International Journal of Operations Research

International Journal of Operations Research Vol. 3, No. 1, 16−22 (2006)

Non-linear Stochastic Optimization Using Genetic Algorithm for Portfolio Selection

S.-M. Wang[1], J.-C. Chen[1], H.-M. Wee[1, *], and K.-J. Wang[2]

[1]Department of Industrial Engineering, Chung-Yuan Christian University , 200, Chung Pei Road, Chung Li, 320, Taiwan, R.O.C.

[2]Department of Business Administration, National Dong Hwa University, No.1, Sec. 2, Da Hsueh Road, Shoufeng, Hualien, 974, Taiwan, R.O.C.

Received January 2005; *Revised October 2005*; *Accepted November 2005*

Abstract**—**Portfolio optimization is an important research field in modern finance. The most important characteristic within this optimization problem is the risk of the returns. In this paper, a non-linear stochastic optimization algorithm named Stochastic Portfolio Genetic Algorithm (SPGA) is proposed to determine a profitable portfolio selection planning plan under risk. The algorithm improves a conventional two-stage stochastic programming by integrating a genetic algorithm into a stochastic sampling procedure to solve this large-scale portfolio selection optimization. The tradeoff between returns and risks is evaluated under different settings of algorithmic and hedging parameters. Finally, the historical data from Taiwan Stock Exchange are used to evaluate SPGA's performance. Results show that a practical problem can be efficiently solved and the expected return of SPGA outperforms the one in the market.

Keywords**—**Genetic algorithm, Portfolio selection, Stochastic programming

1. INTRODUCTION

Stochastic portfolio selection deals with the problem of how to find an optimal portfolio under uncertain demands. Originally proposed by Markowitz (1952), the mean-variance theory for the portfolio selection problem has served as a basis for modern financial theory development during the past decades. Efforts in revising, extending and improving this model have led to a large number research outputs in the forms of textbooks, monographs, and journal papers (Kato and Shillheim, 1985; Konno and Wijayanayake, 1999). They have helped investment agents to measure risk and develop economic strategies.

In general, investors always aim for the highest return on investment and the least risk. However, in most cases the complexity of calculation is too great. For practical purposes, it may be desirable to limit the complexity of the algorithm, and transform the risk function from covariance into mean absolute deviation (MAD). Konno and Yamazaki (1991) proposed a well-known MAD model to solve a large-scale portfolio optimization problem, and modified the MAD models according to the characteristics of stochastic portfolio selection problem.

In terms of stochastic returns, most of the related studies use the scenario optimization technique only. However, the technique cannot precisely reflect the situations of continuous distribution in real returns and other scenarios. This paper presents a modified algorithm for portfolio selection under uncertainty. The algorithm is a revision of an existing algorithm that combines both stochastic sampling procedure and systematic search. The core concept of the proposed algorithm is that the stochastic return of each sampling asset is based on random variables in the historical data (from 1995Q1 to 2003Q3). The resulting portfolio is optimal when convergence is achieved in the sampling procedure. The proposed algorithm is then evaluated by comparing the results from the Taiwan Stock Exchange (TAIEX) market in the forthcoming return (2003Q4 to 2004Q3).

The rest of this paper is organized as follows. In section 2, the modified MAD model and two-stage stochastic program are reviewed. The stochastic portfolio genetic algorithm (SPGA) and corresponding procedure for the nonlinear stochastic optimization portfolio problem are designed in section 3. In section 4, the algorithm is shown to converge quickly to a near optimal solution. Comparing to the results in TAIEX, the proposed algorithm outperforms the expected returns in the market. Section 5 is the conclusions and suggestions for further studies.

2. REVIEW OF MODIFIED MAD MODEL AND TWO-STAGE STOCHASTIC PROGRAM

In this section, the mean absolute deviation (MAD) model and two-stage stochastic program are reviewed.

