Determination of the Control Parameters of Evolutionary Techniques Utilizing Analysis of Variance

Zuzana Čičková, Mária Vojtková, Marian Reiff

University of Economics in Bratislava, Faculty of Economic Informatics, Slovakia

Abstract – **Evolutionary techniques are considered to be an alternative to the effective solution of many, especially non-standard optimization problems. However, these algorithms are controlled by a special set of control parameters according to algorithm type and these settings radically affect the quality of the results. Paper focuses on the possibility of using selected statistical methods to determine the effective values of control parameters, while algorithm of differential evolution has been chosen as an evolutionary technique representative. The analysis has been elucidated on an illustrative example of Sortino ratio maximization problem which is one of the nonlinear measures of portfolio performance.**

Keywords – **Evolutionary techniques, Differential evolution, Sortino ratio, Control parameters, Analysis of variance.**

1. Introduction

Nowadays, different evolutionary techniques are widely and successfully used to solve a wide range of optimization problems. They are generally considered to be effective tools when it is hard to use traditional optimization techniques (e.g. when computability can be difficult due to substandard

structures or in the case of combinatorial optimization). The big advantage they hold over traditional methods is that they are designed to find global extremes (with built-in stochastic components) and their use does not require a priori knowledge of optimized functions (convexity, differential etc.). However, those algorithms are controlled by a special set of parameters according to their type that also causes limitations, whilst improper settings can radically affect the quality of the results obtained. Some of these parameters can be successfully set exogenously based on the philosophy of the algorithm, but there is not a deeper theoretical base whereby certain parameters can be adjusted (e.g. parameters determining the rate stochastics). Recommended values for the parameters are usually derived only "empirically from experiments" [7], [8], [14]. However, statistical methods offer tools that may be helpful in clarifying the control parameters settings of evolutionary techniques [1]. Problems pertaining to parameter control in evolutionary algorithms have not been very well researched and publications in this field are scarce. For example, Eiben, Hinterding and Michalewicz [2] provide a survey covering various forms of parameters control in evolutionary algorithms. This paper is focused on the possibility of using selected statistical methods which can be efficaciously used to determine the effective values of the control parameters of evolutionary techniques. The well-known and widely used differential evolution (DE) has been chosen as representative of evolutionary techniques, but the approach presented can be used generally when one wants to use evolutionary computation. A calculation is provided as an illustrative example. The use of statistical methods has been elucidated by an illustrative example, where one of the nonlinear measure techniques aimed at portfolio performance measurement (Sortino ratio) in financial modelling has been calculated.

The paper is divided into the following interrelated parts. A brief view is presented on the measurement techniques of portfolio performance in finance

theory; in addition, the motivation for the use of evolutionary techniques to calculate portfolio performance is presented in the first part. This part also includes an outline of the basic principle of the algorithm of Differential evolution and its modification to solve the portfolio selection problem based on the afore-mentioned measure. The core part of this paper is devoted to a presentation of relevant statistical methods that are helpful when one wants to provide a setting of the control parameters of differential evolution. The last part comprises the practical results. The computation is provided using the Dow Jones Industrial Average, one of the major market indices, as well as one of the most popular indicators of the U.S. market. The historical data published in the period February 6nd 2017 to January 29th 2018 on a weekly basis have been used.

2. Methods of analysis

2.1. Calculation of Sortino ratio by Differential evolution

Differential evolution was firstly introduced by Storn and Price and Storn [13] and to date it is considered to be a powerful evolutionary technique that has been successfully applied to address a wide variety of optimization problems. For example Mařík, Štěpánková and Lažanský [5] highlight the performance of the algorithm in finding free extremes of several test functions such as Rosenbrock's saddle, Ackley function, Egg holder. Onwubolu and Babu [7] showed the solution of problems with integer and discrete variables for example Design of Gear Train, Design of Pressure Vessel and Design of Coil Spring and also combinatorial problems such as Travelling Salesman Problem and Flow Shop Scheduling. Permutation optimization problems solutions are discussed by Onwubolu and Davendra [8], for example Knapsack Problem, Travelling salesman Problem, Automated Drilling Location and Hit Sequencing, Quadratic Assigment Problem, Flow Shop Scheduling. The economic problem of portfolio selection is addressed in [10]. This algorithm has several features in common with genetic algorithms, but there are a number of differences: for example, the offspring is made up of four (and not two) parents, and stochastics is applied in two ways, namely by crossing and also by creating a "noise" vector. Detailed information about Differential evolution can be found in the afore-mentioned papers, so the subsequent section only presents outline information about the control parameters of the algorithm. The control parameters are as follows:

d – dimensionality. Number of parameters of individual (usually equates to the number of variables of solved problems).

