# Example I: Capital Accumulation

Time $t = 0, 1, \ldots, T < \infty$

Output $y$, initial output $y_0$

Fraction of output invested $a$, capital $k = ay$

Transition (production function) $y' = g(k) = g(ay)$

Reward (utility of consumption) $u(c) = u((1-a)y)$

Discount factor $\beta$

Action rules $a_t = f_t(y_t)$

Policy $\pi = (f_0, f_1, \ldots, f_T)$

State transitions: $(y_0, a_0 = f_0(y_0)), (y_1 = g(a_0 y_0), a_1 = f_1(y_1)), \ldots$

Value of the policy $\pi$ is

$$V_\pi(y_0) = \sum_{t=0}^{T} \beta^t u((1 - f_t(y_t))y_t)$$

where $y_{t+1} = g(f_t(y_t)y_t)$

An *Optimal Policy* maximizes $V$ over $\pi$.

# Example II: Savings and Portfolio Choice

Time $t = 0, 1, \ldots, T < \infty$

Wealth $w$, initial wealth $w_0$

Fraction consumed $\alpha$, consumption $c = \alpha w$

Fraction saved $1 - \alpha$, savings $s = (1 - \alpha)w$

Fraction of savings invested in risky asset $\gamma$

Fraction of savings invested in risk free asset $1 - \gamma$

Gross rate of return on risk free asset is $R$

Gross rate of return on risky asset is $x$ with probability $p$ or $y$ with probability $1 - p$, i.i.d.

Reward $u(c)$ and discount factor $\beta$

Transitions:

$$
\begin{aligned}
& w_{t+1} \\
& \quad = x\gamma_t(1-\alpha_t)w_t + R(1-\gamma_t)(1-\alpha_t)w_t \\
& \qquad \text{with prob } p \\
& \quad = y\gamma_t(1-\alpha_t)w_t + R(1-\gamma_t)(1-\alpha_t)w_t \\
& \qquad \text{with prob } 1-p
\end{aligned}
$$

Action rule $a_t = (\alpha_t, \gamma_t) = f_t(w_t)$

Policy $\pi = (f_0, f_1, \ldots, f_T)$

Value of $\pi$ given $w_0$ is

$$V_\pi(w_0) = E_\pi[\sum_{t=0}^{T} \beta^t u(c_t)]$$

## Finite Horizon, State and Action

Time: $t = 0, 1, \ldots, T < \infty$

States: $S$ a finite set

Actions: $A$ a finite set

$P(a, s)(s')$ the probability of $s'$ given $(a, s)$

Transitions: For each action $a$, $P(a, \cdot)(\cdot)$ is an $S \times S$ matrix. A *transition probability.*

Reward: $u : A \times S \to \mathbf{R}$

Discount factor $\beta$

Action Rule: f:S$\to A$

Policy: $\pi = (f_0, f_1, \ldots, f_T)$.

## Stochastic Process on States

Fix a policy $\pi$.

$Q_t(s)(s') = P(f_t(s), s)(s')$ is the probability of moving from state $s$ at date $t$ to state $s'$ at date $t+1$.

Let $P_t(s, s')$ denote the probability of $s_t = s'$ given that $s_0 = s$.

- $P_0(s, s')$ is 1 if $s' = s$ and 0 otherwise.
- $P_1(s, s') = Q_0(s)(s')$
- $P_2(s, s') = \sum_{x=1}^{S} P_1(s, x)Q_1(x)(s')$ or $P_2 = P_1Q_1$
- Generally, $P_t(s, s') = \sum_{x=1}^{S} P_{t-1}(s, x)Q_{t-1}(x)(s')$ or $P_t = P_{t-1}Q_{t-1}$.

Thus, $P_t = Q_0Q_1 \cdots Q_{t-1}$, for $t > 1$.

# Value

If we begin at state $s_0 = s$ and follow policy $\pi$ how much total expected return do we earn?

At date $t$ earn $\beta^t U_t(s_t) = \beta^t u(f_t(s_t), s_t)$, if in state $s_t$.

