Asset Allocation Using Markov Decision Process

Tahseen Jilani[1*], Faheem Zaidi[2], Syed Khalid Jamal[1], Arfa Maqsood[3], Suboohi Safdar[3]

[1]Department of Computer Science, University of Karachi, Karachi.
[2]School of Mathematics, Federal Urdu University, Karachi.
[3]Department of Statistics, University of Karachi, Karachi.

Abstract
Markov Decision Processes enable researchers to analyze the dynamics of a stochastic process whose transition mechanism is controlled over time. In this study, the solution to a portfolio allocation problem using a Markov Decision Process (MDP) is formulated. The main subject of interest would be to find an optimal policy that minimizes the associated cost. The two challenges faced were, uncertainty about the price of assets which follow a probabilistic model and a large state/action space that creates it difficult to apply orthodox techniques to solve. Stocks data is obtained from Karachi Stock Exchange - 100 index (KSE-100) that have daily cumulative returns about 49.98% in 2012. It is found that all portfolios allocations achieve by Markov Decision Process, have better daily cumulative returns as compared to benchmark KSE-100 index. Therefore, it is concluded that Markov decision process is better approach to calculate assets' allocation in designing stocks portfolios.
Keywords:
Karachi Stock Exchange 100 Index, Markov Decision Process, Wealth fraction.

1. Introduction

Markov decision processes (MDPs) are results partly random and somewhat under the guidance of a decision maker. They are used in situations where modeling decision-making is provided through a mathematical framework. MDPs use dynamic programming and reinforcement learning to solve a wide range of optimization and real world problems.
Bellman [1] was the first who introduced the term Markov Decision Process (MDP). Markov Decision process, with reference to stochastic scenario, was first studied by Shapley [8]. Afterwards, mathematical accurate action of this hypothesis came in view by Dubins & Savage [3], Shiryaev [9], Hinderer [6] and others. Dubins & Savage [3] wrote a book which covered the gambling model; however the fundamental ideas were basically the same. The model which was introduced by Blackwell [2] is used till date. Accurate action of discount problem was also firstly introduced by him for general state space. Heyman &Sobel [4], [5] and Stokey & Lucas [10] published books on application of MDP in the field of economics. Ingersoll [7] examined inter-temporal portfolio choice problem.

The objective of this study is two fold. From a theoretical perspective, this research proposes a novel Markov decision processes method to incorporate uncertainty about the kind of return distribution to get an optimal blend among a risky and a risk-less asset. From an application perspective, this study characterizes the exante asset allocation decisions of investors who factor in distribution ambiguity into their model of portfolio.
In section-2, a review of MDPs is presented, in Section-3, the porposed approach along with experimental results are presented.

2. Data & Methodology:

In order to analyze the model, we consider daily adjusted stock prices from Bloomberg and interest rates are taken from official website of State Bank of Pakistan (*www.sbp.org.pk*) for the period Jan-Dec 2012.
S is the set of all possible states in the model. The number of states in the model depends on how many risky assets (stocks) are present in the model.
Number of states = 2^X
Where, X is the number of risky assets in the model.
α_{k+1} be the decision take at the end of the day k, which will tell what fraction of the wealth will invest on the particular stock for the next day k + 1. These are many possible actions but we have to select such an action for each state which maximizes the wealth. On the basis of that fraction of the wealth $\alpha_{k+1} = (\alpha_{1,k+1}, \alpha_{2,k+1}, ..., \alpha_{M,k+1})$, we have to take the decision for each stock either we buy or sell them for the next day. By using that we can also compute the wealth for the next day.

$$N_{i,k+1} = RoundD\left(\frac{\alpha_{i,k+1} W_k}{S_{i,k}}\right)$$

At the end of that particular day, wealth is computed as

$$W_{k+1} = (W_k - TC_k)\left(1 + r_f + \sum_{i=1}^{M} \alpha_{i,k+1}(r_{i,k+1} - r_f)\right)$$

Manuscript received March 5, 2019
Manuscript revised March 20, 2019

Round D(A), represents round toward negative infinity, rounds the elements of A to the nearest integers less than or equal to A.

