Article

Kelly's Criterion in Portfolio Optimization: A Decoupled Problem

Zachariah Peterson[1,*]

1   Adams State University
2   208 Edgemont Blvd.
3   Alamosa, CO 81101
4   petersonz1@grizzlies.adams.edu
*   Correspondence: petersonz1@grizzlies.adams.edu; Tel.: +1-503-706-2674

Abstract: Kelly's Criterion is well known among gamblers and investors as a method for maximizing the returns one would expect to observe over long periods of betting or investing. These ideas are conspicuously absent from portfolio optimization problems in the financial and automation literature. This paper will show how Kelly's Criterion can be incorporated into standard portfolio optimization models. The model developed here combines risk and return into a single objective function by incorporating a risk parameter. This model is then solved for a portfolio of 10 stocks from a major stock exchange using a differential evolution algorithm. Monte Carlo calculations are used to verify the accuracy of the results obtained from differential evolution. The results show that evolutionary algorithms can be successfully applied to solve a portfolio optimization problem where returns are calculated by applying Kelly's Criterion to each of the assets in the portfolio.

Keywords: portfolio optimization; Kelly criterion; differential evolution; mean-variance

JEL Classification: C6 Mathematical Methods; Programming Models; Mathematical and Simulation Modeling

1. Introduction

In general, portfolio optimization problems aim to determine an optimal allocation of wealth among a pool of candidate assets or securities. Portfolio optimization was first discussed in 1952 by Harry Markowitz in his work on modern portfolio theory (MPT) (Markowitz 1952). According to MPT, an optimum portfolio can be arranged such that return is maximized for a specified level of risk, or vice-versa, where risk is minimized for a specified level of return. Many current formulations of portfolio optimization problems are linear or quadratic, depending on the definition of portfolio risk that is used in the particular problem. The original formulation of Markowitz is known as the Mean Variance (MV) model and treats return on a portfolio of investments using averages of changes in market prices of the individual assets over time. The total portfolio return was shown to be a sum of returns from individual investments, with each return weighted by the fraction of total wealth invested in each security. Risk was defined as variance of returns and is found by taking the inner product of the covariance matrix for the assets in the portfolio. Both risk and return parameters can be calculated from market data. Later models for asset pricing, such as the Capital Asset Pricing Model (CAPM) (Fama and French 2004), would continue to use covariance values among changes in securities values to quantify risk. Other models, for example in (Bichpuriya and Soman 2016, El Ghaoui, et. al. 2003) use value-at-risk (VaR) or conditional value-at-risk (CVaR) to model the variation in portfolio returns; these formulations are purely linear programming problems. Under the MV model of portfolio optimization, the problem of maximizing return (risk) for a constrained risk (return) for a portfolio of N assets forms a linear (quadratic) objective function with a quadratic (linear) constraint in N-2 dimensions, i.e. the sum of wealth fractions and the prescribed risk or return function

determines the wealth fractions placed in 2 investments in terms of the remaining fractions. If neither risk nor return are constrained, then portfolio optimization is a dual objective problem of maximizing risk and specifying return simultaneously.

The MV approach to portfolio optimization, where returns are defined using average changes in market prices of assets over time, over-simplifies the problem and ignores the probabilities of very large movements in values of various securities in the portfolio during the investment horizon. A better reflection of reality is to determine the probability distribution of price changes for each of the assets in the portfolio and reformulate the return function in terms of these probabilities. For example, this is done in (Yang and Liu 2016), where returns are treated as fuzzy numbers. Portfolio optimization based on the tenets of MPT has also been has also been used in electricity generation and distribution, where electricity demand is treated as a random variable. In (Bichpuriya and Soman 2016), probability densities for demand and electricity generation costs are used to optimize portfolios of bilateral forward contracts with generating companies. Both of these examples formulate the portfolio return function in terms of expectation values and covariant risk, generating a linear (Bichpuriya and Soman 2016, El Ghaoui, et. al. 2003) or quadratic (Yang and Liu 2016) objective function problem that is equivalent to the MV formulation. The linear return function in the MV model has been used by many authors in the automation literature (Bichpuriya and Soman 2016, El Ghaoui, et. al. 2003, Yang and Liu 2016, Kamili and Riffi 2016, Chen, et. al. 2012, Zaheer and Pant 2016, Korczak and Roger 2000, Ma, et. al. 2012, Chang, et. al. 2009) and is treated as something of a standard model for portfolio optimization. However, one can show from probability theory that the optimum return on an investment (or portfolio of investments) is not a linear function of the fraction of wealth placed in each investment.

Kelly's Criterion is well known among gamblers and can be used as a betting strategy. Kelly's result is, in its simplest sense, a solution to an optimization problem which maximizes a geometric mean and was originally applied to a technical problem in information theory (Kelly 1956, Kim 2008). The Kelly Criterion has been discussed in contexts outside of gambling, for example, in engineering economics (Kim 2008). These ideas were later embraced by gamblers and were used to maximize the winnings one would expect to see from a large number of bets with well-defined probabilities. A gambler that bets a critical fraction of their wealth per bet can expect to see an average rate of return per bet that maximizes their total return over time (Thorpe 1997). While it is known that serious practitioners of mathematical finance equate investing and gambling (Thorpe 1997), one has yet to see the Kelly method applied to portfolio optimization in the automation literature. In its simplest form, where the outcome of each bet was considered binary with well-defined odds and probabilities, and successive bets are mutually independent, Kelly's result is a simple formula for the critical fraction of wealth that will maximize a gambler's average return over a large number of bets. More complicated outcomes do not have such simple formulae. However, one can show that Kelly's return function is concave and has a maximum solution. Although this was originally applied to a set of discretely distributed outcomes using the Central Limit Theorem, it can also be applied to continuously distributed random variables in the same manner. The intention of this paper is to examine the general case where multiple correlated outcomes are continuously distributed, and we can place fractions of our wealth in a large number of investments. It will be shown that one can use Kelly's Criterion to define a multidimensional nonlinear function for the rate of return from a portfolio of these investments.