np – population size. Number of individuals in population. Recommended setting is $5d$ to $30d$, respectively $100d$, in cases where the optimized function is multimodal.

g – generations. Represents the maximum number of iterations (g is also the stopping criterion).

cr – crossover constant, $cr \in \langle 0,1 \rangle$.

f – mutation constant, $f \in \langle 0,1 \rangle$.

The principle of the fundamental version of Differential evolution can be described using the following pseudocode:

```
BEGIN
  SETTING of control parameters;
  INITALIZATION of population;
      EVALUATION of each individual;
        WHILE (STOPPING CRITERION is not satisfied)
  DO
          FOR (each individual of the population) DO
(REPRODUCTIVE CYCLE):
              CREATE differential vector
              CREATE trial vector
              CREATE test vector
              IF (EVALUATION of test vector) >
(EVALUATION of current selected individual)
              THEN (SUBSTITUTE the selected individual
              with the test vector)
                  ENDIF
      ENDFOR
      ENDWHILE
      EVALUATE process of calculating
  END
```

The following section briefly introduces the steps of the algorithm to calculate this Sortino ratio. The Sortino ratio [11] measures an investment's risk-adjusted returns based on downside semi-variance and penalizes only those returns falling below a user-specified rate. This is a measurement of return deviation below a minimally acceptable rate. By utilizing this value, the Sortino ratio only penalizes for "harmful" volatility. The Sortino ratio calculation assumes the existence of historical data on portfolio returns for T periods, so that for each period t ($t = 1, 2, ... T$) the portfolio return R_t is known. The Sortino ratio uses a specified value of minimal expected return (MAR) that is exogenously set by the investor.

The formulation of the portfolio selection problem based on the Sortino ratio includes the maximization of the performance measurement function and can be provided according to the mathematical model (1) – (3). The model deals with the variables w_i, $i = 1, 2, \ldots n$ (where n represents the number of portfolio assets) that represent the weights of each asset in the portfolio. The corresponding problem can be formulated as follows [12]:

$$\max \ SR(\mathbf{w}) = \frac{\sum_{i=1}^{n} E_i w_i - MAR}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\max\left(MAR - \sum_{i=1}^{n} r_{ti} w_i, 0\right)\right)^2}} \quad (1)$$

$$\sum_{i=1}^{n} w_i = 1 \quad (2)$$

$$w_i \geq 0, i = 1, 2, \ldots n \quad (3)$$

where

T represents the number of periods,

MAR denotes the minimum acceptable return,

r_{ti} denotes the return of the i-th portfolio asset in the t-th period, where $t = 1, 2, \ldots T$, $i = 1, 2, \ldots n$,

E_i denotes the expected returns of the i-th portfolio asset, $i = 1, 2, \ldots n$.

The computational complexity of the problems based on the the (1) – (3) calculation arises from its non-linear structure. Therefore, evolutionary algorithms seem to be a suitable alternative to standard techniques due to their ability to achieve suboptimal solutions in a relatively short period of time. Using any evolutionary technique to solve special types of problems involves some special modification of the algorithm base. In order to apply DE to solve problems of portfolio selection based on performance measurement, it is also necessary to consider the following factors: *selection of an appropriate representation of the individual, transformation of unfeasible solutions, setting of control parameters.*

The population $\mathbf{P}^{(0)}$ was randomly initialized at the beginning of the evolutionary process according to the rule: $w_{i,j}^{begin} = \dfrac{rnd\langle 0,1\rangle}{\sum_{l=1}^{d} w_{i,1}^{begin}}$ $i = 1,2,\ldots np$ $j = 1,2,\ldots d$

that ensure that the total weights of the portfolio is equal to one. Each individual is then evaluated with the *fc* (given by function SR(**w**)).