So the expected (as of date 0) return at date $t$ is $\sum_{s'} \beta^t P_t(s, s') U_t(s')$ or $\beta^t P_t U_t$.

Let $V_\pi^T(s)$ denote the total expected return for our $T$ period problem from following policy $\pi$ if we start in state $s$

$$V_\pi^T(s) = \sum_{t=0}^{T-1} \sum_{s'} \beta^t P_t(s, s') U_t(s')$$

# Optimality

A policy $\pi$ is *Optimal* given initial state $s$ if

$$V_\pi^T(s) \geq V_{\pi'}^T(s)$$

for all policies $\pi'$.

There are only finitely many policies so there is an optimal policy. How do we find it?

**One period problem, $T = 0$:**

Clearly, choose an action to maximize $u(a, s)$ for initial state $s$. Let $f_0^*(s)$ be this action. So $\pi^* = (f_0^*)$ is an optimal policy for the one period problem. The value of this problem is

$$V^{*1}(s) = \max_a u(a, s) = u(f_0^*(s), s)$$

## Two period problem, $T = 1$:

For any policy $\pi = (f_0, f_1)$

$$V_\pi^2(s) = u(f_0(s), s) + \sum_{s'} \beta u(f_1(s'), s') P(f_0(s), s)(s')$$

Clearly choose $f_1(s')$ to maximize $u(f_1(s'), s')$. This is the optimal policy for a one period problem. So choose $f_0(s)$ to maximize

$$u(f_0(s), s) + \sum_{s'} \beta V^{*1}(s') P(f_0(s), s)(s')$$

The value of the two period problem is

$$V^{*2}(s) = \max_{f_0(s)} [u(f_0(s), s) + \sum_{s'} \beta V^{*1}(s') P(f_0(s), s)(s')]$$

This defines the optimal two period policy (and it is optimal for all initial states $s$).

# Optimality Principle

For finite horizon, finite action, finite state problems:

- The value of the problem is given by

$$V^{*T+1}(s) = \max_a [u(a,s) + \sum_{s'} \beta V^{*T}(s') P(a,s)(s')]$$

- There is an optimal policy $\pi^* = (f_0^*, f_1^*, \cdots, f_T^*)$.
- $(f_1^*, \cdots, f_T^*)$ is optimal for the $T$ period problem beginning in period 1 at any state
- $f_0^*$ solves

$$V^{*T+1}(s) = u(f_0^*(s), s) + \sum_{s'} \beta V^{*T}(s') P(f_0^*(s), s)(s')$$

for all $s$.

# Capital Accumulation $T = 1$

Since $f_1$ is optimal for the final period there is no investment in that period, $f_1(y_1) = 0$ and $V^{*1}(y_1) = u(y_1)$. So $f_0(y_0)$ maximizes

$$u((1 - f_0(y_0))y_0) + \beta u(g(f_0(y_0)y_0))$$

Let $f^*(y_0)$ be the optimum. Then

$$V^{*2}(y_0) = u((1 - f_0^*(y_0))y_0) + \beta u(g(f_0^*(y_0)y_0))$$

Since $f_0^*(y_0)$ is an optimum any deviation must reduce the value of the problem. So the following expression is maximized at $\epsilon = 0$ :

$$u((1 - f_0^*(y_0))y_0 - \epsilon) + \beta u(g(f_0^*(y_0)y_0 + \epsilon))$$

Using derivatives to give an approximation to an optimum (I will ignore corner conditions), we have

$$-u'(c_0) + \beta u'(c_1)g'(k_1) = 0$$

Calculation shows that

$$V^{*2'}(y_0) = u'((c_0))$$

# Capital Accumulation T=2

Let $f_0^*(y_0)$ be the optimal first period investment in the $T = 2$ problem. Then

$$V^{*3}(y_0) = u((1 - f_0^*(y_0))y_0) + \beta V^{*2}(g(f_0^*(y_0)y_0))$$

Considering a deviation from $f_0^*(y_0)$ we have

$$-u'(c_0) + \beta V^{*2'}(y_1)g'(k_1) = 0$$
$$-u'(c_0) + \beta u'(c_1)g'(k_1) = 0$$

