Mathematical Methods in Operations Research 2005
Special issue in honor of *Arie Hordijk*

Time Consistent Dynamic Risk Measures

Kang Boda, Jerzy A.Filar

Center for Industrial and Applied Mathematics, School of Mathematics & Statistics, University of South Australia, Mawson Lakes, SA 5095, Australia

Received: January 2005 / Revised version: March 2005

Abstract We introduce the time consistency concept that is inspired by the so-called "principle of optimality" of dynamic programming and demonstrate − via an example − that the conditional value-at-risk (CVaR) need not be time consistent in a multi-stage case. Then, we give the formulation of the target-percentile risk measure which is time consistent and hence more suitable in the multi-stage investment context. Finally, we also generalize the value-at-risk (VaR) and CVaR to multi-stage risk measures based on the theory and structure of the target-percentile risk measure.

Key Words: Time consistency, multi-stage, target-percentile, value-at-risk, conditional value-at-risk, Markov decision process.

1 Introduction

Since Markowitz's seminal work [10] on financial risk, the variance (or, equivalently, standard deviation) of a random return/loss has been frequently used as a measure of risk. However, this measure has also been criticized (see, e.g. [15]), because the standard deviation does not adequately account for the phenomenon of "fat tails" in loss distributions and, moreover, it equally penalizes "ups" and "downs". Consequently, nowadays, the so-called "value-at-risk" (VaR) has become a commonly used measure of risk. The latter measures risk as the maximum loss that might be incurred with respect to a given, and fixed, confidence level.

Nonetheless, even value-at-risk is known to have certain drawbacks such as lack of convexity, monotonicity, as well as of reasonable continuity properties. To address these issues, Rockafellar and Urysaev [13–15] proposed another risk measure "conditional value-at-risk" (CVaR). The latter satisfies four axioms that are deemed desirable, namely, translation invariance, subadditivity, positive homogeneity, and monotonicity and leads to a tractable form of a portfolio optimization problem. That is, the problem of minimizing CVaR, with respect to portfolio allocation variables, is a convex programming problem.

It is, perhaps, important to note that all of the above measures of risk are based on certain characteristics of the distribution of a random loss under a fixed, and static (single-stage), portfolio allocation. However, nowadays, most investors are making portfolio decisions dynamically (over time); usually at discrete times (e.g., once a day or once a week). Hence, a natural question that arises is: *What measure of risk is most appropriate for a dynamic portfolio allocation problem?*

Clearly, for a dynamic multi-stage problem, an investor wants to make a sequence of portfolio allocation decisions − one at each stage − so that his or her risk is at an appropriate level not only when considering the entire time horizon, but also at intermediate stages as the process evolves. This concept is what has led us to define the notion of *time consistency of a dynamic portfolio, with respect to a measure of risk.*

We believe that time consistency is an important property when we consider multi-stage investment problems. In this paper, we propose a definition of time consistency that is inspired by the "principle of optimality" of dynamic programming (see, e.g. [2]). We believe that this is the most natural definition.

However, we shall show that − with respect to most of the commonly used measures of risk (e.g., VaR or CVaR) − there may not exist time consistent optimal dynamic portfolios. In our opinion, this calls into question the appropriateness of these risk measures in multi-stage setting.

To address the problem, we propose an alternative measure of risk which ensures that there always exist time-consistent optimal portfolios. We also demonstrate that the conventional measures of risk, such as VaR or CVaR, can still be used in conjunction with the new, time consistent, measure.

The structure of the paper is as follows. In Section 2, we first define precisely the time consistency concept and demonstrate − via an example − that CVaR need not be time consistent in a multi-stage case. In Section 3, we give the formulation of the target-percentile risk measure which is time consistent and hence more suitable in the multi-stage investment context. We also generalize VaR and CVaR to multi-stage risk measures based on the theory and structure in previous section in Section 4.

2 Time consistency problem & counter examples

2.1 Time consistent risk measures

We begin with an intuitive description of the time consistency concept.

Time consistency of a risk measure means that if a decision-maker uses a risk measure minimizing policy for the $n-$stage problem, then the component of that multi-stage policy from the $t^{th}-$stage to the end should be a risk measure minimizing policy in the remaining $(n-t+1)-$stage problem, for every $t = 1, 2, \cdots, n$. In common sense terms, a decision maker needs to be constantly concerned about optimizing his or her decisions for the remaining portion of the time horizon. That is, current optimal decisions must look to the future, rather than the past.

We shall now make the above concept more precise.

Consider an $n-$stage portfolio optimization planning problem. Let a *decision rule* at time/stage t be denoted by the column vector π_t whose entries represent the fractions of the total portfolio allocated to individual stocks. It is assumed that $\pi_t^T \cdot e = 1$, where e is the column vector of all ones of the dimension equal to the number m of available stocks. That is, we are assuming that all the resources are invested in these m stocks at each time t. A *policy* π will be defined as a sequence of decision rules, that is, $\pi = (\pi_1, \cdots, \pi_n)$.

Let the return/reward at stage t be denoted by the random variable r_t whose probability distribution function depends on the policy π. Let $R_t := g(r_1, \cdots, r_t)$, $t = 1, \cdots, n$ be the aggregated return[1] for the $t-$stage process. For all $t = 1, \cdots, n$, let $Z_t^\pi = Z_t(\pi_1, \cdots, \pi_t)$ be the risk measure of $t-$stage value corresponding to the decision rules $\pi_1, \cdots, \pi_t$ and with respect to g.

We shall say that the risk measure Z is *time-consistent with respect to* g if two conditions are satisfied:

TC1 If the decision rule π_t^* at each stage t, $t = 1, \cdots, n$ is chosen by

$$\pi_t^* \in \text{Argmin}_{\pi_t} Z_{n-t+1}(\pi_t, \pi_{t+1}^*, \cdots, \pi_n^*), \forall t = 1, \cdots, n; \quad (1)$$

then the policy $(\pi_1^*, \cdots, \pi_n^*)$ will be the optimal policy in the problem

$$\min_\pi Z_n(\pi). \quad (2)$$

TC2 If the policy

$$\pi^* = (\pi_1^*, \cdots, \pi_n^*) \in \text{Argmin}_\pi Z_n(\pi), \quad (3)$$

satisfies

$$(\pi_t^*, \cdots, \pi_n^*) \in \text{Argmin}_{\pi_t, \cdots, \pi_n} Z_{n-t+1}(\pi_t, \cdots, \pi_n), \ \forall t = 2, \cdots, n. \quad (4)$$

Remark 1 Clearly, $TC1$ ensures that a policy assembled from risk measure minimizing decision rules, as time evolves, is a risk measure minimizing policy over the entire horizon. On the other hand, $TC2$ ensures that "sub-policies" of an optimal policy π^*, over the remaining (shorter than n) time horizons, are also risk minimizing policies in the corresponding (shorter) sub-problems.

The above definition is inspired by the so-called "principle of optimality" of dynamic programming (see [2]).

[1] In many applications aggregation may simply refer to a summation, however, multiplicative aggregation of compounded interests is also very natural.

2.2 VaR and CVaR need not be time consistent

Here we will take a simple portfolio problem as an example.

As above, we assume that there are m stocks in the market. The investor will invest in these stocks, at the beginning of every time period t, $t = 1, \cdots, n$. The initial capital is given and there is no loss of generality in assuming it to be \$1.

