linearly combining each term leads to difficulties in determining appropriate weights but, as Deb (2014) discusses, multi-objective methods address this issue by treating each objective independently to build a set of Pareto-efficient non-dominated solutions.

We take into consideration the work of Verel, Liefooghe, Jourdan, and Dhaenens (2013) when exploiting the problem structure of the PF problem. Correlated objective functions influence the search space and resulting Pareto fronts, so care must be taken to avoid highly correlated objectives in one model. One can see that the Pareto front's size is inversely related to objective correlations in Fig. 2. This is a natural characteristic since multiple objectives can reduce into a single representative objective as correlations approach $\rho = 1$. Therefore, the design of the PF models should exploit objective correlations to yield diverse Pareto fronts.

4. Literature review

Krauss (2017) provides a comprehensive survey of pairs trading literature which clusters the techniques in the following categories: *Distance, Cointegration, Time-Series, Stochastic Control*, and other approaches. The Distance and Cointegration categories tend towards the PF problem while Time-Series and Stochastic Control concern the PT problem. The latter two usually assume pairs are given a priori and instead focus on issues concerning trading thresholds, regime detection, and optimal trade allocations. We also briefly review some applications of optimization in finance, which have mostly focused on either PT or portfolio optimization problems.



Node	Description	Node	Reference
PF	Pair Formation subproblem	10	Do and Faff (2012)
PT	Pair Trading subproblem	11	Elliott et al. (2005)
PO	Portfolio optimization	12	Galenko et al. (2012)
U	Univariate	13	Gatev et al. (2006)
Q	Quasi-multivariate	14	Huang et al. (2015)
M	Multivariate	15	Huck (2010)
SO	Single objective optimization	16	Huck and Afawubo (2014)
MO	Multi-objective optimization	17	Huck (2015)
EN	Enumeration	18	Lin et al. (2006)
EV	Evolutionary optimization	19	Lin and Cao (2008)
0	This Paper	20	Perlin (2007)
1	Anagnostopoulos and Mamanis (2010)	21	Perlin (2009)
2	Anagnostopoulos and Mamanis (2011)	22	Rad et al. (2016)
3	Burgess (2000)	23	Smith and Xu (2017)
4	Chan (2013)	24	Valle et al. (2014a)
5	Chen et al. (2013)	25	Valle et al. (2014b)
6	Chen et al. (2017)	26	Vidyamurthy (2004)
7	Clegg and Krauss (2018)	27	Woodside-Oriakhi et al. (2011)
8	Do et al. (2006)	28	Woodside-Oriakhi et al. (2013)
9	Do and Faff (2010)	29	Xidonas and Mavrotas (2014)

Fig. 3. A representative network of selected literature generated in R using the package *igraph* from Csardi and Nepusz (2006). Key topics are represented as shaded square nodes while individual papers are circular white nodes. The two shaded regions represent the problem areas of Pairs Trading and Portfolio Optimization. As an example, our paper (highlighted in yellow as node 0) belongs to the pairs trading problem space but has strong connections to problems in portfolio optimization. It connects the multi-objective optimization and evolutionary algorithms to the PF problem and single/quasi/multivariate pairs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 3 displays the relationships between selected literature most relevant to our study[9]. Research in the realms of univariate pair formation and portfolio optimization are evidently crowded but mostly disjoint from one another. Enumeration techniques dominate existing PF approaches, while multivariate pairs literature is almost nonexistent. We connect a new combination of problem areas in an attempt to bridge the gap between the PF problem with multivariate pairs through multi-objective optimization. No work regarding optimization for pair formation, whether single or multi-objective, was encountered throughout our research.

4.1. Distance technique

Distance measures, like SSD, are used to identify suitable pairs under this parameter-free approach. The most prominent study is by Gatev et al. (2006) and Goetzmann, Rouwenhorst et al. (1998), hereafter GGR. Working in the univariate realm, the SSD of cumulative returns is calculated between each component for all possible pair combinations over a 12-month formation period. Only pairs yielding the lowest SSD values are passed on to the trading period.

Perlin (2007) uses quasi-multivariate pairs by forming a synthetic long-only index for each univariate counterpart. However, the procedure only trades the univariate component of each pair so this more representative of a general mean-reversion strategy where the univariate component is expected to revert to its synthetic index.

Do and Faff (2010) review and analyze the work of GGR to find diminishing profitability of distance approaches which are often burdened by non-converging pairs. A consequence of traditional distance approaches, as highlighted by Krauss (2017), is that minimizing for SSD as a single objective yields lower spread volatility and therefore limited profit potential. Do and Faff (2012) later introduce the number of zero-crossings (NZC) which acts as a mean-reversion proxy measure. However, NZC adds limited value in their application because SSD is already minimized, thus implying higher NZC values.

Huck (2010) employs neural network to forecast multi-period returns and then ranks different univariate pair combinations using the multicriteria decision-making technique Electre III. Chen, Chen, Chen, and Li (2017) employ Pearson correlation over a five year formation period of monthly returns for quasi-multivariate pairs. For each stock, they gather the 50 most correlated counterparts into a basket of paired stocks. The procedure relates to Perlin (2007) where only the univariate components trade against the reference basket index which does not necessarily make this a pairs trade.

4.2. Cointegration technique

Cointegration differs from distance methods by testing the resulting spread series for stationarity. Most of the theory from these two approaches draw from the original works of Dickey and Fuller (1979), Engle and Granger (1987), and Johansen (1988). Vidyamurthy (2004) optimizes trading thresholds for pre-selected univariate pairs through a screening process based on statistical and fundamental relationships before testing for cointegration. The screening has a dual purpose of domain reduction for expensive cointegration tests and also a latent selection bias towards inter-industry pairs.

Galenko, Popova, and Popova (2012) present a transparent methodology to form a cointegrating vector from multiple assets for multivariate pairs trading. It does not focus on the search for multivariate pairs but instead develops a framework to establish a cointegrating vector from a given set of constituents. For this reason, and the proposed trading rules for the daily rebalancing, we believe this paper addresses the PT problem. Our methodology would make a natural complement to their approach as a source of generating candidate pair constituents.

Chan (2013) highlights the disadvantages of univariate pairs trading in today's market and suggests employing ETF pairs as a viable alternative that reduces transaction costs and limits risk of permanent regime-shifts. Huck and Afawubo (2014) show that, on a trade by trade basis, pairs formed through cointegration methods are not only more reliable but also yield higher spread volatility than those selected with distance methods.

Rad, Low, and Faff (2016) compare large-scale implementations of Gatev et al. (2006) and Vidyamurthy (2004), where both strategies are backtested from 1962 to 2014 using univariate pairs on the normalized price level. The cointegration method first filters candidate pairs with 12-month SSD before applying an Engle-Granger cointegration test where only the top 20 cointegrated pairs with

[8] An estimated 3×10^{11} stars and 7.5×10^{18} grains of sand exist.

[9] Papers are categorized manually on a subjective manner so this is just an approximation to visualize the current literature space and may not be exact.

minimum SSD are traded. As noted by Krauss (2017), the SSD sorting heuristic likely hinders the returns for the cointegration strategy because only the smallest SSD pairs reach the cointegration tests. Minimizing the SSD as a single objective consequentially reduces spread variance so the procedure could quite possibly be selecting the least-profitable cointegrated pairs. Huck and Afawubo (2014) showed how pairs formed purely on cointegration yield higher spread volatility, so the only benefit from overlaying a SSD sorting procedure is the resulting domain reduction for enumeration. Clegg and Krauss (2018) relax the stationarity condition for cointegration to develop a partial cointegration model which, in their analysis, outperforms the benchmark model of Gatev et al. (2006) and two other cointegration variants.

4.3. Related optimization techniques in finance

Most applications have focused on either portfolio optimization, forecasting, or memetic algorithms (such as optimizing a trading strategy's parameters). Those most related to pairs trading belong under problem areas of arbitrage, portfolio selection, and trading. Aguilar-Rivera, Valenzuela-Rendón, and Rodríguez-Ortiz (2015) provide a recent review of the current literature space for evolutionary algorithms in finance.

Burgess (2000) applied evolutionary algorithms to select a portfolio of strategies that maximize a trade-off between risk and return. Lin and Cao (2008) use a robust genetic algorithm to find actionable sets of trading rules that satisfy different constraints. Lin notes the efficiency of a GA search mechanisms as being only 1% of the execution time compared to enumerating algorithms. Chen, Huang, and Hong (2013) apply a multi-objective GA to maximize historical returns while minimizing risk for the portfolio optimization problem.

Woodside-Oriakhi, Lucas, and Beasley (2011) solve the single-objective cardinality-constrained portfolio optimization problem in a 2-step methodology. Anagnostopoulos and Mamanis (2010) evaluate three different multi-objective evolutionary algorithms for the same problem but with three objectives. Similar work by Anagnostopoulos and Mamanis (2011) compares five multi-objective evolutionary algorithms (NSGA-II, SPEA2, NPGA-II, PESA, e-MOEA) for the cardinality constrained portfolio optimization problem, from which the PF problem draws similarities.

Huang, Hsu, Chen, Chang, and Li (2015) use GA to simultaneously optimize capital allocation between a small set of candidate pairs and trading signals in an overall mean-reverting trading system. The GA runs on training data to optimize the optimization period returns with decision variables of trade allocations and signal parameters. The approach from Huang et al. (2015) certainly shares some commonalities with this paper, such as mixed chromosome types, but some distinctions must be emphasized. Given that only one multivariate pair is optimized within the same ten stocks, some might argue that this reduces to a long/short portfolio optimization problem since it lacks an ensemble of simultaneously traded pairs. This approach would be intractable for larger sets of stocks due to the sparsity of stock weightings in their fixed chromosome length. A variable chromosome length, as outlined in this paper, would help amend that shortcoming.