* Corresponding author's email: weehm@cycu.edu.tw

Finally the conventional concept of two-stage stochastic program is explained.

2.1 Mean absolute deviation model

In the portfolio selection model proposed by Markowitz (1952), statistic mean and variance are used to represent the expected return and risk of the portfolio. Let x_i be the proportion invested in stock i (i = 1, …, k), where $\sum_{i=1}^{k} x_i = 1$. S= (x_1, x_2, …, x_k) is called a portfolio. The random variable P_i denotes the return on stock j. The expected return p_s and the risk σ_ς of portfolio ζ are given by mean and variance of P_j. respectively.

$$p_\zeta = E\left[\sum_{i=1}^{k}(P_i)x_i\right] = \sum_{i=1}^{k}(p_i)x_i \tag{1}$$

and

$$\sigma_\varsigma = \sqrt{E\left[\left(\sum_{i=1}^{k}(p_i)x_i - E\left[\sum_{i=1}^{k}(P_i)x_i\right]\right)^2\right]}$$
$$= \sqrt{\sum_{i=1}^{k}\sum_{j=1}^{k}\sigma_{ij}x_i x_j} \tag{2}$$

where σ_{ij} is covariance representing the expected risk of returns on stock i and j. The mean-variance model can then be formulated as follows:

$$Max \quad (1-\lambda)\left(\sum_{i=1}^{k} p_i x_i\right) - \lambda\left(\sum_{i=1}^{k}\sum_{j=1}^{k}\sigma_{ij}x_i x_j\right)$$
$$s.t. \quad \sum_{i=1}^{k} x_i = 1$$
$$x_i \geq 0 \quad \forall i$$

where $\lambda(0 \leq \lambda \leq 1)$ is the tradeoff factor between return and risk. For instance, when λ = 0, the investor prefers the highest returns without considering the risk of investment. Conversely, when λ = 1, the investor is very conscious of the risk of investment while paying no attention returns.

In order to simplify the complexity of calculation and information requirement of covariance, Konno and Yamazaki (1991) introduced the MAD model to transform the risk function from covariance into MAD. The term $\sum_{i=1}^{k}\sum_{j=1}^{k}\sigma_{ij}x_i x_j$ in the mean-variance model can be replaced by MAD. The risk of a scenario (sample set) is defined as

$$\frac{\sum_{i=1}^{k}(|\, p_i - \overline{p_i}\,|\, x_i)}{k}$$

where $\overline{p_i}$ is the expected return.

Therefore, the objective function of MAD model can be reformulated as

$$(1-\lambda)\left(\sum_{i=1}^{k} p_i x_i\right) - \frac{\lambda}{k}\sum_{i=1}^{k}(|\, p_i - \overline{p_i}\,|\, x_i)$$

However, transaction cost is another important factor for an investor to take into consideration in portfolio selection. Chang et al. (2002) considered the transaction cost of the investor. The modified multiple periods (N period) model for portfolio selection, called the modified MAD model, can be formulated as

$$Max \ (1-\lambda)\left(\sum_{i=1}^{k}\left(\sum_{n=1}^{N} p_{in} - r_i\right)x_i\right) - \frac{\lambda}{kN}\sum_{n=1}^{N}\sum_{i=1}^{k}(|\, p_{in} - \overline{p_{in}}\,|\, x_i)$$
$$s.t. \quad \sum_{i=1}^{k} x_i = 1$$
$$a_i \leq x_i \leq b_i \quad \forall i$$

where a_i and b_i are the lower bound and upper bound on the proportion of stock, respectively. r_i is the transaction cost of stock i.

2.2 Two-stage stochastic program

Solving approaches of portfolio selection optimization can be roughly divided into two categories: mathematical programming and soft-computing methods. Two-stage stochastic program as typical forms of the exact methods are used to model stochastic portfolio selection problems.