The test of the stopping condition in its canonical form covers achieving the maximum number of iterations (represent by parameter *g*). The *reproductive cycle* comprises the process of crossing and mutation to create individuals for the next generation. For each individual $\mathbf{w}_i^g$, $i = 1, 2, \ldots np$, from the population another three different individuals are chosen (vectors **r1**, **r2**, **r3**). The difference between the first two vectors (**r1** and **r2**) gives the *differential vector*, which is multiplied by mutation constant *f* and added to vector **r3** and this gives *trial vector* **v**. After the mutation process comes the formation of a new individual, which is also called *test vector* $\mathbf{w}^{\mathbf{test}}$ so that one element after another is selected from the currently selected individual $\mathbf{w}_i^g$ and from the trial vector **v** and for

every pair a random number from the interval $\langle 0,1 \rangle$ is generated, which is compared with the crossing constant *cr*. If the generated random number is less than or equal to *cr*, to the relevant position of $\mathbf{w}^{\mathbf{test}}$ comes the element of *trial vector* **v,** otherwise the current selected individual is $\mathbf{w}_i^g$.

To ensure the feasibility of our solution, we use the following rule: if $w_j^{test} < 0$, then $w_j^{test} = rnd\langle 0,1 \rangle$ and

$$\mathbf{P}_{(i)}^{test} = \mathbf{w}_{i,j}^{test} = \frac{w_j^{test}}{\sum_{j=1}^{d} w_j^{test}} \qquad i = 1, 2, \ldots np \ \ j = 1, 2, \ldots d$$

, where *k* is a random index, which always ensures a change of at least one parameter in the test vector. The value of the objective function for the test vector is compared to the value of the objective function of the current selected individual and for the next generation the vector with the better objective value is selected, so that process continues in each generation for all individuals. The result is a new generation with the same number of individuals.

The steps concerning how to adjust the algorithm to enable this problem to be solved (1) – (3) are formally described in [1] where a similar problem was presented, but solved by a Self-Organizing Migrating Algorithm, or in the article [9] that was focused on Differential evolution.

2.2. The analysis of control parameters values with the help of selected statistical methods

Some of the control parameters of DE are given directly owing to the nature of the problem and can only be changed by its reformulation. An example of such a parameter is *dimensionality (d)*, which usually equates to the number of variables in the optimized model (in the case of problem (1) – (3), it is equal to the *number of assets (n)*). Parameters *population size (np)* and *generations (g)* determine "the size and length" of the simulation and their settings can adopt the philosophy "more is better" (however, increasing these parameters affects the time needed to make calculations and thus is dependent on the user's hardware). The setting of the remaining parameters *crossover constant (cr)* and *mutation constant (f)* are limited only by their recommended settings. Our further analysis is focused on the *f* and *cr* parameters values.

It can be appropriate to run several simulations with, e.g., smaller population sizes and lower numbers of iterations (which are not time-consuming) with different values of control parameters before making final calculations when one wants to adjust the control parameters *cr* and *f*. Furthermore, statistical methods of analysis of

variance can be applied in order to select values of f and cr parameters values.

Now let us introduce the testing example. The testing example considers the calculation of the Sortino ratio according to (1) - (3) and the portfolio analysis was based on the Dow Jones Industrial (DJI). Data, which can be downloaded from http://finance.yahoo.com, were processed weekly for the period February 6th 2017 to January 29th 2018. The list of 29 companies is as follows: #1. Procter & Gamble (PG), # 2. United Technologies (UTX), #3. Travelers (TRV), #4. Nike (NKE), #5. Verizon (VZ), #6. Apple Inc. (AAPL), #7. Coca-Cola (KO), #8. American Express (AXP), #9. Caterpillar (CAT), #10. Johnson & Johnson (JNJ), #11. Boeing (BA), #12. JPMorgan Chase (JPM), #13. Visa (V), #14. Microsoft (MSFT), #15. Walt Disney (DIS), #16. Goldman Sachs (GS), #17. ExxonMobil (XOM), #18. Cisco Systems (CSCO), #19. The Home Depot (HD), #20. Merck (MRK), #21. Wal-Mart (WMT), #22. 3M (MMM), #23. IBM (IBM), #24. General Electric (GE), #25. Intel (INTC), #26. Chevron (CVX), #27. McDonald's (MCD), #28. Pfizer (PFE), #29. UnitedHealth Group (UNH). The analysis does not include DWDP DowDuPont Inc. because of missing data up until March 2018. The time series consisted of 52 observations (for each company).

