Lecture Notes 10: Dynamic Programming

Peter J. Hammond

Latest minor revisions 2024 September 21st;
typeset from dynProg24.tex

Antecedents: Pierre Massé

Pierre Massé (1946) *Les réserves et la régulation de l'avenir dans la vie économique, I: avenir déterminé*
and — *II: avenir aléatoire* (Paris: Hermann & Cie)

> *En exploitation optimum, il y a à chaque instant égalité entre le profit marginal instantané attaché au flux et le profit marginal futur attaché au stock.*
>
> *vol. I, p. 90*

> *La condition nécessaire et suffisante pour que soit maximum l'espérance totale, considérée comme fonction du flux transmis dans un intervalle de temps élémentaire, est qu'il y ait égalité entre le profit marginal attaché au flux transmis et l'espérance marginale attachée au stock residuel.*
>
> *vol. II, p. 39*

Arrow's Citation of Massé

Kenneth J. Arrow (1957) "Statistics and Economic Policy" *Econometrica* 25: 523–531.

> *It would be easy to show that much of the reasoning used in capital theory has in fact made use of the principle of optimality. The explicit recognition of this principle has stemmed from the work of P. Massé ... (footnote on p. 525)*

Pierre Massé (1898–1987)

- engineer at l'École nationale des ponts et chaussées;
- 1948: the deputy general manager of Electricité de France;
- 1959–1966 Commissaire général du Plan;
- 1965–1969: chairman of the board of directors (CEO) of Electricité de France.

One part of his expertise: hydroelectric power and dam management.

Arrow, Bellman, Howard, Marschak, Blackwell, etc.

Kenneth J. Arrow, Theodore Harris and Jacob Marschak (1951)
"Optimal Inventory Policy" *Econometrica* 19 (3): 250–272.

Richard Bellman (1954) "The Theory of Dynamic Programming"
Bulletin of the American Mathematical Society 60 (6): 503–516.

Richard Bellman (1957) *Dynamic Programming*
Princeton University Press. Dover paperback edition (2003).

Ronald A. Howard (1960)
Dynamic Programming and Markov Processes (MIT Press).

David Blackwell (1965) "Discounted Dynamic Programming"
Annals of Mathematical Statistics 36(1): 226–235.

Source for the saving problem discussed later

David Levhari and T.N. Srinivasan (1969)
"Optimal Savings Under Uncertainty"
Review of Economic Studies 36 (2): 153–163.

Outline

Basic Equation

A simple stochastic linear difference equation of the first order in one variable takes the form

$$x_t = a x_{t-1} + \epsilon_t \quad (t \in \mathbb{N})$$

Here a is a real parameter,
and each ϵ_t is a real random disturbance.

Assume that:

1. there is a given or pre-determined initial state x_0;
2. the random variables ϵ_t
   are independent and identically distributed (IID)
   with mean $\mathbb{E}\epsilon_t = 0$ and variance $\mathbb{E}\epsilon_t^2 = \sigma^2$.

A special case is when the disturbances are all normally distributed — i.e., $\epsilon_t \sim N(0, \sigma^2)$.

Explicit Solution and Conditional Mean

For each fixed outcome $\epsilon^{\mathbb{N}} = (\epsilon_t)_{t \in \mathbb{N}}$ of the random sequence, there is a unique solution which can be written as

$$x_t = a^t x_0 + \sum_{s=1}^{t} a^{t-s} \epsilon_s$$

The main stable case occurs when $|a| < 1$.

Then each term of the sum $a^t x_0 + \sum_{s=1}^{t} a^{t-s} \epsilon_s$ converges to 0 as $t \to \infty$.

This is what econometricians or statisticians call a first-order autoregressive (or AR(1)) process.

In fact, given x_0 at time 0,
our assumption that $\mathbb{E}\epsilon_s = 0$ for all $s = 1, 2, \ldots, t$
implies that the conditional mean of x_t is

$$m_t := \mathbb{E}[x_t | x_0] = \mathbb{E}\left[a^t x_0 + \sum_{s=1}^{t} a^{t-s} \epsilon_s | x_0\right] = a^t x_0$$

Conditional Variance

The conditional variance, however, is given by

$$v_t := \mathbb{E}\left[(x_t - m_t)^2 | x_0\right] = \mathbb{E}[(x_t - a^t x_0)^2 | x_0] = \mathbb{E}\left[\sum_{s=1}^{t} a^{t-s} \epsilon_s\right]^2$$

In the case we are considering
with independently distributed disturbances ϵ_s,
the variance of a sum is the sum of the variances.

Hence, because each $\mathbb{E}\epsilon_s = 0$, one has

$$v_t = \sum_{s=1}^{t} \mathbb{E}\left[a^{t-s} \epsilon_s\right]^2 = \sum_{s=1}^{t} a^{2(t-s)} \mathbb{E}\epsilon_s^2 = \sigma^2 \sum_{s=1}^{t} a^{2(t-s)}$$

Using the rule for summing the geometric series $\sum_{s=1}^{t} a^{2(t-s)}$,
we finally obtain

$$v_t = \frac{1 - a^{2t}}{1 - a^2} \sigma^2$$

Outline

Sums of Normally Distributed Random Variables, I

Recall that if $X \sim N(\mu, \sigma^2)$, then the characteristic function defined by $\phi_X(t) = \mathbb{E}[e^{iXt}]$ takes the form

$$\phi_X(t) = \mathbb{E}[e^{iXt}] = \int_{-\infty}^{+\infty} e^{ixt} \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2 dx$$

This reduces to $\phi_X(t) = \exp\left(it\mu - \frac{1}{2}\sigma^2 t^2\right)$.

Hence, if $Z = X + Y$ where $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$ are independent random, variables, then

$$\phi_Z(t) = \mathbb{E}[e^{iZt}] = \mathbb{E}[e^{i(X+Y)t}] = \mathbb{E}[e^{iXt} e^{iYt}] = \mathbb{E}[e^{iXt}]\mathbb{E}[e^{iYt}]$$

Sums of Normally Distributed Random Variables, II

So

$$\begin{aligned}\phi_Z(t) &= \exp\left(it\mu_X - \tfrac{1}{2}\sigma_X^2 t^2\right)\exp\left(it\mu_Y - \tfrac{1}{2}\sigma_Y^2 t^2\right) \\ &= \exp\left(it(\mu_X + \mu_Y) - \tfrac{1}{2}(\sigma_X^2 + \sigma_Y^2)t^2\right) \\ &= \exp\left(it\mu_Z - \tfrac{1}{2}\sigma_Z^2 t^2\right)\end{aligned}$$

where $\mu_Z = \mu_X + \mu_Y = \mathbb{E}(X + Y)$ is the mean of $X + Y$,

and $\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2$ is the variance of $X + Y$.