Woodside-Oriakhi, Lucas, and Beasley (2013) establish a method of rebalancing established portfolios under transaction costs and investment horizons across a variety of useful constraints. The approaches in both Valle, Meade, and Beasley (2014a) and Valle, Meade, and Beasley (2014b) focus on building portfolios yielding consistent returns irrespective of market conditions. Although entirely different than pair formation, their MIP models share similar characteristics to the PF problem. The model allows for long-short combinations of asset weights, includes cardinality constraints, enforces dollar-neutrality, and places directional position constraints which are important for the PF problem.

4.3.1. Literature summary.

Based on the current state-of-the-art for pairs trading literature, there is an apparent absence in the multivariate and multi-objective space. No technique, other than brute force enumeration, has used any form of optimization to find profitable pair combinations. Table 2 shows existing areas covered by state of the art literature.

Enumeration remains a viable option for just about all existing state-of-the-art because of their constrained univariate solution space. Formation criteria has mostly only varied between some sequence of SSD minimization, NZC, industry classifications, and cointegration tests. Sorting pairs on different criteria is less effective than finding Pareto-optimal pairs through multi-objective optimization. Given the state of existing literature, the following points are clear:

- Diminishing profitability of univariate pairs, possibly due to the limited candidate pool
- Absence of optimization methods applied to the PF problem
- Existing approaches are intractable for multivariate pairs
- There exists conflicting objectives of risk (mean-reversion) and profitability (volatility)

5. Methodology

We solve the PF problem by forming a population of candidate pairs with profitable spread characteristics through the use of a multi-objective genetic algorithm called NSGA-II. GA is a population-based method that advantageously handles both discrete and continuous variables, thus making it better suited for the PF problem as opposed to other common metaheuristics. There is also more flexibility in designing constraints with GA and potential to parallelize for large scale computing as described in Dao, Abhary, and Marian (2017) and Sonmez and Bettemir (2012). As noted in Tahir and Smith (2007), GA has also been shown to effectively handle problems having exponential and noisy search spaces like the PF problem.

A metaheuristic, instead of traditional scalarization techniques, is employed for a number of reasons. The interested reader can refer to section four of Emmerich and Deutz (2018) for a detailed comparison of metaheuristics and scalarization techniques for multi-objective problems. First, care must be taken to normalize multiple objectives to the same unit of measure when combining them into a single weighted linear or non-linear objective function. NSGA-II is more robust to such nuances since candidates are evaluated on Pareto dominance criteria. Second, weighted max-min models require the decision maker to choose what magnitude to assign each weight. Small adjustments to the objective's weights can lead to drastically different solution types. NSGA-II allows the freedom to swap different objective functions to test a range of models with minimal added effort. Lastly, a weighted approach returns just one single solution, but a pairs trading strategy benefits from having an entire set of candidate pairs to trade. NSGA-II returns a Pareto front of solutions by improving a diverse population of candidates, which lends itself nicely to pairs trading due to the diversifying effect. The decision maker has a range of trading opportunities to choose from instead of just one solution.

5.1. Genetic representation for NSGA-II

The genetic algorithm (GA) is a type of evolutionary algorithm belonging to a broader class of metaheuristics. They are derivative-free methods that do not place any convexity requirements on constraints and cost functions, which is useful for non-smooth functions and search spaces. Originating from the notable works of Fogel (1964) and Holland (1992), the algorithm simulates Darwinian natural selection by recursively improving a set (population) of candidate solutions (individuals). The quality of candidate

Table 2
Coverage of PF problem characteristics.

	Univariate	Multivariate	Enumeration	Optimization	Multi-objective
Existing	√		√		
This Paper	√	√		√	√

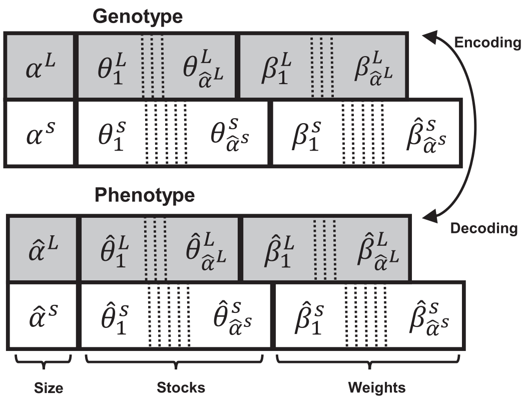


Fig. 4. Solution representation.

solutions depends on its solution representation where the decision variables are encoded in a phenotype and evaluated by the objective (fitness) function. The procedure undergoes a selection phase at the end of each iteration (generation) where only the fittest individuals survive to create new candidates for the next generation. Candidates are improved through the crossover (breeding) and mutation modification procedures.

Definition 6. Genotype: The set of variables on which crossover and mutation operate. Variables of a genotype are called genes. Chromosomes are sets of genes that represent different types of decision variable.

Definition 7. Phenotype: The translated genotype values that represent the true decision variables. The phenotype values are used in fitness evaluation and to ensure constraint feasibility.

Definition 8. Solution encoding/decoding: The process of translating decision variables to a genotype and back to their representative state in the phenotype.

5.1.1. Solution representation

Solution structures play a direct role in the fitness landscape and constraint handling. Fig. 4 shows how each individual solution is structured. Floating vector sizes (or lists) are employed due to the sparse nature of the problem. By doing so, we prevent the algorithm from adjusting the weights on stocks that are not included in the pair.

Parameters α, $\boldsymbol{\theta}$, and $\boldsymbol{\beta}$ are the real-valued decision variables between [0,1] whose values are tweaked through crossover and mutation operators. This α parameter is translated to $\hat{\alpha}$ which is the integer representation of the cardinality of each component, which also dictates the list sizes for $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$. The parameter $\boldsymbol{\theta}$ represents the selected asset indices and $\boldsymbol{\beta}$ contains their respective weights. Similar to α, $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ are translated to their actual evaluation forms through encoding. The following equations show how a candidate's genotypes are translated to their phenotypes:

$$\hat{\alpha}^c = \min[u, \ell + \lfloor \alpha^c \cdot (1 + u - \ell) \rfloor] \tag{12}$$

$$\hat{\boldsymbol{\theta}}^c = \min[n, \lfloor \mathbf{1} + n\boldsymbol{\theta}^c \rfloor] \tag{13}$$

$$\hat{\boldsymbol{\beta}}^c = \frac{\boldsymbol{\beta}^c}{|\boldsymbol{\beta}^c|} \tag{14}$$

Any change in α that translates to a new $\hat{\alpha}$ value has a direct impact on the $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ lengths. Therefore, the entire genotype and phenotype must undergo a problem-specific reconstruction step, detailed in Algorithm 1. This is similar to how Anagnostopoulos and Mamanis (2010, 2011) and Chang, Meade, Beasley, and Sharaiha (2000) employ a repair mechanism to enforce constraint satisfaction in their cardinality-constrained portfolio optimization problem. We treat this issue similarly by dropping those assets corresponding to the smallest weights until the constraint is satisfied. If the component size, $\hat{\alpha}^c$, increases then new stocks are added and assigned weights that match the current average $\boldsymbol{\beta}^c$. If the component size decreases, then assets corresponding to the lowest weights are removed.

5.1.2. Crossover

Crossover is synonymous to "breeding" and creates two new candidate solutions from a combination of two parents. It is meant to create an improved, but similar, combination of decision variables from both parents that ideally carry their best traits. Only one component from each parent is chosen because tweaking both components simultaneously would likely result in random noise. It is not component-constrained so Long-Short crossovers are just as likely as Long-Long and Short-Short, which enables a stock in one candidate's long component to crossover into another candidate's short component.

Parent selection for crossover undergoes an enhanced tournament selection procedure to encourage diversity in the crossover operations. The first parent is selected based on normal tournament of size three, but the second parent has a roulette wheel element to it. Instead of uniformly sampling for tournament competi-

Algorithm 1 Pair Reconstruction.

$\{\} * L_o, S_o \leftarrow P_o$ ▷ Original Long&Short components
$\{\} * \alpha^c_o, \boldsymbol{\theta}^c_o, \boldsymbol{\beta}^c_o, \hat{\alpha}^c_o, \hat{\boldsymbol{\theta}}^c_o, \hat{\boldsymbol{\beta}}^c_o$ where $c \in \{\} * L_o, S_o$

$\{\} * L_n, S_n \leftarrow P_n$ ▷ New Long&Short components
$\{\} * \alpha^c_n, \boldsymbol{\theta}^c_n, \boldsymbol{\beta}^c_n, \hat{\alpha}^c_n, \hat{\boldsymbol{\theta}}^c_n, \hat{\boldsymbol{\beta}}^c_n$ where $c \in \{\} * L_n, S_n$

$\mathbf{A} \leftarrow$ Set of potential stocks

for each $c \in \{\} * L, S$ **do**
  $\delta \leftarrow \hat{\alpha}^c_n - \hat{\alpha}^c_o$
  $m \leftarrow \hat{\alpha}^c_n$
  if $\delta < 0$ **then** ▷ Remove stocks with least weight
    $\boldsymbol{\beta}' \leftarrow \boldsymbol{\beta}^c_o$ sorted by descending $\boldsymbol{\beta}$
    $\boldsymbol{\theta}' \leftarrow \boldsymbol{\theta}^c_o$ sorted by descending $\boldsymbol{\beta}$
    $\boldsymbol{\beta}^c_n \leftarrow \boldsymbol{\beta}'_{1..m}$
    $\boldsymbol{\theta}^c_n \leftarrow [\boldsymbol{\theta}'_1, \ldots, \boldsymbol{\theta}'_m]$
  else if $\delta > 0$ **then** ▷ Add new stocks
    $\boldsymbol{\mu} \leftarrow \boldsymbol{\beta}^c_o / \hat{\alpha}^c_o$
    $\mathbf{B} \leftarrow \{\} * \hat{\boldsymbol{\theta}}^L_o, \hat{\boldsymbol{\theta}}^S_o$ ▷ All stocks in Pair
    $\mathbf{S} \leftarrow \mathbf{A} - \mathbf{B}$
    $\boldsymbol{\theta}' \leftarrow randomSample(\delta, \mathbf{S})$ ▷ Without replacement
    $\boldsymbol{\beta}' \leftarrow \boldsymbol{\mu}$
    $\boldsymbol{\theta}^c_n \leftarrow \{\} * \boldsymbol{\theta}^c_n, \boldsymbol{\theta}'$
    $\boldsymbol{\beta}^c_n \leftarrow \{\} * \boldsymbol{\beta}^c_n, \boldsymbol{\beta}'$
  end if
end for
return $Pair_{new}$.

tors, probabilities are assigned based on the competitor's similarity to the first parent. In particular, Hamming distances are computed between the first parent and each candidate's set of stock indices so that the most distinct candidates are assigned higher probabilities. This biases the procedure to select the fittest, yet diverse, individuals for crossover. It both saves on redundant crossover operations and helps maintain diversity in the resulting population.

