Stock Investment with Stochastic Programming*

Hung-Hsin Chen, and Chang-Biau Yang†
Department of Computer Science and Engineering
National Sun Yat-sen University
Kaohsiung 80424, Taiwan

***Abstract*—In this paper, we propose the stochastic programming (SP) model with risk measure conditional value at risk (CVaR) for investing stocks in Taiwan stock market. In each period of investment, 200 scenarios are generated for solving SP, and the CVaR is utilized to manage the risk. The experiment interval starts from 2005/1/1 and ends on 2013/12/31, which has totally 2235 trading periods. The experimental results show that our method is able to earn positive return. We also perform superior predictive ability test to illustrate that our method can avoid the data snooping problem. Our method achieves the best annualized return 13.26%, which is higher than the buy-and-hold annualized return 12.19%.**

***Keywords*-Stock; Stochastic programming; Conditional value at risk; Superior predictive ability.**

I. Introduction

The portfolio problem is a very import issue in the investment. One investor always prefers to have their return as high as possible, but to have the risk as low as possible at the same time. However, an investment with high return usually accompanies with high risk. Markowitz [9] proposed the mean-variance model, which built the foundation of portfolio theory.

In the investment, it is difficult to evaluate the performance of one asset, and there are many uncertain parameters for modeling the problem. Therefore, to make appropriate decision to earn return requires mathematical analysis. Many researchers were devoted to the study of mathematical models for dealing with uncertainty for making better decisions. Ben-Tal *et al.* [2] proposed robust the multistage model for the portfolio problem. Li and Ng [8] solved the multistage mean-variance model and they provided the analytic solution. Bertsimas and Pachamanova [3] proposed the robust modeling of mutlistage portfolio with transaction cost. Topaloglou *et al.* [15] utilized CVaR for control the risk in the investment, and they utilized stochastic programming (SP) and CVaR [16] for solving the international asset allocation problem. However these results mainly discussed theoretical models, it seems to require more effectiveness in the investment.

Some researchers utilized evolutionary computing techniques for solving particular portfolio problems which are NP-complete or NP-hard. Chang *et al.* [4] discussed portfolio optimization problems with various risk measures by the genetic algorithm. Chen *et al.* [5] utilized the genetic network programming for solving portfolio optimization problems, Soleimani *et al.* [14] utilized the genetic algorithm for solving portfolio optimization problems with cardinality constraints. These techniques are mainly to deal with optimization problems which are difficult to solve efficiently. However, approximate solutions can be obtained by the evolutionary computing techniques.

In this paper, we consider the problem of the active stock portfolio management. In multistage investment, rebalancing capital decisions are required. The transaction fee has a large influence to the portfolio return since frequent transaction will cause large transaction fee and a potential increase in risk exposure. Stochastic programming (SP) is a method for modeling optimization problem involving uncertainty. Deterministic optimization problems are formulated with known parameters [13]. In the investment, investors may not always predict future return of stocks precisely, but it is possible to predict the future return in a confidence interval. Uncertainty in input parameters of SP is represented by discrete scenarios that describe the joint distribution of the random variables, and the expected cost and constraints of these scenarios are added into the model. After the scenarios are generated, the uncertainty problem is transformed into a deterministic problem, which can be solved efficiently.

Risk management aims to avoid portfolios that may suffer large loss in the investment. The basel committee on banking supervision recommended the value at risk (VaR) as the basis for modelling market risk. However, VaR is not a coherent risk measure, it is not satisfy the subadditivity property [1], which is not consistent with the principle that diversification is able to reducing risk . Conditional value at risk (CVaR) [10] is a related risk measure, which computes the expected loss below the VaR, and it is a coherent risk measure. Rockafellar and Uryasev [11] introduced the CVaR for continuous distribution, and they also gave definition of the CVaR for general distribution [12]. Topaloglou *et al.* [16] utilized mutlistage stochastic programming with CVaR for international portfolio.

In this paper, we confine our investment in the Taiwan stock market, and our goal is to earn return in the investment. We also employ superior predictive ability (SPA) test [6] for the wealth process to verify our method without data snooping problem. The experiment interval is from 2005/1/1 to 2013/12/31, which has totally 2235 trading periods. We employ the stochastic programming with CVaR to control the risk in the investment. The best annualized return of our method for portfolio size $n = 5$

*This research work was partially supported by the National Science Council of Taiwan under contract NSC 101-2221-E-110-020-MY3.
†Corresponding author: cbyang@cse.nsysu.edu.tw

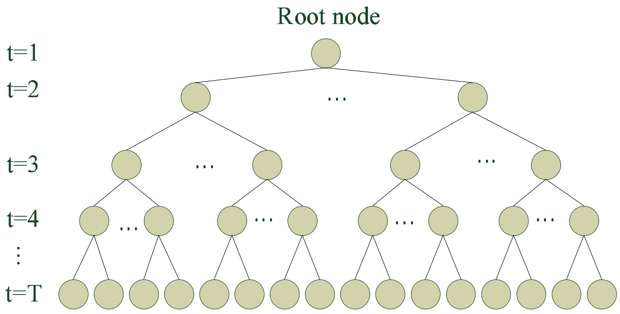

Figure 1. A scenario tree with T stages.

and $n = 10$ achieves 13.26% and 12.9%, respectively.