Solving this type of optimization problem requires an efficient algorithm for nonlinear objective functions in any dimension. One such class of algorithms that is applicable to this type of problem are meta-heuristic techniques. A subclass of these methods comprises evolutionary algorithms, which are intended to mimic the behavior of natural systems and are based on stochastic search methods. Various evolutionary algorithms have been applied to optimization problems in a wide range of fields including circuit design, mechanical engineering (Storn and Price 1997), aerodynamics (Rogalsky, et. al. 1999), traveling salesman problems, and medical imaging (Kamili and Riffi 2016, Qin, et. al. 2009). Several meta-heuristics (swarm optimization (Kamili and Riffi 2016, Chen, et. al. 2012), simulated annealing, differential evolution (Zaheer and Pant 2016, Korczak and Roger 2000, Ma, et. al. 2012), and

genetic algorithms (El Ghaoui, et. al. 2003, Chang, et. al. 2009)) have been applied to solving the portfolio optimization problem as well.

This paper will show how the return function in the portfolio optimization problem can be reformulated using Kelly's Criterion. The results demonstrate the existence of two types of return functions, which are termed here as the decoupled and the coupled return functions. The reformulation of the return function requires that the function defining portfolio risk must be reformulated as well. By following Kelly's process, the return function for a portfolio of N investments can be shown to be a nonlinear function of the fraction of total wealth that is placed in each investment. The result transforms the risk function from a second degree to a fourth degree polynomial objective function in N dimensions. This paper will focus on solving the portfolio optimization problem for the decoupled return function.

The second section will show the derivation of the nonlinear return function for a portfolio of assets with specified distributions of changes in asset values. This is then used to calculate the average return and the variance of returns. The third section discusses the incorporation of these results into some well-known models for portfolio optimization. A cardinality-constrained portfolio optimization problem which transforms the dual objective problem into a single objective problem by incorporating a risk parameter will be presented. The parameters in this problem will be defined using a standard model where individual asset values follow correlated geometric Brownian motion. The fourth section will present and discuss numerical solutions obtained using differential evolution. The results from the model with the decoupled return function will be compared to the results obtained using the objective functions for risk and return in the MV model under the same set of constraints. The accuracy of the results from both models is also verified using a Monte Carlo simulation. The fifth section will present and discuss conclusions, as well as offer a guide to further research in this area. The practical applicability of these results will also be briefly discussed.

2. The nonlinear return function

To understand the necessity of reformulating the return function in portfolio optimization model based on mean-variance models, the original linear return function of Markowitz will first be examined. The linear return function defines the return from a portfolio of N investments with initial value of w_0 as a sum of returns from each of the investments r_i, with each investment return weighted by the fraction of the portfolio that is invested in each asset F_i. Returns on investments are random variables and can be thought of as accruing over discrete time periods similar to compound interest. The value of a portfolio of N investments after n accrual periods is an exponential function of the number of periods, where each accrual multiplies the value of the ith investment by a factor $R_i = 1 + r_i$. If r_i represents a stock, then $r_i = \left(\frac{S_i}{S_0}\right) - 1$, where S_0 is the value of the stock at the time of initial investment. The value of the portfolio after the nth period is

$$w_n = \sum_{i=1}^{N} F_i w_0 R_i^n. \tag{1}$$

In the MV model, the average return rate after a single period is given by the sum of expected values of the fractional changes in asset values as

$$R_{avg} = \sum_{i=1}^{N} F_i E[r_i]. \tag{2}$$

Markowitz's MV model, as well as others working in this area (Bichpuriya and Soman 2016, El Ghaoui, et. al. 2003, Yang and Liu 2016, Kamili and Riffi 2016, Chen, et. al. 2012, Zaheer and Pant 2016, Korczak and Roger 2000, Ma, et. al. 2012, Chang, et. al. 2009), have formulated each of the $E[r_i]$ values as an average over a large number of changes in asset values over discrete time periods. These changes can occur daily, weekly, etc. In some cases, particularly in (Bichpuriya and Soman

2016, Yang and Liu 2016) this is done by first determining a probability density function for changes in asset values between successive points in time and then calculating the value of $E[r_i]$ directly from this distribution. Here, no assumption has been made on the type of investments in the portfolio; the only assumption is that each of the distribution functions for returns on individual investments are known a priori or can be determined from market data. The same assumption applies to the covariance matrix defining risk, which will be discussed more later.

Kelly's formulation can now be used to define the portfolio's return function in terms of the expected value of an exponential rate constant. Here we can derive the form of this expectation value directly from a formulation of the return function for the portfolio. If we consider the fractional change in the value of a security between two discrete time periods to be a random variable, then the value of the portfolio invested in the ith security after a single accrual (i.e. from period j to period $j + 1$) is

$$w_{j+1,i} = w_{j,i} - f_i w_{j,i} + f_i w_{j,i}(1 + X_i). \tag{3}$$

Here, X_i could be any type of investment. If, for example, X_i represents a stock, then $X_i = \left(\frac{S_{j+1,i}}{S_{j,i}}\right) -$ 1, where $S_{j,i}$ is the value of the ith stock at period j. For a single accrual, equation (3) reduces to

$$w_{1,i} = w_{0,i} - f_i w_{0,i} + f_i w_{0,i}(1 + X_i) = w_{0,i}(1 + f_i X_i). \tag{4}$$

In (3) and (4), $w_{j,i} = f_i w_j$ is related to the fraction of the portfolio invested in the ith asset, and the total portfolio value at period j is given by a summation over index i. From equation (3), it is obvious that the portfolio value depends on changes over all previous accrual periods. Therefore, the value of the portfolio after n periods can be founds by induction using permutations on index j in equation (4), followed by a summation over index i. Iterating index j from 0 to n and using the Binomial Theorem, we have the following equation for the value of the portfolio in the ith investment after n periods $w_{n,i}$:

$$w_{n,i} = w_{0,i}(1 + f_i X_i)^n. \tag{5}$$

Using $w_{j,i} = f_i w_j$ and taking summation over index i, the total portfolio value after n periods is

$$w_n = \sum_{i=1}^{N} f_i w_0 (1 + f_i X_i)^n. \tag{6}$$

This process shows that the rate of return for the entire portfolio after n periods depends nonlinearly on the fraction of total wealth that is placed in each investment:

$$R_{avg} = \sum_{i=1}^{N} f_i ((1 + f_i X_i)^n - 1). \tag{7}$$

Equation (7) is the return function for the portfolio after n accrual periods. We now have a function that treats returns as draws from a random variable, and this function can be used to calculate the expected return and the variance in returns (i.e. portfolio risk). In equation (7) the actual fraction of wealth that is invested in the ith asset is equal to f_i^2.