Although α, $\boldsymbol{\theta}$, and $\boldsymbol{\beta}$ are all real-valued between [0,1], it is important to treat each chromosome uniquely. Take $\alpha = 0.1$ and $\boldsymbol{\theta} = [0.9, 0.95]$ with $\alpha = 0.2$ and $\boldsymbol{\theta} = [0.2, 0.2]$ for example. Both parents have small α values indicating small component sizes (in this case two stocks each). Crossover between α and $\boldsymbol{\theta}$ chromosomes could yield drastically different phenotypes who's component sizes nearly doubled, even though both parents had small component sizes. Most inter-chromosome crossovers would result in random noise and be counterproductive to the key ideas behind GA.

Given two parents, P_1 and P_2, the crossover procedure only operates on either α, $\boldsymbol{\theta}$, or $\boldsymbol{\beta}$ with the following probabilities. Note that, if the component sizes to be crossed are the same, then we assign a $p(\alpha) = 0$ to prevent redundancy.

$$\{p(\alpha), p(\boldsymbol{\theta}), p(\boldsymbol{\beta})\} = \begin{cases} \{0, \frac{2}{5}, \frac{3}{5}\}, & \text{if } \hat{\alpha}_{P_1} = \hat{\alpha}_{P_2} \\ \{\frac{1}{5}, \frac{2}{5}, \frac{2}{5}\}, & \text{otherwise} \end{cases}$$

The parameter $\boldsymbol{\theta}$ undergoes Double Point Crossover with a simple adjustment from Luke (2009) to allow for the crossover between different sized vectors. The main caveat being that the maximum segment length to crossover must be no larger than the smallest component size for both pairs. We use Extended Intermediate Recombination for both α and $\boldsymbol{\beta}$ genes. Since $\boldsymbol{\beta}$ has varying size, this operation must also be adjusted to accommodate different sized vectors. An offspring o has its i'th gene derived from two parents in a linear fashion by $o_i = p_{1i} \cdot \lambda + p_{2i} \cdot (1 - \lambda)$, where $\lambda \in [-0.25, 0.25]$ is a bounded random variable chosen uniformly. Implementation is straight forward from Mühlenbein and Schlierkamp-Voosen (1993).

5.1.3. Mutation

The only mechanism that adds exploration to the stock search-space so far is the reconstruction step if a component size increases. Mutation further encourages exploration by tweaking either α, $\boldsymbol{\theta}$, or $\boldsymbol{\beta}$ separately with the same chromosome-specific probabilities as crossover. Standard Gaussian convolution is used for α and $\boldsymbol{\beta}$ with a standard deviation of 0.15. Weighted integer randomization was chosen for θ which aims to substitute stocks of low-weight with new indices and weights as before in Algorithm 1. Since $\boldsymbol{\theta}$ is in non-metric space, a simple re-sampling of all stocks is sufficient, exclusive of those already in the pair.

5.1.4. Pareto dominance

It was necessary to introduce another scenario for Pareto dominance that prevents pairs with identical fitness values and component assets from belonging to the same frontier. A solution $\mathbf{f}^a$ dominates $\mathbf{f}^b$ if both fitness values and component-stocks are identical.

$$\mathbf{f}^a = \mathbf{f}^b \wedge \hat{\boldsymbol{\theta}}_a = \hat{\boldsymbol{\theta}}_b$$

It also prevents the algorithm from converging to a local optima because, if omitted, multiple versions of the same pair but with different component weights might exist in the Pareto front and retained for subsequent generations.

5.1.5. NSGA-II parameters

We use a population size of 50 with crossover percentage of 0.8 and mutation percentage of 0.4. The mutation probability depends on which chromosome is chosen as noted in Section 5. The algorithm terminates when the number of pairs belonging to the first Pareto-front reaches 50 at any point between the minimum (50) and maximum (150) generations. These parameters were chosen either based on common practice or trial-and-error. In particular, most would argue that a mutation rate of 0.4 is too high for a GA but this was found beneficial for the large discrete search space of stocks. Faster convergence was needed to alleviate the burden when backtesting each model in the experiments.

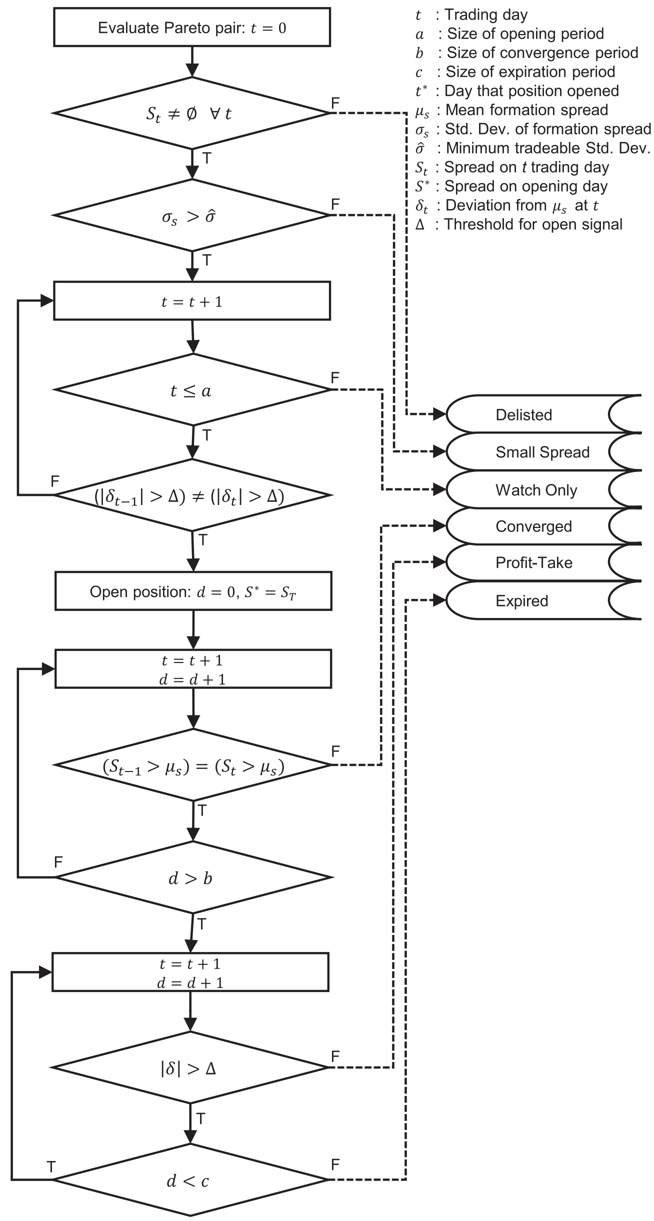


Fig. 5. Trading flowchart.

5.2. Trading rules

Fig. 5 shows the overall trading process for the experiments. All Pareto pairs from the optimization routine enter the trading process and flow into one of the six categories (shown on the right). We discard any pairs missing a price, S_t, during the trading period. This simplifying assumption is justified by the highly liquid and established S&P 500 constituents so missing prices are likely due to an M&A which is equally likely to result in a good or

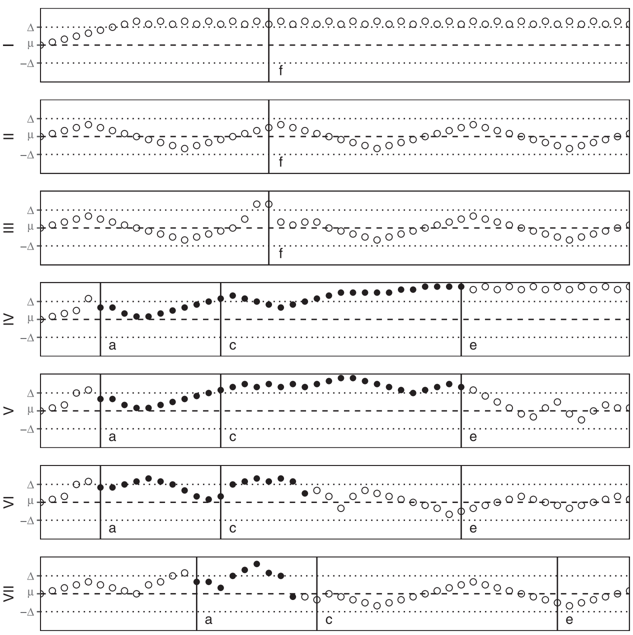


Fig. 6. Trading Scenarios: The upper and lower dotted lines denote opening thresholds while the middle line represents a closing threshold. The vertical lines are (f) final day of the allowable opening period, (a) open signal crossed, (c) last allowable convergence day, and (e) expiration day. Only (f) is static whereas the rest depend on when the position was opened denoted by (a). The filled circles represent days in which the position is open. The first, if any, filled circle corresponds with the day a position is opened and the last corresponds to the closing day.

bad trade. The “Small Spread” category captures any spread whose standard deviation, which represents the profit potential, is less than a threshold chosen to be $\hat{\sigma} = 0.025$. The “Watched Only” category tags any pairs that entered the trading system but never breached an open signal. “Converged” contains pairs that opened and also reached their specified close threshold, whereas “Expired” trades never reached that threshold. The “Profit-Take” category is for any pair that opened but failed to converge within the allowable “convergence period” and therefore closed at the first moment the trade became profitable.