For $t = 1, \cdots, n$, let $Y_t = (Y_{t1}, Y_{t2}, \cdots, Y_{tm})^T$ be the random percentage return vector of the m stocks under consideration by the investor. For notational convenience, we assume that $Y_{tk},\ t = 1, \cdots, n,\ k = 1, \cdots, m$ are all continuous random variables. Now, define the $m \times n$ matrix Y to be the matrix whose columns are the percentage return vectors $Y_t = (Y_{t1}, Y_{t2}, \cdots, Y_{tm})^T$, for every t.

Further, note that a policy π as defined in the previous section can also be regarded as an $m \times n$ matrix whose columns are the decision rules π_t for $t = 1, \cdots, n$.

Of course, if the initial capital were \$1 and an investment policy π were used, then the *total random return* after n years will be simply:

$$R_n(\pi, Y) := \Pi_{t=1}^n (1 + \pi_t^T Y_t).$$

A simple logarithmic transformation converts the above multiplicative return to a more convenient additive one. Clearly, investors are most concerned about the initial capital going down and this corresponds to $R_n(\pi, Y)$ being strictly less than 1 which $-$ after taking a logarithm $-$ becomes negative. Consequently, it is natural to define the *total random loss* resulting from policy π as

$$L_n(\pi, Y) := -\log(R_n(\pi, Y)) = -\log \Pi_{t=1}^n (1 + \pi_t^T Y_t) = -\sum_{t=1}^{n} \log(1 + \pi_t^T Y_t).$$

Since $-\log(\cdot)$ is a convex function, it is easy to check that the total loss function $L_n(\pi, Y)$ is also a convex function of the policy π. Hence, following standard arguments (see, e.g. Rockafellar et.al in [13]) we can define VaR and CVaR; we simply substitute $L_n(\pi, Y)$ in place of $f(x, y)$ in page 23 of [13]. That is, we have the following definitions.

Definition 1 *The $n-$stage value-at-risk, denoted by α-VaR and associated with the total loss $L_n(\pi, Y)$ is defined by:*

$$\zeta_{n,\alpha}^{\pi} = \zeta_{n,\alpha}(\pi) := \min\{\zeta | P_Y(L_n(\pi, Y) \le \zeta) \ge \alpha\}.$$

Definition 2 [2] *The $n-$stage conditional value-at-risk, denoted by α-CVaR and associated with the total loss $L_n(\pi, Y)$ is defined by:*

$$\phi_{n,\alpha}^{\pi} = \phi_{n,\alpha}(\pi) := E[L_n(\pi, Y) | L_n(\pi, Y) \ge \zeta_{n,\alpha}(\pi)],$$

[2] This definition is analogous to that in [13] since we assume $Y_{tk},\ t = 1, \cdots, n,\ k = 1, \cdots, m$ are all continuous random variables.

where $E(\cdot)$ denotes the mathematical expectation.

For fixed α, if we choose $Z_n(\pi) = \zeta_{n,\alpha}(\pi)$, then we use VaR as our measure of risk; otherwise we choose $Z_n(\pi) = \phi_{n,\alpha}(\pi)$, then we use CVaR as our measure of risk. In what follows, we will show that both of VaR and CVaR need not to be time consistent in multi-stage setting.

From Theorem 1 in [13], we know that $\phi_{n,\alpha}(\pi)$ is a convex function of π and can be calculated as

$$\phi_{n,\alpha}(\pi) = \min_{\zeta} F_{n,\alpha}(\pi, \zeta),$$

where $F_{n,\alpha}(\pi, \zeta) = \zeta + \frac{1}{1-\alpha} E_Y\{[L_n(\pi, Y) - \zeta]^+\}$.

Hence it is not surprising that by an argument analogous to that presented in [13,14] we can derive the following algorithm to minimize the $n-$stage CVaR $\phi_{n,\alpha}(\pi)$ by choosing appropriate decision rule at each stage.

For each stage t, we assume the distribution of the random return Y_t is known and given by $p(y)$. Hence, we can generate (vector-valued) samples $y_t^k, \quad k = 1, \cdots, q$ for each t from the distribution $p(y)$. Thus we obtain a corresponding approximation $\tilde{F}_{n,\alpha}(\pi, \zeta)$ for $F_{n,\alpha}(\pi, \zeta)$ as follows:

$$\tilde{F}_{n,\alpha}(\pi, \zeta) = \zeta + \frac{1}{q(1-\alpha)} \sum_{k=1}^{q} [- \sum_{t=1}^{n} \log(1 + \pi_t^T y_t^k) - \zeta]^+ .$$

Based on this, we have the following optimization algorithm for deriving an optimal policy (investment portfolio) in the $n-$stage problem:

$$\begin{cases} \min \ \zeta + \frac{1}{q(1-\alpha)} \sum_{k=1}^{q} u_k \\ \text{s.t.} \\ \pi_{tj} \geq 0, \ t = 1, \cdots, n, j = 1, \cdots, m; \ \sum_{j=1}^{m} \pi_{tj} = 1, \ t = 1, \cdots, n; \\ u_k \geq 0, \ \ u_k + \sum_{t=1}^{n} \log(1 + \pi_t^T y_t^k) + \zeta \geq 0, \ k = 1, \cdots, q. \end{cases} \tag{5}$$

Next, we want to calculate the optimal VaR, CVaR and corresponding optimal portfolios for some given stocks, time periods and distributions.

In this simple example (based on an example given in [13]), we set $m = 3$ and $n = 2$, so that we have three stocks and we invest in them, each year, for three consecutive years. Further, we set $\alpha = 0.99$ and want to choose an optimal portfolio at the beginning of each year so that the 0.99-CVaR of the total random loss L_t, for $(t = 1, 2)$ is as small as possible. Finally, we assume that the distribution $p(y)$ of the return is the multi-normal distribution $N(\mu, \Sigma)$ with the mean vector μ and the variance-covariance matrix Σ. In our example, the numerical values of the latter are given in Tables 1 and 2.

After generating 100 samples – for each of the 2 stages – from the multi-normal distribution $N(\mu, \Sigma)$ we are able to solve the mathematical program (5) to obtain an optimal portfolio and a corresponding value of 0.99-CVaR. This was done with n set to 1 and 2, respectively. Here, our measure of risk is $Z_t(\pi) = \phi_{t,\alpha}(\pi)$ for $t = 1, 2$ and $\alpha = 0.99$. We used Lingo 8.0 to obtain

Table 1 Portfolio mean return μ

Instrument	Mean return
Option 1	0.0101110
Option 2	0.0043532
Option 3	0.0137058

Table 2 Portfolio variance-covariance matrix Σ

	Option 1	Option 2	Option 3
Option 1	0.00324625	0.00022983	0.00420395
Option 2	0.00022983	0.00049937	0.00019247
Option 3	0.00420395	0.00019247	0.00764097

the following numerical results and checked if this $Z_t(\pi)$ satisfied TC1 and TC2.

An optimal policy for the 1-stage problem (set $t = 2$ in Eq.(1)) with respect to the samples, $y_2^k, k = 1, \cdots, q$ is $\pi^* = \pi_2^*$, where,

$$\pi_2^* = (0.101, 0.829, 0.070)^T.$$

An optimal policy for the 2-stage problem ($n = 2$) is $\hat{\pi} = (\hat{\pi}_1, \hat{\pi}_2)$, where,

$$\hat{\pi}_1 = (0.000, 1.000, 0.000)^T, \hat{\pi}_2 = (0.000, 0.994, 0.006)^T.$$

However, setting $t = 1$ in Eq.(1), by direct calculation, it can be checked that if we fix $\pi_2 = \pi_2^*$ in (5) and minimize its objective with respect to π_1 only then we obtain $\pi_1^* = (0.154, 0.846, 0.000)$.