Now that returns are defined in terms of the outcomes from a set of random variables, it is of interest to know the expected value of the return function (7). This defines the expected return from a large number of accruals, i.e. as n approaches infinity. Applying a logarithmic identity to the return function for a large number of accruals allows the products in (7) to be converted to a sum. This converts each of the $(1 + f_i X_i)$ product terms in the series to an exponential function of a summation:

$$(1 + f_i X_i)^n = \exp\left(\sum_{j=1}^{n} \ln(1 + f_{i,j} X_{i,j})\right). \tag{8}$$

Here a natural logarithm has been used, but one can use any base for the logarithm. Kelly's original paper used a base of 2 (Kelly 1956). The subscripts have been modified to $\{i, j\}$, where j denotes the jth outcome for the ith asset. To proceed further requires the Central Limit Theorem, which states that the sum of a large number of outcomes from a random variable approaches the expected value of a random variable multiplied by the number of outcomes. This is used to convert the summation in the argument of the jth exponential function in (8) to the expected value

$$(1 + f_i X_i)^n \rightarrow \exp(nE[\ln(1 + f_i X_i)]). \tag{9}$$

Equation (9) expresses the limit as the number of accruals becomes very large and is effectively an approximation; this has been discussed previously (Samuelson 1971). This is accurate to within a certain confidence interval that depends on the square root of the number of accruals. As long as the number of accruals is large, and the expected value grows faster than the size of the confidence interval, the actual return approaches the expected return in equation (9) with probability 1.

Placing the result in (9) back into (7) and setting $n = 1$ gives the average return for a single accrual:

$$R_{avg} = \sum_{i=1}^{N} f_i (\exp(E[\ln(1 + f_i X_i)]) - 1). \tag{10}$$

This return function can be incorporated into any of the portfolio optimization models discussed in the introduction. It is worth recalling that the actual fraction of wealth invested in the ith asset is f_i^2, while f_i is the coordinate used to perform the optimization. Equation (10) will henceforth be referred to as the *decoupled return function* for reasons that will soon be apparent.

It is important to note that the process used to derive the decoupled return function above could have been applied to equation (1) immediately without splitting the portfolio into N fractions. If we apply this process to (1), we arrive at a return function that couples the random variables and the wealth fractions F_i into a single expectation value in N dimensions. This result for the expected value of a single return can be called the *coupled return function* and is given by

$$R_c = \exp(E[\ln(1 + \sum_{i=1}^{N} F_i X_i)]) - 1. \tag{11}$$

In the general case, this expectation value is an N-dimensional integral with each of the F_i values appearing as parameters. If a single $F_i = 1$, and we take zero for the remaining wealth fractions, the coupled and decoupled equations are identical.

Both return functions require advance knowledge of the joint distribution function $p(\mathbf{X})$ for the assets in the portfolio. As equation (11) involves an N-dimensional integral that is not seperable, it is computationally more complex and may not be analytically solvable (depending on the form of $p(\mathbf{X})$), even in the case where the X_i random variables are independent. By contrast, the decoupled return function (10) will reduce to a marginal distribution for a single asset $p(X_i)$ as the dependence of $N - 1$ of the variables is eliminated via integration over the space of outcomes in $N - 1$ dimensions. Thus the decoupled problem may be preferable both analytically and numerically. If the time required to compute a single expectation value in the decoupled problem is T, then the time required to compute all N expectation values is NT. The time required to compute the expectation value in the coupled problem is T^N. The remainder of this paper will focus on solving the portfolio optimization problem using the decoupled return function.

Equation (7) is a linear combination of random variables and the variance of this function defines the portfolio risk. The variance of a linear combination of random variables can be written as an inner product of the covariance matrix for these random variables (Fisher 1990). Specifically, let Z

be a linear combination of random variables with covariance matrix $[M]$, and let $[a]$ be the column vector defining the coefficents. The variance of Z is given by the following

$$Var[Z] = [a]'[M][a]. \tag{12}$$

In equation (12), $[a]'$ is the transpose of $[a]$. Using (12) we can calculate the variance of a single return (i.e. when $n = 1$ in (7)). The random variables in the return function are $1 + f_i X_i$, and the coefficients are $a_i = f_i$. The entries in the covariance matrix are

$$M_{ij} = Cov[X_i, X_j]. \tag{13}$$

We see in equation (7) that the coefficient vector is $[a] = [f]$. Taking the inner product of the covariance matrix defined by equation (13) gives the risk function for the portfolio

$$Var[R] = \sum_{i=1}^{N}(f_i^4 M_{ii} + 2\sum_{j>i}^{N} f_i^2 f_j^2 M_{ij}). \tag{14}$$

From equation (14) we see that the portfolio risk function is a fourth degree polynomial in terms of the variables f_i. Applying the same operation to equation (1) returns the MV risk function with coefficients equal to $F_i F_j$. The two risk functions are actually equivalent since $F_i = f_i^2$.

3. The decoupled return function in portfolio optimization

The principle results from the previous section can be incorporated into any of the standard models for portfolio optimization. Equations (10) and (14) form a dual objective optimization problem. However, the two results can be combined into a single objective function using a risk parameter. In this model, the risk parameter P measures an investor's risk indifference (large P) or risk aversion (small P) and allows a portfolio to be tailored to an individual investor's risk preferences. In the cardinality-constrained efficient frontier (CCEF) model (Kamili and Riffi 2016), the risk parameter takes values in the interval $[0, 1]$; these are the bounds that will be used in the model that will be presented below. It should be noted that these are not the only bounds that have been used in models based on risk parameters. The CVaR model for portfolio optimization in (Bichpuriya and Soman 2016), for example, used bounds of $[0, 3]$ for the risk parameter in their single objective problem. This section will present the cardinality-constrained single objective problem that will be used for portfolio optimization. To form the objective function for portfolio optimization, the return and risk equations from the MV model are first combined into a single objective problem, then a second model will be presented that combines equations (10) and (14) into a single objective problem.

The imposition of cardinality-constraints in portfolio optimization limits the wealth fractions invested in different assets at some specified range within the interval $[0, 1]$. One can specify different limits for each asset, or the same limit can be specified for all the assets in the portfolio. Formally, this type of constraint can be written using a pair of vectors $[K_{min}]$ and $[K_{max}]$. If we form the wealth fractions into a vector $[f]$, the limiting constraint in the decoupled Kelly problem is

$$[K_{min}] \le [f]'[f] \le [K_{max}]. \tag{15}$$

The sum of components of the $[K_{max}]$ vector must be less than the total number of assets N, which is equivalent to limiting $K_{max,i} < 1$ for all assets.