The rest of this paper is organized as follows. In Section II, we will present some background knowledge, including the stochastic programming, the conditional value at risk, and the superior predictive ability. In Section III, we will propose our method, which is able to earn return in the stock market. In Section IV, we will present the experimental results of our method. Finally, the conclusion of this paper will be given in Section V.

II. Preliminaries

In this section, we will give an introduction to the stochastic programming, the conditional value at risk, and the superior predictive ability, which serve as the background knowledge used in this paper.

A. Stochastic Programming for Investment

How to describe uncertainty is a critical issue for modeling a problem. The key uncertain variables in our method are the future returns of stocks. A major advantage of the *stochastic programming* is that it does not require any assumption for the random variables. It is able to deal with any discrete distribution which is expressed by a scenario tree as shown in Figure 1. One *scenario* is a path starting from the root to the leaf node, a stage is the moment when a decision is making, which is the level of a tree, and the period is the interval between two time points, which is the edge length between two neighboring nodes.

The basic idea of *two-stage tochastic programming* is that the decisions should be made by the data available at the time and they should not depend on future observation, which is called non-anticipativity property [13]. The classical two-stage stochastic programming is formulated as Equation 1.

$$\begin{array}{ll} \min_{\mathbf{x}\in\mathbf{X}} & \{g(\mathbf{x}) := f(\mathbf{x}) + E(Q(\mathbf{x},\xi))\}, \\ s.t. & \mathbf{A}\mathbf{x} \le \mathbf{b}, \mathbf{x} \ge \mathbf{0}, \\ \min_{\mathbf{y}\in\mathbf{Y}} & \mathbf{q}^{\mathbf{T}}\mathbf{y}, \\ s.t. & \mathbf{T}\mathbf{x} + \mathbf{W}\mathbf{y} \le \mathbf{h}, \mathbf{y} \ge \mathbf{0}, \end{array} \tag{1}$$

where $\mathbf{x}$ represents the vector of the first stage decision, $\mathbf{X}$ represents the domain of $\mathbf{x}$, which is defined by a finite number of constraints $\mathbf{A}\mathbf{x} \le \mathbf{b}$ and $\mathbf{x} \ge \mathbf{0}$, $E(\cdot)$ represents the expectation operator, $Q(\mathbf{x}, \xi)$ represents the optimal value of the second stage, $\xi = (\mathbf{q}, \mathbf{T}, \mathbf{W}, \mathbf{h})$ contains the data of the second stage, where $\mathbf{q}$ represents the vector of coefficients for $\mathbf{y}$, $\mathbf{T}$ represents the matrix of constraint values for $\mathbf{x}$, $\mathbf{W}$ represents the matrix of constraint values for $\mathbf{y}$, and $\mathbf{h}$ represents the vector of constraints values; $\mathbf{y}$ represents the vector of the second stage decision, and $\mathbf{Y}$ represents the domain of $\mathbf{y}$, which is defined by a finite number of constraints $\mathbf{T}\mathbf{x} + \mathbf{W}\mathbf{y} \le \mathbf{h}$ and $\mathbf{y} \ge \mathbf{0}$. At the first stage, we have to make a decision $\mathbf{x}$ before the realization of the uncertain data ξ with the cost $f(\mathbf{x})$ plus the expected cost of the optimal second stage problem. At the second stage, after the realization of ξ becomes available, we optimize the decision by solving the second stage problem, which is also called recourse action.

The standard approach for solving stochastic programming is to assume that the random vector ξ has a finite number of possible realization, which is called *scenarios*, denoted as $\xi_1, \cdots, \xi_k$, with respective to probabilities $p_1, \cdots, p_k$. Therefore, the expectation can be written as Equation 2.

$$E(Q(\mathbf{x}, \xi)) = \sum_{k=1}^{K} p_k Q(\mathbf{x}, \xi_k). \tag{2}$$

After generating the scenarios, the original nondeterministic equations of the stochastic programming can be transformed into deterministic ones, which are usually able to be solved efficiently.

B. Conditional Value at Risk

Value at risk (VaR) is a percentile based metric that has become financial standard for risk measurement. It is usually defined as the maximal allowable loss with a certain confidence level $\alpha \times 100\%$. The definition of VaR is given in Equation 3.

$$VaR(x, \alpha) = min\{u : P(x \le u) \ge 1 - \alpha\}, \tag{3}$$

where x represents the random variable to be measured, α represents the confidence level, and $P(\cdot)$ represents the probability measure.

However, VaR is not a coherent risk metric, since it does not satisfy the sub-additivity property [1], which is not consistent with the financial principle that diversification is of help to risk reduction. Conditional value at risk (CVaR) [10] is an improved risk measure, which computes the expected loss below the VaR, and it is a coherent risk measure. Rockafellar and Uryasev [11] introduced the CVaR for continuous distribution as shown in Equation 4, and they also gave definition of the CVaR for general distribution [12] as shown in Equation 5.

$$CVaR(x, \alpha) = E(x | x \le VaR(x, \alpha)), \tag{4}$$

$$\begin{aligned} CVaR(x, \alpha) = \ & (1 - \tfrac{\sum_{\{s\in\Omega | x_s \le z\}} p_s}{1-\alpha}) z + \\ & \tfrac{1}{1-\alpha} \textstyle\sum_{\{s\in\Omega | x_s \le z\}} p_s x_s, \end{aligned} \tag{5}$$

where $z = VaR(x, \alpha)$, x_s is the realized value of x of scenario s, $E(\cdot)$ represents the expectation operator, s represents a scenario, p_s is the probability that the scenario s is realized, and Ω represents the set of all scenarios.