To further clarify the trading scenarios, Fig. 6 displays the different cases a pair might undergo during the trading period. Each depicts a hypothetical pair’s spread during the trading period.

Scenarios I, II, III are all “Watch Only” pairs. In case I, the spread crosses above the upper threshold but never reverts below. Case II shows a spread that never crosses either open limits. The spread in III crosses below the upper threshold but only after the allowable opening period. Scenarios IV and V show expired trades that were temporarily profitable but never reached the close threshold. They also never became profitable within the subsequence between (c) and (e) where it could have closed as a “profit-take”. In case V, however, the spread did mean-revert but after the allotted expiration period. Scenario VI is a “Profit-take” because it was closed after the first profitable day following the convergence period (c). Note that this also crossed the closing threshold but after the position was already closed. Scenario VII is the ideal case where the pair converged within the allowable convergence period.

6. Empirical experiment

6.1. Experiment overview

The goal of the following experiment is to investigate the following two points:

- *Single vs Multi-objectives:* How does pair formation criteria impact profitability?
- *Univariate vs Multivariate:* How does pair cardinality impact profitability?
- *Performance against benchmark:* How do the models perform against a buy-and-hold S&P 500 strategy?

We test the three models (BS,DM,CM) formulated in Section 3 with different perspectives on performance. The profitability is first analyzed on an individual pair basis and we interpret the resulting Pareto fronts across “Convereged” and “Expired” outcomes. Statistical tests are run for individual pair profitability before comparing the overall strategy performance of each model. The strategies are compared for annual and monthly returns on both committed and invested capital.

To examine the efficacy of including multivariate components in a pairs trading strategy, we equate the cardinality constraint in the DM model to $v = u = \ell$ which bounds the pair component sizes to $v \in \{1, 2, ..., 7\}$ for each variant. This creates seven different models, one being a pure univariate case, all sharing the same three objective functions as the DM model.

6.2. Experimental design

Daily close prices for index constituents of the S&P 500 are used for the stock universe which was obtained from Chan (2013). Real-world implementation details like transaction costs, borrow costs, taxes, corporate actions, and trade slippage are excluded since they lack consistent treatment in literature. The model formulations and trading system were developed and debugged using a test sample period between 4/25/2011 and 12/31/2011, prior to the out-of-sample period for the experiments.

The out-of-sample experiments take place between 11/27/2012 and 5/31/2016 where a sequence of 12 equally spaced optimization days are executed, each with their own formation look-back period of 400 days, to provide a sufficient sample size across varying market conditions. The frequency is about one formation period per quarter. This time span mostly falls within a bull market, but there is a pronounced sell-off period between 2015–2016. For each formation period, NSGA-II runs ten different times to supply a sufficient population of candidate pairs to the trading phase. Since some solutions in a formation period might be too similar, a final sorting procedure removes similar pairs within the same period. A pair is removed if at least 75% of its constituents match another pair or if one pair’s long/short component exactly matches another pair’s long/short component. For the entire test period, the number of candidate pairs P for each formation period is $P \leq 500$ which leads to a total of $P \leq 6000$ for the whole backtest. NSGA-II and the various adjustments were programmed in Matlab 2017b on an Intel Core i5-4210U processor with up to 2.40 GHz and 8 GB of RAM.

The trading parameters are uniform across all models. Opening thresholds depend on the pair’s average spread, $\mu_{\mathbf{d}}$, and standard deviation, $\sigma_{\mathbf{d}}$, from the formation period. They are set to $\mu_{\mathbf{d}} \pm 1.5\sigma_{\mathbf{d}}$ and with a closing threshold of $\mu_{\mathbf{d}}$. A pair is allowed 100 days to reach an open signal before being discarded. If opened, the allowable convergence period is set to 20 days and expiration period of 50 days. In other words, an opened pair has 20 days to completely converge plus an additional 30 days to become profitable.

The amount invested for each pair is constant at $1. There are more sophisticated approaches to bet sizing, such as the Kelly criterion, but such decisions belong under the PT problem. The daily return of pair j is given by Equation 15 where t is the pair’s spread as defined earlier in Equation 1.

$$r_{j,t} = \pm(d_t - d_{t-1}) \tag{15}$$

Table 3
BS: pair statistics.

		All	Traded	Converged	Profit take	Expired
	Count	1192	667	124	427	116
	Proportion (%)		56	10.4	35.8	9.7
$\sigma_{\mathbf{d}}$	Min	0.012	0.025	0.025	0.025	0.025
	Max	0.358	0.333	0.133	0.333	0.236
	Median	0.037	0.041	0.034	0.043	0.043
	Mean	0.05	0.053	0.041	0.055	0.057
NZC	Min	1	1	6	1	1
	Max	76	74	71	68	74
	Median	26	23	28.5	22	19.5
	Mean	27.4	24.2	29.4	23.2	22.3
t-stat	Min	−5.29	−5.2	−5.2	−3.7	−4.71
	Max	2.73	2.36	2.36	2.23	2.33
	Median	−1.47	−1.3	−1.81	−1.15	−1.2
	Mean	−1.47	−1.24	−1.76	−1.13	−1.12
Return	Min		−0.263	0.009	0	−0.263
	Max		0.193	0.193	0.14	−0.001
	Median		0.017	0.052	0.017	−0.093
	Mean		0.009	0.059	0.025	−0.103
	Std. Dev.		0.063	0.028	0.025	0.059

The $\pm$ sign is negative only if the spread was shorted. Pair *j*'s total return *tr* (the cumulative gain or loss) can be expressed as the summation of daily returns $r_{j,t}$.

$$tr_{j,t} = \prod_{t=1}^{n}(1 + r_{j,t}) \tag{16}$$

6.3. Results

Tables 3–5 present descriptive statistics for individual pairs in each model. The “All Pairs” column covers all pairs generated, “Opened Pairs” only contains pairs that reached an open-signal in the trading period, and the remaining columns are as described earlier. The number of pairs represented for each category is shown in the top row. Below each table are violin plots for the pair returns across all 12 formation days to show how the returns are distributed. The black line shows the overall mean return for each formation day. The violin plots, which are box-plots that also show the probability density distributions, were generated with the R package *ggplot2* from Wickham (2016).

The marginal scatter plots were created using the R package *ggpubr* from Kassambara (2017). They display regions of the Pareto front that are more profitable by marginalizing the pairs in the ”Converged” and ”Expired” categories. Correlation between objective functions are also calculated, but across all pairs returned for that model.

6.3.1. Benchmark single-objective model (BS)

For the baseline model in Table 3, the optimization routines created 1192 unique pairs across the entire testing period, 667 of which reached an open signal. Out of the traded pairs, 124 fully converged and 116 expired with the most frequent category being 427 partially converged trades under the Profit-Take category. It is clear that the converged trades exhibited lower volatility, $\sigma_{\mathbf{d}}$, and higher NZC than both Profit-Take and Expired categories. The average t-statistics for ADF tests are also lower for converged pairs, indicating a stationary behavior of the spreads. The violin plots in Fig. 7 display relatively uniform return distributions across all formation periods with a limited range between maximum and minimums. Ten out of the 12 days yielded profitable average returns.

The plot of traded Pareto pairs in Fig. 8 indicates a larger presence of converged pairs in the region of low SSD and higher NZC. The marginal distributions peak in the same region for both categories on each axis, but the likelihood of converged trades is visibly

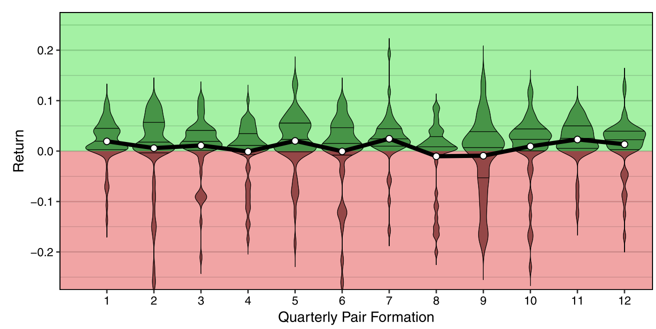


Fig. 7. BS model: violin plots for return distributions across formation periods.