The two-stage stochastic program mentioned originally by Higle and Sen (1996) has been justified to represent a stochastic model with randomness in sampling effectively. The purpose of the decomposition procedure applied in two-stage program is to approximate the expected $\mathrm{MAD}(\eta_T)$ generated by T cutting planes where each cutting plane is obtained by a scenario realization. For the first T realizations, $\eta_T(x)$ is the point estimate of MAD under first T scenarios, the model is solved as

$$Max \quad (1-\lambda)\left(\sum_{i=1}^{k}\sum_{n=1}^{N}(p_{in} - r_i)x_i\right) - \lambda\eta_{\mathrm{T}}(x)$$
$$s.t. \quad \sum_{i=1}^{k} x_i = 1$$
$$a_i \leq x_i \leq b_i \quad \forall i$$

The corresponding "optimal solution" under first T scenarios (sample sets), denoted by x^T, will review whether T+1th cutting plan is necessary through sampling for the next estimated $\mathrm{MAD}\eta_{T+1}$. It can be shown that x^T, as $T \to \infty$, will converge to the optimal portfolio. The decomposition procedure applied in the two-stage program approximates the expected risk generated by cutting planes where each cutting plane is obtained by a return realization (Chang et al., 2002). The original decomposition procedure of the two-stage sampling-based stochastic program is shown in Figure 1.

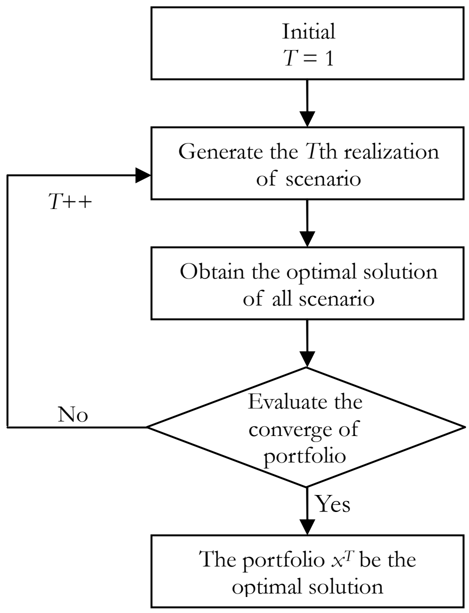


Figure 1. Basic steps of two-stage stochastic program.

To design a proper solution method for the addressed problem, one must consider the trade-off between solution efficiency and quality. Soft computing methods have rapidly emerged to solve the portfolio selection problem. As compared with simulated annealing and tabu search, genetic algorithm is the most popular one in solving the problem. Holland (1975) first proposed a simple genetic algorithm. Certain concerns exist regarding when a genetic algorithm (GA) methodology should be used, including the representation of a chromosome structure, initial population, population size, selection probabilities, genetic operators, and termination conditions. A fitness function is then used to screen for good chromosomes. The survey of a genetic algorithm can be found in numerous studies (e.g., Goldberg, 1989; Mitsuo and Runwei, 2000). With respect to tradeoffs between the risk and return, several scholars proposed heuristic methods for solving stochastic optimization portfolio problems using genetic algorithm (e.g., as Xia et al., 2000; Xia et al., 2001; Chang et al., 2000 and Ehrgott et al., 2004).

In summary, most of the above studies dealing with stochastic optimization portfolio problems are based on the probability of occurrence and scenario optimization. However, when potential scenarios are numerous or of continuous distributions, a sampling-based stochastic programming has higher precision than a scenarios-based one. Furthermore, the sampling-based approach can conduct plan dynamically by continuously updating information. Although soft-computing-based methods can solve the portfolio selection problem more efficiently, much computational effort is needed due to poor algorithm design.

3. STOCHASTIC PORTFOLIO GENETIC ALGORITHM (SPGA)

Genetic algorithm is a systematic search method for optimization problem based on the mechanics of natural selection and natural evolution. In this section, the proposed stochastic portfolio genetic algorithm (SPGA) extends conventional stochastic programming and uses the concept of two-stage stochastic programming to solve the modified MAD model. The SPGA for deriving an optimal portfolio is composed of three components: the structure of chromosomes, the sampling procedure for stochastic return and the operators of genetic algorithm.