The equations in the MV model for risk and return form the following cardinality-constrained portfolio optimization model:

$$\max P \sum_{i=1}^{N} F_i E[r_i] - (1-P) \sum_{i=1}^{N} \left(F_i^2 M_{ii} + 2 \sum_{j>i}^{N} F_i F_j M_{ij} \right)$$

$$\text{subject to } \sum_{i=1}^{N} F_i = 1 \tag{16}$$

$$K_{min,i} \leq F_i \leq K_{max,i} \; \forall \, i.$$

Here, the $[M]$ matrix appears in equation (13), P is the risk parameter bounded in the interval $[0,1]$, and the optimization variables are the set of $\{F_i\}$. The $E[R_i]$ terms are calculated directly from market data. This is a similar formulation to the model in (Kamili and Riffi 2016) and uses the same risk and return terms.

The above model is now reformulated using the decoupled return and risk functions given by equations (10) and (14) respectively. The portfolio optimization problem incorporating Kelly's Criterion is given by the following:

$$\max P \sum_{i=1}^{N} f_i(\exp(E[\ln(1+f_i X_i)]) - 1) - (1-P) \sum_{i=1}^{N} \left(f_i^4 M_{ii} + 2 \sum_{j>i}^{N} f_i^2 f_j^2 M_{ij} \right)$$

$$\text{subject to } \sum_{i=1}^{N} f_i^2 = 1 \tag{17}$$

$$K_{min,i} \leq f_i^2 \leq K_{max,i} \; \forall \, i.$$

This model will henceforth be called the *decoupled Kelly model*. Here, the optimization variables are the set of $\{f_i\}$, and the actual wealth fraction invested in each asset is the set of $\{f_i^2\}$.

The individual terms in equation (10) are known to be concave functions (Vince 2011), and therefore the linear combination of these terms with positive coefficients is also a concave function. Thus a solution exists that will maximize equation (10). The covariance matrix is positive-definite and symmetric; the entries in the covariance matrix are the coefficients for the linear combination of functions in equation (14). Each term is an even degree polynomial function and is convex. The linear combination of these convex functions is also convex and a solution exists that will minimize equation (14). These two functions are combined in equation (17) via linear combination, where the $-(1-P)$ coefficient changes the convex risk function to a concave function. Therefore, the linear combination of the risk and return functions in equation (17) forms a concave function, and a solution must exist that will maximize the portfolio optimization problem in equation (17).

It has been noted in the literature that wealth fractions generated from Kelly's Criterion are only true wealth fractions under certain conditions (Vince 2011). In the discrete case, this requires that the probability of total loss of investment in an asset be nonzero over the entire course of the investment horizon. If this condition is not met, the Kelly Criterion may return optimum F_i values that are greater than 1 (Kim 2008, Vince 2011). A wealth fraction value $F_i > 1$ can be interpreted as a "leveraging factor" (Vince 2011) rather than as a fraction of investment capital. If the asset in question is a stock, changes in its value are continuously log-normally distributed and the probability of total loss is vanishingly small. The probability of an asset's downward movement to zero value approaches zero as the rate of downward movement approaches negative infinity. This issue has been addressed for the discrete case in (Thorpe 1997, Vince 2011), and it will be shown that a similar result holds in the continuous case.

For an asset following geometric Brownian motion, where the fractional changes in asset values are distributed log-normally, the expectation values in the return function in (10) involve functions of correlated normally distributed random variables with a known joint distribution function $p(\mathbf{X})$. To derive the bounds on the wealth fractions, one can treat each of the terms in equation (10) individually. For an individual term in (10), the expectation value $E[\ln(1 + fX)]$ in N dimensions reduces to a one-dimensional expectation involving only the marginal distribution function $p(X)$. Taking a derivative of the expectation value with respect to f and setting the result to zero give the result $E\left[\frac{X}{1+fX}\right] = 0$. Expanding this result as a Taylor series in f gives the following condition on the solution for f:

$$\sum_{m=0}^{\infty}(-1)^m f^m E[X^{m+1}] = 0. \tag{18}$$

Up to second order, the solution is $f = \frac{E[X]}{E[X^2]}$. A similar result is shown in the discrete case in (Kim 2008, Thorpe 1997, Vince 2011). In the case of continuously distributed changes in the value of the asset, the optimal wealth fraction to invest in the asset is bounded in the interval $(0,1)$ as long as $E[X^2] > E[X]$. As long as f is not too large this approximation is valid. If the solution to equation (18) is greater than 1 the value of f^2 should be interpreted as a leveraging factor. In the models presented above, the imposition of cardinality constraints can still limit the leveraging at some pre-determined level for each asset in the portfolio, and the maximum leverage allowed for the entire portfolio can be limited to some multiple of the capital available for investment.

The portfolio optimization models in (16) and (17) are solved for a portfolio of 10 stocks from the Mumbai National Stock Exchange based on data from the period April 1, 2007 to March 31, 2008. Data for the assets in this portfolio appears in (Zaheer and Pant 2016). The month-to-month return data for the 10 companies can be used to calculate the parameters for the MV model and the decoupled Kelly model, and the results from the both models will be compared. The monthly return data in (Zaheer and Pant 2016) is used to determine average monthly returns and a sample covariance matrix for the MV model. The data set is instructive as the original authors attempted to solve the problem by using the MV return function as a constraint and the quadratic risk function as the objective function for minimization. Their differential evolution algorithm only generated a single feasible portfolio. Their remaining solutions required more than 100% of the capital be leveraged, which violated the constraints of their problem. The results here will show that the problem can indeed be solved using a risk parameter formulation as defined in equations (16) and (17).

For the decoupled Kelly model, one must first determine the drift and volatility parameters using the solution to the stochastic differential equation for geometric Brownian motion for each of the assets in the portfolio. The stochastic differential equation for the ith stock is based on an Ito process given by

$$dS_t = S_t(\mu dt + \sigma dW). \tag{19}$$

The solution to this stochastic equation is well known (Joshi 2008) and defines a log-normal distribution for changes in stock values over a single time period. The X_i random variable as a function of time in (17) is defined as

$$X_i(t + \Delta t) = \frac{S_i(t+\Delta t)}{S_i(t)} - 1 = e^{\left(\mu_i - \frac{1}{2}\sigma_i^2\right)\Delta t + \sigma_i\sqrt{\Delta t}\, y} - 1, \tag{20}$$

where y is a standard normally-distributed random variable that defines the magnitude of an up or down movement in the stock value over a specified time period, and $\Delta t = 1$ month.