Unfortunately, it is easy to check that,

$$\phi_{2,\alpha}(\pi_1^*, \pi_2^*) = 0.0394 > 0.0374 = \phi_{2,\alpha}(\hat{\pi}_1, \hat{\pi}_2),$$

which contradicts (TC1).

Conversely we verify that with respect to the samples, $y_2^k, k = 1, \cdots, q$

$$\phi_{1,\alpha}(\hat{\pi}_2) = 0.0547 > 0.0312 = \phi_{1,\alpha}(\pi_2^*),$$

so, $\hat{\pi}_2$ is not the optimal solution for $\phi_{1,\alpha}(\pi)$, that is Eq.(4) doesn't hold when $t = 2$ and this contradicts (TC2).

The above calculation shows that the policy chosen from optimal action for each stage is not optimal for the total horizon. That means conditional value-at-risk is not a time consistent risk measure.

Corollary 9 in [14] shows that for suitably chosen probability threshold α and sample size, value-at-risk and conditional value-at-risk coincide. In our example, $\alpha = 0.99$ and sample size equal to 100 were chosen so as to satisfy the conditions of that corollary. Thus the above example also shows that value-at-risk $\zeta_{n,\alpha}(\pi)$ is not a time consistent risk measure.

3 Time consistent target percentile risk measures

We shall propose a new risk measure that is not only time-consistent in the multi-stage case but will also consider the decision-maker's target. We will use Markov decision processes with probability criteria (e.g., see [3] and [17]) to model such a risk measure.

3.1 Model description

We consider the following discrete-time and stationary *Markov decision process*:

$$\Gamma = (S, A, R, P, \beta), \tag{6}$$

where the *state space* S is countable, the *action space* $A(i)$ in each state i is finite and the *overall action space* $A = \cup_{i \in S} A(i)$ is countable. The *return set* R is a bounded countable subset of $\mathcal{R} = (-\infty, +\infty)$. For each t from $1, \cdots$, let i_t, a_t and r_t denote the state of the system, the action taken by the decision maker, and the return received at stage t, respectively. The stationary, single-stage, conditional *transition probabilities* are defined by

$$p^a_{ijr} := P(i_{t+1} = j, r_t = r | i_t = i, a_t = a), \ i, j \in S, a \in A(i), r \in R, n \geq 1. \tag{7}$$

$$\sum_{j \in S, r \in R} p^a_{ijr} = 1, i \in S, a \in A(i). \tag{8}$$

We shall also assume that future costs are discounted by the *discount factor* $\beta \in (0, 1]$.

In our formulation, when making a decision and taking an action at each stage, the decision maker considers not only the state of the original system but also his *target*. Effectively, this means that a new *hybrid state* $(i, x) \in S \times \mathcal{R}$ is introduced. Hence we expand MDP Γ by enlarging the state space. We refer to (i, x) as the *hybrid state* of the decision maker to distinguish it from the system's state i, where x is the target value. Note that if the initial state of the decision maker is (i, x) and an action a is taken according to (7), the decision-maker's new hybrid state transits from (i, x) to $(j, (x - r)/\beta)$ with probability p^a_{ijr}.

Thus, if we denote E as the extended (hybrid) state space, then the extended MDP $\tilde{\Gamma}$ has the following structure:

$$\tilde{\Gamma} = (E, A, R, P, \beta), \tag{9}$$

where the state space $E = S \times \mathcal{R}$, the action space $A = \cup_{(i,x) \in E} A(i, x) = \cup_{i \in S} A(i)$. Note that $A(i, x) = A(i), (i, x) \in E$, the extended transition probabilities are simply P: $P(e_{t+1} = (j, \frac{x-r}{\beta}) | e_t = (i, x), a_t = a) = p^a_{ijr}, \ i, j \in S, a \in A(i), r \in R, x \in \mathcal{R}$. The return set R and the discount factor β are the same as in MDP Γ.

Since in the model (9), the target is important when making decisions we must define policies which depend both on the system's state and the target, that is on the hybrid state.

Let Π_m, Π_m^d, Π_s, Π_s^d denote the set of all Markov policies, all deterministic Markov policies, all stationary policies and all deterministic stationary policies in $\tilde{\Gamma}$ defined in the usual way (e.g., see [11]).

Definition 3 *A policy $\pi = \{\pi_t, t = 1, 2, \cdots\} \in \Pi$ is said to be a* **TI-policy** *if the policy π is independent of all targets x_t $(t \geq 1)$. Let Π_0 denote the set of all TI-policies.*

Note that a transition law P and a policy π determine the conditional probability measure P_π on the space of all possible histories of the process. Let R_n^π denote the random variable that is the sum of discounted returns generated by policy π for the n-stage finite horizon problem. That is, $R_n^\pi = \sum_{t=1}^n \beta^{t-1} r_t$, for $n \geq 1$. To simplify the notation, we will use R_n instead of R_n^π when the choice of the policy is clear in the context.

Note that for any $\pi \in \Pi$, the functions

$$F_n^\pi(i,x) = P_\pi(R_n^\pi \leq x | e_1 = (i,x)),\ (i,x) \in E, n \geq 1,$$

are the objective functions that the decision maker wishes to minimize if he or she is interested in achieving the minimal risk of not attaining the target x in the return. Consequently, $F_n^\pi(i,x)$ is called the **objective function** generated by π.

Definition 4 *The following functions $F_n^*(i,x) = \inf_{\pi \in \Pi}\{F_n^\pi(i,x)\},\ (i,x) \in E, n \geq 1$ are called the* **optimal value functions**.

Definition 5 *If the policy $\pi^* \in \Pi$ is such that $F_n^{\pi^*}(i,x) = F_n^*(i,x), \forall\ (i,x) \in E, n \geq 1$, then π^* is called an* n**-stage optimal policy**.

Remark 2 It can be checked that with the above definitions, π^* is an n-stage optimal policy if and only if π^* is the policy that minimizes the probability that the cumulative discounted return at the stage n does not exceed x.

3.2 $F_n^\pi(i,x)$ is time consistent

In this section, we invoke the results in Section 3 in Wu and Lin [17], to demonstrate that $F_n^\pi(i,x)$ is a time consistent risk measure. Since, many steps in our argument depend on the proofs in [17], we omit the technical details and, instead, refer the reader to [17].

In fact, for fixed $(i,x) \in E$, we let $Z_n(\pi) = Z_n^\pi = F_n^\pi(i,x)$. We select $\delta_t, t = 1, \cdots, n$ from Theorem 2 in [17] and denote the decision rules π_t^* in the policy $\pi^* = (\pi_1^*, \pi_2^*, \cdots, \pi_n^*)$ by $\pi_t^* = \delta_{n-t+1}, t = 1, \cdots, n$. Then from the definition of $A_t^*(e)$ we know for $t = 1, \cdots, n$

$$\pi_t^* = \delta_{n-t+1} \in \mathrm{Argmin}_\delta F_{n-t+1}^{\delta_1, \cdots, \delta_{n-t}, \delta} = \mathrm{Argmin}_\delta Z_{n-t+1}(\delta, \pi_{t+1}^*, \cdots, \pi_n^*),$$

and Theorem 2 in [17] concludes that

$$\pi^* = (\pi_1^*, \pi_2^*, \cdots, \pi_n^*) = (\delta_n, \delta_{n-1}, \cdots, \delta_1)$$

is n-stage optimal. Clearly, F_n^π follows TC1.