Alternatively, suppose that we are on an optimal path $f_0^*(y_0), f_1^*(y_1)$, and $f_2^*(y_2) = 0$. Then any deviation must reduce the value of the problem. So the following expression is maximized at $\epsilon = 0$

$$u((1 - f_0^*(y_0))y_0 - \epsilon)$$
$$+\beta u\Big(g((1 - f_1^*(y_1))y_1) + g(f_0^*(y_0)y_0 + \epsilon) - g(f_0^*(y_0)y_0)\Big)$$
$$+\beta^2 u(g(f_1^*(y_1)y_1))$$

Differentiating with respect to $\epsilon$ yields

$$-u'(c_0) + \beta u'(c_1) g'(k_1) = 0$$

Note that this is also result we obtained when we considered deviations in the value function approach.

For any $T$ period problem along an optimal path we have

$$-u'(c_t) + \beta u'(c_{t+1}) g'(k_t) = 0$$

## Savings and Portfolio Choice

The deviations from an optimal policy approach also works with uncertainty. The following expression is maximized at deviations $\epsilon_1 = 0, \epsilon_2 = 0$:

$$V^{*T}(w_0) = \cdots + u(\alpha_t w_t - \epsilon_1 - \epsilon_2)$$

$$+\beta[pu\Big(x\gamma_t(1-\alpha_t)w_t + R(1-\gamma_t)(1-\alpha_t)w_t + x\epsilon_1 + R\epsilon_2\Big)$$

$$+(1-p)u\Big(y\gamma_t(1-\alpha_t)w_t + R(1-\gamma_t)(1-\alpha_t)w_t$$

$$+y\epsilon_1 + R\epsilon_2\Big)] + \cdots$$

# Infinite Horizon

We will maintain the assumptions of finite action and state spaces.

A policy is an infinite sequence $\pi = (f_0, f_1, \dots)$. For any truncated horizon $T$ the value of the policy $\pi$ is

$$V_\pi^T(s) = \sum_{t=0}^{T-1} \sum_{s'} \beta^t P_t(s, s') U_t(s')$$

What happens as we let $T \to \infty$?

Assume $0 \leq \beta < 1$. Then the limit exists and we call it $V_\pi$.

- What can we say about $V_\pi$?
- Does an optimal policy exist?
- Is it stationary?
- Can we characterize it?

# Operators

For $V : S \to \mathbf{R}$ define the operator $W$ by

$$(WV)(s) = \max_a [u(a, s) + \beta \sum_{s'} V(s')P(a, s)(s')]$$

The finite horizon optimality principle is that

$$V^{*T+1} = WV^{*T}$$

Applying this repeatedly we get ($V^{*0}(s) = 0$ for all $s$)

$$V^{*T+1} = W^{T+1}V^{*0}$$

where $W^n$ is $W$ iterated $n$-times.

For any $V : S \to \mathbf{R}$ define $||V|| = \max |V(s)|$.

**Contraction Mapping Theorem:** For any $V : S \to \mathbf{R}$ and $\hat{V} : S \to \mathbf{R}$,

$$||WV - W\hat{V}|| \leq \beta ||V - \hat{V}||$$

Fix a state $s$ and let $a$ solve the optimization problem in $WV$.

$$
\begin{aligned}
& WV(s) - W\hat{V}(s) \\
& \leq WV(s) - [u(a,s) + \beta \sum_{s'} \hat{V}(s') P(a,s)(s')] \\
& = \beta \sum_{s'} P(a,s)(s')(V(s') - \hat{V}(s')) \\
& \leq \beta \sum_{s'} P(a,s)(s')|V(s') - \hat{V}(s')| \\
& \leq \beta \sum_{s'} P(a,s)(s')||V - \hat{V}|| \\
& = \beta ||V - \hat{V}||
\end{aligned}
$$

Reversing the roles of $V$ and $\hat{V}$ we have

$$W\hat{V}(s) - WV(s) \leq \beta ||\hat{V} - V||$$

This holds for all $s$ so we have the contraction result.