It follows that $t \mapsto \phi_Z(t)$
is the characteristic function of a random variable $Z \sim N(\mu_Z, \sigma_Z^2)$
where $\mu_Z = \mu_X + \mu_Y = \mathbb{E}(X + Y)$ and $\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2$.

That is, the sum $Z = X + Y$
of two independent normally distributed random variables X and Y
is also normally distributed, with:

1. mean equal to the sum of the means;
2. variance equal to the sum of the variances.

The Gaussian Case and the Asymptotic Distribution

In the particular case when each ϵ_t is normally distributed as well as IID, then x_t is also normally distributed with mean m_t and variance v_t.

As $t \to \infty$, the conditional mean $m_t = a^t x_0 \to 0$ and the conditional variance

$$v_t = \frac{1 - a^{2t}}{1 - a^2} \sigma^2 \to v := \frac{\sigma^2}{1 - a^2}$$

In the case when each ϵ_t is normally distributed, this implies that the asymptotic distribution of x_t is also normal, with mean 0 and variance $v = \sigma^2/(1 - a^2)$.

Stationarity

Now suppose that x_0 itself has this asymptotic normal distribution — that is, suppose that $x_0 \sim N(0, \sigma^2/(1-a^2))$.

This is what the distribution of x_0 would be
if the process had started at $t = -\infty$ instead of at $t = 0$.

Then the unconditional mean of each x_t is $\mathbb{E}x_t = a^t \mathbb{E}x_0 = 0$.

On the other hand, because $x_{t+k} = a^k x_t + \sum_{s=1}^{k} a^{k-s} \epsilon_{t+s}$,
the unconditional covariance of x_t and x_{t+k} is

$$\mathbb{E}(x_{t+k} x_t) = \mathbb{E}[a^k x_t^2] = a^k v = \frac{a^k}{1-a^2}\sigma^2 \quad (k = 0, 1, 2 \ldots)$$

In fact, given any t, the joint distribution
of the r random variables $x_t, x_{t+1}, \ldots, x_{t+r-1}$
is multivariate normal with variance–covariance matrix
having elements $\mathbb{E}(x_{t+k} x_t) = a^k \sigma^2/(1-a^2)$, independent of t.

Because of this independence, the process is said to be stationary.

Outline

Intertemporal Utility

Consider a household which at time s is planning its intertemporal consumption stream $\mathbf{c}_s^T := (c_s, c_{s+1}, \ldots, c_T)$ over periods t in the set $\{s, s+1, \ldots, T\}$.

Its intertemporal utility function $\mathbb{R}^{T-s+1} \ni \mathbf{c}_s^T \mapsto U_s^T(\mathbf{c}_s^T) \in \mathbb{R}$ is assumed to take the additively separable form

$$U_s^T(\mathbf{c}_s^T) := \sum\nolimits_{t=s}^{T} u_t(c_t)$$

where the one period felicity functions $c \mapsto u_t(c)$ are differentiably increasing and strictly concave (DISC) — i.e., $u_t'(c) > 0$, and $u_t''(c) < 0$ for all t and all $c > 0$.

Suppose the household faces:

1. fixed initial wealth w_s;
2. a terminal wealth constraint $w_{T+1} \geq 0$.

Risky Wealth Accumulation

We assume a wealth accumulation equation $w_{t+1} = \tilde{r}_t(w_t - c_t)$, where $\tilde{r}_t$ is the household's gross rate of return on its wealth in period t.

It is assumed that:

1. the return $\tilde{r}_t$ in each period t is a random variable with positive values;
2. the return distributions for different times t are stochastically independent;
3. starting with predetermined wealth w_s at time s, the household seeks to maximize the expectation $\mathbb{E}_s[U_s^T(\mathbf{c}_s^T)]$ of its intertemporal utility.

Outline

Two Period Case

We work backwards from the last period, when $s = T$.

In this last period the household will obviously choose $c_T = w_T$, yielding a maximized utility equal to $V_T(w_T) = u_T(w_T)$.

Next, consider the penultimate period, when $s = T - 1$.
The consumer will want to choose c_{T-1} in order to maximize

$$\underbrace{u_{T-1}(c_{T-1})}_{\text{period } T-1} \quad + \quad \underbrace{\mathbb{E}_{T-1} V_T(w_T)}_{\text{result of an optimal policy in period } T}$$

subject to the wealth constraint

$$w_T = \underbrace{\tilde{r}_{T-1}}_{\text{random gross return}} \underbrace{(w_{T-1} - c_{T-1})}_{\text{saving}}$$

First-Order Condition

Substituting both the function $V_T(w_T) = u_T(w_T)$
and the wealth constraint into the objective reduces the problem to

$$\max_{c_{T-1}} \{ u_{T-1}(c_{T-1}) + \mathbb{E}_{T-1} [u_T(\tilde{r}_{T-1}(w_{T-1} - c_{T-1}))] \}$$

subject to $0 \leq c_{T-1} \leq w_{T-1}$ and $\tilde{c}_T := \tilde{r}_{T-1}(w_{T-1} - c_{T-1})$.

Assume we can differentiate under the integral sign,
and that there is an interior solution with $0 < c_{T-1} < w_{T-1}$.

Then the first-order condition (FOC) is

$$0 = u'_{T-1}(c_{T-1}) + \mathbb{E}_{T-1}[(-\tilde{r}_{T-1}) u'_T(\tilde{r}_{T-1}(w_{T-1} - c_{T-1}))]$$

The Stochastic Euler Equation

Rearranging the first-order condition while recognizing that $\tilde{c}_T := \tilde{r}_{T-1}(w_{T-1} - c_{T-1})$, one obtains

$$u'_{T-1}(c_{T-1}) = \mathbb{E}_{T-1}[\tilde{r}_{T-1} u'_T(\tilde{r}_{T-1}(w_{T-1} - c_{T-1}))]$$

Dividing by $u'_{T-1}(c_{T-1})$ gives the stochastic Euler equation

$$1 = \mathbb{E}_{T-1}\left[\tilde{r}_{T-1}\frac{u'_T(\tilde{c}_T)}{u'_{T-1}(c_{T-1})}\right] = \mathbb{E}_{T-1}\left[\tilde{r}_{T-1}\mathrm{MRS}^T_{T-1}(c_{T-1};\tilde{c}_T)\right]$$

involving the marginal rate of substitution function

$$\mathrm{MRS}^T_{T-1}(c_{T-1};\tilde{c}_T) := \frac{u'_T(\tilde{c}_T)}{u'_{T-1}(c_{T-1})}$$

The CES Case

For the marginal utility function $c \mapsto u'(c)$,
its elasticity of substitution is defined
for all $c > 0$ by $\eta(c) := d \ln u'(c) / d \ln c$.

Then $\eta(c)$ is both the degree of relative risk aversion,
and the degree of relative fluctuation aversion.

A constant elasticity of substitution (or CES) utility function
satisfies $d \ln u'(c) / d \ln c = -\epsilon < 0$ for all $c > 0$.

