Dynamic programming and portfolio optimization

Shuenn-jyi Sheu
Institute of Mathematics, Academia Sinica

1. Introduction

The portfolio selection for an investor is to choose investment policy in different financial assets. That is, an investor has to decide the assets and the shares he likes to hold on the assets so that he can optimize his utility of consumption during the time interval $[0, T]$ and/or his total wealth at time T. This is a very practical problem. One of the major difficulty involved is to model the financial market which may consist of the following: A good understanding of mechanism of the economics; the statistical estimation of the parameters in the model. Searching for good model is far from complete and probably will never end. We shall introduce some approach developed and widely used in the portfolio theory: mean-variance analysis and dynamic programming approach. Concerning the usefulness of these ideas, we believe in order to use them effectively, each individual has to choose his own model based on his believe(or experience). Then the analysis provided in the approach could be useful since they in general can be applied to a class of models(in some sense, it is model independent). The mean-variance approach is probably the first known example to be used in the portfolio problem. This was initiated by H. Markowitz. It attracts a lot of attention and it is widely used by practitioner. The algorithm uses the expected values and covariances, a minimum knowledge about randomness of the prices of the assets. It is easy to implement even it is applied only for one-period model, a very simplified financial market. We have to mention the idea here. The idea of dynamic programming comes from control theory. Later, we shall introduce dynamic programmin approach for multi-period model and continuous time model.

2. Mean-Variance analysis

Assume that an investor wants to trade in d different securities at time $t = 0$. He decide to allocate his wealth to these d securities at time $t = 0$ and holds them until the time $t = T$. His goal is to maximize his wealth at time $t = T$. Such a model is called one-period model. We assume that there are no transaction costs and the securities are perfectly divisible. The future prices of the securities are modelled via their returns $R_i, i = 1, \cdots, d$ at time $t = T$. That is, denote $P_i(0), P_i(T)$ are the prices of i-th security at times $t = 0, T$, then

$$R_i = P_i(T)/P_i(0).$$

The returns are modeled as random variables with expected values and covariances,

$$\mu_i = E(R_i),\ \sigma_{ij} = Cov(R_i, R_j).$$

Let x be the initial wealth and ψ_i be the number of shares of security i held by the investor at time $t = 0$(hence also at time $t = T$). The π_i is the fraction of the initial

wealth in security i. Then

$$\pi_i = \frac{\psi_i P_i(0)}{x}.$$

The portfolio $\pi = (\pi_1, \cdots, \pi_d)$ by the investor results in a total return R at time $t = T$,

$$R = \sum \pi_i R_i.$$

The total wealth at time $t = T$ is given by

$$X^\pi = xR = \sum \psi_i P_i(T).$$

The expect value and the variance of total return for the portfolio π are given by

$$E(R) = \sum \pi_i \mu_i = \mu(\pi)$$
$$Var(R) = \sum \pi_i \sigma_{ij} \pi_j = \sigma^2(\pi).$$

The portfolio constraint U has to be specified by the investor such that $\pi \in U$. For example,

- $U = R^d$ is no constraint;
- $U = \{(\pi_1, \cdots, \pi_d); \pi_i \geq 0, i = 1, \cdots, d\}$ is non-short-selling constraint.

If $\pi \in U$, then π is called admissible.

We have the mean-variance principle following the work of Markowitz:

- For a given upper bound σ^2 for the variance of the portfolio return, choose an admissible portfolio π such that $\mu(\pi)$ is maximal.
- For a given lower bound μ for the mean of the portfolio return choose an admissible portfolio π such that $\sigma^2(\pi)$ is minimal.

The mathematic formulation for the first is

$$\max\{\mu(\pi); \sigma^2(\pi) \leq \sigma^2, \pi \in U\}. \tag{1}$$

The second becomes

$$\min\{\sigma^2(\pi); \mu(\pi) \geq \mu, \pi \in U\}. \tag{2}$$

For the case of non-short-selling, (2) is a quadratic programming problem which the algorithm exists to find the solution. (1) can not be directly solved, but an iterative scheme can be used to get the solution.