3.1 Design of chromosome structure in SPGA

Chromosome structure is crucial to solving the optimal simultaneous resource portfolio planning problem when using GA. Each valid chromosome represents a unique solution to the problem given a set of returns. The chromosome of SPGA, as shown in Figure 2, is composed of all decision variables. Given a set of decision variables, each valid gene x_i represents a proportion of the investment in stock i.

x_1	x_2	…	x_k

Figure 2. Structure of chromosome.

In the initialization, a set of chromosomes with each component has a value within the domain of x_i ($a_i \le x_i \le b_i$). Note that the values of chromosome genes are normalized to satisfy the constraint of $\sum_{i=1}^{k} x_i = 1$.

According to the stochastic returns, the evaluation procedure of fitness is applied to calculate the objective values. A suitable portfolio to fit T scenarios (sample set) is found to determine the expected returns. The fitness function of SPGA may unsettle initially due to small scenario, but will converge to a steady state as the size of sample increases. In the evaluation and design of chromosome, we ensure that SPGA will handle sampling procedure and respond to the stochastic return.

3.2 Sampling procedure for stochastic return

In order to reduce the search effort of the two-stage program in real nonlinear stochastic portfolio problems, SPGA performs the search systematically. That is, in each sample, SPGA finds a suitable portfolio for all realized scenarios under an uncertain environment through "finite solutions" with perspicacity. The stochastic search procedure is explained as follows:

For each sample of the realized return set (1, 2, …, T), a total of $i \times N$ random variables of P_{in} (a realized return of stock i in period n) represent the realization of the returns scenario. The objective value of a portfolio $S = (x_1, x_2, \ldots, x_k)$ is evaluated when computing the expected returns of all scenarios (i.e., the fitness). Usually, the chromosome with higher fitness for all sampled scenarios has more chance to produce offspring by using roulette wheel selection. Following selection, crossover and mutation, the new population is ready for evaluation as new generation.

The sampling procedure will be triggered after Y generations. SPGA generates the next scenario, T+1, to realize the stochastic return of nth period in ith stocks. On the ground, the new objective function is derived to calculate the fitness of chromosome.

It is reasonable for the same portfolio to produce a different fitness value due to the increased scenario in different generation. Although the initially achieved portfolio is still not optimal (due to an instable objective function and insufficient sample), SPGA can guide the search along the direction to derive a suitable portfolio for all scenarios. Finally, the results will converge as the size of samples increase. Thus, the optimal portfolio can be obtained. The procedure of SPGA is summarized in Figure 3.

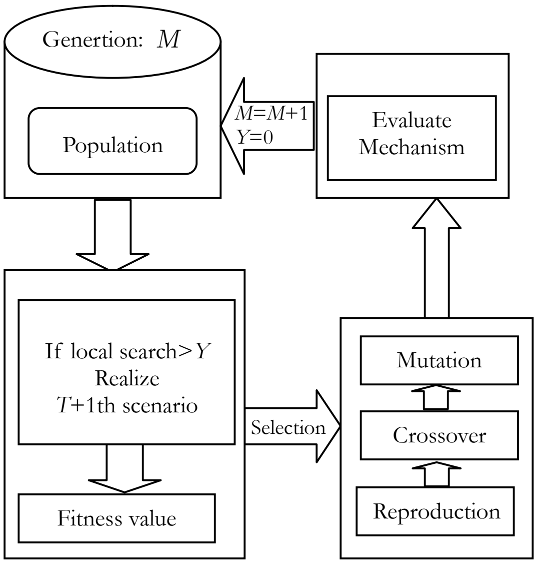


Figure 3. Overview of stochastic portfolio genetic algorithm.