Equation (20) can now be used to relate the drift and volatility to the average value and sample variance of monthly returns from the market data (see table 1 in (Zaheer and Pant 2016)). Taking the expected value and variance of (20) gives the following values for the drift and volatility of the ith stock over a single month. Let $Avg[R]$ and $Var[R]$ be the sample mean and sample variance calculated from the data respectively:

$$\mu_i = \ln(1 + Avg[R])$$
$$\sigma_i = (\ln(Var[R]e^{-2\mu_i} + 1))^{\frac{1}{2}} \tag{21}$$

The values for mean return, the sample covariance matrix, drift, volatility, are summarized in tables 1 and 2 below. Table 1 also shows the values of the first-to-second moment ratios for each stock calculated directly from equation (18).

Table 1. Parameters for the portfolio optimization models in (16) and (17).

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
$Avg[R]$	0.1750	0.0995	0.3398	0.2366	0.1149	0.2799	0.2158	0.2593	0.2686	0.4405
μ_i	0.1613	0.0949	0.2925	0.2123	0.1087	0.2468	0.1954	0.2305	0.2379	0.3650
σ_i	0.3516	0.3277	0.3804	0.2660	0.2205	0.4398	0.2750	0.2167	0.3477	0.1991
$\dfrac{E[X]}{E[X^2]}$	1.5521	1.6268	1.5447	1.0347	0.8814	1.9151	1.0856	0.8129	1.4232	0.7285

Table 2. Sample covariance matrix from the data in (Zaheer and Pant 2016).

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
X_1	0.1817	0.0978	0.1403	0.0962	0.0481	0.1745	0.0752	0.0574	0.1326	0.0040
X_2	0.0978	0.1370	0.1246	0.0696	0.0288	0.1823	0.0924	0.0438	0.1121	0.0402
X_3	0.1403	0.1246	0.2778	0.1352	0.0582	0.1756	0.0914	0.0953	0.1621	0.0703
X_4	0.0962	0.0696	0.1352	0.1121	0.0443	0.1440	0.0617	0.0572	0.1198	0.0321
X_5	0.0481	0.0288	0.0582	0.0443	0.0619	0.0970	0.0338	0.0537	0.0690	-0.0141
X_6	0.1745	0.1823	0.1756	0.1440	0.0970	0.3495	0.1543	0.1055	0.2195	0.0291
X_7	0.0752	0.0924	0.0914	0.0617	0.0338	0.1543	0.1161	0.0581	0.1160	0.0484
X_8	0.0574	0.0438	0.0953	0.0572	0.0537	0.1055	0.0581	0.0763	0.0844	0.0216
X_9	0.1326	0.1121	0.1621	0.1198	0.0690	0.2195	0.1160	0.0844	0.2068	0.0466
X_{10}	0.0040	0.0402	0.0703	0.0321	-0.0141	0.0291	0.0484	0.0216	0.0466	0.0839

Equations (20) and (21) define a marginal distribution for changes in each of the stock values over a single time period (in this case, over a single month). When taken together with the entries in the covariance matrix, one can define a joint distribution function for changes in the asset values for all 10 stocks in the portfolio. The changes in stock values are jointly normally distributed with the coupling determined by the covariance matrix. However, as was noted previously, the expectation value $E[\ln(1 + f_i X_i)]$ reduces to a one-dimensional integral involving only the marginal distribution. Thus the ith expectation value in the argument of the exponential function in equations (10) and (17) is given by

$$E[\ln(1 + f_i X_i)] = \frac{1}{\sqrt{2\pi}} \int \ln\left(1 + f_i\left(e^{\mu_i - \frac{1}{2}\sigma_i^2 + \sigma_i y} - 1\right)\right) e^{-\frac{y^2}{2}}\, dy. \tag{22}$$

4. Results from differential evolution

The portfolio optimization problems in equations (16) and (17) are solved using differential evolution. This approach is a robust stochastic searching technique that is ideally suited to solving nonlinear problems such as those shown above. The original algorithm for differential evolution is presented in (Storn and Price 1997). The algorithm used here was run on a PC with a 2.4 GHz dual-core processor.

Differential evolution proceeds via three principle steps: mutation, crossover, and greedy selection. During mutation, candidate values for the objective variables in the optimization problem are generated from an initial population. The mutation step generates candidate solutions by combining components from the previously acceptable solution and compares these candidates with the constraints on the variables in the problem. The crossover step randomly selects candidates that were generated in the mutation step into a trial solution. The greedy selection step compares the previous acceptable solution with the trial solution. If the trial solution is favorable and satisfies the constraints of the problem then it is accepted as the current best solution. Otherwise it is rejected. The process repeats until particular stopping criteria have been reached, at which point the algorithm terminates. The stopping criteria used elsewhere in the literature typically forces termination of the algorithm after a specified number of iterations have been performed. Here, the stopping criteria are based on the elapsed processing time. The algorithm starts a timer, and once the timer reaches 30 seconds the algorithm terminates. If the algorithm generates an acceptable new solution during the greedy selection step, the timer is reset to zero and the algorithm repeats itself. The selection process in the mutation step used here was taken from Storn and Price's original C-code in (Storn and Price 1997).

The two critical parameters that control the progress of differential evolution algorithms are the crossover rate C and the scaling factor F. The crossover rate is equivalent to the probability that a candidate component generated in the mutation operation crosses over to the trial solution. $C = 0.75$ in all the calculations performed in this study. The scaling factor scales the random combinations generated during the mutation step and is critical to generating new candidate solutions. A number of trial generation strategies can be found in the literature (Storn and Price 1997, Qin, et. al. 2009), all of which use a scaling factor to amplify or diminish mutation. The original formulation of differential evolution restricted $F \in [0, 2]$.

The algorithm used here is a variant on this original formulation (known in the literature as DE/rand/1/bin (Storn and Price 1997)). The generation strategy in DE/rand/1/bin has been shown to be effective in optimizing quadratic and higher even degree polynomial and other nonlinear objective functions in many dimensions (Storn and Price 1997, Qin, et. al. 2009). Scaling factors on the order of $F \sim 0.5$ were shown to be effective in solving these problems with nearly 100% reproducibility. Here the value of F is taken to be the average of $\frac{1}{2}(K_{max,i} - K_{min,i})$ for all N assets in the portfolio. A noise term $N(0, 0.01)$ is also added to the trial solutions generated during the mutation step, where $N(0, 0.01)$ represents a normally-distributed random variable with 0 mean and 0.01 standard deviation. This ensures that there is always some slight perturbation to the solutions that are generated during mutation. C-style pseudo-code for this algorithm can be found in the Appendix.