The following is an alternative formulation,

$$\max\{\mu(\pi) - \alpha\sigma^2(\pi); \pi \in U\}. \tag{3}$$

For large α, the investor aims at choosing admissible portfolio such that the variance of the portfolio return is small and the mean is large. For small α, he is willing to tolerate large uncertainty. We can also consider

$$\max\{E(u(R)); (\pi) \in U\},$$

where $u(\cdot)$ is an utility function (a concave and increasing function).

Consider one particular case,

$$u(R) = \frac{1}{\gamma}\exp(\gamma R),$$

$\gamma < 0$. For $-\gamma$ small, then

$$\log E(\exp(\gamma R)) \sim \gamma\mu(\pi) + \frac{1}{2}\gamma^2\sigma^2(\pi).$$

That is,

$$\frac{1}{\gamma}\log E(\exp(\gamma R)) \sim \mu(\pi) + \frac{1}{2}\gamma\sigma^2(\pi)$$

which is closely related to (3) with $\alpha = -\frac{1}{2}\gamma$. Another example which is frequently use is HARA,

$$u(R) = \frac{1}{\gamma}R^\gamma$$

$\gamma < 1$.

It is more realistic to consider multi-period model in which the investor is allowed to change his holding of assets at some given trading dates $t = 0, 1, \cdots, T-1$ before the time $t = T$. He needs to choose a portfolio process $\pi(t), t = 0, 1, \cdots, T-1$; $\pi(t)$ has to be decided before the prices of the securities at later trading dates are revealed. And also we assume that the investor behaves in a self-financing way, that is , there is no money flows in or out. For each $\{\pi(t); t = 0, 1, \cdots, T-1\}$, in principle we can calculate

$$R^\pi(T) = X^\pi(T)/x$$

the portfolio return and solve the problems (1)(2). This is an impossible task: the relation between $\mu(\pi(t), t = 0, 1, \cdots, T-1), \sigma^2(\pi(t), t = 0, 1, \cdots, T-1)$ and the portfolio process becomes very complicate. It is tempting to solve the problem step by step starting from the time $t = 0$. This resembles the local search for optimum could lead to a very poor performance. Under the Markovian structure of the price dynamics, indeed, we can solve the problem step by step, however backward in time. This is the dynamic programming which is widely used in control theory. The idea is simple can be illustrated as follows.

3. Dynamic programming:discrete model

Assume there are only two assets, one is riskless and another one is risky. The riskless asset has a constant return and is given by r. That is, denote $P_0(t)$ the prices of the riskless asset at time t, then

$$P_0(t+1) = rP_0(t), t = 0, 1, \cdots.$$

The price of the risky asset is binomial. That is, denote $P_1(t)$ the price of the risky asset at time t, then

$$P_1(t+1) = \begin{cases} aP_1(t) & \text{with probability } p \\ bP_1(t) & \text{with probability } 1-p \end{cases}$$

for some $a, b > 0$. We assume that $P_1(t+1)/P(t), t = 0, 1, \cdots$ are independent. The portfolio process is given by $\mathbf{\Pi} = \{\pi(t) = (\pi_0(t), \pi_1(t)), t = 0, 1, \cdots, T-1\}$ such that $\pi(t)$ only depends on $P_1(0), P_1(1), \cdots, P_1(t)$. Let x be the initial wealth and $X^{\mathbf{\Pi}}(t)$ be the wealth at time t by choosing the portfolio process $\mathbf{\Pi}$. The goal is to optimize

$$\max\{E(u(X^{\mathbf{\Pi}}(T)); \pi(t) \in U, X^{\mathbf{\Pi}}(0) = x, \mathbf{\Pi} \text{ is self-financing }\}.$$

For $t \leq T, x > 0$ and $P \in \{a^n b^{t-n} P_1(0), n = 0, \cdots, t\}$, denote

$$\begin{aligned} v(t,x,P) = \quad & \max\{E(u(X^{\mathbf{\Pi}}(T)|P_1(t) = P); \pi(s) \in U \text{ for all } s \ , \\ & X^{\mathbf{\Pi}}(t) = x, \mathbf{\Pi} \text{ is self-financing }\}. \end{aligned} \tag{4}$$