The pseudo-code of SPGA is such that $F(M)$ and $S(T)$ in the SPGA are parents and offspring in the current generation M, respectively.

Pseudo-Code: Stochastic Portfolio Genetic Algorithm (SPGA)

Preliminary

$M \leftarrow 0$;
$T \leftarrow 1$; //the number of sample
Initialize $F(M)$ *and realization* T^{st} p_{in};
Evaluate $F(M)$;
While (not a termination condition) ***do***
If (sampling condition is True)
$T \leftarrow T+1$
realize T^{th} p_{in}
end If
Crossover: recombine $F(M)$ to yield $S'(M)$;
Mutation: alter the values of the genes of $F(M)$ to yield $S''(M)$, and $S(M) = S'(M) + S''(M)$;
Evaluate the best solution of $S(M)$ for all scenarios $(1\ldots T)$;
Repair: repair the infeasible solutions $S(M)$ (chromosomes) to be feasible solutions;
Select $F(M+1)$ from $F(M)$ and $S(M)$;
$M \leftarrow M+1$;
End
Termination

3.3 The operator of SPGA

The SPGA applies a uniform crossover (Chambers, 1995), and a uniform mutation to diversify the chromosomes (Haupt and Haupt, 1998; Gen and Cheng, 2000). Chromosomes are reproduced by roulette wheel method. The fitness is calculated using the objective function

$$(1-\lambda)\left(\sum_{i=1}^{k}\left(\sum_{n=1}^{N} p_{in} - r_i\right)x_i\right) - \frac{\lambda}{kN}\sum_{n=1}^{N}\sum_{i=1}^{k}(|\, p_{in} - \overline{p_{in}}\,|\, x_i)$$

of the modified MAD model, where p_{in} is a random variable obtained from the scenarios. SPGA is implemented by using C++ language.

4. RESULTS

In this section, computational results are demonstrated with 48 risky and one non-risky asset in the TAIEX. The historical data from 1995 Q1 to 2003 Q3 for these stocks are used to realize the returns. Furthermore, the performances of portfolios are then evaluated in 2003 Q4 to 2004 Q3. N is set to four to represent the four quarters under consideration.

The effect of different SPGA parameters on objective values is reported in Table 1. As can be seen, a low population, low crossover rate and low mutation rate design can obtain the highest objective value. Based on the results shown in Table 1, the best parameters are applied in the experiments of model parameters. Figures 4 to 8 show the SPGA performances with respect to different risk factors, λ = 0.00, 0.25, 0.50, 0.75 and 1.00 (note that a lower λ represents an investor's higher preference in return). A double P4 CPU 512 RAM personal computer is used; it took about 600 CPU-SEC for each run. Thus, the SPGA we developed outperforms the one in the market. To illustrate the convergence, consider the left of those figures which show a continuous converge toward the steady-state in a short time. On the right-hand side, performances of the portfolios are demonstrated from 2003 Q4 to 2004 Q3. Worthy to mention, the portfolio obtained under risk factor λ = 1.00 suggests that the investor should select the unique non-risky asset (displayed in Figure 8)

Table 1. Sensitivity experiments of SPGA parameters to the expected maximal objective value

SPGA Parameters		Crossover Rate (0.95)		Crossover Rate (0.75)	
		Mutation Rate (0.10)	Mutation Rate (0.03)	Mutation Rate (0.10)	Mutation Rate (0.03)
Population Size (30)	CPU Time (300 sec.)	0.461642	0.442809	0.423245	0.513318
	CPU Time (600 sec.)	0.431952	0.488149	0.431807	0.522145
Population Size (60)	CPU Time (300 sec.)	0.506313	0.463790	0.466505	0.525794
	CPU Time (600 sec.)	0.498577	0.482853	0.463684	0.514346