The DE algorithm was implemented in MATLAB 8.3. All the experiments were run on a PC with Intel ® Core ™ i7-3770 CPU with a frequency of 3.40 GHz and 8 GB of RAM under MS Windows 8. The input parameter of MAR (the target of the required rate of return) was set to 0.005. Some of the control and termination parameters were set as follows: Parameter d was equal to 29 (the number of assets), parameter np was set to 250 and parameter g was set to 100 (the small size of the instance enabled all iterations to be achieved in a relatively short period of time).

The settings of parameters f and cr on the basis of statistical methods are described in the next section. The analysis of variance (ANOVA) was used in practice as a separate technique or procedure that permitted the analysis of the sources variability in statistical models [3], [6]. Its main function was to assess the impact of one or more independent variables (factors) on the quantitative variable. The most frequent case of analysis of variance is a One-way analysis of variance: this assumes that the observed dependent variable is affected by only one factor. When examining the impact of several factors, we are dealing with a Multifactor analysis of variance. Where we distinguish models that analyse the impact of major effects (direct factor influence) and models considering iterative effects of factors (the resulting effect of the joint action of two or more factors). In practical tasks, first, the use of the parametric analysis of variance option is verified, which requires the fulfilment of certain initial prerequisites. If we cannot confirm their validity or the validity of an assumption is rejected, the conclusions obtained by applying parametric methods may lead to inaccurate conclusions. More reliable conclusions in this case can provide non-parametric methods of analysis of variance, despite their lower test strength. In this paper, the Kruskal-Wallis non-parametric sequence test [4] was used, for which the probability of type II error is greater than for the ANOVA parametric method. Using this test, the distribution of k random samples may be unknown or known, the size of k random samples may be small, and the variance of k random samples may be significantly different. The condition of their application is only the continuity of the dependent random variable. The advantage of using the Kruskal-Wallis test is its lesser sensitivity to extreme values. On the other hand, it has to be noted that the nonparametric Kruskal-Wallis test uses less information from the data because it does not take into account the differences between the values of the continuous random variables, just the position from the values order, so it is less efficient.

3. Setting the control parameters of the DE algorithm

One of the commonly used statistical methods is analysis of variance (ANOVA), which serves to compare the mean values of more than two basic files. This method results from the fact that determining whether or not the mean values of a few basic sets are different is based on an analysis of the various forms of variance associated with random choices. The goal of the method can also be formulated as verifying the effect of one or more factors on a quantitative variable. In this paper, we decided to use the results of this method to set the control parameters of the DE algorithm.

The aim of the analysis (experiment) was to determine the effect of parameters (factors) cr and f on the variable fc, which is the Sortino ratio (1). To set the control parameters cr and f, we followed the recommended intervals for setting these parameters, namely $cr, f \in \langle 0,1 \rangle$. To test the efficacy of parameters cr and f for both input parameters, the following levels were selected: 0.1; 0.2; 0.3; 0.4; 0.5; 0.6; 0.7; 0.8 and 0.9. Border points 0 and 1 were purposefully excluded from the analysis due to their undesirable behaviour (extremely "bad" values of the variable fc). Because of this, to achieve a balanced experiment plan (for each pair we carried out the same number of experiments), we attained more reliable data than for the unbalanced experiment plan (for each pair we carried out a different number of

experiments), the same number of simulations were made - ten experiments for each pair of factor cr and f. In this way, a total of 810 simulations were made. An initial projection of the influence of the control parameters can be seen in Figure 1. It shows the average value of SR results from ten simulations in relation to the setting of the control parameters f and cr.
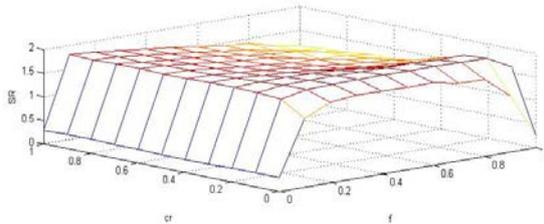

Figure 1. Influence of the control parameters on performance measure fc - Sortino ratio (SR) Source: Own construction

To determine whether the levels of factor cr and f statistically affect - significantly or insignificantly - the variable fc, one-way ANOVA was used first. The results of the analysis are shown in Tables 1. and 2. in ANOVA tables. Since the p-value is less than α in both cases, there is justification to reject the hypothesis of an equivalence of mean values at different levels of factor cr and f. The fraction of variability of the variable fc explained by factor cr is 34.63% and for factor f it is 44%. Based on this result, it can be stated that the levels of factor cr and f statistically significantly affect the variable fc.