Then intuitively, for $t < t + s \leq T$, the following relation holds,

$$\begin{aligned} v(t,x,P) = \quad & \max\{E(v(t+s, X^{\mathbf{\Pi}}(t+s), P_1(t+s))|P_1(t) = P); \\ & \pi(n) \in U \text{ for all } n \ , X^{\mathbf{\Pi}}(t) = x, \mathbf{\Pi} \text{ is self-financing }\}. \end{aligned} \tag{5}$$

Indeed, we can prove that if $v(t,x,P)$ satisfies (5), then (4) holds. This is referred as Verification Theorem in control theory. Moreover, (5) also suggests an algorithm to calculate $v(t,x,P)$: First, v(T, x, P)= u(x); knowing $v(t+1, \cdot, \cdot)$, the relation (5) with $s = 1$ is a portfolio problem in one-period model, which can be solved without much difficulty. This gives an iterative schemes to solve (5). The algorithm is implemented backward in time. It is called dynamic programming approach. Let consider a particular case such that we have an analytic solution: $U = R^2$ and $u(\cdot)$ is HARA, $u(x) = \frac{1}{\gamma}x^\gamma$. Then we have the following result.

Theorem 1. Assume that $a < r < b$ and $U = R^2$ and $u(\cdot)$ is HARA and given by $u(x) = \frac{1}{\gamma}x^\gamma$. Then

$$v(t,x,P) = \frac{1}{\gamma}x^\gamma c^{T-t}$$

for $t = 0, 1, \cdots, T$, $x > 0$ and $P \in \{a^n b^{t-n} P_1(0), n = 0, \cdots, t\}$. Here c is given by

$$c = \max\{(r + (a-r)\pi)^\gamma p + (r + (b-r)\pi)^\gamma (1-p); \pi\} \tag{6}$$

if $0 < \gamma < 1$ and

$$c = \min\{(r + (a-r)\pi)^\gamma p + (r + (b-r)\pi)^\gamma (1-p); \pi\} \tag{7}$$

if $\gamma < 0$. The optimal portfolio is given by $\pi_0^*(t) = \pi_0^*$ and $\pi_1^*(t) = \pi_1^*$, $\pi_0^* = 1 - \pi_1^*$ which is time independent and π_1^* is equal to the *arg max* in (6)(and *arg min* in (7)).

Proof: Let $t = T-1$. Then

$$\begin{aligned} X^{\mathbf{\Pi}}(T) = & \ \frac{\pi_0(T-1)x}{P_0(T-1)}P_0(T) + \frac{\pi_1(T-1)x}{P_1(T-1)}P_1(T) \\ = & \begin{cases} r\pi_0(T-1)x + a\pi_1(T-1)x \text{ with probability } p \\ r\pi_0(T-1)x + b\pi_1(T-1)x \text{ with probability } 1-p \end{cases} \end{aligned}$$

If $X^{\mathbf{\Pi}}(T-1) = x$, then

$$E(u(X^{\mathbf{\Pi}}(T))|P_1(T-1) = P) = x^\gamma(r + (a-r)\pi)^\gamma p + (r + (b-r)\pi)^\gamma(1-p).$$

Therefore, it is easy to see

$$v(T-1, x, P) = \frac{1}{\gamma}x^\gamma c.$$

By induction, we can prove the result.

4. Continuous time model

In the following, we consider a continuous time model and discuss the dynamic programming approach for the portfolio problem.

We assume that there are two assets, one is riskless and one is risky. The price of the riskless asset is given by

$$P_0(t) = P_0(0)\exp(rt);$$

the price of risky asset is given by

$$P_1(t) = P_1(0)\exp(\alpha t + \sigma w(t)),$$

where $w(t)$ is the one dimensional standard Brownian motion. Then $P_1(t)$ satisfies

$$dP_1(t) = P_1(t)(bdt + \sigma dw(t))$$

where $b = \alpha + 1/2\sigma^2$.