According to the study of Topaloglou *et al.* [15], the CVaR can be reformulated as Equation 6.

$$\begin{aligned} y_s &= max(0, z - x_s), \\ z - \sum_{s\in\Omega} p_s y_s &= (1 - \frac{\sum_{\{s\in\Omega|x_s\leq z\}} p_s}{1-\alpha})z + \\ & \frac{1}{1-\alpha}\sum_{\{s\in\Omega|x_s\leq z\}} p_s x_s. \end{aligned} \tag{6}$$

Then the CVaR can be optimized by using linear programming (LP). The goal function of LP is the optimal CVaR measure at a confidence level $\alpha \times 100\%$ as shown Equation 7.

$$\begin{aligned} \max \quad & z - \frac{1}{1-\alpha}\sum_{s=1}^{S} p_s y_s \\ s.t. \quad & x_s \in X, z \in R, \\ & y_s \geq z - x_s, \\ & y_s \geq 0, \\ & s = 1, 2, \cdots, S \end{aligned} \tag{7}$$

C. Superior Predictive Ability Test

Some researchers applied statistical and machine learning models to return maximization in the investment by adjusting different parameters. And some models were found to have ability to earn positive returns. However, the approach to search profitable models has a serious issue, which is called the *data snooping* problem. Data snooping is the inappropriate use of data mining or statistics to induce misleading relationships in data. When such data are repeatedly tested, there exists the possibility that any satisfactory results obtained may simply be due to luck rather than to any merit inherent in the method yielding the results [18].

Hansen [6] proposed the *superior predictive ability* (SPA) test, which is a multiple test for testing if there is any model in a given set whose performance is better than the benchmark model without data snooping.

Given M models, let $d_{m,t}$ represent the performance measure of model m related to the benchmark model at period t . The null hypothesis is to determine whether a model outperforms the benchmark model, and it is shown in Equation 8. If no any model can beat the benchmark model, the mean of the relative performance $d_{m,t}$ should be less than or equal to zero.

$$H_0^m : \mu_m \leq 0, \text{where } 1 \leq m \leq M, \tag{8}$$

where the μ_m represents the expectation of $d_{m,t}$.

The SPA statistics is given as follows.

$$SPA_T = \max(\max_{1\leq m\leq M} \sqrt{T}\frac{\bar{d_m}}{\hat{\omega}_m}, 0), \tag{9}$$

where $\hat{\omega}_m$ represents the consistent estimator of the standard deviation of performance measure of the mth model. When the performance of many models is inferior to the benchmark model, the situation that the power of the test is decreased should be avoid. Hence in Equation 9, the statistics set the statistics to zero if the value $\sqrt{T}\frac{d_m}{\hat{\omega}_m}$ is negative.

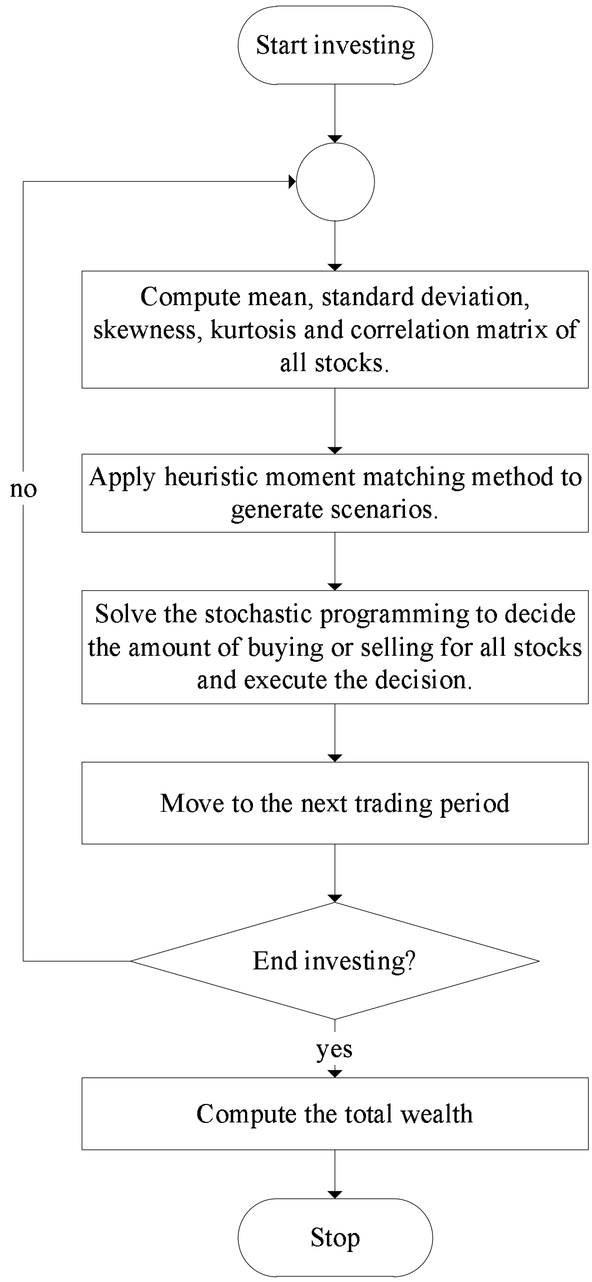

Figure 2. Flowchart of our method.