# Relation between finite and infinite horizon values

Let $C$ be the bound on the reward function.

**Claim 1:** $||V_\pi - V_\pi^T|| \leq \beta^T C/(1-\beta)$.

Why?

$V_\pi(s) = V_\pi^T(s) + \beta^T \sum P_T(s,s')V_\pi(s')$.

So $|V_\pi - V_\pi^T| = \beta^T \sum P_T(s,s')V_\pi(s') \leq \beta^T C/(1-\beta)$.

So for any policy $\pi$ its finite horizon value converges to its infinite horizon value.

Infinite horizon values are bounded by $C/(1-\beta)$ so the infinite horizon value function is well defined:

$$V^*(s) = \sup_\pi V_\pi(s)$$

We do not know yet that there is a policy that attains the sup.

**Claim 2:** For every $T$, $||V^* - V^{T*}|| \leq \beta^T C/(1-\beta)$.

- For every $s$ and $\epsilon > 0$, there is a policy $\pi$ such that $V_\pi(s) \geq V^*(s) - \epsilon$.
- For every $T$, $V^{T*}(s) \geq V_\pi^T(s)$.
- By Claim 1, $V_\pi^T(s) \geq V_\pi(s) - \beta^T C/(1-\beta)$.
- So $V^{T*}(s) \geq V^*(s) - \beta^T C/(1-\beta) - \epsilon \geq V^*(s) - \beta^T C/(1-\beta)$.

- By Claim 1, for every $T$ and $\pi$, $V_\pi^T(s) \leq V_\pi(s) + \beta^T C/(1-\beta) \leq V^*(s) + \beta^T C/(1-\beta)$.
- So, $V^{T*}(s) = \max_\pi V_\pi^T(s) \leq V^*(s) + \beta^T C/(1-\beta)$

So finite horizon values converge to infinite horizon value.

# Optimality Principle

The infinite horizon value function is the unique solution of

$$V^* = WV^*$$

Why?

- By the contraction theorem and the previous claim $||WV^* - WV^{T*}|| \leq \beta||V^* - V^{T*}|| \leq \beta^{T+1}C/(1-\beta)$
- By the finite horizon optimality principle $WV^{T*} = V^{(T+1)*}$
- So $||WV^* - V^{(T+1)*}|| \leq \beta^{T+1}C/(1-\beta)$.
- By the previous claim $||V^{(T+1)*} - V^*|| \leq \beta^{T+1}C/(1-\beta)$.
- By the triangle inequality $||WV^* - V^*|| \leq \beta^{T+1}C/(1-\beta) + \beta^{T+1}C/(1-\beta)$.
- Letting $T \to \infty$ shows that $V^*$ solves the functional equation.
- Uniqueness? Use contraction property.

## Computation of Value $V^*$

The argument for the optimality principle gives a procedure for calculating $V^*$ to any desired degree of accuracy.

Begin with any guess $V$.

Apply the operation $W$, $T$-times.

The resulting function is within $\beta^T 2C/(1-\beta)$ of $V^*$.

# Optimal Policies

A policy $\pi^*$ is *optimal* if $V_{\pi^*} = V^*$.

A policy $\pi = (f_0, f_1, \dots)$ is *stationary* if there is an action rule $f$ such that $f_t = f$ for all $t$.

**Theorem:** There is an optimal policy that is stationary.

For any $s$ there is an action, $f(s)$ that solves

$$\max_a [u(a, s) + \beta \sum_{s'} V^*(s') P(a, s)(s')]$$

So $WV^*$ is attained using the action rule $f(s)$. Define the operator $W_f$ by

$$W_f V(s) = u(f(s), s) + \beta \sum_{s'} V(s') P(f(s), s)(s')$$

So $W_f V^* = V^*$ and for every $T$, $W_f^T V^* = V^*$.

Then $\lim W_f^T V^* = V_{(f,f,\dots)} = V^*$.