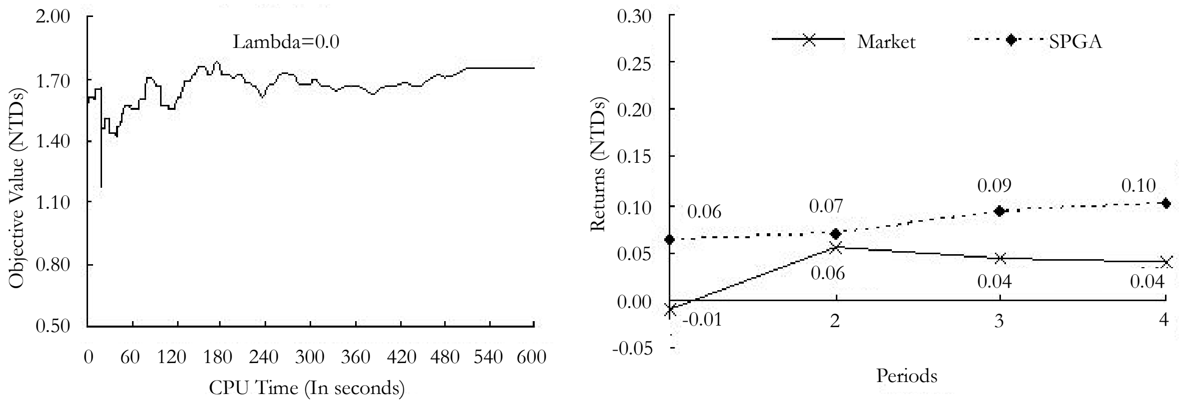


Figure 4. Convergence and performance of portfolios with $\lambda = 0.00$.

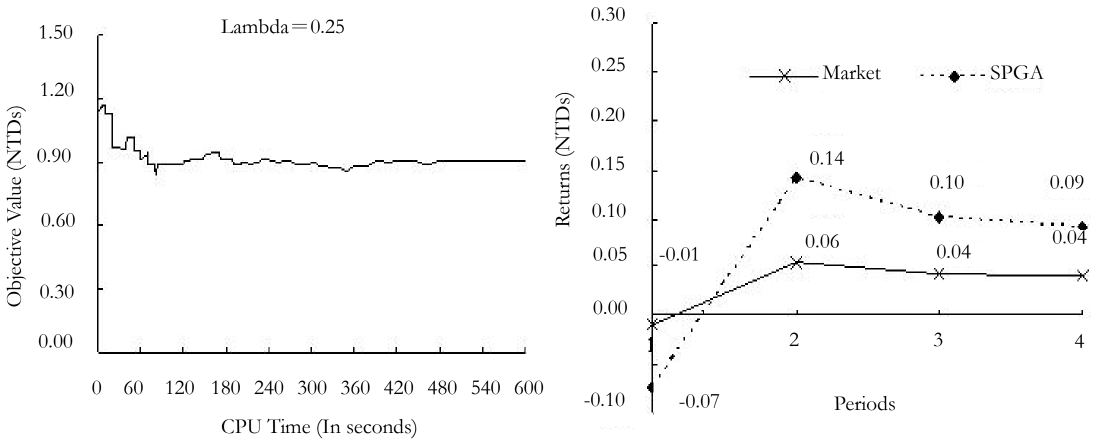


Figure 5. Convergence and performance of portfolios with $\lambda = 0.25$.