The marginal rate of substitution
satisfies $u'(c)/u'(\bar{c}) = (c/\bar{c})^{-\epsilon}$ for all $c, \bar{c} > 0$.

Normalized Utility

Normalize by putting $u'(1) = 1$, implying that $u'(c) \equiv c^{-\epsilon}$.

Then integrating gives

$$
\begin{aligned}
u(c;\epsilon) &= u(1) + \int_1^c x^{-\epsilon} dx \\
&= \begin{cases} u(1) + \dfrac{c^{1-\epsilon} - 1}{1-\epsilon} & \text{if } \epsilon \neq 1 \\ u(1) + \ln c & \text{if } \epsilon = 1 \end{cases}
\end{aligned}
$$

Introduce the final normalization

$$
u(1) = \begin{cases} \dfrac{1}{1-\epsilon} & \text{if } \epsilon \neq 1 \\ 0 & \text{if } \epsilon = 1 \end{cases}
$$

The utility function is reduced to

$$
u(c;\epsilon) = \begin{cases} \dfrac{c^{1-\epsilon}}{1-\epsilon} & \text{if } \epsilon \neq 1 \\ \ln c & \text{if } \epsilon = 1 \end{cases}
$$

The Stochastic Euler Equation in the CES Case

Consider the CES case when $u'_t(c) \equiv \delta_t c^{-\epsilon}$,
where each δ_t is the discount factor for period t.

Definition

The one-period discount factor in period t
is defined as $\beta_t := \delta_{t+1}/\delta_t$.

Then the stochastic Euler equation takes the form

$$1 = \mathbb{E}_{T-1}\left[\tilde{r}_{T-1}\beta_{T-1}\left(\frac{\tilde{c}_T}{c_{T-1}}\right)^{-\epsilon}\right]$$

Because c_{T-1} is being chosen at time $T-1$, this implies that

$$(c_{T-1})^{-\epsilon} = \mathbb{E}_{T-1}\left[\tilde{r}_{T-1}\beta_{T-1}(\tilde{c}_T)^{-\epsilon}\right]$$

The Two Period Problem in the CES Case

In the two-period case, we know that

$$\tilde{c}_T = \tilde{w}_T = \tilde{r}_{T-1}(w_{T-1} - c_{T-1})$$

in the last period, so the Euler equation becomes

$$\begin{aligned}(c_{T-1})^{-\epsilon} &= \mathbb{E}_{T-1}\left[\tilde{r}_{T-1}\beta_{T-1}(\tilde{c}_T)^{-\epsilon}\right] \\ &= \beta_{T-1}(w_{T-1} - c_{T-1})^{-\epsilon}\mathbb{E}_{T-1}\left[(\tilde{r}_{T-1})^{1-\epsilon}\right]\end{aligned}$$

Take the $(-1/\epsilon)$ th power of each side and define

$$\rho_{T-1} := \left(\beta_{T-1}\mathbb{E}_{T-1}\left[(\tilde{r}_{T-1})^{1-\epsilon}\right]\right)^{-1/\epsilon}$$

This reduces the Euler equation to $c_{T-1} = \rho_{T-1}(w_{T-1} - c_{T-1})$.

Its solution is evidently $c_{T-1} = \gamma_{T-1} w_{T-1}$ where

$$\gamma_{T-1} := \rho_{T-1}/(1 + \rho_{T-1}) \quad \text{and} \quad 1 - \gamma_{T-1} = 1/(1 + \rho_{T-1})$$

are respectively the optimal consumption and savings ratios.

It follows that $\rho_{T-1} = \gamma_{T-1}/(1 - \gamma_{T-1})$

is the consumption/savings ratio.

Optimal Discounted Expected Utility

The optimal policy in periods T and $T-1$
is $c_t = \gamma_t w_t$ where $\gamma_T = 1$ and γ_{T-1} has just been defined.

In this CES case, the discounted utility of consumption in period T
is $V_T(w_T) := \delta_T u(w_T; \epsilon)$.

The discounted expected utility at time $T-1$
of consumption in periods T and $T-1$ together is

$$V_{T-1}(w_{T-1}) = \delta_{T-1} u(\gamma_{T-1} w_{T-1}; \epsilon) + \delta_T \mathbb{E}_{T-1}[u(\tilde{w}_T; \epsilon)]$$

where $\tilde{w}_T = \tilde{r}_{T-1}(1 - \gamma_{T-1}) w_{T-1}$.

Discounted Expected Utility in the Logarithmic Case

In the logarithmic case when $\epsilon = 1$, one has

$$V_{T-1}(w_{T-1}) = \delta_{T-1} \ln(\gamma_{T-1} w_{T-1}) + \delta_T \mathbb{E}_{T-1}[\ln(\tilde{r}_{T-1}(1 - \gamma_{T-1}) w_{T-1})]$$

It follows that

$$V_{T-1}(w_{T-1}) = \alpha_{T-1} + (\delta_{T-1} + \delta_T) u(w_{T-1}; \epsilon)$$

where

$$\alpha_{T-1} := \delta_{T-1} \ln \gamma_{T-1} + \delta_T \{\ln(1 - \gamma_{T-1}) + \mathbb{E}_{T-1}[\ln \tilde{r}_{T-1}]\}$$

Discounted Expected Utility in the CES Case

In the CES case when $\epsilon \neq 1$, one has

$$(1-\epsilon)V_{T-1}(w_{T-1}) = \delta_{T-1}(\gamma_{T-1}w_{T-1})^{1-\epsilon} + \delta_T[(1-\gamma_{T-1})w_{T-1}]^{1-\epsilon}\,\mathbb{E}_{T-1}[(\tilde{r}_{T-1})^{1-\epsilon}]$$

so $V_{T-1}(w_{T-1}) = v_{T-1}u(w_{T-1};\epsilon)$ where

$$v_{T-1} := \delta_{T-1}(\gamma_{T-1})^{1-\epsilon} + \delta_T(1-\gamma_{T-1})^{1-\epsilon}\,\mathbb{E}_{T-1}[(\tilde{r}_{T-1})^{1-\epsilon}]$$

In both cases,
one can write $V_{T-1}(w_{T-1}) = \alpha_{T-1} + v_{T-1}u(w_{T-1};\epsilon)$
for a suitable additive constant α_{T-1} (which is 0 in the CES case)
and a suitable multiplicative constant v_{T-1}.

Outline

The Time Line

In each period t, suppose:

- the consumer starts with known wealth w_t;
- then the consumer chooses consumption c_t, along with savings or residual wealth $w_t - c_t$;
- there is a cumulative distribution function $F_t(r)$ on $\mathbb{R}$ that determines the gross return $\tilde{r}_t$ as a positive-valued random variable.

After these three steps have been completed,
the problem starts again in period $t + 1$,
with the consumer's wealth known to be $w_{t+1} = \tilde{r}_t(w_t - c_t)$.