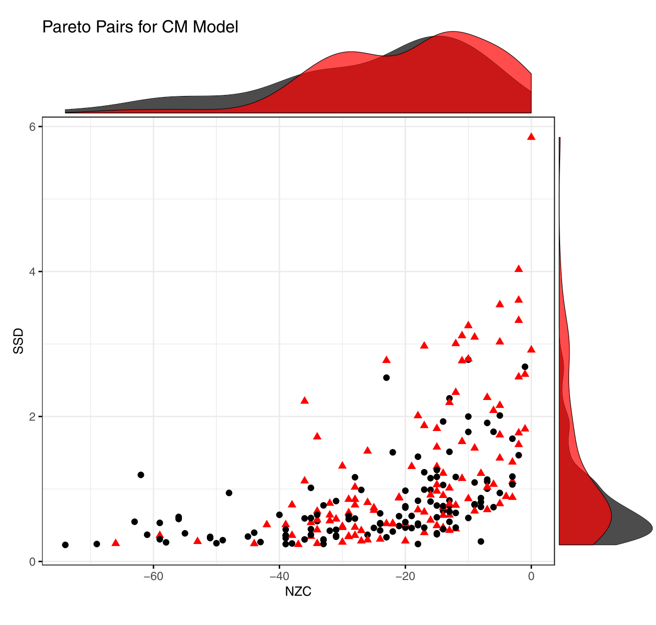


Fig. 8. Resulting Pareto pairs for the BS model against each formation objective. Only traded pairs from the "Converged" and "Expired" categories are displayed along with their marginal density distributions.

Table 4
DM: return statistics.

		All	Traded	Converged	Profit take	Expired
	Count	2529	1471	131	1069	271
	Proportion (%)		58.2	5.2	42.3	10.7
$\sigma_{\mathbf{d}}$	Min	0.018	0.025	0.026	0.025	0.028
	Max	0.826	0.826	0.337	0.826	0.689
	Median	0.152	0.124	0.058	0.135	0.134
	Mean	0.188	0.169	0.068	0.181	0.167
NZC	Min	1	1	5	1	1
	Max	70	70	70	56	53
	Median	9	11	28	9	11
	Mean	12.7	14.3	28	12.6	14.4
t-stat	Min	−4.8	−4.34	−4.34	−3.9	−3.47
	Max	3.59	3.48	1.85	3.48	3.29
	Median	−0.184	−0.403	−1.73	−0.238	−0.385
	Mean	−0.141	−0.364	−1.73	−0.184	−0.417
Return	Min		−0.727	−0.024	0	−0.727
	Max		0.568	0.568	0.495	−0.001
	Median		0.026	0.085	0.036	−0.139
	Mean		0.022	0.099	0.058	−0.156
	Std. Dev.		0.116	0.077	0.068	0.109

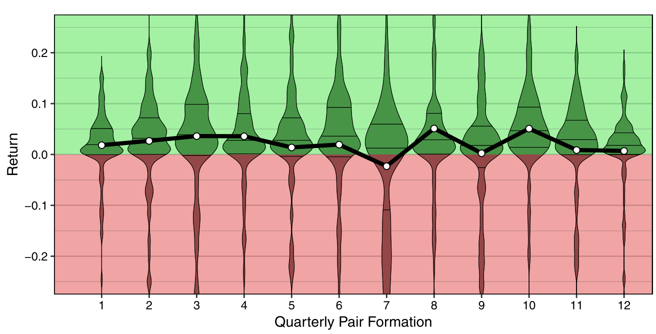


Fig. 9. DM model: violin plots for return distributions across formation periods.

greater than the expired category as NZC exceeds 30 in the formation period. The correlation coefficient for all optimized pairs, regardless of trading outcome, is $\rho = 0.59$. The high correlation between objectives supports how this model is used as a single-objective benchmark.

6.3.2. Distance multi-objective model (DM)

The DM model yields the highest proportion of traded pairs from minimizing the final formation spread. Table 4 shows that about 80% of traded pairs are profitable even though the amount of expired trades more than doubles the converged category. The spread variance and zero-crossing behavior are similar to the other models where converged pairs generally exhibit lower volatility and more zero-crossings during their formation period. We see a very high average return of 9.9% for converged trades and an impressive 2.2% for all trades. Fig. 9 shows all but one formation period had positive average returns and with higher averages than the other models.

Two plots of Pareto pairs are displayed in Fig. 10 to visualize the three objectives for DM. Both plots suggest that converged pairs are somewhat clustered along the same regions of the Pareto fronts. Converged pairs show more zero-crossings in the formation period and a smaller final formation spread $|d_m|$. As for volatility, there is a stronger concentration of converged pairs with lower volatility than expired pairs. For all optimized pairs, the correlations between $\sigma_{\mathbf{d}}$ and *NZC* is $\rho = -0.43$, and between $|d_m|$ and *NZC* is $\rho = 0.39$, which confirms how the chosen PF objectives are indeed conflicting criteria.

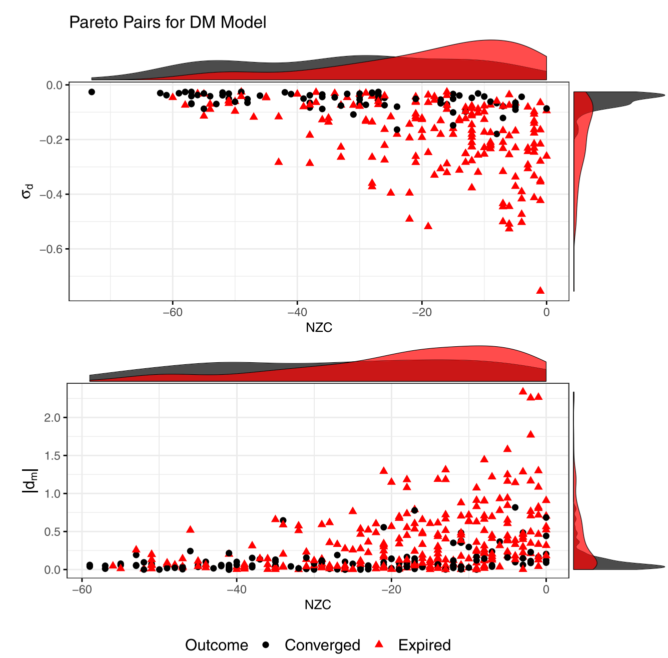


Fig. 10. Resulting Pareto pairs for the DM model against each formation objective. Only traded pairs from the "Converged" and "Expired" categories are displayed along with their marginal density distributions.

6.3.3. Cointegrated multi-objective model (CM)

The spread volatility and ADF *t*-statistic for converged pairs is much lower than both profit-take and expired categories in Table 5. ADF *t*-statistics of expired pairs are slightly lower than the partially-converged profit-take class which is counterintuitive. There is a substantial difference in NZC for converged pairs against the other two categories as well. The return distributions in Fig. 11 suggest another wide dispersion of returns like in the DM model but with only one formation period yielding a negative average return.

Fig. 12 shows how most pairs that fully converged had a *t*-stat less than −2 and volatility under $\sigma_d < 0.1$. Once again, there is clear region in the full Pareto front that is more desirable and has a higher presence of converged pairs. The correlation coefficient

Table 5
CM model: return statistics.

		All	Traded	Converged	Profit take	Expired
	Count	1860	1138	132	787	($n = 219$)
	Proportion (%)		61.2	7.1	42.3	11.8
$\sigma_{\mathbf{d}}$	Min	0.016	0.025	0.025	0.025	0.026
	Max	0.825	0.825	0.258	0.825	0.529
	Median	0.148	0.126	0.054	0.149	0.112
	Mean	0.185	0.172	0.07	0.195	0.151
NZC	Min	1	1	6	1	1
	Max	67	67	67	66	54
	Median	12.5	15	31.5	12	18
	Mean	16.2	18	31.3	15.5	18.7
t-stat	Min	−4.98	−4.65	−4.65	−4.38	−4.23
	Max	5.01	4.15	0.845	4.15	3.46
	Median	−0.793	−1.02	−2.41	−0.747	−1.22
	Mean	−0.747	−0.925	−2.37	−0.618	−1.16
Return	Min		−0.784	0.026	0	−0.784
	Max		0.585	0.362	0.585	−0.009
	Median		0.032	0.078	0.04	−0.154
	Mean		0.022	0.098	0.065	−0.178
	Std. Dev.		0.133	0.068	0.081	0.126

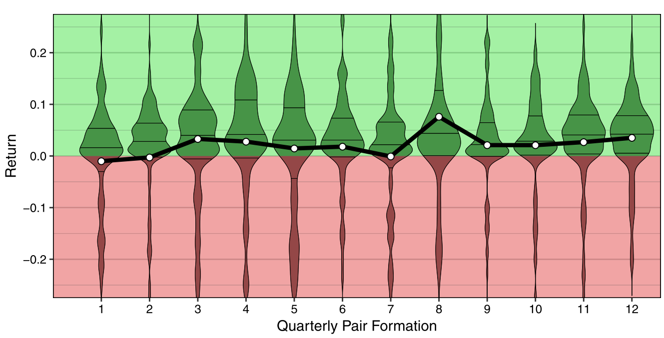


Fig. 11. CM Model: Violin plots for return distributions across formation periods.

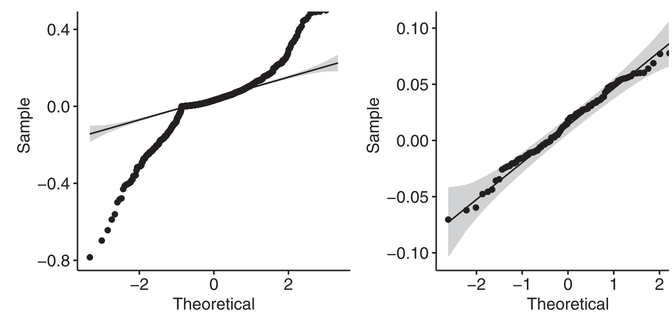


Fig. 13. Normal Q-Q plots before (left) and after (right) transforming trades by bins of 10. This suggests the transformed data is approximately normally distributed.

