M.I.T. **15.450-Fall 2010**

Sloan School of Management **Professor Leonid Kogan**

Dynamic Programming: Justification of the Principle of Optimality

This handout adds more details to the lecture notes on the Principle of Optimality. As in the notes, we are dealing with the case of IID return distribution.

Recall that the portfolio policy ϕ_t is a flexible function of the entire observed history up to time t. We suppress this dependence in our notation. We will use notation

$$\mathrm{E}_t(U(W_T)|\phi_{t,\ldots,T-1})$$

to denote the time-t conditional expectation of terminal utility under the portfolio policy $(\phi_t, \phi_{t+1}, \ldots, \phi_{T-1})$.

Recall the definition of the value function

$$\mathrm{E}_t\left[U(W_T)|_{(t)}\phi^{\star}_{t,\ldots,T-1}\right] = J(t, W_t) \tag{1}$$

We assume it has been established that at t, $t+1$, ..., T, the value function depends only on W, and we verify below that the same is true at $t-1$.

By Law of Iterated Expectations, for any portfolio policy $\phi_{t-1,\ldots,T-1}$,

$$\mathrm{E}_{t-1}\left[U(W_T)|\phi_{t-1,\ldots,T-1}\right] = \mathrm{E}_{t-1}\left[\mathrm{E}_t\left[U(W_T)|\phi_{t,\ldots,T-1}\right]|\phi_{t-1}\right] \tag{2}$$

The expected utility achieved by any portfolio policy cannot exceed the expected utility under the optimal policy, and therefore

$$\mathrm{E}_t\left[U(W_T)|\phi_{t,\ldots,T-1}\right] \leq \mathrm{E}_t\left[U(W_T)|_{(t)}\phi^{\star}_{t,\ldots,T-1}\right] = J(t, W_t) \tag{3}$$

for any policy $\phi_{t-1,\ldots,T-1}$. We have used the definition of the value function in the last equality above.

Substitute the upper bound from (3) into the right hand side of (2). We are replacing $\mathrm{E}_t\left[U(W_T)|\phi_{t,\ldots,T-1}\right]$ with something at least as large, $J(t, W_t)$, to obtain an inequality:

$$\mathrm{E}_{t-1}\left[U(W_T)|\phi_{t-1,\ldots,T-1}\right] \leq \mathrm{E}_{t-1}\left[J(t, W_t)|\phi_{t-1}\right], \tag{4}$$

Now, maximize the right hand side of (4) with respect to ϕ_{t-1}. This clearly produces a valid upper bound on expected utility:

$$\mathrm{E}_{t-1}\left[U(W_T)|\phi_{t-1,\dots,T-1}\right] \le \mathrm{E}_{t-1}\left[J(t,W_t)|\phi_{t-1}\right] \le \max_{\widetilde{\phi}_{t-1}} \mathrm{E}_{t-1}\left[J(t,W_t)|\widetilde{\phi}_{t-1}\right] \tag{5}$$

Let's denote the optimal choice of $\widetilde{\phi}_{t-1}$ in (5) by ${}^{J}\phi^{\star}_{t-1}$:

$$\max_{\widetilde{\phi}_{t-1}} \mathrm{E}_{t-1}\left[J(t,W_t)|\widetilde{\phi}_{t-1}\right] = \mathrm{E}_{t-1}\left[J(t,W_t)|{}^{J}\phi^{\star}_{t-1}\right]$$

The upper bound on the expected terminal utility in (5) is achievable. Consider the policy

$${}^{(J)}\phi^{\star}_{t-1},\ {}_{(t)}\phi^{\star}_{t},\ \dots,\ {}_{(t)}\phi^{\star}_{T-1}.$$

Under this policy, according to (5),

$$\begin{aligned}
&\mathrm{E}_{t-1}\left[U(W_T)|{}^{(J)}\phi^{\star}_{t-1}, {}_{(t)}\phi^{\star}_{t}, \dots, {}_{(t)}\phi^{\star}_{T-1}\right] \overset{\text{Iterated Expectations}}{=} \\
&\mathrm{E}_{t-1}\left[\mathrm{E}_t\left[U(W_T)|{}_{(t)}\phi^{\star}_{t}, \dots, {}_{(t)}\phi^{\star}_{T-1}\right]|{}^{(J)}\phi^{\star}_{t-1}\right] \overset{(1)}{=} \mathrm{E}_{t-1}\left[J(t,W_t)|{}^{(J)}\phi^{\star}_{t-1}\right] \overset{(5)}{\ge} \\
&\mathrm{E}_{t-1}\left[U(W_T)|\phi_{t-1,\dots,T-1}\right], \quad \forall \phi_{t-1,\dots,T-1}
\end{aligned} \tag{6}$$

Again, we have used the definition of the value function.

Thus, we have shown that the policy

$${}^{(J)}\phi^{\star}_{t-1},\ {}_{(t)}\phi^{\star}_{t},\ \dots,\ {}_{(t)}\phi^{\star}_{T-1}$$

is optimal at time $t-1$, and it generates the time-$(t-1)$ value function, which equals

$$J(t-1,W_{t-1}) = \max_{\widetilde{\phi}_{t-1}} \mathrm{E}_{t-1}\left[J(t,W_t)|\widetilde{\phi}_{t-1}\right]$$

The above equation is the Bellman equation. Moreover, by inspecting the optimal policy, we see that

$${}_{(t-1)}\phi^{\star}_{s} = {}_{(t)}\phi^{\star}_{s}, \quad s = t, t+1, \dots, T-1$$

and

$${}_{(t-1)}\phi^{\star}_{t-1} = {}^{(J)}\phi^{\star}_{t-1} = \arg\max_{\widetilde{\phi}_{t-1}} \mathrm{E}_{t-1}\left[J(t,W_t)|\widetilde{\phi}_{t-1}\right]$$

The optimal portfolio policy is time-consistent, and can be computed from the Bellman equation.

Note that, as a part of our argument, we have shown that the value function indeed depends only on time and portfolio value:

$$J(t-1, W_{t-1}) = \max_{\widetilde{\phi}_{t-1}} \mathrm{E}_{t-1}\left[J(t, W_t)|\widetilde{\phi}_{t-1}\right]$$

The above equality is meaningful because returns are IID, and therefore $\max_{\widetilde{\phi}_{t-1}} \mathrm{E}_{t-1}\left[J(t, W_t)|\widetilde{\phi}_{t-1}\right]$ depends on the past history only through W_{t-1}. A related important implication is that the optimal control policy, $\phi^{\star}_{t-1}$, also depends on the past history only through W_{t-1}. We call such $\phi^{\star}$ a feedback control policy.

15.450 Analytics of Finance
Fall 2010