In SPA test, $\hat{\mu_m}$ is estimated as follows.

$$\hat{\mu_m} = \bar{d_m} 1(\sqrt{T}\bar{d_m} \leq -\hat{\omega}_m\sqrt{2\log\log T}), \tag{10}$$

where $\bar{d_m}$ represents the mean of the relative performance measure $d_{m,t}$ of model m, and $1\{\cdot\}$ is the indicator function.

III. Stock Investment with Stochastic Programming

In this section, we propose our method for investment. The flow chart of our method is shown in Figure 2.

The stochastic programming for modeling an investor's wealth with CVaR as risk management in the investment is given in Equation 11, which is inspired from the paper proposed by Bertsimas and Pachamanov [3], however they only considered maximize return and the robust formulation of the problem.

$$
\begin{array}{ll}
\max & z - \frac{1}{1-\alpha}\sum_{s=1}^{S} p_s y_s \\
s.t. & y_s \geq z - \sum_{m=0}^{M} w_t^m (1 + r_{t+1,s}^m), \\
& w_t^m = (1 + r_t^m) w_{t-1}^m + b_t^m - s_t^m, \\
& w_t^0 = (1 + r_t^0) w_{t-1}^0 - \\
& \quad \sum_{m=1}^{M} (1 + c_{b,t}^m) b_t^m + \\
& \quad \sum_{m=1}^{M} (1 - c_{s,t}^m) s_t^m, \\
& y_s \geq 0, \\
& w_t^m \geq 0, w_t^0 \geq 0, \\
& b_t^m \geq 0, s_t^m \geq 0, \\
& t = 1, \cdots, T, m = 1, \cdots, M, \\
& s = 1, \cdots, S,
\end{array}
\quad (11)
$$

where z represents the variable of CVaR defined in Equation 5, which is equal to VaR at the optimal solution, s represents a scenario for the uncertain joint returns of assets of period $t+1$, y_s represents the portfolio wealth loss in excess of VaR, p_s represents the probability of a scenario s, w_t^m represents the wealth invested in risky asset m in period t, w_t^0 represents the wealth invested in riskless asset in period t, b_t^m represents the amount of money that an investor buys risky asset m in period t, s_t^m represents the amount of money that an investor sells risky asset m in period t, r_t^m represents that the return of asset m in period t, $r_{t+1,s}^m$ represents that the return of asset m in period $t+1$ in the scenario s, $c_{b,t}^m$ represents that the buying transaction fee of asset m in period t, $c_{s,t}^m$ represents that the selling transaction fee of asset m in period t, M denotes the number of risky assets, T denotes the number of investment periods, and S denotes the number of scenarios.

The decision variables in Equation 11 are z, y_s, w_t^m, w_t^0, b_t^m and s_t^m. Other variables r_t^m, $c_{b,t}^m$, and $c_{s,t}^m$ are known at time t, and all the variables $r_{t+1,s}^m$ of the scenario s are also generated , hence they are parameters in the equation.

We view each decision in period t as a second stage stochastic programming, and the decision of each stage is independent. In each stage, we generate scenarios by the heuristic moment matching method proposed by [7]. This method requires five statistics for generating scenarios, which are the mean, the standard deviation, the skewness, the kurtosis, and the correlation matrix of return of all stocks in period t in the portfolio. For computing these statistics, we utilized past h periods returns (including current trading period) to estimate these statistics. After generating scenarios of period $t+1$ returns, we solve the Equation 11 to get optimal value of decision variables z, y_s, w_t^m, w_t^0, b_t^m and s_t^m. We apply the b_t^m and s_t^m to all the stocks in the portfolio and updating the wealth. Those steps are repeated until the last period T.

For solving the stochastic programming, we utilize the software python-based stochastic programming (PySP) [17], which is an open source software for solving the stochastic programming optimization problem.

Table I
THE STOCKS WITH THE TEN LARGEST MARKET VALUE AT 2013/12/31.

Rank	Symbol	Company name
1	2330	Taiwan Semiconductor Manufacturing Company
2	2317	Hon Hai Precision Industry Company
3	6505	Formosa Petrochemical Corporation
4	2412	Chunghwa Telecom Company
5	2454	Mediatek Incorporation
6	2882	Cathay Financial Holding Company
7	1303	Nan Ya Plastics Corporation
8	1301	Formosa Plastics Corporation
9	1326	Formosa Chemicals & Fibre Corporation
10	2881	Fubon Financial Holding Company

IV. EXPERIMENTAL RESULTS

The dataset is collected from the Taiwan Economic Journal database and it contains the adjusted return. We choose the ten stocks with the top ten largest market value on 2013/12/31 as our investment targets, and these stocks are listed in Table I.

In the experiments, we assume that we can always buy and sell the shares of these stocks at the after-hour trading in Taiwan stock market and the wealth which is invested in a stock is fractional. The experiment interval starts from 2005/1/1 and ends on 2013/12/31, which has totally 2235 trading periods in the interval. In each period t, we apply the heuristic moment matching method [7] to generating 200 scenarios of the joint returns of the portfolio of period $t+1$ by five statistics, the mean , the standard deviation, the skewness, the kurtosis, and the correlation matrix. These statistics are estimated by using the returns of each stock in the portfolio of past h periods, including the current period. The buying and selling transaction fee, $c_{b,t}^m$ and $c_{s,t}^m$ are set to 0.1425% and 0.4425% , respectively, and all the risk-free returns r_t^0 are set to 0, for all periods and for all stocks in Equation 11.