Table 1. ANOVA Table for fc by cr

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Between groups	7.64831	8	0.956039	53.05	0.0000
Within groups	14.4345	801	0.0180206		
Total (Corr.)	22.0828	809			

Source: Own construction

Table 2. ANOVA Table for fc by f

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Between groups	9.71711	8	1.21464	78.68	0.0000
Within groups	12.3657	801	0.0154379		
Total (Corr.)	22.0828	809			

Source: Own construction

In Table 3., the basic descriptive statistics of variable fc are summarised for the individual levels factor cr. The Minimum and Maximum columns allow the variation range of fc variable values to be specified, which gives us an initial idea about the variability of its values. The Average and Median columns specify the central tendency of the fc response variable values. The column with values of standard deviations and variation coefficients also provides a basic idea of the variability of the fc values. For homogeneous groups $cr = 0.1$ and $cr = 0.2$, the variability of the variable fc is the lowest whereas the highest variability is at $cr = 0.9$, which is a consequence of the increasing influence of the mutation (by crossing) on the population's diversity.

Table 3. Summary statistics of performance measure fc at different values of cr

cr	Count	Average	Median	Stand. Dev.	Coeff. of var. %	Min	Max
0.1	90	1.790	1.794	0.022	1.215	1.709	1.832
0.2	90	1.799	1.792	0.040	2.222	1.718	1.865
0.3	90	1.773	1.744	0.071	3.979	1.666	1.887
0.4	90	1.741	1.706	0.096	5.500	1.602	1.896
0.5	90	1.706	1.671	0.118	6.941	1.535	1.900
0.6	90	1.660	1.611	0.142	8.542	1.411	1.902
0.7	90	1.616	1.570	0.167	10.345	1.362	1.903
0.8	90	1.561	1.507	0.194	12.433	1.227	1.903
0.9	90	1.515	1.454	0.215	14.194	1.192	1.892
Total	810	1.684	1.726	0.165	9.808	1.192	1.903

Source: Own construction

In the case of the parametric analysis of variance, it is important to verify the conditions of its use. Several tests were performed to verify homoscedasticity, Table 4. contains the results of the Leven test of variance equality at various levels of factor cr.

Table 4 Variance Check

	Test	P-Value
Levene's	57.576	0

Source: Own construction

The results of the test show that at the 0.05 level of significance there is justification to reject the hypothesis about the equality of variances, i.e. the variability of the fc; in response to different levels of the factor cr, the fc variability is not the same, which could lead to a distortion of the results obtained by the one-way ANOVA. Therefore, the Kruskal-Wallis nonparametric median test was used for testing (see Table 5.). The Kruskal-Wallis test also confirmed the statistically significant differences between medians at each cr level.

Table 5. Kruskal-Wallis Test for performance measure fc at different values of cr

Cr	Sample Size	Average Rank
0.1	90	545.511
0.2	90	559.878
0.3	90	513.456
0.4	90	466.3
0.5	90	416.422
0.6	90	354.889
0.7	90	308.444
0.8	90	259.311
0.9	90	225.289

Test statistic = 205.071   P-Value = 0
Source: Own construction

The summary results of our analysis are also indicated by the graphical presentation displayed by the Box-plot charts shown in Figure 2., for each individual *cr* level. Each individual Box-plot chart displays the minimum, lower quartile, the median, the average, upper quartile and maximum values. If there are remote observations in the data, the values are represented by points. It is obvious from the box-plots in Figure 2. that the level of the variable *fc* differs at different levels of factor *cr*. Thus, we see that increasing the number of crossings in the population increases the variability of response variable, resulting from an increase in the formation of new individuals. At a lower value of *cr*, the algorithm responds more according to deterministic rules, so the variability decreases. Therefore, it is likely to be more suitable to use smaller values of the crossing constant, because in the case of any excessive overrun, there is a disturbance of the parts of the individuals.
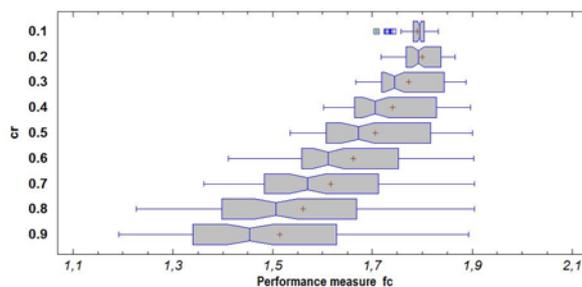

Figure 2. Box and Whisker plots for values of performance measure fc at different values of cr Source: Own construction

For verification of homoscedasticity, an assumption of the test of variance equality at various levels of factor *f* was conducted - see Table 6. The results of the Leven test show that at the 0.05 level of significance, there is reason to reject the variance equality hypothesis, so that the variability of the values of the variable *fc* for the different levels of the factor *f* is not the same, which could lead to a distortion in the results obtained by the one-way ANOVA. Therefore, the Kruskal-Wallis median difference test was again used to test the median values equality, see Table 7.