On the other hand, for

$$\pi^* = (\pi_1^*, \cdots, \pi_n^*) \in \text{Argmin}_\pi F_n^\pi(i,x) = \text{Argmin}_\pi Z_n(\pi),$$

Theorem 3 in [17] shows that

$$(\pi_2^*, \cdots, \pi_n^*) \in \text{Argmin}_{{}^1\pi^{(i,x,a)}} F_{n-1}^{{}^1\pi^{(i,x,a)}} = \text{Argmin}_{\pi_2,\cdots,\pi_n} Z_{n-1}(\pi_2, \cdots, \pi_n),$$

that means Eq.(4) holds when $t = 2$, if we substitute n by $n-1$ in Theorem 3 in [17], we will know that Eq.(4) holds when $t = 3$, continuing to do this, we can see that Eq.(4) holds for all $t = 2, \cdots, n$. This means that $Z_n(\pi) = F_n^\pi$ also satisfies TC2.

3.3 Complete Stochastic Order Optimization

Here, we consider the following discrete time and stationary Markov decision process,

$$\Gamma^0 = (S, A, p, R).$$

The state space S and the action space $A = \bigcup_{i\in S} A(i)$ are both countable and for each $i \in S$, the set of admissible actions $A(i)$ when the system is in state i is finite. The stationary conditional transition probabilities $p = (p_{ij}^a;\ i,j \in S, a \in A(i))$ will satisfy (i) $p_{ij}^a \geq 0,\ \ \forall i,j \in S, a \in A(i)$, and (ii) $\sum_{j\in S} p_{ij}^a = 1,\ \ \forall i,j \in S, a \in A(i)$. Finally, we have the return function $r = r(i,a,j), i,j \in S, a \in A(i)$ that is bounded.

After letting i_n, a_n and r_n denote by the system's state, the action taken by the decision maker and the return the decision maker will receive at the stage n respectively, the system will evolve as follows, starting from the state $i_n = i \in S$ and following an action $a_n = a \in A(i)$, the system transits to the next state $i_{n+1} = j \in S$ with probability p_{ij}^a and receives a return $r_n = r(i,a,j)$. The set of all possible returns is denoted by R and is bounded. It is obvious that the return here is a degenerate distribution which is a special case of the model discussed in Section 2 in [17].

Similarly to Section 2 in [17] we can define histories and policies that will now depend on the system's state and the action only but will not depend on the target. Hence the present model closely resembles the classical MDP model. However, we are still interested in minimizing the probability that the total return does not exceed a target x at some fixed stage.

For a given policy $\pi \in \Pi_0$, the $n-$stage total discounted return is defined by $R_n^\pi = \sum_{t=1}^n \beta^{n-1} r_t$. Based on above MDP structure, we have following definitions,

Definition 6 *The objective function and optimal value function for the Complete Stochastic Order Optimization problem are as follows,*

$$V_n^\pi(i,x) := P_\pi(R_n^\pi \le x | i_1 = i), \forall i \in S, x \in \mathcal{R}, \pi \in \Pi_0, n \ge 1; \quad (10)$$

$$V_n^*(i,x) := \inf_{\pi \in \Pi_0} \{V_n^\pi(i,x)\}, i \in S, x \in \mathcal{R}, n \ge 1. \quad (11)$$

Remark 3 Note that the difference between the process Γ_1 introduced in Section 2 in [17] and the current process Γ^0 is that we do not include the target x as part of the description of the state. So, recalling that Π_0 is the set of policies that are independent of the target x, if we select $\pi \in \Pi_0$, $V_n^\pi(i,x)$ becomes simply a probability distribution function of x. In the next section, we can use this distribution function to define the concepts of multi-stage value-at-risk and conditional value-at-risk.

Definition 7 *A policy $\pi^* \in \Pi_0$ will be called an $n-$stage optimal policy, if,*

$$V_n^{\pi^*}(i,x) = V_n^*(i,x), \ \forall i \in S, x \in \mathcal{R}.$$

At first sight, it may seem very difficult to establish the existence of such an n stage optimal policy. However with the help of Section 3 in [17] a number of results can be easily derived.

It is easy to see that, for any $\pi \in \Pi_0$,

$$F_n^\pi(i,x) = V_n^\pi(i,x), \ \forall i \in S, x \in \mathcal{R}, n \ge 1. \quad (12)$$

The following lemma illustrates the relationship between $F_n^\pi(i,x)$ and $V_n^\pi(i,x)$ for a general policy π and the fact that there is no loss of generality in restricting attention to policies in Π_0.

Lemma 1 *Let $x \in \mathcal{R}$ be given, then for each $\pi \in \Pi$, there exists a policy $\sigma(x) \in \Pi_0$ such that*

$$P_{\sigma(x)}(\cdot|i) = P_\pi(\cdot|i,x), \quad \forall i \in S.$$

Hence, we have

$$F_n^\sigma(i,x) = F_n^\pi(i,x), \quad \forall i \in S, n \ge 1,$$

and

$$\inf_{\pi \in \Pi}\{F_n^\pi(i,x)\} = \inf_{\pi \in \Pi_0}\{F_n^\pi(i,x)\}, \ \forall i \in S, x \in \mathcal{R}, n \ge 1.$$

Proof. Since x is fixed we can construct a policy $\sigma \in \Pi_0$ that "imitates" the policy π. In particular, for $k = 1, 2, \cdots$, define

$$\begin{aligned}
&\sigma_1(\cdot|i_1) = \pi_1(\cdot|i_1, x), \quad \forall i_1 \in S,\\
&\sigma_k(\cdot|i_1, a_1, i_2, a_2, \cdots, i_{k-1}, a_{k-1}, i_k)\\
&\qquad = \pi_k(\cdot|i_1, x, a_1, i_2, x_2, a_2, \cdots, i_{k-1}, x_{k-1}, a_{k-1}, i_k, x_k),\\
&\text{where } x_2 = (x - r(i_1, a_1, i_2))/\beta,\\
&\qquad x_3 = (x_2 - r(i_2, a_2, i_3))/\beta, \cdots, x_k = (x_{k-1} - r(i_{k-1}, a_{k-1}, i_k))/\beta,\\
&\qquad\qquad \forall i_l \in S, a_l \in A(i_l), l = 1, \cdots, k-1, \ i_k \in S, k \ge 2.
\end{aligned}$$

So $\sigma = (\sigma_n, n \geq 1) \in \Pi_0$. Obviously, starting from the initial state (i, x) in $\tilde{\Gamma}$ and using policy π is equivalent to using the imitating policy σ in Γ^0, since they have same decision rules. That is, $P_{\sigma(x)}(\cdot|i) = P_\pi(\cdot|i, x), \quad \forall i \in S$, and $F_n^\sigma(i, x) = F_n^\pi(i, x), \quad \forall i \in S, n \geq 1$. □

It easily follows from Lemma 1 that,

$$F_n^*(i, x) = V_n^*(i, x), \ \forall i \in S, x \in \mathcal{R}, n \geq 1. \tag{13}$$

which can explain the property that the optimal value function $F_n^*(i, x)$ is a distribution function of some random variable X taking values x.