Our method contains three parameters (n, h, α), which represent the portfolio size, the historical interval for estimating statistics, and the confidence level of CVaR. In the experiments, we consider $n \in \{5, 10\}$, $h \in \{10, 20, \cdots, 80\}$ and $\alpha \in \{0.5, 0.55, \cdots, 0.95, 0.99\}$, hence there are total $2 \times 8 \times 11 = 176$ trials. For getting robust results, we run each trial three times.

The experiment results are shown in Tables II and III. In these tables, we observe that our method is able to earn cumulative return while the confidence level $0.5 \leq \alpha \leq 0.95$ and the historical interval $h \geq 20$. Besides, the longer h is, the higher return is. By observing the cumulative returns of the same historical interval h but with different confidence level α of CVaR , our method is consistent to the idea that the investment is a tradeoff between return and risk, that is, the higher risk is, the higher return of our method is. When $n=5$ and $h=10$, we cannot earn return, because that a short historical interval for estimating statistics may be biased and it may cause the model to make inappropriate decision in the investment. In addition, we cannot earn positive return when $\alpha = 0.99$, because that it is too conservative to buy promising stock while the stock prices are rising. Some

Table II
THE AVERAGE AND STANDARD DEVIATION OF CUMULATIVE RETURNS OF PORTFOLIO SIZE $n = 5$, WHERE ALL THE VALUES ARE SHOWN IN PERCENTAGE, AND THE VALUES INSIDE PARENTHESES ARE STANDARD DEVIATION.

α	h =10	20	30	40
0.50	4.26(3.16)	109.43(4.37)	126.15(5.54)	162.44(5.11)
0.55	3.04(4.92)	112.88(0.46)	130.61(2.60)	151.83(4.34)
0.60	-1.75(3.57)	125.12(5.73)	127.70(3.16)	146.01(3.48)
0.65	3.08(2.51)	125.26(2.37)	133.17(2.74)	144.32(3.36)
0.70	-2.66(1.42)	123.45(3.84)	137.83(4.87)	137.22(3.36)
0.75	-13.58(3.50)	120.04(4.27)	140.78(5.90)	138.12(6.69)
0.80	-18.76(1.51)	117.05(7.11)	139.01(2.16)	134.62(2.73)
0.85	-31.23(2.41)	90.92(5.04)	143.48(3.13)	146.98(3.29)
0.90	-37.51(1.68)	77.27(4.89)	129.48(4.85)	137.47(5.11)
0.95	-50.43(1.62)	27.51(3.01)	80.02(10.74)	101.47(2.77)
0.99	-65.45(3.77)	-46.90(5.62)	-32.13(2.91)	-41.58(7.23)
α,	h =50	60	70	80
0.50	174.25(2.85)	175.06(1.77)	189.01(1.82)	192.76(6.48)
0.55	172.03(7.64)	180.73(4.94)	188.96(1.56)	190.36(2.75)
0.60	163.95(4.19)	173.85(2.86)	187.08(6.82)	200.13(9.21)
0.65	161.23(1.85)	168.25(3.49)	181.13(3.47)	206.58(5.25)
0.70	161.81(4.03)	163.82(4.67)	182.57(3.09)	200.34(1.33)
0.75	143.71(2.40)	167.75(1.11)	178.31(0.41)	196.24(7.25)
0.80	155.89(4.63)	156.41(2.82)	172.42(8.27)	192.22(2.68)
0.85	149.49(2.32)	150.11(2.10)	151.85(6.76)	178.87(5.14)
0.90	153.68(6.60)	146.36(1.31)	152.68(8.97)	153.06(3.75)
0.95	115.72(1.46)	112.18(9.21)	111.31(11.45)	127.11(0.65)
0.99	-41.55(3.34)	-40.87(1.12)	-48.99(7.39)	-48.62(7.79)

Table III
THE AVERAGE AND STANDARD DEVIATION OF CUMULATIVE RETURNS OF PORTFOLIO SIZE $n = 10$, WHEN ALL THE VALUES ARE SHOWN IN PERCENTAGE, AND THE VALUES INSIDE PARENTHESES ARE STANDARD DEVIATION.

α	h =10	20	30	40
0.50	-	80.22(4.20)	94.70(4.42)	146.40(3.02)
0.55	-	70.08(3.17)	99.53(4.16)	137.77(6.62)
0.60	-	77.48(1.98)	92.85(1.14)	132.48(3.88)
0.65	-	73.80(3.29)	91.42(5.29)	135.41(2.68)
0.70	-	72.03(4.67)	89.96(6.79)	120.88(3.81)
0.75	-	70.39(7.33)	88.36(3.55)	112.85(3.79)
0.80	-	54.79(6.63)	90.31(3.67)	108.90(3.98)
0.85	-	47.21(4.70)	84.56(5.72)	101.77(1.77)
0.90	-	36.59(3.60)	63.55(3.76)	87.78(9.89)
0.95	-	7.52(0.59)	32.93(10.33)	48.97(10.66)
0.99	-	-57.73(4.04)	-58.57(0.86)	-65.59(5.11)
α	h =50	60	70	80
0.50	167.41(3.38)	175.23(5.26)	171.19(4.51)	179.67(6.19)
0.55	159.57(6.08)	167.77(2.56)	175.87(4.57)	184.56(6.59)
0.60	155.08(2.70)	165.99(6.49)	187.36(10.77)	195.40(5.37)
0.65	149.57(10.56)	162.07(5.08)	180.52(4.81)	188.11(1.71)
0.70	138.19(0.49)	154.00(4.56)	177.66(4.83)	193.44(8.50)
0.75	133.56(5.81)	138.17(2.85)	170.62(7.64)	179.70(4.93)
0.80	126.72(6.92)	133.44(5.81)	150.43(2.34)	164.45(13.07)
0.85	119.51(2.21)	127.24(7.09)	142.57(6.87)	152.93(2.08)
0.90	116.24(4.06)	114.12(3.66)	115.86(4.71)	139.09(3.61)
0.95	70.24(2.27)	75.96(14.30)	70.77(2.16)	87.33(8.55)
0.99	-64.03(1.72)	-69.35(3.09)	-71.68(3.39)	-72.20(4.20)