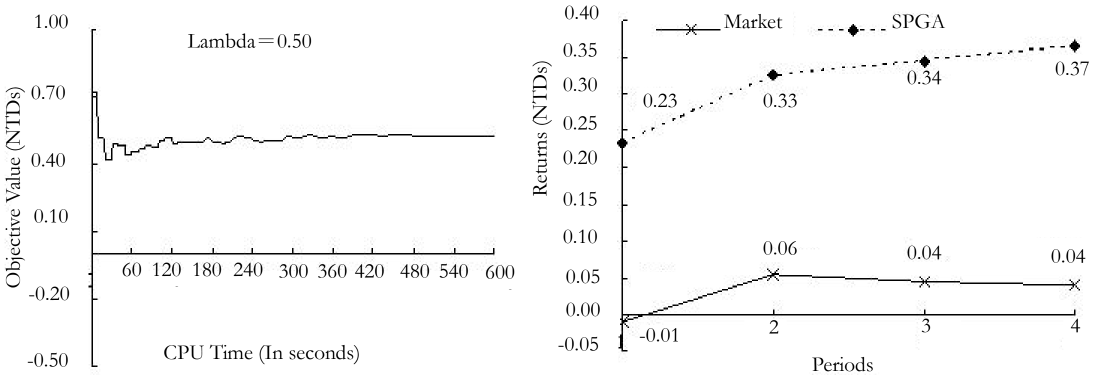
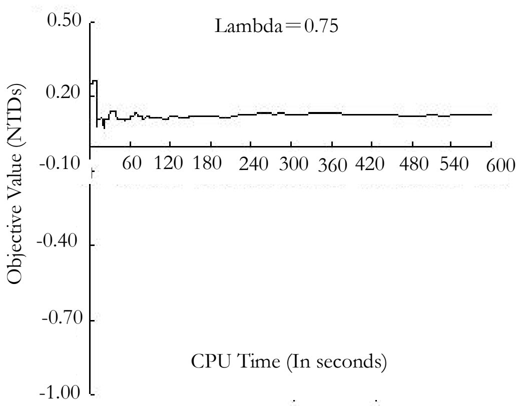


Figure 6. Convergence and performance of portfolios with $\lambda = 0.50$.



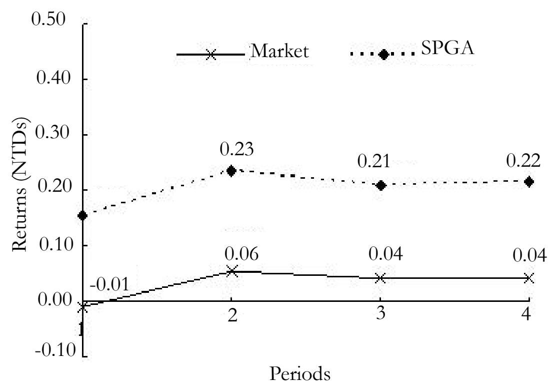
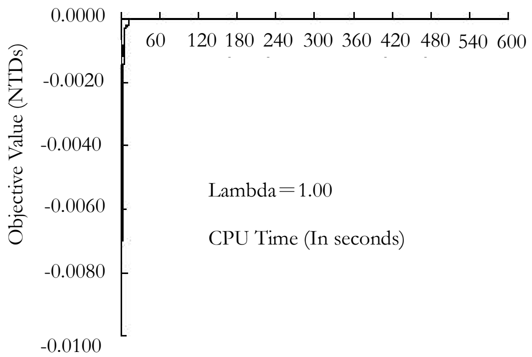


Figure 7. Convergence and performance of portfolios with $\lambda = 0.75$.



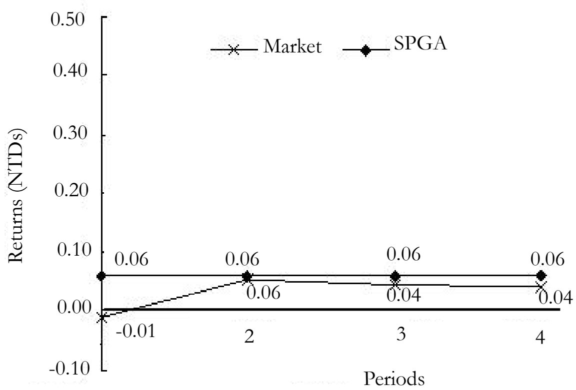


Figure 8. Convergence and performance of portfolios with $\lambda = 1.00$.