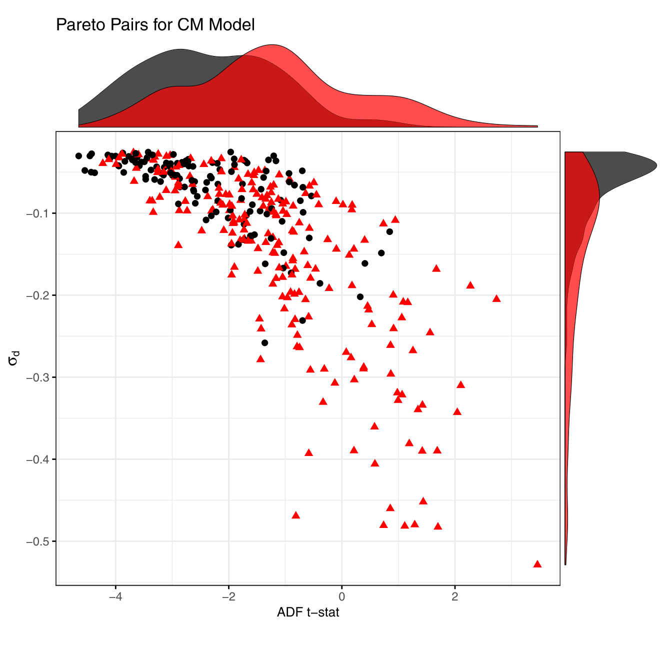


Fig. 12. Resulting Pareto pairs for the CM model against each formation objective. Only traded pairs from the "Converged" and "Expired" categories are displayed along with their marginal density distributions.

between the two objectives is $\rho = -0.7$ so, naturally, there is a strong negative correlation between stationarity and volatility.

6.4. Pair performance

In order to verify whether the optimized pairs from each model are profitable, we conduct Student's *t*-test with a null hypothesis of the average return being 0. The returns must first undergo a transformation to provide more Gaussian-like distributions for the *t*-test. We randomly sample all trades into subgroups of size 10, where those average return distributions are used for the *t*-test samples. The quantile-quantile plot in Fig. 13 demonstrates the before and after comparison of the transformed return data. All tests were conducted in R and the Q-Q plot was generated with the *ggpubr* package from Kassambara (2017).

The original returns are undoubtedly non-normal, where the profit-take rule in our trading system probably causes the sharp departure from normality. Results from a Shapiro–Wilk test for normality are included in Table 6 to demonstrate how this simple transformation yields approximately normal distributions. The purpose of this is to show how the transformed data is reasonable enough for the normality assumption in Student's *t*-test. Table 6 combines results from both tests along with the estimated confidence interval from the *t*-test.

The BS model presents the weakest case for generating profitable pairs since a *p*-value of 0.03 barely rejects the null at the $\alpha = 0.05$ significance level. The multi-objective models are both significantly different from an average return $\mu_r = 0$ with their

Table 6
The null hypotheses for the Shapiro–Wilk test (left) and t-test (right) are $H_o: \mathcal{N}(\mu, \sigma^2)$ and $H_o: \mu_r = 0$, respectively. Smaller values of test statistic, W, indicate a departure from normality.

	Shapiro–Wilk		Student's t-test			
Model	W	p-value	t-stat	df	p-value	99% C.I.
BS	0.977	0.243	2.17	65	0.0334	[−0.0008, 0.0082]
DM	0.988	0.249	5.62	146	9×10^{-8}	[0.0061, 0.0167]
CM	0.993	0.846	4.42	112	2×10^{-5}	[0.0047, 0.0183]

entire 99% confidence intervals in the profitable range. Given the above results, it is reasonable to infer that the use of multi-objective optimization is a valuable contribution to pair formation routines.

6.5. Strategy performance

We compare daily, monthly, and annualized returns on both committed and fully-invested capital to evaluate the overall performance of each model's trading strategy. These two return measures are used in other studies like Gatev et al. (2006), Do and Faff (2010), and Rad et al. (2016). Return on committed capital is more conservative and offers a better comparison between model variants whereas the fully-invested returns provide an adequate comparison against the S&P 500 benchmark.

Equation 17 calculates a strategy's daily return on committed capital. We observed that no strategy exceeded 100 open positions on any given day, so committed capital is set as $C = \$100$.

$$RCC_t = \frac{1}{C} \sum_{j \in \boldsymbol{P_t}} r_{j,t} \tag{17}$$

Equation 18 calculates a strategy's daily return on fully-invested capital, where $\boldsymbol{P_t}$ is the set of open positions and $|\boldsymbol{P_t}|$ represents the total number of open positions on day t.

$$RIC_t = \frac{1}{|\boldsymbol{P_t}|} \sum_{j \in \boldsymbol{P_t}} r_{j,t} \tag{18}$$

Fig. 14 shows the performance of each optimization model as a trading strategy. For comparison, the highlighted blue line represents the benchmark S&P 500 index. Since the system is placing $1 trades for each pair, this bottom figure can also be viewed as the proportion of committed capital invested throughout the test period.

The buy-and-hold benchmark strategy has a distinct advantage over the alternative models because it remains fully invested throughout the testing period. Return on invested capital highlights the outperformance of multi-objective models over both benchmark S&P strategy and the single-objective model. Not only does the BS strategy suffer from a lower return on invested capital, but it even lacks sufficient trading opportunities as indicated by the low level of invested capital. We can attribute this to the correlated objectives and also the trading filter that removes any trades with low formation variance.

The multi-objective models DM and CM grow by over 150% on fully-invested cumulative returns over the four year period, while the BS model underperforms the S&P. When observing the returns on committed capital, the DM model appears to perform best due to both the higher average returns and trading volume. The bottom chart shows that this model generally has the most capital invested at any given time and the BS has the least. There are recurring peaks across all models due to the quarterly pair formation routine. Running the optimization on a more frequent basis (perhaps monthly) could have increased the amount of invested capital.

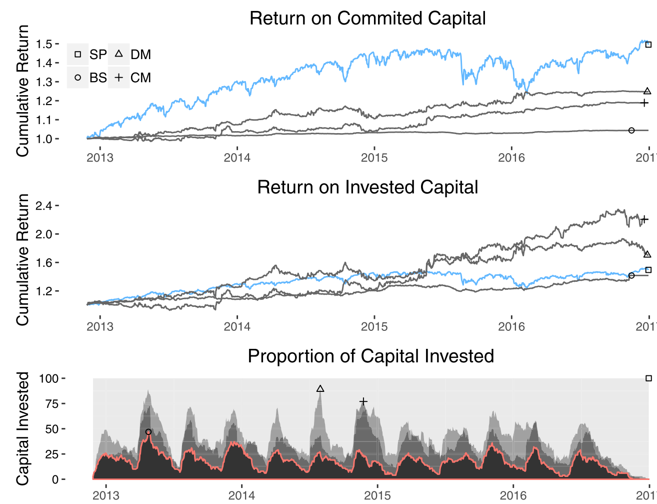


Fig. 14. Comparison of Strategy Performances. The blue line labeled SP represents the S&P buy-and-hold benchmark strategy. The top plot shows cumulative excess returns on committed capital, the middle plot shows the same but for fully-invested returns, and the bottom plot displays the amount of open trades for each strategy. The red line outlines the BS model's trades, which has the least positions, and the largest number of open positions corresponds to the DM model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 7
Monthly returns for fully-invested capital.

	S&P	BS	DM	CM
Mean	0.0085	0.0071	0.011	0.0172
SE (Newey-West)	0.003	0.002	0.004	0.005
t-Statistic (Newey-West)	2.72	3.07	2.64	3.22
Return Distribution				
Median	0.001	0.006	0.015	0.006
Std. Dev.	0.03	0.018	0.039	0.051
Skewness	−0.13	0.18	0.15	1.38
Kurtosis	−0.030	0.778	−0.10	2.89
Minimum	−0.068	−0.041	−0.069	−0.064
Maximum	0.081	0.062	0.18	0.29
Neg. Ret. %	0.38	0.36	0.38	0.42
Relative to S&P				
Avg Excess Return		−0.0013	0.0029	0.0086
SE (Newey-West)		0.003	0.0043	0.0053
t-statistic (Newey-West)		−0.455	0.576	1.163
Beta		0.17	−0.19	−0.03

6.5.1. Monthly returns

Monthly returns are presented in a similar format as Gatev et al. (2006) where each model's monthly return series is tested using Newey-West to account for autocorrelation and heteroskedasticity (Table 7). Only fully-invested returns are reported by compounding daily returns within each month and then regressing on an intercept-only linear model for the Newey-West test in R using the *sandwich* package from Zeileis (2004).

All models are profitable based on the t-statistics with the highest monthly average returns generated by the DM model. The row labeled "Neg. Ret. %" displays the proportion of months that experienced losses, which appears to be fairly constant across all four strategies. The BS model is the only one that does not beat the S&P benchmark in the relative return section. The CM model outperforms by an average of 0.86% per month. Beta is a popular measure that indicates correlation with overall market returns.[10] The low magnitudes support a market-neutrality characteristic for all

[10] We used the S&P to represent market returns

Table 8
Strategy comparisons on performance of Employed Capital.

	Year	S&P	BS	CM	DM
Return	1	27.20%	2.62%	12.15%	6.88%
	2	13.20%	1.48%	8.60%	1.25%
	3	−1.37%	−0.21%	5.56%	10.55%
	4	3.90%	2.19%	6.35%	6.26%
	All	**10.28%**	**1.56%**	**8.23%**	**6.38%**
Volatility	1	11.24%	1.35%	7.08%	7.74%
	2	10.80%	1.22%	7.18%	5.62%
	3	15.21%	1.07%	9.18%	7.16%
	4	13.97%	0.81%	5.98%	3.29%
	All	**12.95%**	**1.13%**	**7.45%**	**6.18%**
Sharpe Ratio	1	2.20	1.92	1.65	0.90
	2	1.20	1.20	1.18	0.25
	3	−0.01	−0.19	0.63	1.43
	4	0.34	2.66	1.06	1.86
	All	**0.82**	**1.37**	**1.10**	**1.03**

Table 9
Strategy comparisons on performance of Invested Capital.