Let $\mathcal{F}_t$ be the σ-algebra generated by $w(s), s \le t$. The portfolio process $\mathbf{\Pi} = \{(\pi_0(t), \pi_1(t)); 0 \le t \le T\}$ is adapted to the filtration $\{\mathcal{F}_t\}$. We assume $(\pi_0(t), \pi_1(t)) \in U$, $U \subset R^2$ is the investment constraint. For example,

- $U = R^2$(no investment constraint);
- $U = \{(\pi_0, \pi_1); \pi_0 \ge 0, \pi_1 \ge 0\}$(non-short-selling constraint) or
- $U = \{(\pi_0, \pi_1); a_0 \le \pi_0 \le A_0, a_1 \le \pi_1 \le A_1\}$.

Let $X^{\mathbf{\Pi}}(t)$ be the wealth process by choosing the investment policy $\mathbf{\Pi}$. We assume the investor is self-financing and there is no transaction cost, this means,

$$\begin{aligned} dX^{\mathbf{\Pi}}(t) &= \frac{\pi_0(t)X^{\mathbf{\Pi}}(t)}{P_0(t)}dP_0(t) + \frac{\pi_1(t)X^{\mathbf{\Pi}}(t)}{P_1(t)}dP_1(t) \\ &= X^{\mathbf{\Pi}}(t)(r\pi_0(t)dt + \pi_1(t)(bdt + \sigma dw(t))) \\ &= X^{\mathbf{\Pi}}(t)((r + \pi_1(t)(b-r))dt + \pi_1(t)\sigma dw(t)). \end{aligned} \tag{8}$$

Here $b = \alpha + \frac{1}{2}\sigma^2$. The goal is to maximize the following

$$\max\{E(u(X^{\mathbf{\Pi}}(T)); (\pi_0(t), \pi_1(t)) \in U, \\ \mathbf{\Pi} \text{ is progressive measurable with respect to } \{\mathcal{F}_t\}\ \}.$$

Parallel to the multi-period model, we describe the dynamic programming approach as follows. Denote

$$v(t,x,P) = \max\{E(u(X^{\mathbf{\Pi}}(T))|P_1(t) = P, X^{\mathbf{\Pi}}(t) = x); (\pi_0(t), \pi_1(t)) \in U \\ , \mathbf{\Pi} \text{ is progressive measurable with respect to } \{\mathcal{F}_t\}\ \}.$$

Then we have the dynamic programming principle:
(9)

$$v(t,x,P) = \max\{E(v(t+s, X^{\mathbf{\Pi}}(t+s), P_1(t+s))|P_1(t) = P, X^{\mathbf{\Pi}}(t) = x); \\ (\pi_0(a), \pi_1(a)) \in U, t \le a \le t+s, \ \mathbf{\Pi} \text{ is progressive measurable w.r.t } \{\mathcal{F}_t\}\ \}.$$

for $0 \le t < t+s \le T$. Different from discrete model, we can not solve (9) step by step. We need to derive an infinitesimal relation for $v(\cdot,\cdot,\cdot)$. This can be done formally as follows. (9) implies

$$v(t,x,P) \ge E(v(t+s, X^{\mathbf{\Pi}}(t+\delta), P_1(t+\delta))|P_1(t) = P, X^{\mathbf{\Pi}}(t) = x) \tag{10}$$

for any *admissible* $\mathbf{\Pi}$. Assume $v(\cdot,\cdot,\cdot)$ is smooth. Then by Ito's rule for $t \le s \le t+\delta$,

$$dv(s, X^{\mathbf{\Pi}}(s), P_1(s)) = (\tfrac{\partial}{\partial t} + L^{\pi_1(s)})v(s, X^{\mathbf{\Pi}}(s), P_1(s))ds \\ +(\tfrac{\partial v}{\partial x} X^{\mathbf{\Pi}}(s)\pi_1(s) + \tfrac{\partial v}{\partial P} P_1(s))\sigma dw(s),$$

where

$$L^{\pi} f(x,P) = \tfrac{1}{2}\sigma^2 x^2 \pi^2 \tfrac{\partial^2 f}{\partial x^2}(x,P) + \sigma^2 \pi x P \tfrac{\partial^2 f}{\partial x \partial P}(x,P) + \tfrac{1}{2}\sigma^2 P^2 \tfrac{\partial^2 f}{\partial P^2}(x,P) \\ +(r + (b-r)\pi)\tfrac{\partial f}{\partial x}(x,P) + bP\tfrac{\partial f}{\partial P}(x,P).$$