experimental results for the column $h = 10$ in Table III are not available since the historical interval is too short so that the correlation matrix cannot be performed by the Cholesky decomposition in the heuristic moment matching method for generating scenarios.

The superior predictive ability (SPA) test is applied to the parameters of our method to check if there exist some parameters which have the ability to earn portfolio return without data snooping. The number of samples for SPA is set to 5000 to get precise p-values. The performance measure for SPA test is described as follows.

Table IV
THE P-VALUE OF SUPERIOR PREDICTIVE ABILITY TEST OF PORTFOLIO SIZE$n = 5$. ***, **, AND * REPRESENTS SIGNIFICANCE AT 0.01 (1%), 0.05 (5%), AND 0.1 (10 %), RESPECTIVELY. THE STANDARD DEVIATION IS SHOWN INSIDE PARENTHESES.

α	h =10	20	30
0.50	0.3787 (0.0181)	**0.0308 (0.0011)	**0.0284 (0.0034)
0.55	0.3483 (0.0343)	**0.0273 (0.0006)	**0.0262 (0.0007)
0.60	0.4241 (0.0081)	**0.0208 (0.0035)	**0.0253 (0.0034)
0.65	0.3808 (0.0162)	**0.0208 (0.0000)	**0.0195 (0.0012)
0.70	0.3959 (0.0092)	**0.0197 (0.0008)	**0.0162 (0.0012)
0.75	0.5091 (0.0405)	**0.0242 (0.0041)	**0.0161 (0.0014)
0.80	0.5753 (0.0053)	**0.0313 (0.0052)	**0.0129 (0.0023)
0.85	0.6515 (0.0173)	**0.0431 (0.0096)	**0.0201 (0.0043)
0.90	0.7591 (0.0029)	*0.0801 (0.0098)	**0.0249 (0.0021)
0.95	0.8735 (0.0201)	0.2045 (0.0217)	*0.0680 (0.0117)
0.99	0.9360 (0.0138)	0.8651 (0.0171)	0.6993 (0.0089)
α	h =40	50	60
0.50	**0.0113 (0.0024)	***0.0027 (0.0008)	***0.0032 (0.0000)
0.55	**0.0151 (0.0007)	***0.0027 (0.0008)	***0.0032 (0.0000)
0.60	**0.0165 (0.0015)	***0.0027 (0.0008)	***0.0037 (0.0008)
0.65	**0.0156 (0.0006)	***0.0032 (0.0000)	***0.0037 (0.0008)
0.70	**0.0189 (0.0019)	***0.0027 (0.0008)	***0.0032 (0.0000)
0.75	**0.0192 (0.0026)	***0.0037 (0.0008)	***0.0037 (0.0008)
0.80	**0.0173 (0.0015)	***0.0037 (0.0008)	***0.0053 (0.0015)
0.85	**0.0125 (0.0006)	***0.0064 (0.0013)	***0.0064 (0.0013)
0.90	**0.0167 (0.0019)	***0.0048 (0.0000)	***0.0080 (0.0013)
0.95	**0.0451 (0.0074)	**0.0149 (0.0030)	**0.0191 (0.0077)
0.99	0.7689 (0.0339)	0.7743 (0.0189)	0.7768 (0.0184)
α	h = 70	80	
0.50	***0.0032 (0.0000)	***0.0032 (0.0000)	
0.55	***0.0032 (0.0000)	***0.0032 (0.0000)	
0.60	***0.0032 (0.0000)	***0.0032 (0.0000)	
0.65	***0.0032 (0.0000)	***0.0016 (0.0013)	
0.70	***0.0037 (0.0015)	***0.0016 (0.0000)	
0.75	***0.0021 (0.0008)	***0.0011 (0.0008)	
0.80	***0.0037 (0.0008)	***0.0016 (0.0000)	
0.85	***0.0043 (0.0015)	***0.0016 (0.0000)	
0.90	***0.0059 (0.0027)	***0.0037 (0.0008)	
0.95	**0.0208 (0.0039)	**0.0112 (0.0013)	
0.99	0.9013 (0.0351)	0.8730 (0.0764)	

$$\bar{d_m} = \frac{1}{T}\sum_{t=1}^{T} R_{m,t}, \tag{12}$$

where $\bar{d}_m$ represents the average performance measure of the mth model in the test, T represents the number of trading periods, and $R_{m,t}$ represents the portfolio return of the mth model in period t.