Expected Conditionally Expected Utility

Starting at any t, suppose the consumer's choices, together with the random returns, jointly determine a cdf F_t^T over the space of intertemporal consumption streams $\mathbf{c}_t^T$.

The associated expected utility is $\mathbb{E}_t \left[U_t^T(\mathbf{c}_t^T) \right]$, using the shorthand $\mathbb{E}_t$ to denote integration w.r.t. the cdf F_t^T.

Then, given that the consumer has chosen c_t at time t, let $\mathbb{E}_{t+1}[\cdot|c_t]$ denote the conditional expected utility.

This is found by integrating w.r.t. the conditional cdf $F_{t+1}^T(\mathbf{c}_{t+1}^T|c_t)$.

The law of iterated expectations allows us to write the unconditional expectation $\mathbb{E}_t \left[U_t^T(\mathbf{c}_t^T) \right]$

as the expectation $\mathbb{E}_t[\mathbb{E}_{t+1}[U_t^T(\mathbf{c}_t^T)|c_t]]$ of the conditional expectation.

The Expectation of Additively Separable Utility

Our hypothesis is that the intertemporal von Neumann–Morgenstern utility function takes the additively separable form

$$U_t^T(\mathbf{c}_t^T) = \sum\nolimits_{\tau=t}^{T} u_\tau(c_\tau)$$

The conditional expectation given c_t must then be

$$\mathbb{E}_{t+1}[U_t^T(\mathbf{c}_t^T)|c_t] = u_t(c_t) + \mathbb{E}_{t+1}\left[\sum\nolimits_{\tau=t+1}^{T} u_\tau(c_\tau)|c_t\right]$$

whose expectation is

$$\mathbb{E}_t\left[\sum\nolimits_{\tau=t}^{T} u_\tau(c_\tau)\right] = u_t(c_t) + \mathbb{E}_t\left[\mathbb{E}_{t+1}\left[\sum\nolimits_{\tau=t+1}^{T} u_\tau(c_\tau)\right]|c_t\right]$$

The Continuation Value

Let $V_{t+1}(w_{t+1})$ be the state valuation function
expressing the maximum of the continuation value

$$\mathbb{E}_{t+1}\left[U_{t+1}^T(\mathbf{c}_{t+1}^T)|w_{t+1}\right] = \mathbb{E}_{t+1}\left[\sum_{\tau=t+1}^T u_\tau(c_\tau)|w_{t+1}\right]$$

as a function of the wealth level or state w_{t+1}.

Assume this maximum value is achieved
by following an optimal policy from period $t+1$ on.

Then total expected utility at time t will then reduce to

$$\begin{aligned}
\mathbb{E}_t\left[U_t^T(\tilde{\mathbf{c}}_\mathbf{t}^\mathbf{T})|c_t\right] &= u_t(c_t) + \mathbb{E}_t\left[\mathbb{E}_{t+1}\left[\sum_{\tau=t+1}^T u_\tau(c_\tau)|w_{t+1}\right]|c_t\right] \\
&= u_t(c_t) + \mathbb{E}_t[V_{t+1}(\tilde{w}_{t+1})|c_t] \\
&= u_t(c_t) + \mathbb{E}_t[V_{t+1}(\tilde{r}_t(w_t - c_t))]
\end{aligned}$$

The Principle of Optimality

Maximizing $\mathbb{E}_s \left[U_s^T(\mathbf{c}_s^T) \right]$ w.r.t. c_s, taking as fixed the optimal consumption plans $c_t(w_t)$ at times $t = s+1, \ldots, T$, therefore requires choosing c_s to maximize

$$u_s(c_s) + \mathbb{E}_s[V_{s+1}(\tilde{r}_s(w_s - c_s))]$$

Let $c_s^*(w_s)$ denote a solution to this maximization problem.

Then the value of an optimal plan $(c_t^*(w_t))_{t=s}^T$ that starts with wealth w_s at time s is

$$V_s(w_s) := u_s(c_s^*(w_s)) + \mathbb{E}_s[V_{s+1}(\tilde{r}_s(w_s - c_s^*(w_s)))]$$

Together, these two properties can be expressed as

$$\left. \begin{array}{rcl} V_s(w_s) & = & \\ c_s^*(w_s) & = & \arg \end{array} \right\} \max_{0 \leq c_s \leq w_s} \{ u_s(c_s) + \mathbb{E}_s[V_{s+1}(\tilde{r}_s(w_s - c_s))] \}$$

which can be described as the the principle of optimality.

An Induction Hypothesis

Consider once again the case when $u_t(c) \equiv \delta_t u(c; \epsilon)$
for the CES (or logarithmic) utility function
that satisfies $u'(c; \epsilon) \equiv c^{-\epsilon}$ and, specifically

$$u(c; \epsilon) = \begin{cases} c^{1-\epsilon}/(1-\epsilon) & \text{if } \epsilon \neq 1; \\ \ln c & \text{if } \epsilon = 1. \end{cases}$$

Inspired by the solution we have already found
for the final period T and penultimate period $T-1$,
we adopt the induction hypothesis that there are
constants α_t, γ_t, v_t $(t = T, T-1, \ldots, s+1, s)$ for which

$$c_t^*(w_t) = \gamma_t w_t \quad \text{and} \quad V_t(w_t) = \alpha_t + v_t u(w_t; \epsilon)$$

In particular, the consumption ratio γ_t and savings ratio $1 - \gamma_t$
are both independent of the wealth level w_t.

Applying Backward Induction

Under the induction hypotheses that

$$c_t^*(w_t) = \gamma_t w_t \quad \text{and} \quad V_t(w_t) = \alpha_t + v_t u(w_t; \epsilon)$$

the maximand

$$u_s(c_s) + \mathbb{E}_s[V_{s+1}(\tilde{r}_s(w_s - c_s))]$$

takes the form

$$\delta_s u(c_s; \epsilon) + \mathbb{E}_s[\alpha_{s+1} + v_{s+1} u(\tilde{r}_s(w_s - c_s); \epsilon)]$$

The first-order condition for this to be maximized w.r.t. c_s is

$$0 = \delta_s u'(c_s; \epsilon) - v_{s+1} \mathbb{E}_s[\tilde{r}_s u'(\tilde{r}_s(w_s - c_s); \epsilon)]$$

or, equivalently, that

$$\delta_s (c_s)^{-\epsilon} = v_{s+1} \mathbb{E}_s[\tilde{r}_s (\tilde{r}_s(w_s - c_s))^{-\epsilon})] = v_{s+1}(w_s - c_s)^{-\epsilon} \mathbb{E}_s[(\tilde{r}_s)^{1-\epsilon}]$$

Solving the Logarithmic Case

When $\epsilon = 1$ and so $u(c; \epsilon) = \ln c$,
the first-order condition reduces to $\delta_s (c_s)^{-1} = v_{s+1}(w_s - c_s)^{-1}$.