Thus $\pi = (f, f, \dots)$ is an optimal policy and it is stationary.

# Computing Optimal Policies

Consider any stationary policy $\pi = (f, f, \dots)$.

Let $g(s)$ solve

$$\max_a [u(a, s) + \beta \sum_{s'} V_\pi(s') P(a, s)(s')]$$

Let $g = (g(1), g(2), \dots, g(S))$, $\pi' = (g, g, \dots)$ and $W_g V_\pi$ be the value of the problem above. The operator $W_g$ is just another way to write the operator $W$ applied to $V_\pi$.

**Claim:** If $\pi$ is not optimal then $V_{\pi'} > V_\pi$.

- If $\pi$ is not optimal then $W_g V_\pi > V_\pi$.
- The operator $W_g$ is monotone so $W_g^n V_\pi \geq W_g^{n-1} V_\pi \geq \cdots \geq V_\pi$.
- The limit of $W_g^n V_\pi$ is $V_{\pi'}$.
- So $V_{\pi'} > V_\pi$.

There are only a finite number of stationary policies so this improvement method finds an optimal stationary policy.

## Countable State Space

- Time: $t = 0, 1, \ldots$
- States: $S$ non-empty, countable set
- Actions: $A$ a subset of $\mathbf{R}^n$
- Histories: $H_t = S \times A \times S \times \cdots \times S$ with element $h_t = (s_0, a_0, \ldots, a_{t-1}, s_t)$
- Constraint: $\psi : S \to A$ a (non-empty valued) correspondence from $S$ to $A$, $\psi(s)$ describes the set of all actions feasible at state $s$
- Transition probability: For each action $a$ and state $s$, $P(a, s)(\cdot)$ is a probability on $S$. If at time $t$ the state is $s_t$ and action $a_t$ is chosen the distribution of states at time $t + 1$ is $P(a_t, s_t)$.

- Reward: $u : A \times S \to \mathbf{R}$
- Discount factor: $\beta$
- Action Rule: $f_t : H_t \to A$ such that $f_t(h_t) \in \psi(s_t)$
- Policy: $\pi = (f_0, f_1, \dots)$.

Begin at $s_0$, take action $f_0(s_0)$, move to state $s_1$ selected according to $P(f_0(s_0), s_0)$, take action $f_1(h_1)$, and so on.

So any policy $\pi$ defines a distribution $P_t(s_0, s_t)$ giving the probability of $s_t$ when $\pi$ is used and $s_0$ is the initial state.

# Optimality

The value of the problem when policy $\pi$ is used is

$$V_\pi(s_0) = E[\sum_{t=0}^{\infty} \beta^t u(f_t(h_t), s_t)]$$

where the expectation is computed using the $\{P_t\}$ induced by $\pi$.

For any probability $p_0$ on $S$ and any $\epsilon > 0$ a policy $\pi^*$ is $(p_0, \epsilon)$-optimal if $p_0\{s : V_{\pi^*}(s) > V_\pi(s) - \epsilon\} = 1$ for every policy $\pi$.

A policy $\pi^*$ is $\epsilon$-optimal if it is $(p_0, \epsilon)$-optimal for all probabilities $p_0$.

A policy is *optimal* if it is $\epsilon$-optimal for every $\epsilon > 0$, or equivalently if $V_{\pi^*}(s) \geq V_\pi(s)$ for all policies $\pi$ and initial states $s$.

## Assumptions

1. The Reward function $u$ is bounded (there is a number $c < \infty$ such that $||u|| < c$) and for each $s \in S$ the reward function $u(\cdot, s)$ is a continuous function of actions.
2. The discount factor is non-negative and less than 1, $0 \leq \beta < 1$.
3. The action space $A$ is compact.
4. The constraint sets $\psi(s)$ are closed for all $s \in S$.
5. For each pair of states $s, s'$ the transition probability $P(\cdot, s)(s')$ is a continuous function of actions.