The objective of this research is to find such a policy which maximizes the expected reward of performing that action in the state. A policy is optimal, if it makes nonetheless as much total rewards all other possible policies. The value function is defined as

$$V_T = E[\sum_{k=0}^{T-1} R_k(s,\alpha) + U(W_0)]$$

Where $R_k(s,\alpha)$ is the cost function or the belief state which defined a

$$R_k(s,\alpha) = \sum_{i \in S} C_k(s,\alpha_k,i)\Pr(i,s,\alpha_k)$$

With $C_k(s,\alpha_k,i)$ is the cost if the next state is *"i"* and probability distribution over s is $\Pr(i,s,\alpha_k)$. α_k belongs to action space belong to action A in state S at decision epoch k, the decision maker will receive the cost

$$C_k(s,\alpha) = U(W_{k+1}) - U(W_k)$$

Where U(x) represents the utility function

After applying each action, we have to select such action for the state which gives maximum value to V_T. In short, we have to achieve $\max_{\alpha_1,\alpha_2,\alpha_3...,\alpha_n} V_T$. To get the maximum value, value iterations algorithms is used

3. Application with Evaluation

We take into consideration two different cases and evaluate their results (see TABLE 1). In the first scenario, two risky assets are taken into account for the period Jan-Dec 2012, in which we consider three portfolios.

Table 1(a): Portfolio with listed companies. Names of the companies are available at www.kse100.net

Portfolio name	Companies
Portfolio 2a	FCCL SNBL
Portfolio 2b	NETSOL POL
Portfolio 2c	APL HBL
Portfolio 3a	FCCL LPCL LUCK
Portfolio 3b	SNBL MEBL AGTL
Portfolio 3c	BAHL MEBL OGDC
Portfolio 3d	AHCL APL BAHL

Table 1(b): Company names with abbreviations

Company Name	Symbol
AL-Ghazi Tractors Ltd.	AGTL
Arif Habib Corporation Limited	AHCL
Attock Petroleum Ltd	APL
Bank AL-Habib Limited	BAHL
Fauji Cement Company Ltd.	FCCL
Habib Bank Ltd	HBL
Jahangir Siddiqui Co.Ltd.	JSCL
Lafarge Pakistan Cement Ltd	LPCL
Lucky Cement Limited	LUCK
Meezan Bank Ltd.	MEBL
Netsol Technologie	NETSOL
Oil & Gas Dev.Co	OGDC
Pakistan Oilfields Ltd.	POL
Sonari Bank Limited	SNBL

For simplicity we name these portfolios as 2*a*, 2*b* and 2*c*. For portfolio 2*a*, it was noticed that it had higher cumulative returns as compare to benchmark *KSE*-100 index during this period. Portfolio 2*b* resulted in one higher asset and the other lower in comparison with benchmark. While in 2*c*, we found that both risky assets were lower than *KSE* - 100 index. When working on the second scenario, with three risky assets, we make four portfolios and label them as 3*a*, 3*b*, 3*c* and 3*d*. In first portfolio 3*a*, the cumulative return of all three risky assets is higher than benchmark. Portfolio 3*b* contains the assets in which two of the stocks have higher cumulative return than *KSE* - 100, while one stock has lower. In portfolio 3*c*, it was considered that two stocks have less cumulative return but one has higher. The last portfolio 3*d* consists of stocks which have lower cumulative return compare to benchmark.

We assign the state for each day by observing the closing price of the risky asset on that particular day. If we consider the case of two risky assets then we have 4 states (see TABLE 2).

Table 2: State allocation for the case of two risky assets

S1	S2	S3	S4
DD	DU	UD	UU

If prices of both risky assets fall on any particular day, then state will be *S*1, *S*2 state represents that the first asset has fallen down while second has risen; *S*3 represents first asset moves upwards, while seconds goes down. *S*4 represents both assets move upwards.

In the case of three risky assets we have 8 states (see table 3).

Table 3: State allocation for the case of three risky assets

S1	S2	S3	S4	S5	S6	S7	S8
DDD	DDU	DUD	DUU	UDD	UDU	UUD	UUU

If the prices of all of the three risky assets fall on any particular day, then its state will be *S*1. *S*2 state represents that first and second assets have fallen down while third has raised, *S*3 represents that first and third assets move upwards while second goes down. *S*4 represents that the first asset has fallen, while both the second and third assets move upwards and so on.

If we have *n* risky asset then dimension of each action is *n*+1 for this model. By using these weights we can assign how many share we have to buy or sell of each stocks for the next day.