Its solution is indeed $c_s = \gamma_s w_s$ where $\delta_s (\gamma_s)^{-1} = v_{s+1}(1 - \gamma_s)^{-1}$,
implying that $\gamma_s = \delta_s / (\delta_s + v_{s+1})$.

The state valuation function then becomes

$$\begin{aligned} V_s(w_s) &= \delta_s u(\gamma_s w_s; \epsilon) + \alpha_{s+1} + v_{s+1} \mathbb{E}_s [u(\tilde{r}_s (1 - \gamma_s) w_s; \epsilon)] \\ &= \delta_s \ln(\gamma_s w_s) + \alpha_{s+1} + v_{s+1} \mathbb{E}_s [\ln(\tilde{r}_s (1 - \gamma_s) w_s)] \\ &= \delta_s \ln(\gamma_s w_s) + \alpha_{s+1} + v_{s+1} \{\ln(1 - \gamma_s) w_s + \ln R_s\} \end{aligned}$$

where we define the geometric mean certainty equivalent return R_s
so that $\ln R_s := \mathbb{E}_s [\ln(\tilde{r}_s)]$.

The State Valuation Function

The formula

$$V_s(w_s) = \delta_s \ln(\gamma_s w_s) + \alpha_{s+1} + v_{s+1}\{\ln(1-\gamma_s)w_s + \ln R_s\}$$

reduces to the desired form $V_s(w_s) = \alpha_s + v_s \ln w_s$
provided we take $v_s := \delta_s + v_{s+1}$, which implies that $\gamma_s = \delta_s / v_s$, and also

$$\begin{aligned}\alpha_s &:= \delta_s \ln \gamma_s + \alpha_{s+1} + v_{s+1}\left\{\ln(1-\gamma_s) + \ln R_s\right\} \\ &= \delta_s \ln(\delta_s/v_s) + \alpha_{s+1} + v_{s+1}\{\ln(v_{s+1}/v_s) + \ln R_s\} \\ &= \delta_s \ln \delta_s + \alpha_{s+1} - v_s \ln v_s + v_{s+1}\{\ln v_{s+1} + \ln R_s\}\end{aligned}$$

This confirms the induction hypothesis for the logarithmic case.

The relevant constants v_s are found by summing backwards, starting with $v_T = \delta_T$, implying that $v_s = \sum_{\tau=s}^{T} \delta_\tau$.

The Stationary Logarithmic Case

In the stationary logarithmic case:

- the felicity function in each period t is $\beta^t \ln c_t$, so the one period discount factor is the constant β;
- the certainty equivalent return R_t is also a constant R.

Then $v_s = \sum_{\tau=s}^{T} \delta_s = \sum_{\tau=s}^{T} \beta^\tau = (\beta^s - \beta^{T+1})/(1-\beta)$, implying that $\gamma_s = \beta^s / v_s = \beta^s(1-\beta)/(\beta^s - \beta^{T+1})$.

It follows that

$$c_s = \gamma_s w_s = \frac{(1-\beta)w_s}{1-\beta^{T-s+1}} = \frac{(1-\beta)w_s}{1-\beta^{H+1}}$$

when there are $H := T - s$ periods left before the horizon T.

As $H \to \infty$, this solution converges to $c_s = (1-\beta)w_s$, so the savings ratio equals the constant discount factor β.

Remarkably, this is also independent of the gross return to saving.

First-Order Condition in the CES Case

Recall that the first-order condition in the CES Case is

$$\delta_s(c_s)^{-\epsilon} = v_{s+1}(w_s - c_s)^{-\epsilon}\mathbb{E}_s[(\tilde{r}_s)^{1-\epsilon}] = v_{s+1}(w_s - c_s)^{-\epsilon}R_s^{1-\epsilon}$$

where we have defined the certainty equivalent return R_s as the solution to $R_s^{1-\epsilon} := \mathbb{E}_s[(\tilde{r}_s)^{1-\epsilon}]$.

The first-order condition indeed implies that $c_s^*(w_s) = \gamma_s w_s$, where $\delta_s(\gamma_s)^{-\epsilon} = v_{s+1}(1-\gamma_s)^{-\epsilon}R_s^{1-\epsilon}$.

This implies that

$$\frac{\gamma_s}{1-\gamma_s} = \left(v_{s+1}R_s^{1-\epsilon}/\delta_s\right)^{-1/\epsilon}$$

or

$$\gamma_s = \frac{\left(v_{s+1}R_s^{1-\epsilon}/\delta_s\right)^{-1/\epsilon}}{1+\left(v_{s+1}R_s^{1-\epsilon}/\delta_s\right)^{-1/\epsilon}} = \frac{\left(v_{s+1}R_s^{1-\epsilon}\right)^{-1/\epsilon}}{(\delta_s)^{-1/\epsilon}+\left(v_{s+1}R_s^{1-\epsilon}\right)^{-1/\epsilon}}$$

Completing the Solution in the CES Case

Under the induction hypothesis that $V_{s+1}(w) = v_{s+1} w^{1-\epsilon}/(1-\epsilon)$, one also has

$$(1-\epsilon)V_s(w_s) = \delta_s(\gamma_s w_s)^{1-\epsilon} + v_{s+1}\mathbb{E}_s[(\tilde{r}_s(1-\gamma_s)w_s)^{1-\epsilon}]$$

This reduces to the desired form $(1-\epsilon)V_s(w_s) = v_s(w_s)^{1-\epsilon}$, where

$$\begin{aligned} v_s &:= \delta_s(\gamma_s)^{1-\epsilon} + v_{s+1}\mathbb{E}_s[(\tilde{r}_s)^{1-\epsilon}](1-\gamma_s)^{1-\epsilon} \\ &= \frac{\delta_s(v_{s+1}R_s^{1-\epsilon})^{1-1/\epsilon} + v_{s+1}R_s^{1-\epsilon}(\delta_s)^{1-1/\epsilon}}{[(\delta_s)^{-1/\epsilon} + \left(v_{s+1}R_s^{1-\epsilon}\right)^{-1/\epsilon}]^{1-\epsilon}} \\ &= \delta_s v_{s+1} R_s^{1-\epsilon} \frac{(v_{s+1}R_s^{1-\epsilon})^{-1/\epsilon} + (\delta_s)^{-1/\epsilon}}{[(\delta_s)^{-1/\epsilon} + \left(v_{s+1}R_s^{1-\epsilon}\right)^{-1/\epsilon}]^{1-\epsilon}} \\ &= \delta_s v_{s+1} R_s^{1-\epsilon}[(\delta_s)^{-1/\epsilon} + \left(v_{s+1}R_s^{1-\epsilon}\right)^{-1/\epsilon}]^{\epsilon} \end{aligned}$$

This confirms the induction hypothesis for the CES case.

Again, the relevant constants are found by working backwards.