The following theorem illustrates the existence of an optimal policy.

Theorem 1 *For a given target x and any $n \geq 1$, there exists $\pi \in \Pi_0$ such that,*

$$V_n^\pi(i, x) = V_n^*(i, x), \quad \forall i \in S.$$

Proof. This follows from Theorem 1 in [17] and Lemma 1.

For notational convenience, we will still use $A_n^*(i, x)$, $A_n^*(i)$ to denote the optimal action sets but will use the optimal value function $V_n^*(i, x)$ in Γ^0 instead of using the optimal value function $F_n^*(i, x)$ in Γ_1 in [17]. The structure of optimal policies is described by the next theorem.

Theorem 2 *1. A policy $\pi = (\pi_k, k \geq 1) \in \Pi_0$ is an optimal policy for the $n-$stage problem if and only if $\pi_1(A_n^*(i)|i) = 1, \forall i \in S$, and when $\pi_1(a|i)p_{ij}^a > 0$, we have $V_{n-1}^{\bar{\pi}^{(i,a)}}(j, x) = V_{n-1}^*(j, x), \forall x \in \mathcal{R}$.*

2. If $A_k^*(i) \neq \emptyset, \forall i \in S, k = 1, \cdots, n$, then there exists an $n-$stage optimal policy. In fact let $f_k(i) \in A_k^*(i), \forall i \in S, k = 1, \cdots, n, \pi = (\pi_k, k \geq 1) \in \Pi_0$, if $\pi_k = f_{n-k+1}, k = 1, \cdots, n$, namely $\pi(n) = (f_n, f_{n-1}, \cdots, f_1)$, then π is an $n-$ stage optimal policy.

Proof. The following, typical, recursive equations of dynamic programming can now be derived directly from definitions.

1. For all $\pi = (\pi_k, k \geq 1) \in \Pi_0$, we have

$$\begin{aligned} &V_0^\pi(i, x) = I_{[0,+\infty)}(x), \quad \forall i \in S, x \in \mathcal{R}, \\ &V_n^\pi(i, x) = \sum_{a \in A(i)} \pi_1(a|i) \sum_{j \in S} p_{ij}^a V_{n-1}^{\bar{\pi}^{(i,a)}}(j, (x - r(i, a, j))/\beta), \\ &\qquad\qquad\qquad\qquad\qquad\qquad \forall i \in S, x \in \mathcal{R}, n \geq 1, \end{aligned} \tag{14}$$

2. The optimal value functions will satisfy the following,

$$\begin{aligned} &V_0^*(i, x) = I_{[0,+\infty)}(x), \quad \forall i \in S, x \in \mathcal{R}, \\ &V_n^*(i, x) = \min_{a \in A(i)} \Big\{ \sum_{j \in S} p_{ij}^a V_{n-1}^*(j, (x - r(i, a, j))/\beta) \Big\}, \\ &\qquad\qquad\qquad\qquad\qquad\qquad \forall i \in S, x \in \mathcal{R}, n \geq 1. \end{aligned} \tag{15}$$

The result now follows from the above equations, Theorem 3, Theorem 4 in [17] and equations (12) and (13).

4 Multi-stage VaR and CVaR

Even though in the preceding section we had implied that, multi-stage problems, the target-percentile risk measure $Z_n(\pi) = F_n^\pi$ is preferable to either VaR or CVaR, we recognize that the latter are so well-established that decision makers may wish to compute them even while using an optimal policy constructed as in Section 3 in [17]. The purpose of this section is to demonstrate that this is still possible but with respect to probability distributions induced by the above policy.

Continuing from Section 3.3, we consider the MDP model Γ^0 and define the concepts of value-at-risk and conditional value-at-risk in this multi-stage context.

Definition 8 *For a given policy $\pi \in \Pi_0$, initial state $i \in S$ and the probability threshold $\alpha \in [0,1]$, the value-at-risk ($\zeta_{n,\alpha}^\pi(i)$) for the n-stage return R_n^π is denoted by:*

$$\zeta_{n,\alpha}^\pi(i) := -\sup\{\zeta | V_n^\pi(i,\zeta) \leq \alpha\}, \forall i \in S, \pi \in \Pi_0, \alpha \in [0,1], n \geq 1. \quad (16)$$

Remark 4 The definition here is equivalent to $\zeta_n^{\pi,\alpha}(i) = -\inf\{\zeta | V_n^\pi(i,\zeta) \geq \alpha\}$ when R_n^π has a continuous distribution. However, if R_n^π is a discrete r.v., then we always have $\zeta_n^{\pi,\alpha}(i) \leq \zeta_{n,\alpha}^\pi(i)$, so from an optimization point of view, minimizing the latter will force the former to be small as well. In following multi-stage setting, we shall always use Definition 8.

The above definition can, perhaps, be explained as follows. Starting from an initial state i and continuing to use the policy π for $n-$stages, the decision-maker wants to minimize the loss associated with the best of the $100\alpha\%$ worst cases of the $n-$stage total return R_n^π, namely, $\zeta_{n,\alpha}^\pi(i)$. Consequently, we have following definitions.

Definition 9 *The* **optimal value functions** *are defined as follows,*

$$\zeta_{n,\alpha}^*(i) := \inf_{\pi \in \Pi_0} \{\zeta_{n,\alpha}^\pi(i)\}, \forall i \in S, \alpha \in [0,1], n \geq 1. \quad (17)$$

Definition 10 *1. A policy $\pi^\alpha \in \Pi_0$ is said to be an $\alpha-$***optimal policy** *for $n-$stage value-at-risk ($\zeta_{n,\alpha}^\pi(i)$) if,*

$$\zeta_{n,\alpha}^{\pi^\alpha}(i) = \zeta_{n,\alpha}^*(i), \forall i \in S, n \geq 1. \quad (18)$$

2. A policy $\pi^* \in \Pi_0$ will be **optimal** *for $n-$stage value-at-risk ($\zeta_{n,\alpha}^\pi(i)$) if,*

$$\zeta_{n,\alpha}^{\pi^*}(i) = \zeta_{n,\alpha}^*(i), \forall i \in S, \alpha \in [0,1], n \geq 1. \quad (19)$$

We shall show that, with the help of our target percentile formulation of Section 3, the optimization problem implied by (17) is tractable. In particular, we exploit the fact that, by equation (13), the optimal value function $V_n^*(i,x)$ is a probability distribution function. Hence it is possible to define α-VaR, denoted by $x_{n,\alpha}^*(i)$, with respect to that particular distribution. Namely,

$$x_{n,\alpha}^*(i) := -\sup\{x | V_n^*(i,x) \leq \alpha\}, i \in S, \alpha \in [0,1], n \geq 1. \quad (20)$$

4.1 Existence of Multi-stage VaR

The following Theorem 3 explains the relationship between $x^*_{n,\alpha}(i)$ and the optimal value-at-risk $\zeta^*_{n,\alpha}(i)$.