	Year	S&P	BS	DM	CM
Return	1	27.2%	9.5%	27.8%	15.6%
	2	13.2%	6.8%	14.5%	9.8%
	3	−1.4%	3.3%	13.6%	44.2%
	4	3.9%	16.5%	11.7%	20.6%
	All	**10.3%**	**9.1%**	**16.9%**	**22.1%**
Volatility	1	11.2%	6.2%	15.6%	18.6%
	2	10.8%	6.4%	15.6%	16.2%
	3	15.2%	6.3%	19.1%	21.4%
	4	14.0%	7.3%	13.6%	14.7%
	All	**13.0%**	**6.6%**	**16.1%**	**17.9%**
Sharpe Ratio	1	2.20	1.50	1.65	0.87
	2	1.20	1.04	0.94	0.66
	3	−0.01	0.55	0.76	1.81
	4	0.34	2.12	0.88	1.34
	All	**0.82**	**1.36**	**1.05**	**1.21**

models, with the CM strategy exhibiting almost no market correlation at all.

6.5.2. Annual returns

Tables 8-9 display annualized performance measures for each year within the testing period. No annualized tests are conducted due to the limited sample size, but annualized volatility and Sharpe ratio are provided. The Sharpe ratio (1994) is the ratio of excess return to volatility which represents the trade-off between risk and return for a strategy. We consider excess return over zero instead of the risk-free rate since it was negligible during the timeframe.[11]

The risk-adjusted returns (Sharpe ratios) outperform the S&P benchmark for the entire testing period. The multi-objective models generate positive returns over each year while the single-objective model experiences one down-year. We see that the BS model has the lowest volatility while DM has the largest annualized return of 8.23% over the period. No models surpass the S&P benchmark return of 10.28% per year, although they do achieve a higher risk-adjusted return, which can be attributed to the lower volatility within each model.

When considering annualized returns on invested capital, almost all models exceed the benchmark S&P with the maximum returns corresponding to the DM model. The invested return calculation, however, yields higher volatility for each strategy leading to similar Sharpe ratios as before. Unlike the S&P's performance in the sell-off period of years 3 and 4, the multi-objective models do not experience a year of less than 7% returns, thus further supporting their market-neutrality.

[11] 3 Month U.S. T-Bills were historically low, remaining below 0.1% for most of the period and eventually climbing to 0.51% by the end of 2016

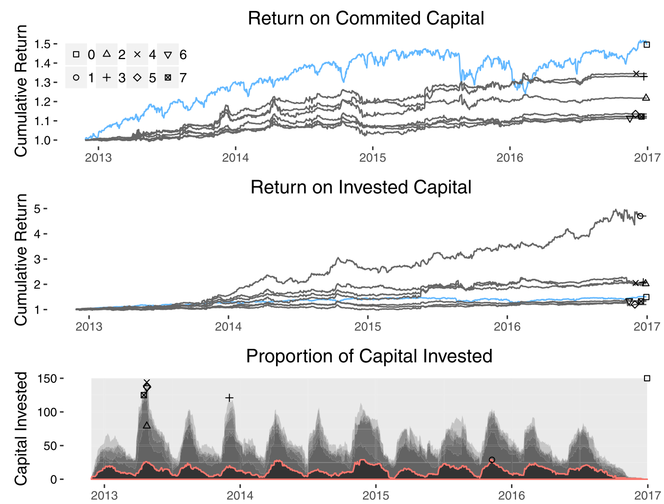


Fig. 15. Comparison between cardinality-constrained variants. The blue line labeled “0” represents the S&P buy-and-hold benchmark strategy. The label “1” represents the univariate strategy which is constrained to only one asset per component. The top plot shows cumulative excess returns on committed capital, the middle plot shows the same but for fully-invested returns, and the bottom plot displays the number of open trades for each strategy. The red line was added for contrast and corresponds to trades of the univariate version, while the largest number of open positions are clustered among variants 3–7. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

6.6. Cardinality comparison

Fig. 15 and Tables 10 & 11 compares cardinality variants of the DM model against the same S&P benchmark. Note that instead of $C = \$100$, we use $C = \$144$ in the returns on committed capital because that was the maximum number of open positions on one of the variants.

The top performers on committed capital are variant sizes $v = 2, 3, 4$. As indicated by the bottom plot, the $k = 2$ variant produced the second least tradeable pairs, only behind the univariate case. The low volume of trades for univariate pairs explains the poor performance on committed capital due to the limited trading opportunities. When looking at fully-invested returns, however, the univariate case trends strongly above the rest of the strategies. Increasing component cardinality has a diminishing benefit since the larger variants start underperforming the benchmark across both performance measures.

Higher cardinality benefits return on committed capital, but same does not hold for fully-invested returns. As indicated by the risk-adjusted performance, there is no clear benefit to include pairs of component sizes $v > 4$. Increasing the cardinality reduces volatility for both committed and invested returns but yields lower annualized returns, not to mention the higher transaction costs that are ignored. The univariate framework outperforms all multivariate cases for fully-invested performance, thus suggesting that univariate pairs are of higher quality with volatile and mean-reverting spreads. The univariate case yields an impressive Sharpe ratio but, due to the limited trading opportunities, suffers in performance on committed capital. Multivariate pairs can therefore offer more trading opportunities, albeit of lesser quality.

6.6.1. Main results

The main results can be summarized as follows:

Table 10
Performance on committed capital across cardinality-constrained variants.

	Year	1	2	3	4	5	6	7
Return	1	5.6%	17.6%	18.3%	21.1%	7.4%	8.7%	6.4%
	2	7.4%	10.4%	4.6%	10.7%	-0.2%	3.7%	5.0%
	3	0.2%	1.4%	15.4%	7.1%	6.6%	-1.5%	1.7%
	4	4.0%	0.3%	4.3%	6.0%	3.5%	3.7%	3.0%
	All	**4.2%**	**7.2%**	**10.6%**	**11.1%**	**4.6%**	**3.9%**	**4.2%**
Volatility	1	2.5%	6.7%	9.0%	9.1%	7.3%	7.3%	5.9%
	2	2.7%	8.8%	11.5%	9.4%	8.9%	8.2%	6.6%
	3	2.8%	9.7%	14.0%	9.9%	9.1%	6.5%	6.9%
	4	2.1%	4.3%	6.2%	5.2%	4.7%	4.2%	3.5%
	All	**2.6%**	**7.7%**	**10.6%**	**8.6%**	**7.7%**	**6.7%**	**5.9%**
Sharpe Ratio	1	2.20	2.46	1.91	2.14	1.03	1.18	1.09
	2	2.62	1.16	0.45	1.12	0.02	0.48	0.77
	3	0.08	0.19	1.09	0.74	0.75	-0.20	0.28
	4	1.85	0.10	0.70	1.14	0.75	0.89	0.86
	All	**1.62**	**0.95**	**1.01**	**1.27**	**0.62**	**0.60**	**0.73**

Table 11
Performance on invested capital across cardinality-constrained variants.

	Year	1	2	3	4	5	6	7
Return	1	51.5%	43.4%	27.4%	34.2%	8.1%	12.3%	10.9%
	2	91.3%	32.5%	15.9%	21.1%	-4.5%	5.5%	9.1%
	3	18.1%	8.1%	35.4%	13.2%	14.0%	-2.7%	3.5%
	4	29.1%	-3.8%	1.1%	12.0%	0.8%	15.1%	5.9%
	All	**44.5%**	**18.7%**	**19.4%**	**19.9%**	**4.7%**	**7.7%**	**7.5%**
Volatility	1	20.4%	17.9%	13.8%	13.5%	10.4%	10.1%	8.9%
	2	20.3%	20.5%	18.7%	16.4%	14.4%	13.7%	12.2%
	3	21.0%	23.5%	25.6%	16.5%	15.5%	10.2%	11.3%
	4	20.6%	14.4%	13.6%	9.9%	15.8%	9.0%	11.0%
	All	**20.6%**	**19.4%**	**18.6%**	**14.3%**	**14.2%**	**10.9%**	**10.9%**
Sharpe Ratio	1	2.14	2.10	1.82	2.25	0.80	1.20	1.21
	2	3.29	1.47	0.88	1.24	-0.25	0.46	0.77
	3	0.89	0.45	1.31	0.83	0.92	-0.22	0.36
	4	1.34	-0.20	0.15	1.20	0.13	1.60	0.57
	All	**1.89**	**0.98**	**1.05**	**1.34**	**0.40**	**0.73**	**0.72**

- *Single vs Multi-objectives:* Multi-objective models (DM, CM) outperformed the benchmark single-objective (BS) model across all measures
- *Univariate vs Multivariate:* Univariate pairs are more profitable but multivariate pairs present greater trading opportunities
- *Performance against benchmark:* For returns on invested capital, multi-objective pair formation significantly outperforms the S&P 500 benchmark

In addition to the above findings, we have demonstrated a new methodology for creating profitable combinations of assets for pairs trading through a multi-objective optimization framework. The pairs are not restricted to univariate size so the system is capable of recommending more trading opportunities where traditional enumeration techniques fail. The larger volume of trading opportunities is beneficial for realistic applications of pairs trading because it offers higher capacity to invest committed capital across pairs. This is clear when analyzing the returns on committed capital across cardinality variants, although the return on invested capital for the univariate case appears far superior.