The results of SPA test are shown in Tables IV and V. In the two tables, we can see that almost all the results with positive returns are significant in statistics, which means that our method can indeed earn positive returns without data snooping.

The wealth process of the highest return of the best parameters for portfolio size $n = 5$ and $n = 10$ and the buy-and-hold rule is shown in Figure 3. In this figure, we can observe that the process of our method is more smooth than the buy-and-hold rule, since our method considers the risk management, which decrease the volatility of wealth process in the investment.

Compared with the buy-and-hold strategy with capital allocation to each stock equally, the cumulative return R_{cum} and the annualized return AR_{cum} are shown in Table VI. As we can see, our method get better annualized return (13.26%) than the buy-and-hold strategy (12.19%).

Table V
THE P-VALUE OF SUPERIOR PREDICTIVE ABILITY TEST OF PORTFOLIO SIZE $n = 10$. ***, **, AND * REPRESENTS SIGNIFICANCE AT 0.01 (1%), 0.05 (5%), AND 0.1 (10 %), RESPECTIVELY. THE STANDARD DEVIATION IS SHOWN INSIDE PARENTHESES.

α	h =10	20	30
0.50	-	*0.0674 (0.0065)	*0.0576 (0.0025)
0.55	-	*0.0816 (0.0085)	*0.0538 (0.0050)
0.60	-	*0.0736 (0.0009)	*0.0627 (0.0069)
0.65	-	*0.0795 (0.0025)	*0.0589 (0.0064)
0.70	-	*0.0837 (0.0067)	*0.0580 (0.0045)
0.75	-	*0.0835 (0.0079)	*0.0623 (0.0057)
0.80	-	0.1123 (0.0105)	*0.0553 (0.0028)
0.85	-	0.1220 (0.0177)	*0.0653 (0.0068)
0.90	-	0.1684 (0.0199)	0.1032 (0.0191)
0.95	-	0.3331 (0.0099)	0.2019 (0.0513)
0.99	-	0.8917 (0.0301)	0.9190 (0.0092)
α	h =40	50	60
0.50	**0.0176 (0.0000)	***0.0032 (0.0000)	***0.0032 (0.0000)
0.55	**0.0261 (0.0038)	***0.0027 (0.0008)	***0.0032 (0.0000)
0.60	**0.0237 (0.0026)	***0.0027 (0.0008)	***0.0032 (0.0000)
0.65	**0.0239 (0.0010)	***0.0032 (0.0013)	***0.0032 (0.0000)
0.70	**0.0287 (0.0028)	***0.0059 (0.0008)	***0.0037 (0.0008)
0.75	**0.0357 (0.0048)	***0.0069 (0.0008)	***0.0059 (0.0015)
0.80	**0.0378 (0.0029)	***0.0096 (0.0023)	***0.0080 (0.0013)
0.85	**0.0431 (0.0020)	***0.0091 (0.0020)	***0.0085 (0.0030)
0.90	*0.0612 (0.0104)	**0.0101 (0.0007)	**0.0165 (0.0042)
0.95	0.1303 (0.0249)	*0.0697 (0.0078)	*0.0626 (0.0304)
0.99	0.9707 (0.0095)	0.9647 (0.0055)	0.9819 (0.0099)
α	h =70	80	
0.50	***0.0032 (0.0000)	***0.0032 (0.0000)	
0.55	***0.0032 (0.0000)	***0.0032 (0.0000)	
0.60	***0.0027 (0.0008)	***0.0032 (0.0000)	
0.65	***0.0032 (0.0000)	***0.0032 (0.0000)	
0.70	***0.0032 (0.0000)	***0.0032 (0.0000)	
0.75	***0.0032 (0.0000)	***0.0032 (0.0000)	
0.80	***0.0043 (0.0008)	***0.0032 (0.0000)	
0.85	***0.0048 (0.0013)	***0.0037 (0.0008)	
0.90	**0.0181 (0.0042)	***0.0048 (0.0000)	
0.95	*0.0557 (0.0175)	**0.0344 (0.0045)	
0.99	0.9862 (0.0028)	0.9871 (0.0108)	

Table VI
THE CUMULATIVE RETURN COMPARISON OF OUR METHOD, WHERE ALL THE VALUES ARE SHOWN IN PERCENTAGE.

method	n	R_{cum}	AR_{cum}
Buy & hold	5	181.56	12.19
Buy & hold	10	160.84	11.24
Our method ($\alpha = 0.65, h = 80$)	5	206.58	13.26
Our method ($\alpha = 0.60, h = 80$)	10	195.40	12.79

V. CONCLUSION

In this paper, we propose one effective model for investment in stocks. Our method consists of two phases: (1) generating scenarios of joint returns of period $t + 1$ by five statistics, the mean, the standard deviation, the skewness, the kurtosis, and the correlation matrix in period t; (2) solving Equation 11 to get optimal values of b_t^m and s_t^m, and apply the decisions to our portfolio; By the results of the superior predictive ability test shown in Tables IV and V, our method can effectively avoid the data snooping problem.

In the future work, it is worth finding a more precise way to generate scenarios, because the heuristic moment matching method only considers the static statistical properties, but the time series has been proven the existence of autocorrelation properties. In addition, our method can involve more constraints for approaching the real investment.