Theorem 3 *For all $i \in S, \alpha \in [0,1]$, and $n \geq 1$, we have*

$$x^*_{n,\alpha}(i) = \zeta^*_{n,\alpha}(i) = \inf_{\pi} \zeta^{\pi}_{n,\alpha}(i),$$

and there exists an $\alpha-$optimal policy $\hat{\pi}_\alpha \in \Pi_0$ such that

$$\zeta^{\hat{\pi}_\alpha}_{n,\alpha}(i) = \zeta^*_{n,\alpha}(i).$$

Proof. According to the definition of $V^*_n(i,x)$, we have,

$$V^*_n(i,x) \leq V^{\pi}_n(i,x), \forall i \in S, x \in \mathcal{R}, \pi \in \Pi_0,$$

so for any $\alpha \in [0,1]$,

$$\{x | V^{\pi}_n(i,x) \leq \alpha\} \subset \{x | V^*_n(i,x) \leq \alpha\}.$$

Now, from the definition of a set supremum function, we have

$$-\zeta^{\pi}_{n,\alpha}(i) = \sup\{x | V^{\pi}_n(i,x) \leq \alpha\} \leq \sup\{x | V^*_n(i,x) \leq \alpha\} = -x^*_{n,\alpha}(i).$$

From the above, we have, $\forall \pi \in \Pi_0$,

$$x^*_{n,\alpha}(i) \leq \zeta^{\pi}_{n,\alpha}(i). \tag{21}$$

Now, to prove that $x^*_{n,\alpha}(i)$ is exactly the infimum of $\zeta^{\pi}_{n,\alpha}(i)$ over all $\pi \in \Pi_0$, it suffices to prove that, $\forall\ \varepsilon > 0$, $\exists\ \pi \in \Pi_0$ such that $x^*_{n,\alpha}(i) > \zeta^{\pi}_{n,\alpha}(i) - \varepsilon$.

Suppose, by contradiction, that the above statement is false. Hence, there must exist $\varepsilon > 0$, such that $\forall\ \pi \in \Pi_0$, we have

$$x^*_{n,\alpha}(i) \leq \zeta^{\pi}_{n,\alpha}(i) - \varepsilon. \tag{22}$$

By the definition of $x^*_{n,\alpha}(i)$ and the property of the set function sup, we have that - for this ε - there exists $x_\varepsilon \in \{x | V^*_n(i,x) \leq \alpha\}$ such that $-x^*_{n,\alpha}(i) < x_\varepsilon + \varepsilon$, namely

$$x^*_{n,\alpha}(i) > -x_\varepsilon - \varepsilon. \tag{23}$$

After combing inequalities (22) and (23), we see that $-x_\varepsilon - \varepsilon < x^*_{n,\alpha}(i) \leq \zeta^{\pi}_{n,\alpha}(i) - \varepsilon$, which implies that $\forall\ \pi \in \Pi_0$

$$-\zeta^{\pi}_{n,\alpha}(i) < x_\varepsilon. \tag{24}$$

From Theorem 1 we know that for the above x_ε, there exists a $\pi^*_\varepsilon \in \Pi_0$ such that,

$$V^{\pi^*_\varepsilon}_n(i, x_\varepsilon) = V^*_n(i, x_\varepsilon) \leq \alpha,$$

so

$$x_\varepsilon \in \{x | V_n^{\pi^*_\varepsilon}(i,x) \le \alpha\}.$$

Using Definition 8, we now obtain

$$-\zeta^{\pi^*_\varepsilon}_{n,\alpha}(i) \ge x_\varepsilon,$$

which contradicts the inequality (24) that must hold for π^*_ε as well.

Now to prove the existence of the required $\alpha-$optimal policy we shall use $-x^*_{n,\alpha}(i)$ as the decision-maker's target in Theorem 1. From that theorem we know that there exists a policy $\hat{\pi}_\alpha \in \Pi_0$ such that

$$V_n^{\hat{\pi}_\alpha}(i, -x^*_{n,\alpha}(i)) = V_n^*(i, -x^*_{n,\alpha}(i)) \le \alpha, \ \forall i \in S, \alpha \in [0,1].$$

Also, it follows from the definition that

$$\zeta^{\hat{\pi}_\alpha}_{n,\alpha}(i) = -\sup\{x | V_n^{\hat{\pi}_\alpha}(i,x) \le \alpha\},$$

so we have

$$-\zeta^{\hat{\pi}_\alpha}_{n,\alpha}(i) \ge -x^*_{n,\alpha}(i) = -\zeta^*_{n,\alpha}(i),$$

and

$$\zeta^{\hat{\pi}_\alpha}_{n,\alpha}(i) \le \zeta^*_{n,\alpha}(i).$$

But according to the definition of $\zeta^*_{n,\alpha}(i)$ a strict inequality is impossible in the above, and hence

$$\zeta^{\hat{\pi}_\alpha}_{n,\alpha}(i) = \zeta^*_{n,\alpha}(i),$$

which shows that $\hat{\pi}_\alpha$ is an $\alpha-$optimal policy. □

The preceding theorem demonstrated that an $\alpha-$optimal policy always exists. This may not be the case for an optimal policy, that is, a policy which is $\alpha-$optimal for every α between 0 and 1. The following theorem which provides a sufficient condition for the existence of such a "uniformly optimal" policy now follows almost immediately.

Theorem 4 *If $A^*_k(i) \ne \emptyset, \forall i \in S, k = 1, \cdots, n$, then there exists an $n-$stage optimal policy $\hat{\pi} \in \Pi_0$ such that*

$$\zeta^{\hat{\pi}}_{k,\alpha}(i) = \zeta^*_{k,\alpha}(i), \ \forall k, \alpha.$$

In fact, if $f_k(i) \in A^*_k(i), \forall i \in S, k = 1, \cdots, n$, $\hat{\pi} = (\hat{\pi}_k, k \ge 1) \in \Pi_0$, and we define $\hat{\pi}_k := f_{n-k+1}, k = 1, \cdots, n$, namely $\hat{\pi}(n) := (f_n, f_{n-1}, \cdots, f_1)$, then $\hat{\pi}(n)$ is an $n-$ stage optimal policy that is time consistent.

Proof. Based on the above selection procedure for the policy $\hat{\pi}(n)$, the result follows from part 2 of Theorem 2 and the argument in the proof of Theorem 3. □

4.2 Computation of Multi-stage VaR

Assume the (S, A, p, R) components of the Markov Decision Process Γ^0 are known. We shall outline a procedure to calculate the optimal multi-stage value-at-risk and a corresponding α-optimal, or n-stage optimal policy.

1. Use dynamic programming to calculate the optimal value functions $V_n^*(i,x)$ and optimal action sets $A_k^*(i,x), k = 1, \cdots, n$ for all $i \in S$ and $x \in \mathcal{R}$. Note that, in practice, this needs to be done numerically on a suitable grid of x values (e.g., see Wu and Lin [17]).
2. For a given probability threshold $\alpha \in [0,1]$, calculate the value-at-risk $x_{n,\alpha}^*(i)$ that corresponds to the optimal value function $V_n^*(i,x)$, that is,
$$x_{n,\alpha}^*(i) := -\sup\{x|V_n^*(i,x) \leq \alpha\}, i \in S, n \geq 1.$$
Note that the optimal value-at-risk $\zeta_{n,\alpha}^*(i) = x_{n,\alpha}^*(i)$.
3. If there exists $k \in \{1, \cdots, n\}$ such that $A_k^*(i) = \emptyset, \forall i \in S$, use $-\zeta_{n,\alpha}^*(i)$ as a target to find a corresponding optimal action $f_k^*(i,x) \in A_k^*(i,x)$, at each stage $k = 1, \cdots, n$. Now, the policy $(f_n^*(i_1,x_1), f_{n-1}^*(i_2,x_2), \cdots, f_1^*(i_n,x_n))$, where $i_1 = i, x_1 = \zeta_{n,\alpha}^*(i)$ is an $\alpha-$optimal policy for the value-at-risk $\zeta_{n,\alpha}^*(i)$.
4. If $A_k^*(i) \neq \emptyset, \forall i \in S, k = 1, \cdots, n$, following Theorem 4, set $\hat{\pi}(n) := (f_n, f_{n-1}, \cdots, f_1)$ where $f_k(i) \in A_k^*(i), \forall i \in S, k = 1, \cdots, n$. The policy $\hat{\pi}(n)$ so constructed is $n-$stage optimal and time consistent.