Optimization problems solved via differential evolution typically start with an initial solution consisting of randomly generated initial values that fall within the feasible set of solutions (Qin, et. al. 2009). If one examines the drift and volatility values in table 1, the drift-to-volatility and Sharpe ratios are largest for the 10th stock. One would reasonably expect that the solver will produce solutions that place a large fraction of investment capital in the 10th stock for any value of risk parameter. The Sharpe ratio for this asset is ~8 (assuming risk-free interest rates are zero), while the Sharpe ratios for the remaining stocks are all significantly smaller. Therefore, the initial wealth fraction for each stock is set equal the drift-to-volatility ratio for each asset divided by the total of drift-to-volatility ratios for all the stocks. This type of initialization is useful in preventing premature convergence at a local maximum. The limits from cardinality are $K_{min} = 0.05$ and $K_{max} = 0.95$ for all stocks in the portfolio. These limits force at least a small fraction of wealth to be invested in each of the assets and prevent all of the wealth from being invested in a single asset. The precision on the equality constraint is set to ±0.01 in the greedy selection step; this allows some flexibility on the selection criteria applied to mutated solutions. Since the total wealth fractions are given by an equality constraint, one of the wealth fractions is selected randomly to be determined by the equality constraint after the first 9 wealth fractions are generated from mutation and crossover.

The portfolio optimization models in equations (16) and (17) were solved for various values of the risk parameter. The solution algorithm was run 20 times in succession for each value of risk parameter so that convergence and sensitivity of the algorithm can be evaluated. Results from the algorithm for each model are shown in the figures below. These figures show how the return and risk values converge over multiple runs for the MV model (figure 1) and the decoupled Kelly model (figure 2). Tables 3 and 4 show the final portfolios generated after the 20th run for each value of risk parameter in each of the models.
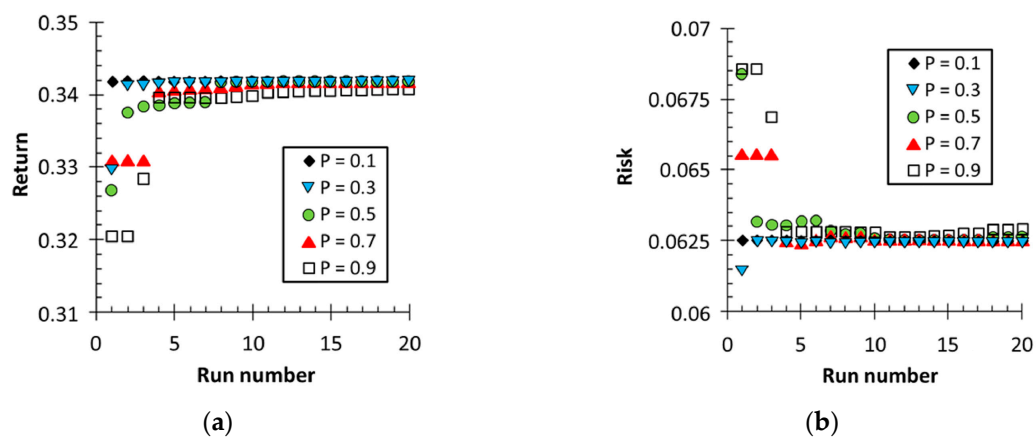
Figure 1. (a) Return and (b) risk values for each run of the solution algorithm for the MV model at various values of risk parameter P.

Table 3. Final portfolios generated for the MV model.

	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}
$P = 0.1$	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.5500
$P = 0.3$	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.5498
$P = 0.5$	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.5499
$P = 0.7$	0.0500	0.0500	0.0500	0.0501	0.0500	0.0501	0.0501	0.0500	0.0503	0.5494
$P = 0.9$	0.0500	0.0500	0.0514	0.0504	0.0500	0.0539	0.0500	0.0507	0.0506	0.5430
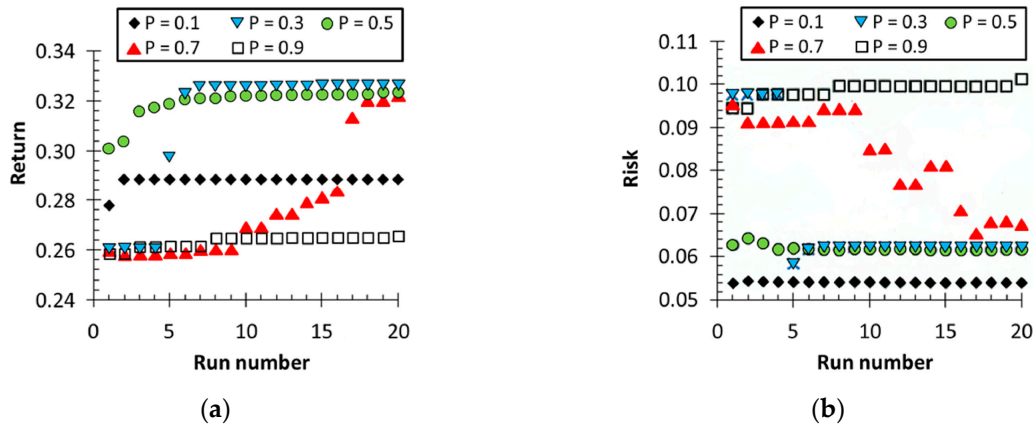
Figure 2. (a) Return and (b) risk values for each run of the solution algorithm for the decoupled Kelly model at various values of risk parameter P.

Table 4. Final portfolios generated for the decoupled Kelly model.

|	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}
$P = 0.1$	0.0500	0.0500	0.0500	0.0501	0.1438	0.0500	0.0502	0.0532	0.0500	0.4339
$P = 0.3$	0.0500	0.0500	0.0500	0.0500	0.0501	0.0500	0.0500	0.0501	0.0500	0.5498
$P = 0.5$	0.0501	0.0500	0.0503	0.0503	0.0509	0.0500	0.0505	0.0638	0.0500	0.5341
$P = 0.7$	0.0522	0.0517	0.0720	0.0544	0.0519	0.0698	0.0500	0.0608	0.0500	0.5039
$P = 0.9$	0.0500	0.0500	0.1451	0.1045	0.0500	0.1182	0.0939	0.1096	0.1143	0.1840

The results in figure 1 show that the risk and return functions in the MV model converge to the same values for all values of risk parameter. A typical run requires ~1000's of iterations and lasts ~100 seconds. The results in table 3 show that the solver generates the same portfolio at all values of risk parameter. The solver generates the expected result that is clearly biased towards stock 10 due to its large drift-to-volatility and Sharpe ratios. The Kelly model, in comparison, converges to the same portfolio as the MV model for only 3 of the 5 values of risk parameter used in this study. The convergence rate in the decoupled Kelly model for $P = 0.3$ and $P = 0.5$ is similar to the convergence rate in the MV model. For $P = 0.7$, the Kelly model eventually converges to the same portfolio but the convergence rate is significantly slower.