Table 6. Variance Check

	Test	P-Value
Levene's	49.1223	0

Source: Own construction

Table 7. Kruskal-Wallis Test for performance measure fc at different values of f

f	Sample Size	Average Rank
0.1	90	727.778
0.2	90	680.356
0.3	90	488.989
0.4	90	372.489
0.5	90	314.067
0.6	90	286.911
0.7	90	271.756
0.8	90	262.1
0.9	90	245.056

Test statistic = 450.613   P-Value = 0
Source: Own construction

The Kruskal-Wallis test also confirmed the statistically significant differences between medians at each *f* level. The afore-mentioned results of our analysis are shown in Box-plot charts in Figure 3. Based on the summary statistics shown in Table 8. and the Box-plot, we can see that the smallest variability of the variable *fc* is related to the low value *f* (0.2 and 0.1), which can also be explained by the decrease in the stochastic component of the DE algorithm. The variability characterized by the coefficient of variation at other levels of *f* can be considered to be approximately the same ranges from 5% to 11.6%.

Table 8. Summary Statistics of performance measure fc at different values of f

| f	Count	Average	Median	Stand. Dev.	Coeff. of var. %	Min.	Max.
0,1	90	1.873	1.892	0.047	2.5	1.709	1.903
0,2	90	1.854	1.863	0.026	1.4	1.768	1.879
0,3	90	1.752	1.799	0.087	5.0	1.472	1.854
0,4	90	1.682	1.714	0.111	6.6	1.411	1.828
0,5	90	1.640	1.654	0.121	7.4	1.354	1.812
0,6	90	1.615	1.627	0.137	8.5	1.306	1.816
0,7	90	1.593	1.625	0.161	10.1	1.281	1.820
0,8	90	1.589	1.603	0.154	9.7	1.275	1.832
0,9	90	1.562	1.591	0.181	11.6	1.192	1.803
Total	810	1.684	1.726	0.165	9.8	1.192	1.903

Source: Own construction
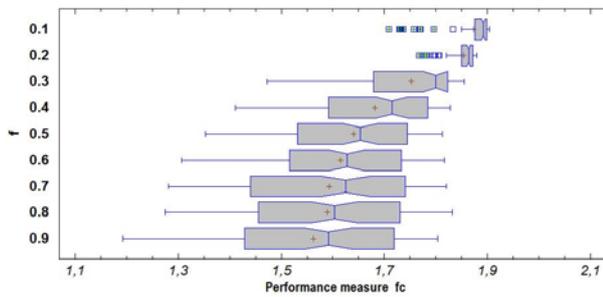
Figure 3. Box and Whisker plots for performance measure fc at different values of f Source: Own construction

Since both factors f and cr act on the variable simultaneously, the influence of their interaction alone was studied and a two-way analysis of the variance (Two-Way ANOVA) was conducted, the results being shown in Table 9. The test also confirmed the statistical significance of the interaction effects between cr and f, with the proportion of variability fc explained by an interaction effect of 17.05%. Overall, the variability of variable fc can be explained at 95.68% by both factors and their interactions at the same time.

Table 9. Analysis of Variance for performance measure fc - Type III Sums of Squares

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Main effects					
A:f	9.71711	8	1.2146388	930.54	0.000
B:cr	7.64831	8	0.9560388	732.42	0.000
Interactions					
AB	3.76584	64	0.0588413	45.08	0.000
Residual	0.95157	729	0.0013053		
Total (corrected)	22.0828	809			

All F-ratios are based on the residual mean square error. Source: Own construction