4.3 Multi-stage Conditional VaR

Next, we will consider the multi-stage conditional value-at-risk (CVaR) in above framework.

Definition 11 *For a given policy $\pi \in \Pi_0$, $\alpha \in [0,1]$ and the initial state i, the $n-$stage conditional value-at-risk, α-CVaR, can be defined as:*

$$\phi_{n,\alpha}^\pi(i) = -[\textit{mean of the } \alpha - \textit{tail distribution of } R_n^\pi].$$

In the above, the distribution in question is defined by

$$V_{n,\alpha}^\pi(i,\zeta) = \begin{cases} 1, & \zeta \geq -\zeta_{n,\alpha}^\pi(i); \\ \frac{V_n^\pi(i,\zeta)}{\alpha}, & \zeta < -\zeta_{n,\alpha}^\pi(i). \end{cases}$$

The $\phi_{n,\alpha}^\pi(i)$ defined above is the mean loss in the $100\alpha\%$ worst cases of the $n-$stage total return R_n^π when the system starts from i and is controlled by the policy π up to n^{th} stage. We can now prove a proposition connecting $\phi_{n,\alpha}^\pi(i)$ with $\zeta_{n,\alpha}^\pi(i)$.

Proposition 1 *For a given policy $\pi \in \Pi_0$, the α-CVaR can be expressed by:*

$$\phi^{\pi}_{n,\alpha}(i) = \frac{1}{\alpha}\int_0^{\alpha} \zeta^{\pi}_{n,p}(i)dp. \tag{25}$$

Proof. Following the definition of $\phi^{\pi}_{n,\alpha}(i)$, we know that,

$$\phi^{\pi}_{n,\alpha}(i) = -\left[\frac{1}{\alpha}\int_{-\infty}^{-\zeta^{\pi}_{n,\alpha}(i)} \zeta dV^{\pi}_n(i,\zeta) + (-\zeta^{\pi}_{n,\alpha}(i)) \times \left(\frac{\alpha - V^{\pi}_n(i,-\zeta^{\pi}_{n,\alpha}(i))}{\alpha}\right)\right] = -\left[I + II\right], \tag{26}$$

since $V^{\pi}_n(i,\zeta)$ is a distribution function of ζ. Now, we can change variable to let $p := V^{\pi}_n(i,\zeta) \in [0,1]$, then the ζ corresponding to p will be $\sup\{\zeta | V^{\pi}_n(i,\zeta) \le p\} = -\zeta^{\pi}_{n,p}(i)$, and we know that p will range from 0 to $V^{\pi}_n(i,-\zeta^{\pi}_{n,\alpha}(i))$ when ζ ranges from $-\infty$ to $-\zeta^{\pi}_{n,\alpha}(i)$. Hence, I in Eq. (26) will be given by

$$I = \frac{1}{\alpha}\int_0^{V^{\pi}_n(i,-\zeta^{\pi}_{n,\alpha}(i))} (-\zeta^{\pi}_{n,p}(i))dp.$$

We now have to consider the following two cases.

1. If $V^{\pi}_n(i,-\zeta^{\pi}_{n,\alpha}(i)) = \alpha$, then it is easy to see II in Eq.(26) is equal to 0, so
$$\phi^{\pi}_{n,\alpha}(i) = -I = -\frac{1}{\alpha}\int_0^{\alpha}(-\zeta^{\pi}_{n,p}(i))dp = \frac{1}{\alpha}\int_0^{\alpha}\zeta^{\pi}_{n,p}(i)dp,$$
which shows Eq.(25) holds.
2. If $V^{\pi}_n(i,-\zeta^{\pi}_{n,\alpha}(i)) < \alpha$, then it follows from the definition of value-at-risk that
$$-\zeta^{\pi}_{n,p}(i) = -\zeta^{\pi}_{n,\alpha}(i), \forall p \in [V^{\pi}_n(i,-\zeta^{\pi}_{n,\alpha}(i)),\alpha], n \ge 1, i \in S.$$
Indeed, for $p \le \alpha$ we know that $-\zeta^{\pi}_{n,p}(i) \le -\zeta^{\pi}_{n,\alpha}(i)$. However, $-\zeta^{\pi}_{n,\alpha}(i) \in \{\zeta | V^{\pi}_n(i,\zeta) \le p\}$, which forces the equality $-\zeta^{\pi}_{n,p}(i) = -\zeta^{\pi}_{n,\alpha}(i)$. Hence we can rewrite II in Eq (26) as follows,
$$II = (-\zeta^{\pi}_{n,\alpha}(i)) \times \left(\frac{\alpha - V^{\pi}_n(i,-\zeta^{\pi}_{n,\alpha}(i))}{\alpha}\right) = \frac{1}{\alpha}\int_{V^{\pi}_n(i,-\zeta^{\pi}_{n,\alpha}(i))}^{\alpha}(-\zeta^{\pi}_{n,p}(i))dp.$$
Now the conditional value-at-risk becomes
$$\begin{aligned}\phi^{\pi}_{n,\alpha}(i) &= -\left[I + II\right]\\ &= -\left[\frac{1}{\alpha}\int_0^{V^{\pi}_n(i,-\zeta^{\pi}_{n,\alpha}(i))}(-\zeta^{\pi}_{n,p}(i))dp + \frac{1}{\alpha}\int_{V^{\pi}_n(i,-\zeta^{\pi}_{n,\alpha}(i))}^{\alpha}(-\zeta^{\pi}_{n,p}(i))dp\right]\\ &= -\frac{1}{\alpha}\int_0^{\alpha}(-\zeta^{\pi}_{n,p}(i))dp = \frac{1}{\alpha}\int_0^{\alpha}\zeta^{\pi}_{n,p}(i)dp,\end{aligned}$$
which shows that Eq.(25) also holds in this case. □

Remark 5 To understand this formulation we note that $(-\zeta^{\pi}_{n,p}(i))$ is exactly a quantile, namely the total return R^{π}_{n} value corresponding to the probability level p. So $1/\alpha$ times the integral from 0 to α is just the average of all possible returns in this range. The minus sign commutes returns into losses.

Similarly we can define $\alpha-$CVaR, namely $y^{*}_{n,\alpha}(i)$, in terms of $\alpha-$VaR $x^{*}_{n,\alpha}(i)$ based on the optimal value function $V^{*}_{n}(i,x)$. That is

$$y^{*}_{n,\alpha}(i) = \frac{1}{\alpha}\int_{0}^{\alpha} x^{*}_{n,p}(i)dp.$$

Ideally, we would like to prove that the infimum of $\phi^{\pi}_{n,\alpha}(i)$ coincides with $y^{*}_{n,\alpha}(i)$. Unfortunately, we are able to do so only under the strong conditions of the following theorem.