For $P = 0.1$, the decoupled Kelly model converges to a different portfolio than the MV model at a fast rate. The portfolio return and risk are both lower for $P = 0.1$, which is to be expected as risk and return should scale in proportion. This would also be expected for low values of risk parameter, which are meant to represent an investor's aversion to risk. When the portfolio generated from the MV model is used to calculate the value of the objective function in the decoupled Kelly model, the objective function in equation (17) has a smaller value using the MV portfolio. This shows that the decoupled Kelly model converges to a better maximum for $P = 0.1$, and the model is sensitive to the risk parameter.

When the risk parameter is set to $P = 0.9$, the decoupled Kelly model returns a counter-intuitive result. Returns are lower and risk is higher. This indicates misconvergence; it is likely that the solution algorithm gets stuck at a local maximum in the decoupled Kelly model for $P = 0.9$.

To validate the accuracy of the objective functions used in each model, the average returns were calculated for the stocks in each portfolio from equation (20) using a Monte Carlo simulation. 10^4 samples were taken for each asset. The ratio of return to risk is then calculated from the portfolio

generated for each value of risk parameter. These results allow for a direct comparison of the ratios predicted by the objective functions in equations (16) and (17). These results are shown in figure 3.
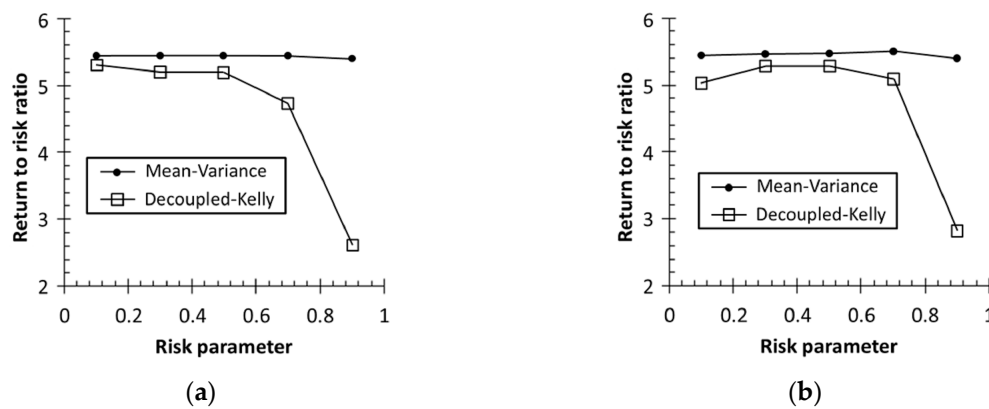
Figure 3. (a) Return-to-risk ratios for the portfolios generated from differential evolution for each model. Ratios are shown as a function of risk parameter. (b) Return-to-risk ratios generated using Monte Carlo for the portfolios use in (a).

The pairs of curves shown in figure 3(a) and 3(b) are very similar and the calculations of return and risk from the objective functions are consistent. The curves in each graph follow the same trend. The Monte Carlo model also shows the anamolous very low value for risk-to-return ratio in the decoupled Kelly model for $P = 0.9$. These resutls confirm the accuracy of the returns and risk predicted under the decoupled Kelly model.

5. Discussion and conclusions

Although the algorithm can return what is known to be the optimized solution for both models (depending on the value of risk parameter), there is a clear difference in the convergence rates. To improve the convergence rate in the decoupled Kelly model, the mutation step in the solution algorithm needs to be modified. A number of mutation strategies can be found in [19]. It is well documented in the differential evolution literature that different objective functions have different responsivity to different mutation strategies. A mutation strategy may need to be tailored specifically to the decoupled Kelly problem for different values of risk parameter. This may also help to prevent the mis-convergence that occurs at $P = 0.9$.

It has also been noted in (Thorpe 1997, Samuelson 1971) that the properties of the predictions under Kelly's Criterion are asymptotic. This means that the observed risk and return for the portfolio will only converge to the value predicted from the decoupled Kelly model as the number of time periods approaches infinity. The model presented here generates results that are most appropriate for making long term investment decisions. In reality the mean return rate is bounded in a confidence interval, however the size of this interval decreases with increasing investment horizons.

In conclusion, the results in this paper show how Kelly's Criterion can be implemented into a portfolio optimization model that combines risk and return into a single objective function using a risk parameter. The two models tested in this study return the optimized portfolios for moderate values of risk parameter, however, the returns predicted by the two portfolios are slightly different. The solutions to these models were found using a differential evolution algorithm. The rate of convergence of the algorithm was slower for the decoupled Kelly model than for the MV model. The

decoupled Kelly model mis-converges for one of the values of risk parameter tested in this study. The accuracy of the objective functions in these models was confirmed using a Monte Carlo simulation that calculated return-to-risk ratios using the portfolios generated from the optimization problem. The results are a proof of concept and show that the decoupled Kelly model is a valid portfolio optimization model that can produce an optimized portfolio as well as accurate calculations of return and risk. Further work in this area should focus on improving the solution algorithm to more effectively and quickly reach the solution to the decoupled Kelly problem.