As λ increases, the profitability also changes (as shown in Figures 4 to 8); a poor performance results if the investment is considered to be without any risk (as shown in Figures 5 and 6). An analysis of two-stage program and SPGA show that a portfolio using two-stage program causes an exponential computational complexity, while in SPGA, only a polynomial computational complexity is needed. The complexity analysis implies that the efficiency of SPGA is better than that of the two-stage program in practice.

5. CONCLUSION

In this article, an effective Stochastic Portfolio Genetic Algorithm is proposed for solving a non-linear stochastic portfolio optimization problem. Based on two-stage program, an algorithm is developed to support the sampling procedure for portfolio selection. To facilitate stochastic evaluation procedures, SPGA derives the solution applying a sampling procedure. As compared with the two-stage program with many what-if analyses in a large-scale nonlinear stochastic optimization problem, this algorithm reduces the computational complexity. The convergence of SPGA is also demonstrated using a real numerical example. The proposed algorithm can guide the search to derive a suitable portfolio for all sampled scenarios. Finally, the performance of SPGA is investigated using historical data obtained from the Taiwan Stock Exchange. The results show that SPGA outperforms the procedures in the market. This algorithm is currently under investigation for its application to the other domains with similar uncertain environments.

REFERENCES

1. Chambers, L. (1995). Applications. In *Practical Handbook of Genetic Algorithms*, 1: 106-113.
2. Chang, K.H., Chen, H.J., and Liu, C.H. (2002). A stochastic programming model for portfolio selection. *Journal of the Chinese Institute of Industrial Engineers*, 19: 31-41.
3. Chang, T.J., Meade, N, Beasley, J.E., and Sharaiha, Y.M. (2000). Heuristic for cardinality constrained portfolio optimization. *Computer & Operation Research*, 27: 1271-1302.
4. Goldberg, D, E. (1989). Genetic Algorithm in Search, In *Optimization and Machine Learning*, Addison Wesley Publishing Company, Inc.

5. Ehrgott, M., Klamroth, K., and Schwehm, C. (2004). Decision aiding an MCDM approach to portfolio optimization. *European Journal of Operational Research*, 155: 752-770.
6. Gen, M. and Cheng, R. (2000). *Genetic Algorithms and Engineering Optimization.* John Wiley & Sons, Inc.
7. Haupt, R.L. and Haupt, S.E. (1998). *Practical Genetic Algorithms*. John Wiley & Sons, Canada.
8. Higle, J.L. and Sen, S. (1996). *Stochastic Decomposition.* Kluwer Academic Publishers, Dordrecht.
9. Holland, J.H. (1975). *Adaptation in Natural and Artificial Systems*, Detroit MI: University of Michigan Press.
10. Kato, K. and Schallheim, J.S. (1985). Seasonal and size anomalies in Japanese stock market. *Journal of Financial and Quantitative Analysis*, 20: 243-260.
11. Konno, H. and Wijayanayake, A. (1999). Mean-absolute deviation portfolio optimization model under transaction costs. *Journal of the Operations Research Society of Japan*, 42: 422-435.
12. Konno, H. and Yamazaki, H. (1991). Mean-absolute deviation portfolio optimization model and its application to Tokyo stock market. *Management Science*, 37: 519-531.
13. Markowitz, H. (1952). Portfolio Selection. *Journal of Finance*, 7: 77-91.
14. Mitsuo, G. and Runwei, C. (2000). *Genetic Algorithms and Engineering Optimization.* A Wiley-Interscience Publication.
15. Xia, Y., Liu, B., Wang, S., and Lai, K.K. (2000). A model for portfolio selection with order of expected returns. *Computers & Operations Research*, 27: 409-422.
16. Xia, Y., Wang, S., and Deng, X. (2001). Theory and methodology: a compromise solution to mutual funds portfolio selection with transaction costs. *European Journal of Operation Reasearch*, 134: 564-581.