The results presented in this section demonstrate the statistically significant influence of DE control parameters on the quality of the results achieved. It can be stated that the high set-up of the two monitored control parameters leads to a high variability of the results achieved and also has an adverse effect on the quality of the results achieved and, therefore, for the given type of problem, we recommend setting the values at lower levels. On the other hand, setting the parameters f and cr to 0.1 was excluded because in both cases the occurrence of remote observations - outliers - could be observed. Next, the results of the solution are stated for the

problem with an increased number of parameters np = 3000 (number of individuals in the population) and g = 1500 (number of iterations), with the parameters f and cr also set to a value of 0.2. Altogether, we performed 5 simulations, and in each case approximately the same value of the Sortino (SR = 1.91) ratio was achieved. The weights of the individual shares are thus: #9 (Caterpillar (CAT)) - 0.0829, #11 (Boeing (BA)) - 0.4145, #14 (Microsoft (MSFT)) - 0.0309, #19 (Home Depot) – 0.0829, #21(Wal-Mart) - 0.3888. The other companies listed in part 1.2 have zero weighting. The value of the Sortino ratio is 1.91.

4. Conclusion

Evolutionary algorithms are considered to be a universal and effective tool for solving various optimization problems, especially when it is difficult to use standard optimization methods, for example in the case of non-linear fumigations. Their effectiveness is limited by the fact that they are generally controlled by special sets of parameters. Although some of parameters can be successfully set exogenously based on the philosophy of the algorithm or according to the type of problem being solved, there is no deeper theoretical basis to adjust all the parameters. This paper has focused on the possibility of using some statistical methods that may be helpful in determining the effective values of some control parameters. As a representative of evolutionary techniques, we chose the algorithm of Differential evolution and, as an illustrative example, we chose the maximization of the Sortino ratio in the field of financial modelling. The setting of control parameters can be supported by statistical methods aimed especially at determining whether the level of some parameters achieves any difference in results.

In the first part of the paper we briefly characterized the stock selection problem solved by modifying the differential algorithm. Based on the results obtained, we presented the possibility of using statistical methods to analyse the setting of two control parameters of this algorithm, which is the main part of the article. At the end of this section we presented a possible solution to the illustrative example.

References

[1]. Čičková, Z., & Lukáčik, M. (2016). *Setting of Control Parameters of SOMA on the Basis of Statistics*. In: D. Davendra and I. Zelinka (eds.), Self-Organizing Migrating Algorithm, Studies in Computational Intelligence 626. Springer International Publishing Switzerland.

[2]. Eiben, Á. E., Hinterding, R., & Michalewicz, Z. (1999). Parameter control in evolutionary algorithms. *IEEE Transactions on evolutionary computation*, *3*(2), 124-141.

[3]. Hendl, J., (2006). *Přehled Statistických Metod Zpracování Dat: Analýza a Metaanalýza Dat.* Praha: Portál.

[4]. Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, *47*(260), 583-621.

[5]. Mařík, V., Štěpánková, O., & Lažanský, J. (2003). *Umělá Inteligence 4*. Academia Praha.

[6]. Meloun, M., & Militký, J. (2012). *Interaktivní Statistická Analýza dat.* Praha: Korolinum.

[7]. Onwubolu, G. C., & Babu, B. V. (2004). *New Optimization Techniques in Engineering*. Springer-Verlag, Berlin-Heidelberg.

[8]. Onwubolu, G. C., & Davendra, D. (2009). *Differential Evolution: a Handbook for Global Permutation-Based Combinatorial Optimization*. Springer-Verlag, Berlin-Heidelberg.

[9]. Pekár, J., Brezina, I., Cicková, Z., & Reiff, M. (2013). Portfolio selection by maximizing omega function using differential evolution. *Technology and Investment*, *4*(1B), 73-77.

[10]. Pekár, J., Čičková, Z., & Brezina, I. (2016). Soma in Financial Modeling. In: D. Davendra and I. Zelinka eds., *Self-Organizing Migrating Algorithm, Studies in Computational Intelligence*. Springer International Publishing Switzerland.

[11]. Sortino, F. A., & Van Der Meer, R. (1991). Downside risk. *Journal of Portfolio Management*, *17*(4), 27-31.

[12]. Sortino, F. A., & Price, L. N. (1994). Performance measurement in a downside risk framework. *the Journal of Investing*, *3*(3), 59-64.

[13]. Storn, R., & Price, K. (1997). Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *Journal of Global Optimization, 11*(4), 341–359.

[14]. Zelinka, I. (2002). *Umělá inteligence v problémech globální optimalizace*. BEN-technická literatura, Praha.