**Example 1:**

$S = \{0\}$, $\psi(0) = A = \{1, 2, 3, \dots\}$

$u(a, 0) = (a - 1)/a$

$\sup_\pi V_\pi = 1/(1 - \beta)$

Is there an $\epsilon$-optimal policy? Is there an optimal policy?

There is no policy $\pi$ with $V_\pi = 1/(1 - \beta)$

**Example 2:**

$S = \{0\}$, $\psi(0) = A = [0, 1]$

$u(a, 0) = a$ if $0 \leq a < 1/2$, $u(1/2, 0) = 0$, $u(a, 0) = 1 - a$ if $1/2 < a \leq 1$

$\sup_\pi V_\pi = (1/2)/(1 - \beta)$

Is there an $\epsilon$-optimal policy? Is there an optimal policy?

There is no policy $\pi$ with $V_\pi = (1/2)/(1 - \beta)$

# Optimality of Stationary, Markov Policies

A policy $\pi = (f_0, f_1, \dots)$ is *Markov* if for each $t$, $f_t$ does not depend on $(s_0, a_0, \dots, a_{t-1})$, i.e. if $f_t(h_t) = f_t(s_t)$.

A Markov policy $\pi = (f_0, f_1, \dots)$ is *stationary* if there is an action rule $f$ such that $f_t = f$ for all $t$.

**Theorem** There is an optimal policy which is Markov and stationary. Let $\pi^* = (f, f, \dots)$ be this policy and $V^*$ be the value of $\pi^*$. Then $V^*$ is the unique solution to the optimality equation

$$V^*(s) = \max_{a \in \psi(s)} [u(a, s) + \beta \sum_{s'} P(a, s)(s') V^*(s')]$$

and for each $s$,

$$f(s) \in \operatorname*{argmax}_{a \in \psi(s)} [u(a, s) + \beta \sum_{s'} P(a, s)(s') V^*(s')].$$

## General State Spaces

Same as previous setup except

**States:** $S$ a non-empty, Borel subset of $\mathbf{R}^m$

**Constraint:** $\psi : S \to A$ a (non-empty valued) correspondence from $S$ to $A$ that admits a measurable selection.

**Action Rule:** a measurable function $f_t : H_t \to A$ such that $f_t(h_t) \in \psi(s_t)$

## Assumptions

1. The Reward function $u$ is a bounded, continuous function.
2. The discount factor is non-negative and less than 1, $0 \leq \beta < 1$.
3. The action space $A$ is compact.
4. The constraint $\psi$ is a continuous correspondence from $S$ to $A$.
5. The transition probability is continuous in $(a, s)$, i.e. for any bounded, continuous function $f : S \to \mathbf{R}$, $\int f(s')dP(a, s)(s')$ is a continuous function of $(a, s)$.

# Optimality of Stationary, Markov Policies

**Theorem** There is an optimal policy which is Markov and stationary. Let $\pi^* = (f, f, \dots)$ be this policy and $V^*$ be the value of $\pi^*$. Then $V^*$ is the unique solution to the optimality equation

$$V^*(s) = \max_{a \in \psi(s)} [u(a, s) + \beta \int_{s'} V^*(s') dP(a, s)(s')].$$

and for each $s$,

$$f(s) \in \operatorname*{argmax}_{a \in \psi(s)} [u(a, s) + \beta \int_{s'} V^*(s') dP(a, s)(s')].$$

Further, the value function $V^*$ is continuous and the action rule $f$ is upper hemi-continuous (and if the solution to the optimization problem is unique for all $s$ then $f$ is a continuous function).

# Application to Savings and Consumption

- States: Wealth $S = \mathbf{R}^1_+$, $s_0 > 0$
- Actions: Fraction to save $\delta$ and fraction of savings to invest in each of two assets $(\alpha^1, \alpha^2) = (\alpha^1, 1 - \alpha^1)$, $A = [0, 1]^2$.
- Constraint: constant, $\psi(s) = A$ for all $s$.
- Transition probability:

$$\begin{aligned} s' &= (s\delta)\alpha^1 R_1 \quad \text{with prob } p^1 \\ s' &= (s\delta)\alpha^2 R_2 \quad \text{with prob } p^2 \end{aligned}$$

- Reward: $u(a, s) = \log((1 - \gamma)s)$
- Discount factor: $0 \leq \beta < 1$
- Assume $R_1 > 0$, $R_2 > 0$ and $\beta < \max\{R_1^{-1}, R_2^{-1}\}$

The reward function is not bounded (above or below), but no policy yields infinite value and there is a policy that gives a value bounded from below.