We define log(*X*) as our utility function, so it means reward is log return of the wealth. For a day *k*, if the state is Si and we select action αj , then the cost function is represent as;

$$C_k(S_i,\alpha_j)=\log\left(\frac{W_{k+1}}{W_k}\right)$$

4. Discussion of Results

An emerging objective function (policy) has been defined which maximizes the expected reward of performing that action in the state. This policy is optimal if it makes nonetheless as much total reward as all other possible policies. After applying value iteration we get the action for each state which maximizes the cumulative return of the particular portfolio. In other words, it shows with which proportion of wealth will be divided for next day, if currently it is on any particular state (see TABLE 4 to 10). The proportion of wealth allocated to each stock in all portfolios which were calculated by using *MDP* has shown outstanding results. It evaluates that the all portfolios have better daily cumulative returns as compared to the benchmark, which is approximately 48.98% for the period Jan-Dec 2012. The cumulative return of the portfolios 2*a*, 2*b* and 2*c* redundant the cumulative return of benchmark by 162.01%, 92.3% and 3.84% respectively. Whereas portfolios 3*a*, 3*b*, 3*c* and 3*d* dominate benchmark by 323.34%, 133.54%, 83.38% and 29.8% respectively. Histograms of returns distribution of all portfolios show that they are positively skewed. It is also analyzed that returns distribution exhibits Leptokurtic properties i.e. sharper than normal distribution and bears fat tails.

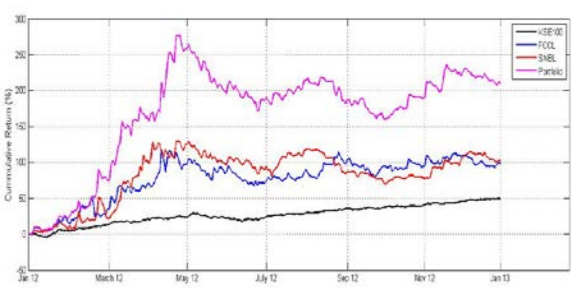

Fig. 1 Portfolio 2a

Table 4: Wealth allocation for each state for Portfolio 2a. With Cumulative return is 162.01%.

	FCCL	SNBL	Saving Account
S1	0	1	0
S2	1	0	0
S3	1	0	0
S4	0	1	0

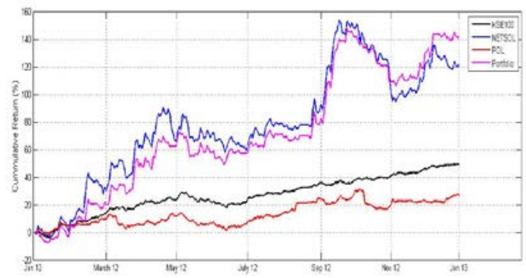

Fig. 2 :Portfolio 2b

Table 5: Wealth allocation for each state for Portfolio 2b, With Cumulative return 92.3%.

	NETSOL	POL	Saving Account
S1	1	0	0
S2	0	0	1
S3	1	0	0
S4	1	0	0

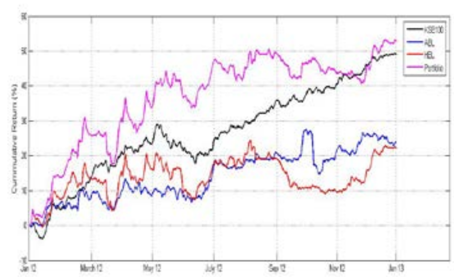

Fig 3:Portfolio 2c

Table 6 Wealth allocation for each state for Portfolio 2c. With Cumulative return 3.84%.

	APL	HBL	Saving Acc
S1	1	0	0
S2	0	1	0
S3	0	1	0
S4	0	1	0

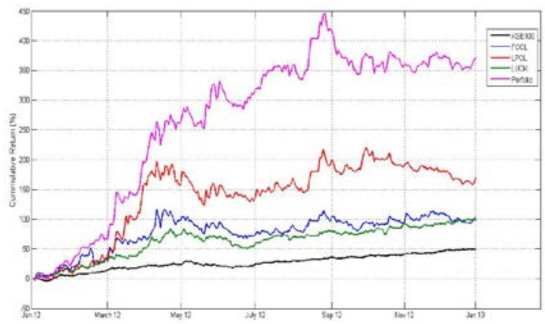

Fig. 4 :Portfolio 3a

Table 7: Wealth allocation for each state for Portfolio 3a, with Cumulative return 323.34%.