Then (10) implies

$$\int_t^{t+\delta} E((\frac{\partial}{\partial t} + L^{\pi_1(s)})v(s, X^{\mathbf{\Pi}}(s), P_1(s))|P_1(t) = P, X^{\mathbf{\Pi}}(t) = x)ds \le 0.$$

Divide this by δ and let δ tends to 0, then we have

$$\frac{\partial v}{\partial t}(t,x,P) + L^{\pi} v(t,x,P) \le 0$$

for all π such that $(1-\pi, \pi) \in U$. The existence of an optimal portfolio implies

$$\frac{\partial v}{\partial t}(t,x,P) + L^{\pi} v(t,x,P) = 0$$

for some π. Therefore,

$$\frac{\partial v}{\partial t}(t,x,P) + \max\{L^{\pi} v(t,x,P); (1-\pi,\pi) \in U\} = 0. \tag{11}$$

This is the dynamic programming equation we want to solve. However, this a very compicated nonlinear partial differential equation and is not easy to solve. We present an particular example that there exists analytic solution. We assume $U = R^2$ and $u(x) = \frac{1}{\gamma}x^\gamma$.

Theorem 2. Assume that $U = R^2$ and $u(x) = \frac{1}{\gamma}x^\gamma$. Then,

$$v(t,x,P) = \frac{1}{\gamma}x^\gamma v(t) \tag{12}$$

where

$$v(t) = \exp(\gamma(r - \frac{(b-r)^2}{2(\gamma-1)\sigma^2 1})(T-t)).$$

An optimal portfolio process is given by

$$\pi_0^*(t) = 1 - \pi_1^*,\ \pi_1^*(t) = \pi_1^* = -\frac{b-r}{\sigma^2(1-\gamma)}.$$

Proof: This can be verified by plugging (12) in the equation (11).

It is in general difficult to solve the equation (11) analytically. See more examples in the following papers where analytic solution can be obtained or properties of the solution can be described,

- Fleming, W. H. and Sheu, S. J. (1999) Optimal long term growth rate of expected utility of wealth, Ann. Appl. Probab.
- Fleming, W. H. and Sheu, S. J. (2000) Risk sensitive control and an optimal investment model, Math. Finance.
- Bielecki, T. R. and Pliska, S. R. (1999) Risk sensitive dynamic asset management, Appl. math. Optim**39**, 337-360.
- Bielecki, T. R., Pliska, S. R. and Sherris, M. (1999) Risk sensitive asset allocation, J. Economic Dynamics and Control.

The result in Theorem 2 says that the investor has to rebalance the portfolio all the time if the optimal portfolio is applied to keep $(\pi_0^*(t), \pi_1^*(t))$ constant. When there is a transaction cost it is very expansive to do the rebalancing all the time. In the following, we shall consider the same model with transaction cost. Although we will not be able to present the full details of the analysis, we will see how dynamic programming approach can give some useful information of the optimal portfolio.