Outline

Histories and Strategies

For each time $t = s, s+1, \ldots, T$
between the start s and the horizon T,
let h^t denote a known history $(w_\tau, c_\tau, \tilde{r}_\tau)_{\tau=s}^t$
of the triples $(w_\tau, c_\tau, \tilde{r}_\tau)$
at successive times $\tau = s, s+1, \ldots, t$ up to time t.

A general policy the consumer can choose
involves a measurable function $h^t \mapsto \psi_t(h^t)$
mapping each known history up to time t,
which determines the consumer's information set,
into a consumption level at that time.

The collection of successive functions $\psi_s^T = \langle \psi_t \rangle_{t=s}^T$
is what a game theorist would call the consumer's strategy
in the extensive form game "against nature".

Markov Strategies

We found an optimal solution
for the two-period problem when $t = T - 1$.

It took the form of a Markov strategy $\psi_t(h^t) := c_t^*(w_t)$,
which depends only on w_t as the particular state variable.

The following analysis will demonstrate in particular
that at each time $t = s, s + 1, \ldots, T$,
under the induction hypothesis that the consumer will follow
a Markov strategy in periods $\tau = t + 1, t + 2, \ldots, T$,
there exists a Markov strategy that is optimal in period t.

It will follow by backward induction
that there exists an optimal strategy $h^t \mapsto \psi_t(h^t)$
for every period $t = s, s + 1, \ldots, T$
that takes the Markov form $h^t \mapsto w_t \mapsto c_t^*(w_t)$.

This treats history as irrelevant, except insofar as it determines
current wealth w_t at the time when c_t has to be chosen.

A Stochastic Difference Equation

Accordingly, suppose that the consumer pursues a Markov strategy taking the form $w_t \mapsto c_t^*(w_t)$.

Then the Markov state variable w_t will evolve over time according to the stochastic difference equation

$$w_{t+1} = \phi_t(w_t, \tilde{r}_t) := \tilde{r}_t(w_t - c_t^*(w_t)).$$

Starting at any time t, conditional on initial wealth w_t, this equation will have a random solution $\tilde{\mathbf{w}}_{t+1}^T = (\tilde{w}_\tau)_{\tau=t+1}^T$ described by a unique joint conditional cdf $F_{t+1}^T(\mathbf{w}_{t+1}^T | w_t)$ on $\mathbb{R}^{T-s}$.

Combined with the Markov strategy $w_t \mapsto c_t^*(w_t)$, this generates a random consumption stream $\tilde{\mathbf{c}}_{t+1}^T = (\tilde{c}_\tau)_{\tau=t+1}^T$ described by a unique joint conditional cdf $G_{t+1}^T(\mathbf{c}_{t+1}^T | w_t)$ on $\mathbb{R}^{T-s}$.

Outline

General Finite Horizon Problem

Consider the objective of choosing
the sequence $(y_s, y_{s+1}, \ldots, y_{T-2}, y_{T-1})$ of controls
in order to maximize

$$\mathbb{E}_s \left[\sum\nolimits_{t=s}^{T-1} u_s(x_s, y_s) + \phi_T(x_T) \right]$$

subject to the law of motion $x_{t+1} = \xi_t(x_t, y_t, \epsilon_t)$,
where the random shocks ϵ_t
at different times $t = s, s+1, s+2, \ldots, T-1$
are conditionally independent given x_t, y_t.

Here $x_T \mapsto \phi_T(x_T)$ is the terminal state valuation function.

The stochastic law of motion can also be expressed
through successive conditional probabilities $\mathbb{P}_{t+1}(x_{t+1}|x_t, y_t)$.

The choices of y_t at successive times determine
a controlled Markov process governing the stochastic transition
from each state x_t to its immediate successor x_{t+1}.

Backward Recurrence Relation

To find the optimal solution, solve the backward recurrence relation

$$\left.\begin{array}{rl} V_s(x_s) & = \\ y_s^*(x_s) & = \arg \end{array}\right\} \max_{y_s \in F_s(x_s)} \{u_s(x_s, y_s) + \mathbb{E}_s [V_{s+1}(x_{s+1}) | x_s, y_s]\}$$

where, for each start time s,

1. x_s denotes the "inherited state" at time s;
2. $V_s(x_s)$ is the current value in state x_s of the state value function $X \ni x \mapsto V_s(x) \in \mathbb{R}$;
3. $X \ni x \mapsto\!\!\!\to F_s(x) \subset Y$ is the feasible set correspondence, with graph $G_s := \{(x, y) \in X \times Y \mid y \in F_s(x)\}$;
4. $G_s \ni (x, y) \mapsto u_s(x, y)$ denotes the immediate return function;
5. $X \ni x \mapsto y_s^*(x) \in F_s(x_s)$ is the optimal "strategy" or policy function;
6. The relevant terminal condition is that $V_T(x_T)$ is given by the exogenously specified function $\phi_T(x_T)$.

Outline

An Infinite Horizon Savings Problem

Game theorists speak of the "one-shot" deviation principle.

This states that if any deviation
from a particular policy or strategy improves a player's payoff,
then there exists a one-shot deviation that improves the payoff.

We consider the infinite horizon extension
of the consumption/investment problem already considered.

Given the initial time s and initial wealth w_s,
this takes the form of choosing a consumption policy $c_t(w_t)$
at the infinite sequence $t = s, s+1, s+2, \ldots$ of times,
in order to maximize the discounted sum of total utility, given by

$$\sum_{t=s}^{\infty} \beta^{t-s} u(c_t)$$

subject to the accumulation equation $w_{t+1} = \tilde{r}_t(w_t - c_t)$
as well as the inequality constraint $w_t \geq 0$ for $t = s+1, s+2, \ldots$.

Some Assumptions

The parameter $\beta \in (0, 1)$ is the constant discount factor.

Note that utility function $\mathbb{R} \ni c \mapsto u(c)$ is independent of t; its first two derivatives are assumed to satisfy the inequalities $u'(c) > 0$ and $u''(c) < 0$ for all $c \in \mathbb{R}_+$.

The investment returns $\tilde{r}_t$ in successive periods are assumed to be i.i.d. random variables.

It is assumed that w_t in each period t is known at time t, but not before.

Terminal Constraint

There has to be an additional constraint
that imposes a lower bound on wealth at some time t.

Otherwise there would be no optimal policy
— the consumer can always gain by increasing debt
(negative wealth), no matter how large existing debt may be.

In the finite horizon,
there was a constraint $w_T \geq 0$ on terminal wealth.

But here T is effectively infinite.

One might try an alternative like

$$\liminf_{t \to \infty} \beta^t w_t \geq 0$$

But this places no limit on wealth at any finite time.

We use the alternative constraint requiring that $w_t \geq 0$ for all time.