Theorem 5 *If $A^{*}_{k}(i) \neq \emptyset, \forall i \in S, k = 1, \cdots, n$, then for all $i \in S, \alpha \in [0,1]$, and $n \geq 1$, we have*

$$y^{*}_{n,\alpha}(i) = \phi^{*}_{n,\alpha}(i) := \inf_{\pi\in\Pi_0} \phi^{\pi}_{n,\alpha}(i),$$

and there exists a policy $\hat{\pi} \in \Pi_0$ such that

$$\phi^{\hat{\pi}}_{n,\alpha}(i) = \phi^{*}_{n,\alpha}(i),$$

that is time-consistent.

Proof. We know that $\forall \pi \in \Pi_0$ we have,

$$\zeta^{*}_{n,p}(i) \leq \zeta^{\pi}_{n,p}(i), \forall i \in S, n \in \mathcal{R}, p \in [0,1].$$

Thus

$$\int_{0}^{\alpha} \zeta^{*}_{n,p}(i)dp \leq \int_{0}^{\alpha} \zeta^{\pi}_{n,p}(i)dp, \forall i \in S, n \geq 1, \alpha \in [0,1].$$

Since $A^{*}_{k}(i) \neq \emptyset, \forall i \in S, k = 1, \cdots, n$, according to the definition of $\zeta^{*}_{n,p}(i)$ and Theorem 4, we have, $\forall \varepsilon > 0, n \geq 1$ there exists $\pi \in \Pi_0$ (we can choose $\pi = \hat{\pi}$ here) such that

$$\zeta^{\pi}_{n,p}(i) \leq \zeta^{*}_{n,p}(i) + \varepsilon/\alpha, \forall i \in S, p \in [0,1].$$

After integrating both sides of the above inequality, we have:

$$\int_{0}^{\alpha} \zeta^{\pi}_{n,p}(i)dp \leq \int_{0}^{\alpha} \zeta^{*}_{n,p}(i)dp + \varepsilon, \forall i \in S, \alpha \in [0,1].$$

Now, we obtain:

$$\inf_{\pi\in\Pi_0}\int_{0}^{\alpha} \zeta^{\pi}_{n,p}(i)dp = \int_{0}^{\alpha} \zeta^{*}_{n,p}(i)dp = \int_{0}^{\alpha}\left(\inf_{\pi\in\Pi_0} \zeta^{\pi}_{n,p}(i)\right)dp, \forall i \in S, n \geq 1, \alpha \in [0,1].$$

Hence,

$$\phi^{*}_{n,\alpha}(i) = \inf_{\pi\in\Pi_0} \phi^{\pi}_{n,\alpha}(i) = \inf_{\pi\in\Pi_0} \frac{1}{\alpha}\int_{0}^{\alpha} \zeta^{\pi}_{n,p}(i)dp = \frac{1}{\alpha}\int_{0}^{\alpha}\left(\inf_{\pi\in\Pi_0} \zeta^{\pi}_{n,p}(i)\right)dp =$$

$$\frac{1}{\alpha}\int_0^\alpha \left(\zeta_{n,p}^{\hat{\pi}}(i)\right) dp = \frac{1}{\alpha}\int_0^\alpha x_{n,p}^*(i)dp = y_{n,\alpha}^*(i),$$

where the second equality follows from Proposition 1 and the fourth equality follows from the condition $A_k^*(i) \neq \emptyset, \forall i \in S, k = 1, \cdots, n$ and Theorem 4. The policy $\hat{\pi}$ here is now constructed in the same way as in Theorem 4. So, $\hat{\pi}$ is the optimal policy for conditional value-at-risk $\phi_{n,\alpha}^{\pi}$. That is,

$$\phi_{n,\alpha}^{\hat{\pi}}(i) = \inf_{\pi \in \Pi_0} \phi_{n,\alpha}^{\pi}(i).$$

The time-consistency part follows from Theorem 4 immediately. □

The procedure to calculate the optimal multi-stage conditional value-at-risk is analogous to the procedure in Section 4.2 for the value-at-risk.

5 Conclusion

We introduced a, possibly novel, *time consistency property* of measures of risk that is applicable to multi-stage investment problems. This property is inspired by the “principle of optimality” of dynamic programming.

We showed (see Section 2) that - with respect to the most of the commonly used measures of risk (e.g., VaR or CVaR) - there need not exist time consistent optimal dynamic portfolios.

To address the above problem, we proposed an alternative *target percentile risk measure* for which there always exist time-consistent optimal portfolios (see Section 3.2).

We also demonstrated that the conventional measures of risk, such as VaR or CVaR, can still be used in conjunction with the new, time consistent, measure (see Section 4).

References

1. Artzner P, Delbaen F, Eber JM, Heath D (1999) Coherent Measures of Risk. Mathematical Finance, 9:203–227
2. Bellman R, Dreyfus S (1962) Applied Dynamic Programming. Princeton Univ. Press, Princeton, N.J.
3. Boda K, Filar JA, Lin YL, Spanjers L (2004) Stochastic Target Hitting Time and the Problem of Early Retirement. IEEE Transactions on Automatic Control, 49(3):409-419
4. Bouakiz M, Kebir Y (1995) Target-level criterion in Markov decision processes. J.Optim. Theory Appl., 86:1-15
5. Filar JA, Kallenberg LCM, Lee HM (1989) Variance-Penalized Markov Decision processes. Math. of Oper. Res., 14:147-161
6. Filar JA, Krass D, Ross KW (1995) Percentile Performance Criteria For Limiting Average Markov Decision Processes. IEEE AC, 40:2-10
7. Howard RA, Matheson JE (1972) Risk-sensitive Markov decision processes. Management Sciences, 18:356-369

8. Larsen H, Mausser H, Uryasev S (2002) Algorithms for Optimization of value-at-risk. Pardalos P, Tsitsiringos VK (Eds.) Financial Engineering, e-Commerce and Supply Chain, Kluwer Academic Publishers, 129-157
9. Lin YL, Wu CB, Kang BD (2003) Optimal models with maximizing probability of first achieving target value in the preceding stages. Science in China Series A-Mathematics 46(3):396-414
10. Markowitz HM (1959) Portfolio Selection: Efficient Diversification of Investment. J.Wiley and Sons
11. Puterman ML (1994) Markov Decision Processes: Discrete Stochastic Dynamic Programming. A Wiley, New York
12. Rockafellar RT (1970) Convex Analysis. Princeton University Press, Princeton, N.J.
13. Rockafellar RT, Uryasev S (2000) Optimization of conditional value-at-risk. Journal of Risk, 2:21–42
14. Rockafellar RT, Uryasev S (2002) Conditional value-at-risk for general loss distributions. Journal of Banking & Finance, 26:1443–1471
15. Rockafellar RT, Uryasev S (2002) Deviation Measures in Risk Analysis and Optimization. Research Report 2002-7, Dept. of Industrial and Systems Engineering, University of Florida
16. Whittle P (1990) Risk-sensitive optimal control. Wiley, New York
17. Wu CB, Lin YL (1999) Minimizing Risk Models in Markov Decision Processes With Policies Depending on Target Values. J.Math.Anal.Appl., 231:47–67
18. Yu SX, Lin YL and Yan PF (1998) Optimization Models for the First Arrival Target Distribution Function in Discrete Time. J.Math.Anal.Appl., 225:193–223