We assume the transaction costs are proportional to the amound traded. The investor has to pay the fraction $\lambda > 0$ of the value of the amount of risky asset purchased and the fraction of $\mu > 0$ of the value of the risky asset sold. Let L_t, U_t be the cumulative amount

of the purchase and sale of the risky asset up to time t and $B(t), S(t)$ are the amount of the wealth in riskless asset (bank account) and risky asset. Then

$$dB(t) = rB(t)dt - (1+\lambda)dL_t + (1-\mu)dU_t,\ B(0) = x,$$
$$dS(t) = S(t)[bdt + \sigma dw(t)] + dL_t - dU_t.$$

Let $\mathcal{P} = \{(L_t, U_t); 0 \le t \le T\}$ which is decided by the investor. The wealth $X^{\mathcal{P}}(t)$ at time t is given by

$$X^{\mathcal{P}}(t) = \begin{cases} B(t) + (1-\mu)S(t), \text{ if } S(t) > 0, \\ B(t) + (1+\lambda)S(t), \text{ if } S(t) < 0. \end{cases}$$

A $\mathcal{P}$ is called admissible if $X^{\mathcal{P}}(t) > 0$ for all $t \le T$ and (L_t, U_t) is progressively measurable w.r.t $\{\mathcal{F}_t\}$. The goal is to maximize

$$\max\{E(u(X^{\mathcal{P}}(T))); \mathcal{P} \text{ is admissible }\}.$$

To apply dynamic programming approach, we denote

$$v(t,x,y) = \max\{E(u(X^{\mathcal{P}}(T)))|B(t) = x, S(t) = y); \mathcal{P} \text{ is admissible }\}.$$

Again, formally we have dynamic programming principle,

$$v(t,x,y) = \max\{E(v(t+s, B(t+s), S(t+s))|B(t) = x, S(t) = y); \mathcal{P} \text{ is admissible }\}$$

for $0 \le t < t+s \le T$. We need to derive an equation for $v(t,x,y)$. For this, we temporarily assume

$$L_t = \int_0^t \ell(t)dt,\ U_t = \int_0^t u(t)dt$$

such that

$$0 \le \ell(t) \le M,\ 0 \le u(t) \le M.$$

We will let M tends to ∞ eventually. By assuming the smoothness of $v(\cdot,\cdot,\cdot)$ as before we can derive the following equation,

$$v_t + \max\{\tfrac{1}{2}\sigma^2 y^2 v_{yy} + byv_y + \ell(-(1+\lambda)v_x + v_y) + u((1-\mu)v_x - v_y); 0 \le \ell \le M, 0 \le u \le M\} = 0. \tag{13}$$

The *arg max*, denoted by $\ell^*(t,x,y), u^*(t,x,y)$, gives a candidate of optimal portfolio $\ell^*(t) = \ell^*(t, B(t), S(t)), u^*(t) = \ell^*(t, B(t), S(t))$, where

$$dB(t) = (rB(t) - (1+\lambda)\ell^*(t, B(t), S(t)) + (1-\mu)u^*(t, B(t), S(t)))dt,\ B(0) = x,$$
$$dS(t) = S(t)[bdt + \sigma dw(t)] + (\ell^*(t, B(t), S(t)) - u^*(t, B(t), S(t)))dt,\ S(0) = y.$$

From (13), we have

$$\ell^*(t,x,y) = \begin{cases} M, & \text{if } -(1+\lambda)v_x + v_y > 0\ , \\ 0, & \text{if } -(1+\lambda)v_x + v_y < 0 \end{cases}$$

and

$$u^*(t,x,y) = \begin{cases} M, & \text{if } (1-\mu)v_x - v_y > 0 \ , \\ 0, & \text{if } (1-\mu)v_x - v_y < 0 \end{cases}$$

Therefore, we see that there is a region $\mathcal{NT}$,

$$\mathcal{NT} = \{(t,x,y); -(1+\lambda)v_x + v_y < 0, 1-\mu)v_x - v_y < 0\},$$

where there is no trading; in the region $\mathcal{S}$,

$$\mathcal{S} = \{(t,x,y); (1-\mu)v_x - v_y > 0\},$$

where he sells the risky asset as much as possible to bring the holding to the region $\mathcal{NT}$; in the region $\mathcal{B}$,

$$\mathcal{B} = \{(t,x,y); -(1+\lambda)v_x + v_y > 0\},$$

where he buys the risky asset as much as possible to bring the holding to the region $\mathcal{NT}$. These three regions should be disjoint.

The material discussed here is taken from

Ralf Korn(1997) Optimal Portfolio, World Scientific, Singapore.