7. Conclusions and future research

Pairs trading presents a variety of important decisions to evaluate. However, any improvements in the selection and trading phases are always bounded by the quality and quantity of candidate pairs generated within the PF problem. This research addresses the issue of subset selection to form pairs that yield volatile and mean-reverting spreads for a multivariate pairs trading strategy. The proposed framework is capable of forming both univariate and multivariate pairs on conflicting criteria through multi-objective optimization. Necessary modifications to NSGA-II and solution representation are described and implemented on S&P 500 market data to empirically evaluate different model designs. The results indicate that multivariate pairs formed on competing objectives significantly outperform existing single-objective techniques.

This new application of intelligent systems directly benefits the pairs trading community and should also extend to different problem areas in finance and other industries. Evolutionary multi-objective optimization has proved particularly useful for the PF problem structure. Most realistic problems typically share the same traits of conflicting criteria, mixed variable types, and non-convex constraints. This application demonstrates the usefulness of intelligent systems for addressing real-world problems without sacrificing practicality for simplifying assumptions.

7.1. Future research directions

The work done in this paper lends itself to many problems of pairs trading, both new and old. The most explicit direction is the further evaluation of returned Pareto pairs. As shown in Figs. 8, 10, and 12, there is a number of ways in which classification algorithms can help discriminate between pair types, whether it be mixture models, support vector machines, or clustering techniques.

Researchers are also encouraged to explore new objective functions and optimal component sizes now that we have provided a framework for finding such complex combinations. New constraints, such as a restricted covariance for same-component stocks, trade lot sizing, and minimum component weights would all be practical extensions to the model.

The methodology only employs NSGA-II but there might be other algorithms that are better suited for the problem. A simplification of the PF problem so that different algorithms can be tested without too many adjustments would be beneficial. For example, a combined approach of Tabu Search for subset selection and then traditional QP to address the asset weight sub-problem would likely improve the solution quality and convergence of the overall algorithm. Furthermore, we solve an inherently stochastic problem using a deterministic technique, thus ignoring any notion of uncertainty in the problem structure. Stochastic and continuous time methods should therefore be considered in future pair formation alternatives.

The proposed system is incomplete without an intelligent trading decision phase. A static set of rules are used to determine which pairs to trade and when. There is much to be done on this end of a pairs trading strategy that expert and intelligent systems can address. Either a unified framework between formation and trading or separate research specifically on the trading phase introduces a new set of problems for researchers to overcome.

Lastly, this methodology only applied an intelligent system to one problem. Other application areas like transportation, inventory control, and marketing should see similar characteristics to the core problem of forming combinations of co-moving sequential data. This application should hopefully inspire readers to pursue similar uses of intelligent systems to solve other hard optimization problems.

Declaration of Competing Interest

The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Credit authorship contribution statement

Jeffrey Goldkamp: Conceptualization, Data curation, Software, Formal analysis, Investigation, Methodology, Resources, Validation, Visualization, Writing - original draft, Writing - review & editing. **Mohammad Dehghanimohammadabadi:** Methodology, Project administration, Software, Supervision, Writing - review & editing.

References

Aguilar-Rivera, R., Valenzuela-Rendón, M., & Rodríguez-Ortiz, J. (2015). Genetic algorithms and darwinian approaches in financial applications: A survey. *Expert Systems with Applications, 42*(21), 7684–7697.

Anagnostopoulos, K. P., & Mamanis, G. (2010). A portfolio optimization model with three objectives and discrete variables. *Computers & Operations Research, 37*(7), 1285–1297.

Anagnostopoulos, K. P., & Mamanis, G. (2011). The mean–variance cardinality constrained portfolio optimization problem: An experimental evaluation of five multiobjective evolutionary algorithms. *Expert Systems with Applications, 38*(11), 14208–14217.

Burgess (2000). *A computational methodology for modelling the dynamics of statistical arbitrage*. University of London Ph.D. thesis..

Chan, E. (2013). *Algorithmic trading: Winning strategies and their rationale*: 625. John Wiley & Sons.

Chang, T.-J., Meade, N., Beasley, J. E., & Sharaiha, Y. M. (2000). Heuristics for cardinality constrained portfolio optimisation. *Computers & Operations Research, 27*(13), 1271–1302.

Chen, H., Chen, S., Chen, Z., & Li, F. (2017). Empirical investigation of an equity pairs trading strategy. *Management Science*.

Chen, S.-S., Huang, C.-F., & Hong, T.-P. (2013). A multi-objective genetic model for stock selection. In *Proceedings of the 27th annual conference of the japanese society for artificial intelligence*.

Clegg, M., & Krauss, C. (2018). Pairs trading with partial cointegration. *Quantitative Finance, 18*(1), 121–138.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695.

Dao, S. D., Abhary, K., & Marian, R. (2017). An innovative framework for designing genetic algorithm structures. *Expert Systems with Applications, 90*, 196–208.

Deb, K. (2014). Multi-objective optimization. In *Search methodologies* (pp. 403–449). Springer.

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation, 6*(2), 182–197.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association, 74*(366a), 427–431.

Do, B., & Faff, R. (2010). Does simple pairs trading still work? *Financial Analysts Journal, 66*(4), 83–95.

Do, B., & Faff, R. (2012). Are pairs trading profits robust to trading costs? *Journal of Financial Research, 35*(2), 261–287.

Emmerich, M. T., & Deutz, A. H. (2018). A tutorial on multiobjective optimization: fundamentals and evolutionary methods. *Natural computing, 17*(3), 585–609.

Engle, R. F., & Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, 251–276.

Fogel, L. J. (1964). *On the organization of intellect*. University of California, Los Angeles-Engineering Ph.D. thesis..

Galenko, A., Popova, E., & Popova, I. (2012). Trading in the presence of cointegration. *The Journal of Alternative Investments, 15*(1), 85–97.

Gatev, E., Goetzmann, W. N., & Rouwenhorst, K. G. (2006). Pairs trading: Performance of a relative-value arbitrage rule. *The Review of Financial Studies, 19*(3), 797–827.

Goetzmann, W., Rouwenhorst, K. G., et al. (1998). Pairs trading: Performance of a relative value arbitrage rule. *Technical Report*. Yale School of Management.

Holland, J. H. (1992). *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*. MIT press.

Huang, C.-F., Hsu, C.-J., Chen, C.-C., Chang, B. R., & Li, C.-A. (2015). An intelligent model for pairs trading using genetic algorithms. *Computational intelligence and neuroscience, 2015*, 16.

Huck, N. (2010). Pairs trading and outranking: The multi-step-ahead forecasting case. *European Journal of Operational Research, 207*(3), 1702–1716.

Huck, N., & Afawubo, K. (2014). Pairs trading and selection methods: is cointegration superior? *Applied Economics*.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control, 12*(2–3), 231–254.

Kassambara, A. (2017). ggpubr: 'ggplot2' based publication ready plots. R package version 0.1.6.999.

Krauss, C. (2017). Statistical arbitrage pairs trading strategies: Review and outlook. *Journal of Economic Surveys, 31*(2), 513–545.

Lin, L., & Cao, L. (2008). Mining in-depth patterns in stock market. *International journal of intelligent systems technologies and applications, 4*(3–4), 225–238.

Luke, S. (2009). *Essentials of metaheuristics*: 113. Lulu Raleigh.

Markowitz, H. (1952). Portfolio selection. *The journal of finance, 7*(1), 77–91.

Moral-Escudero, R., Ruiz-Torrubiano, R., & Suárez, A. (2006). Selection of optimal investment portfolios with cardinality constraints. In *Evolutionary computation, 2006. cec 2006. ieee congress on* (pp. 2382–2388). IEEE.

Mühlenbein, H., & Schlierkamp-Voosen, D. (1993). Predictive models for the breeder genetic algorithm i. continuous parameter optimization. *Evolutionary computation, 1*(1), 25–49.

Perlin, M. (2007). M of a kind: A multivariate approach at pairs trading,.

Rad, H., Low, R. K. Y., & Faff, R. (2016). The profitability of pairs trading strategies: distance, cointegration and copula methods. *Quantitative Finance, 16*(10), 1541–1558.

Sharpe, W. F. (1994). The sharpe ratio. *Journal of portfolio management, 21*(1), 49–58.

Sonmez, R., & Bettemir, Ö. H. (2012). A hybrid genetic algorithm for the discrete time–cost trade-off problem. *Expert Systems with Applications, 39*(13), 11428–11434.

Tahir, M. A., & Smith, J. E. (2007). Feature selection for heterogeneous ensembles of nearest-neighbour classifiers using hybrid tabu search. In *Advances in metaheuristics for hard optimization* (pp. 69–85). Springer.

Valle, C., Meade, N., & Beasley, J. (2014a). Absolute return portfolios. *Omega, 45*, 20–41.

Valle, C., Meade, N., & Beasley, J. E. (2014b). Market neutral portfolios. *Optimization Letters, 8*(7), 1961–1984.

Verel, S., Liefooghe, A., Jourdan, L., & Dhaenens, C. (2013). On the structure of multiobjective combinatorial search space: Mnk-landscapes with correlated objectives. *European Journal of Operational Research, 227*(2), 331–342.

Vidyamurthy, G. (2004). Pairs trading. *Quantitative Methods and Analysis*.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Woodside-Oriakhi, M., Lucas, C., & Beasley, J. E. (2011). Heuristic algorithms for the cardinality constrained efficient frontier. *European Journal of Operational Research, 213*(3), 538–550.

Woodside-Oriakhi, M., Lucas, C., & Beasley, J. E. (2013). Portfolio rebalancing with an investment horizon and transaction costs. *Omega, 41*(2), 406–420.

Zeileis, A. (2004). Econometric computing with hc and hac covariance matrix estimators. *Journal of Statistical Software, 11*(10), 1–17.