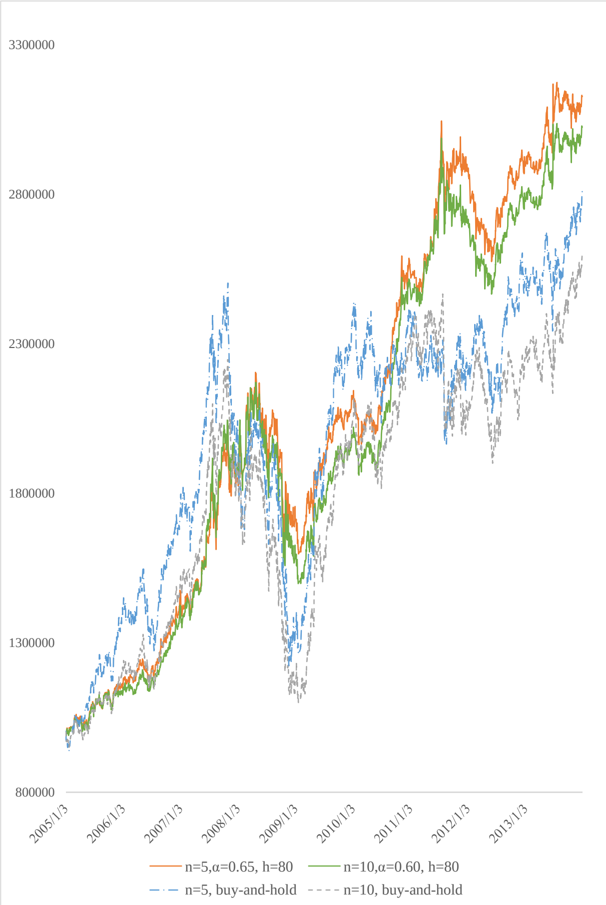

Figure 3. Wealth process of portfolio size $n = 5$ and $n = 10$.

REFERENCES

[1] P. Artzner, F. Delbaen, J. Eber, and D. Heath, “Coherent measures of risk,” *Mathematical finance*, Vol. 9, pp. 203–228, 1999.

[2] A. Ben-Tal, T. Margalit, and A. Nemirovski, *Robust modeling of multi-stage portfolio problems*. Springer, 2000.

[3] D. Bertsimas and D. Pachamanova, “Robust multiperiod portfolio management in the presence of transaction costs,” *Computers & Operations Research*, Vol. 35, pp. 3–17, 2008.

[4] T.-J. Chang, S.-C. Yang, and K.-J. Chang, “Portfolio optimization problems in different risk measures using genetic algorithm,” *Expert Systems with Applications*, Vol. 36, pp. 10529–10537, 2009.

[5] Y. Chen, E. Ohkawa, S. Mabu, K. Shimada, and K. Hirasawa, “A portfolio optimization model using genetic network programming with control nodes,” *Expert Systems with Application*, Vol. 36, pp. 10735–10745, 2009.

[6] P. R. Hansen, “A test for superior predictive ability,” *Journal of Business and Economic Statistics*, Vol. 23, pp. 365–380, 2005.

[7] K. Hyland, M. Kaut, and S. W. Wallace, “A heuristic for moment-matching scenario generation,” *Computational Optimization and Applications*, Vol. 24, pp. 169–185, 2003.

[8] D. Li and W.-L. Ng, “Optimal dynamic portfolio selection: Multiperiod mean-variance formulation,” *Mathematical Finance*, Vol. 10, pp. 387–406, 2000.

[9] H. M. Markowitz, “Portfolio selection,” *Journal of Finance*, Vol. 7, pp. 77–91, 1952.

[10] G. Pflug, *Some remarks on the value-at-risk and the conditional value-at-risk In: Uryasev, S. (Ed.) Probablilstic Constrained Optimization: Methodology and Applications*. Kluwer Academic Publishers, 2000.

[11] R. T. Rockafellar and S. Uryasev, “Optimization of conditional value at risk,” *Journal of Risk*, Vol. 2, pp. 21–42, 2000.

[12] R. T. Rockafellar and S. Uryasev, “Conditional value-at-risk for general loss distribution,” *Journal of Banking & Finance*, Vol. 26, pp. 1443–1471, 2002.

[13] A. Shapiro and A. Philpott, “A tutorial on stochastic programming.” Manuscript. Available at http://www2. isye. gatech. edu/ ashapiro/publications. html, 2007.

[14] H. Soleimani, H. R. Golmakani, and M. H. Salimi, “Markowitz-based portfolio selection with minimum transaction lots, cardinality constraints and regarding sector capitalization using genetic algorithm,” *Expert Systems with Application*, Vol. 36, pp. 5058–5063, 2009.

[15] N. Topaloglou, H. Vladimirou, and S. A. Zenios, “CVaR models with selective hedging for international asset allocation,” *Journal of Banking & Finance*, Vol. 26, pp. 1535–1561, 2002.

[16] N. Topaloglou, H. Vladimirou, and S. A. Zenios, “A dynamic stochastic programming model for international portfolio management,” *European Journal of Operational Research*, Vol. 185, pp. 1501–1524, 2008.

[17] J.-P. Watson, D. L. Woodruff, and W. E. Hart, “PySP: modeling and solving stochastic programs in python,” *Mathematical Programming Computation*, Vol. 4, pp. 109–149, 2012.

[18] H. White, “A reality check for data snooping,” *Econometrica*, Vol. 68, pp. 1097–1126, 2000.