Outline

The Stationary Problem

Our modified problem can be written
in the following form that is independent of s:

$$\max_{c_0, c_1, \ldots, c_t, \ldots} \sum_{t=0}^{\infty} \beta^t u(c_t)$$

subject to the constraints $0 \leq c_t \leq w_t$ and $w_{t+1} = \tilde{r}_t(w_t - c_t)$
for all $t = 0, 1, 2, \ldots$, with $w_0 = w$, where w is given.

Because the starting time s is irrelevant,
this is a stationary problem.

Define the state valuation function $w \mapsto V(w)$
as the maximum value of the objective,
as a function of initial wealth w.

It is independent of s because the problem is stationary.

Bellman's Equation

For the finite horizon problem, the principle of optimality was

$$\left.\begin{array}{rcl} V_s(w_s) & = & \\ c_s^*(w_s) & = & \arg \end{array}\right\} \max_{0 \le c_s \le w_s} \{u_s(c_s) + \mathbb{E}_s[V_{s+1}(\tilde{r}_s(w_s - c_s))]\}$$

For the stationary infinite horizon problem, however,
the time starting time s is irrelevant.

So the principle of optimality can be expressed as

$$\left.\begin{array}{rcl} V(w) & = & \\ c^*(w) & = & \arg \end{array}\right\} \max_{0 \le c \le w} \{u(c) + \beta \mathbb{E}[V(\tilde{r}(w - c))]\}$$

The state valuation function $w \mapsto V(w)$ appears
on both left and right hand sides of this equation.

Solving it therefore involves finding a fixed point, or function,
in an appropriate function space.

Isoelastic Case

We consider yet again the isoelastic case
with a CES (or logarithmic) utility function
that satisfies $u'(c;\epsilon) \equiv c^{-\epsilon}$ and, specifically

$$u(c;\epsilon) = \begin{cases} c^{1-\epsilon}/(1-\epsilon) & \text{if } \epsilon \neq 1; \\ \ln c & \text{if } \epsilon = 1. \end{cases}$$

Recall the corresponding finite horizon case,
where we found that the solution to the corresponding equations

$$\left.\begin{array}{rcl} V_s(w_s) & = & \\ c_s^*(w_s) & = & \arg \end{array}\right\} \max_{0 \leq c_s \leq w_s} \{u_s(c_s) + \beta \mathbb{E}_s[V_{s+1}(\tilde{r}_s(w_s - c_s))]\}$$

takes the form: (i) $V_s(w) = \alpha_s + v_s u(w;\epsilon)$
for suitable real constants α_s and $v_s > 0$, where $\alpha_s = 0$ if $\epsilon \neq 1$;
(ii) $c_s^*(w_s) = \gamma_s\, w_s$ for a suitable constant $\gamma_s \in (0,1)$.

First-Order Condition

Accordingly, we look for a solution to the stationary problem

$$\left.\begin{array}{rcl} V(w) & = & \\ c^*(w) & = & \arg \end{array}\right\} \max_{0 \leq c \leq w} \{u(c;\epsilon) + \beta \mathbb{E}[V(\tilde{r}(w-c))]\}$$

taking the isoelastic form $V(w) = \alpha + vu(w;\epsilon)$
for suitable real constants α and $v > 0$, where $\alpha = 0$ if $\epsilon \neq 1$.

The first-order condition for solving
this concave maximization problem is

$$c^{-\epsilon} = \beta \mathbb{E}[\tilde{r}(\tilde{r}(w-c))^{-\epsilon}] = \zeta^\epsilon (w-c)^{-\epsilon}$$

where $\zeta^\epsilon := \beta R^{1-\epsilon}$ with R as the certainty equivalent return
defined by $R^{1-\epsilon} := \mathbb{E}[\tilde{r}^{1-\epsilon}]$.

Hence $c = \gamma w$ where $\gamma^{-\epsilon} = \zeta^\epsilon (1-\gamma)^{-\epsilon}$,
implying that $\zeta = (1-\gamma)/\gamma$, the savings–consumption ratio.

Then $\gamma = 1/(1+\zeta)$, so $1 - \gamma = \zeta/(1+\zeta)$.

Solution in the Logarithmic Case

When $\epsilon = 1$ and so $u(c; \epsilon) = \ln c$, one has

$$\begin{aligned} V(w) &= u(\gamma w; \epsilon) + \beta\{\alpha + v\mathbb{E}[u(\tilde{r}(1-\gamma)w; \epsilon)]\} \\ &= \ln(\gamma w) + \beta\{\alpha + v\mathbb{E}[\ln(\tilde{r}(1-\gamma)w)]\} \\ &= \ln\gamma + (1+\beta v)\ln w + \beta\{\alpha + v\ln(1-\gamma) + \mathbb{E}[\ln\tilde{r}]\} \end{aligned}$$

This is consistent with $V(w) = \alpha + v\ln w$ in case:

1. $v = 1 + \beta v$, implying that $v = (1-\beta)^{-1}$;
2. and also $\alpha = \ln\gamma + \beta\{\alpha + v\ln(1-\gamma) + \mathbb{E}[\ln\tilde{r}]\}$, which implies that

$$\alpha = (1-\beta)^{-1}\left[\ln\gamma + \beta\left\{(1-\beta)^{-1}\ln(1-\gamma) + \mathbb{E}[\ln\tilde{r}]\right\}\right]$$

This confirms the solution for the logarithmic case.

Solution in the CES Case

When $\epsilon \neq 1$ and so $u(c; \epsilon) = c^{1-\epsilon}/(1-\epsilon)$, the equation

$$V(w) = u(\gamma w; \epsilon) + \beta v \mathbb{E}[u(\tilde{r}(1-\gamma)w; \epsilon)]$$

implies that

$$(1-\epsilon)V(w) = (\gamma w)^{1-\epsilon} + \beta v \mathbb{E}[(\tilde{r}(1-\gamma)w)^{1-\epsilon}] = vw^{1-\epsilon}$$

Hence $v = \gamma^{1-\epsilon} + \beta v(1-\gamma)^{1-\epsilon} R^{1-\epsilon}$, so with $\zeta^\epsilon = \beta R^{1-\epsilon}$ one has

$$v = \frac{\gamma^{1-\epsilon}}{1 - \beta(1-\gamma)^{1-\epsilon} R^{1-\epsilon}} = \frac{\gamma^{1-\epsilon}}{1 - (1-\gamma)^{1-\epsilon}\zeta^\epsilon}$$

But optimality requires $\gamma = 1/(1+\zeta)$, implying finally that

$$v = \frac{(1+\zeta)^{\epsilon-1}}{1 - \zeta(1+\zeta)^{\epsilon-1}} = \frac{1}{(1+\zeta)^{1-\epsilon} - \zeta}$$

This confirms the solution for the CES case.

Outline

Uniformly Bounded Returns

Suppose that the stochastic transition from each state x to the immediately succeeding state $\tilde{x}$ is specified by a conditional probability measure $B \mapsto \mathbb{P}(\tilde{x} \in B | x, u)$ on a σ-algebra of the state space.