# Solution

$$\delta(s) = \beta$$

$$\alpha^i(s) = p^i$$

**Derivation**

$$F(\epsilon) = \cdots + \beta^{t-1} \log\big((1-\delta_t)s_t - \epsilon\big) +$$

$$\beta^t p^1 \log\Bigg(\big(1-\delta_{t+1}(\delta_t s_t \alpha_t^1 R^1)\big)\delta_t s_t \alpha_t^1 R^1 + \epsilon\alpha_t^1 R^1\Bigg) +$$

$$\beta^t p^2 \log\Bigg(\big(1-\delta_{t+1}(\delta_t s_t \alpha_t^2 R^2)\big)\delta_t s_t \alpha_t^2 R^2 + \epsilon\alpha_t^2 R^2\Bigg) + \cdots$$

where $\delta_t$, $\alpha_t^1$, $\alpha_t^2$ and $\delta_{t+1}(s)$ are all optimal.

Optimality implies that $F(\epsilon)$ is maximized at $\epsilon = 0$, so $F'(0) = 0$.

$$F'(0) = \frac{-\beta^{t-1}}{1-\delta_t} + \frac{\beta^t p^1}{\big(1-\delta_{t+1}(1)\big)\delta_t} + \frac{\beta^t p^2}{\big(1-\delta_{t+1}(2)\big)\delta_t}$$
$$= 0$$

So $\delta_t = \delta_{t+1}(1) = \delta_{t+1}(2) = \beta$ is a solution.

$$G(\epsilon) = \cdots + \beta^{t-1} p^1 \log\Big((1-\beta)\beta s_t \alpha_t^1 R^1 - \beta s_t \epsilon R^1\Big) +$$
$$\beta^{t-1} p^2 \log\Big((1-\beta)\beta s_t \alpha_t^2 R^2 + \beta s_t \epsilon R^2\Big) + \cdots$$

where the $\alpha_t^i$ are optimal. $G(0)$ is an optimum.

$$G'(0) = -\frac{p^1}{\alpha_t^1} + \frac{p^2}{\alpha_t^2} = 0$$

so $\alpha_t^1 = p^1$ and $\alpha_t^2 = p^2$ is the solution.

# Value of the Problem

Guess that the value function is of the form $\frac{\log(s)}{1-\beta} + K$ where $K$ is a constant. Check the optimality equation:

$$\begin{aligned} V(s) &= \max[\log((1-\delta)s) \\ &\quad +\beta[p^1 \frac{\log(\delta s \alpha^1 R^1)}{1-\beta} + p^2 \frac{\log(\delta s \alpha^2 R^2)}{1-\beta} + K]] \\ &= \log(s)(1 + \frac{\beta}{1-\beta}) + \max[\log(1-\delta) + \\ &\quad \beta[p^1 \frac{\log(\delta \alpha^1 R^1)}{1-\beta} + p^2 \frac{\log(\delta \alpha^2 R^2)}{1-\beta} + K]] \\ &= \frac{\log(s)}{1-\beta} + \log(1-\beta) + \\ &\quad \beta[p^1 \frac{\log(\beta p^1 R^1)}{1-\beta} + p^2 \frac{\log(\beta p^2 R^2)}{1-\beta} + K] \end{aligned}$$

As $\beta$, $p^1$ and $p^2$ solve the optimization problem.

So the conjecture is correct and we could solve for $K$.

# Reference

K. Hinderer, *Foundations of Non-Stationary Dynamic Programming with Discrete Time Parameters*, Springer-Verlag, 1970