Appendix A

Pseudocode for the differential evolution algorithm used in this study.

```
C ∈ [0, 1];                                          // Crossover probability
F ∈ [0, 2];                                          // Scale factor
P ∈ [0, 1];                                          // Risk parameter
m = 0;                                               // Integer, counts iterations
Initialize N;                                        // Number of assets
Initialize [f]_0 = [f]_initial;                      // Initial [f]_m values for m = 0
Initialize K_min, K_max;                             // Min and max values for f_k

timer = StartTimer();                                // Initialize timer as external object

while (timer < 30)                                   // Main loop
    for (k = 1; k < N; k++)
        while (v_k > K_max or v_k < K_min)           // Generate trial vectors
            do p1 = rand[1, N]/{i}; while (p1 = k);
            do p2 = rand[1, N]/{i}; while (p2 = k or p2 = p1);
            do p3 = rand[1, N]/{i}; while (p3 = k or p3 = p2 or p3 = p1);
            v_k = f_p1 + F(f_p2 − f_p3) + N(0, 0.01); // Trial solution with added noise
        end while
        J = rand[1, NP]/{i};                         // Automatic crossover index
        R = rand[0, 1];
        if (R ≤ C or J ≠ k)
            u_k = v_k;
        else if (R > C and J ≠ k)
            u_k = f_k;
    end for
    if (Obj([u]_m) > Obj([f]_m) and ‖u‖_m = 1 ± 0.001)
        [f]_m+1 = [u]_m;
        m++;
        timer = 0;
    else
        [f]_m+1 = [f]_m;
        m++;
end while
```

References

1. (Bichpuriya and Soman 2016) Bichpuriya, Yogesh K. and S. A. Soman. 2016. Application of Probability Density   Forecast of Demand in Short Term Portfolio Optimization. Presented at 2016 IEEE International   Conference on Power System Technology (POWERCON), Wollongong, NSW, Australia, 28 Sept.-1 Oct.   2016.

2. (Chang, et. al. 2009) Chang, Tun-Jen, Sang-Chin Yang, and Kuang-Jung Chang. 2009. Portfolio optimization   problems in different risk measures using genetic algorithm. *Expert Systems with Applications*, 36(7),   10529-10537.

3. (Chen, et. al. 2012) Chen, Angela H., Yun-Chia Liang, and Chia-Chien Liu. 2012. An artificial bee colony   algorithm for the cardinality constrained portfolio optimization problems. In *IEEE Congress on Evolutionary   Computing*. Paper presented at 2012 IEEE Congress on Evolutionary Computation (CEC), Brisbane, QLD,   Australia, 10-15 June 2012.

4. (El Ghaoui, et. al. 2003) El Ghaoui, Laurent, Maksim Oks, and Francois Oustry. 2003. Worst-case value-at-risk   and robust portfolio optimization: A conic programming approach. *Oper. Res. Journal*, Vol. 51, 543-556.

5. (Fama and French 2004) Fama, Eugene F. and Kenneth R. French. 2004. The Capital Asset Pricing Model: Theory and Evidence. *Journal of Economic Perspectives*, Vol. 18, No. 3, pp. 25-46.

6. (Fisher 1990) Fisher, L. 1990. Probability and Statistics. In *Handbook of Applied Mathematics*, 2nd ed. Edited by   Carl E. Pearson. New York: Van Nostrand Reinhold, pp. 1248-1252. 0-442-00521-0.

7. (Joshi [2003] 2008) Joshi, Mark S. 2008. *The Concepts and Practice of Mathematical Finance*. Cambridge, UK: Cambridge University Press, pp. 111-112. First published in 2003. 978-0-521-51408-8.

8. (Kamili and Riffi 2016) Kamili, Hamza and Mohammed Essaid Riffi. 2016. A comparative study on portfolio   optimization problem. Paper presented at International Conference on Engineering & MIS (ICEMIS),   Agadir, Morocco, 22-24 Sept. 2016.

9. (Kelly 1956) Kelly, J. L. A new interpretation of information rate. *Bell Systems Technical Journal*, vol. 35, pp.   917-926.

10. (Kim 2008) Kim, G. T. On the Applicability of Kelly Criterion to Engineering Economics. In *Proceedings of the 2008 IEEE IEEM*, pp. 267-272. Paper presented at 2008 IEEE International Conference on Industrial   Engineering and Engineering Management, Singapore, Singapore, 8-11 Dec. 2008.

11. (Korczak and Roger 2000) Korczak, Jerzy, and Patrick Roger. 2000. Portfolio Optimization using Differential   Evolution. *Prace Nankowe Akademii Ekonomicznej. AASRI Procedia*.

12. (Ma, et. al. 2012) Ma, X., Yuelin Gao, and Bo Wang. 2012. Portfolio Optimization with Cardinality Constraints   Based on Hybrid Differential Evolution. *AASRI Procedia*, 1, 311-317.

13. (Markowitz 1952) Markowitz, Harry. 1952. Portfolio Selection. *Journal of Finance*, Vol. 7, #1, pp. 77-91.

14. (Qin, et. al. 2009) Qin, A. K., V. L. Huang, and P. N. Suganthan. 2009. Differential Evolution Algorithm with   Strategy Adaptation for Global Numerical Optimization. *IEEE Transactions on Evolutionary Computation*,   Vol. 13, No. 2.

15. (Rogalsky, et. al. 1999) Rogalsky, T., R. W. Derksen, and S. Kocabiyik. 1999. Differential evolution in   aerodynamic optimization. *Proc. 46th Annual Conf. on Canadian Aeronautical Space Inst.*, Montreal, QC,   Canada, pp. 29-36.

16. (Samuelson 1971) Samuelson, Paul A. 1971. The 'Fallacy' of Maximizing the G Mean in Long Sequences of   Investing or Gambling. *Proceedings of the National Academy of Sciences*, Vol. 68, No. 10, 2493-2496.

17. (Storn and Price 1997) Storn, Rainer, and Kenneth Price. 1997. Differential Evolution - A Simple and Efficient   Heuristic for Global Optimization over Continuous Spaces. *Journal of Global Optimization*, Vol. 11, 341-359.

18. (Thorpe 1997) Thorpe, Edward O. 1997. The Kelly criterion in blackjack, sports betting, and the stock market.   *Proc. 10th Inter. Conf. on gambling and risk taking*.

19. (Vince 2011) Vince, Ralph. 2011. Optimal f and the Kelly Criterion. *IFTA Journal (2011 Edition)*, pp. 21-28.

20. (Yang and Liu 2016) Yang, Guang, and Xinwang Liu. 2016. A Improved Algorithm for Fuzzy Multistage   Portfolio Optimization Model. Paper presented at 2016 IEEE International Conference on Fuzzy Systems   (FUZZ-IEEE), Vancouver, BC, Canada, 24-29 July 2016.

21. (Zaheer and Pant 2016) Zaheer, Hira, and Millie Pant. 2016. Solving Portfolio Optimization Problem Through   Differential Evolution. In *IEEE ICEEOT Proceedings*, 978-1-4673-9939-5/16. Paper presented

at International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), Chennai, India, 3-5 March 2016.