	FCCL	LPCL	LUCK	Saving Account
S1	1	0	0	0
S2	0	1	0	0
S3	0	0	1	0
S4	0	0	0	1
S5	0.1	0.9	0	0
S6	0	1	0	0
S7	0	1	0	0
S8	0	0	1	0

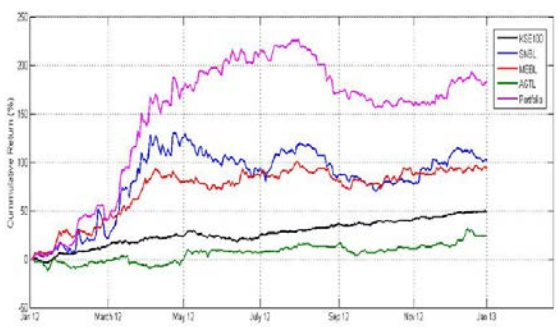

Fig. 5 :Portfolio 3b

Table 8: Wealth allocation for each state for Portfolio 3b, with Cumulative return 133.54%

	JSCL	MEBL	AGTL	Saving Account
S1	1	0	0	0
S2	0	1	0	0
S3	1	0	0	0
S4	0	0	0	1
S5	0.1	0	0	0.9
S6	0	1	0	0
S7	1	0	0	0
S8	0.7	0.3	0	0

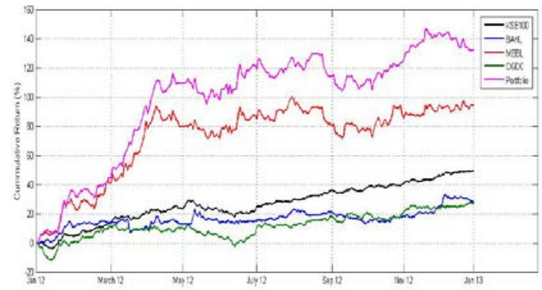

Fig. 6 :Portfolio 3c

Table 9: Wealth allocation for each state for Portfolio 3c, with Cumulative return 83.38%.

	BAHL	MEBL	OGDC	Saving Account
S1	0	1	0	0
S2	0	0	1	0
S3	0	1	0	0
S4	0	0	1	0
S5	0	1	0	0
S6	0	1	0	0
S7	1	0	0	0
S8	0	1	0	0

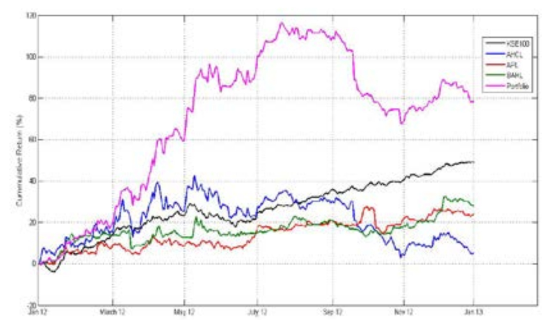

Fig 7:Portfolio 3d

Table 10: Wealth allocation for each state for Portfolio 3d, with Cumulative return 29.8%

	AHCL	APL	BAHL	Saving Account
S1	1	0	0	0
S2	0	0	0	1
S3	1	0	0	0
S4	1	0	0	0
S5	0	1	0	0
S6	0	0	1	0
S7	0	1	0	0
S8	1	0	0	0

5. Conclusion:

It has been found that, all portfolios obtained by Markov Decision Process (*MDP*), results in better daily cumulative returns as compared to benchmark (*KSE*-100 index). Histograms of the returns of all Portfolio assets are positively skewed and follow Leptokurtic distribution.

To out-perform benchmark is a tough task. Therefore, considering the results, we conclude that the Markov decision process is one of best approach which decides the next day portfolio allocation by observing the current state.

References

[1] Bellman, R., 1954. The theory of dynamic programming. Bulletin of the American Mathematical Society, 60(6), pp.503-515.

[2] Blackwell, D., 1965. Discounted dynamic programming. The Annals of Mathematical Statistics, 36(1), pp.226-235.

[3] Dubins, L.E., Savage, L.J., Sudderth, W. and Gilat, D., 2014. How to gamble if you must: Inequalities for stochastic processes. Courier Corporation.

[4] Heyman, D.P. and Sobel, M.J., 2004. Stochastic models in operations research: stochastic optimization (Vol. 2). Courier Corporation.

[5] Hinderer, K., 1970. Foundations of non-stationary dynamic programming with discrete time parameter. Lecture Notes. Operations Research and Mathematical Systems, 33.

[6] Ingersoll, J.E., 1987. Theory of financial decision making (Vol. 3). Rowman & Littlefield.

[7] Shapley, L.S., 1953. Stochastic games. Proceedings of the national academy of sciences, 39(10), pp.1095-1100.

[8] Shiryaev, A.N., 1965. Some new results in the theory of controlled random processes. In Transactions of the Fourth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes (pp. 131-203).

[9] Stokey, N.L., 1989. Recursive methods in economic dynamics. Harvard University Press.

[10] Bookstaber, D., 2005. Using Markov Decision Processes to Solve a Portfolio Allocation Problem. Undergraduate Thesis, Brown University.