Consider the stationary problem of choosing a policy $x \mapsto u^*(x)$ in order to maximize the infinite discounted sum of utility

$$\mathbb{E} \sum_{t=1}^{\infty} \beta^{t-1} f(x_t, u_t)$$

where $0 < \beta < 1$, with x_1 given and subject to $u_t \in U(x_t)$ for $t = 1, 2, \ldots$.

The return function $(x, u) \mapsto f(x, u) \in \mathbb{R}$ is uniformly bounded provided there exist a uniform lower bound M_* and a uniform upper bound M^* such that

$$M_* \leq f(x, u) \leq M^* \quad \text{for all } (x, u) \text{ with } u \in U(x)$$

Bounds on Discounted Total Returns

The boundedness assumption $M_* \leq f(x, u) \leq M^*$ for all (x, u)

ensures that for all $T \in \mathbb{N}$, the finite sum

$$W_T := \mathbb{E} \sum_{t=1}^{T} \beta^{t-1} f(x_t, u_t)$$

satisfies $\underline{W}_T \leq W_T \leq \bar{W}_T$ where

$$\underline{W}_T := \sum_{t=1}^{T} \beta^{t-1} M_* = \frac{1 - \beta^T}{1 - \beta} M_*$$

$$\bar{W}_T := \sum_{t=1}^{T} \beta^{t-1} M^* = \frac{1 - \beta^T}{1 - \beta} M^*$$

Then, because $0 < \beta < 1$ and so $\sum_{t=1}^{\infty} \beta^{t-1} = \dfrac{1}{1 - \beta}$,

the infinite discounted sum of utility

$$W := \mathbb{E} \sum_{t=1}^{\infty} \beta^{t-1} f(x_t, u_t)$$

satisfies $(1 - \beta) W \in [M_*, M^*]$.

The Function Space

This makes it natural to consider the linear (Banach) space $\mathcal{V}$ of all bounded functions $X \ni x \mapsto V(x) \in \mathbb{R}$ equipped with its sup norm defined by $\|V\| := \sup_{x \in X} |V(x)|$.

We will pay special attention to the subset

$$\mathcal{V}_M := \{V \in \mathcal{V} \mid x \in X \Longrightarrow (1-\beta)V(x) \in [M_*, M^*]\}$$

of state valuation functions whose values $V(x)$ all lie within the range of the possible values of W.

Existence and Uniqueness

Theorem

Consider the Bellman equation system

$$\left.\begin{array}{rcl} V(x) & = & \\ u^*(x) & \in & \arg \end{array}\right\} \max_{u \in F(x)} \{f(x,u) + \beta \mathbb{E}[V(\tilde{x})|x,u]\}$$

Under the assumption of uniformly bounded returns satisfying $M_* \leq f(x,u) \leq M^*$ for all (x,u):

1. *among the set $\mathcal{V}_M$*
   of state valuation functions $X \ni x \mapsto V(x) \in \mathbb{R}$,
   that satisfy the inequalities $M_* \leq (1-\beta)V(x) \leq M^*$ for all x,
   there is a unique state valuation function $x \mapsto V(x)$
   that satisfies the Bellman equation system.
2. *any associated policy solution $X \ni x \mapsto u^*(x) \in U(x)$*
   determines an optimal policy that is stationary
   — i.e., independent of time.

Two Functionals

Given any measurable policy function $X \ni x \mapsto u(x) \in U$ denoted by $\mathbf{u}$, define the mapping $\mathcal{V}_M \ni V \mapsto T^{\mathbf{u}} V \in \mathcal{V}$ so that for all $x \in X$ one has

$$[T^{\mathbf{u}} V](x) := f(x, u(x)) + \beta \mathbb{E}\left[V(\tilde{x}) | x, u(x)\right]$$

When the state is x, this gives the value $[T^{\mathbf{u}} V](x)$ of choosing the policy $u(x)$ for one period, and then experiencing a future discounted return $V(\tilde{x})$ after reaching each possible subsequent state $\tilde{x} \in X$.

Define also the mapping $\mathcal{V}_M \ni V \mapsto T^* V \in \mathcal{V}$ so that for all $x \in X$ one has

$$[T^* V](x) := \max_{u \in F(x)} \{f(x, u) + \beta \mathbb{E}\left[V(\tilde{x}) | x, u\right]\}$$

Compressed Bellman Equation

The Bellman equation system is

$$\left.\begin{array}{rcl} V(x) & = & \\ u^*(x) & \in & \arg \end{array}\right\} \max_{u \in F(x)} \{f(x,u) + \beta \mathbb{E}[V(\tilde{x})|x,u]\}$$

We have defined the two functionals that map $\mathcal{V}_M$ into $\mathcal{V}$ so that, for all $x \in X$, one has

$$[T^{\mathbf{u}}V](x) := f(x, u(x)) + \beta \mathbb{E}[V(\tilde{x})|x, u(x)]$$
$$[T^*V](x) := \max_{u \in F(x)} \{f(x,u) + \beta \mathbb{E}[V(\tilde{x})|x,u]\}$$

The definitions of these two functionals
allow the Bellman equation system to be rewritten as

$$\begin{array}{rcl} V(x) & = & [T^*V](x) \\ u^*(x) & \in & \arg\max_{u \in F(x)} [T^{\mathbf{u}}V](x) \end{array}$$

In particular, the state valuation function $X \ni x \mapsto V(x)$ should be a fixed point of T^* in the space $\mathcal{V}_M$.

Two Mappings of $\mathcal{V}_M$ into Itself

For all $V \in \mathcal{V}_M$, policies $\mathbf{u}$, and $x \in X$, we have defined

$$\begin{aligned} [T^{\mathbf{u}} V](x) &:= f(x, u(x)) + \beta\, \mathbb{E}\,[V(\tilde{x})|x, u(x)] \\ \text{and} \quad [T^* V](x) &:= \max_{u \in F(x)} \{f(x, u) + \beta\, \mathbb{E}\,[V(\tilde{x})|x, u]\} \end{aligned}$$

Recall the uniform boundedness condition $M_* \leq f(x, u) \leq M^*$, together with the assumption that V belongs to the domain $\mathcal{V}_M$ of functions satisfying $M_* \leq (1 - \beta) V(\tilde{x}) \leq M^*$ for all $\tilde{x}$.

So these two definitions jointly imply that, for all $x \in X$, one has

$$\begin{aligned} [T^{\mathbf{u}} V](x) &\geq M_* + \beta\,(1 - \beta)^{-1}\, M_* = (1 - \beta)^{-1}\, M_* \\ \text{and} \quad [T^{\mathbf{u}} V](x) &\leq M^* + \beta\,(1 - \beta)^{-1}\, M^* = (1 - \beta)^{-1}\, M^* \end{aligned}$$

Similarly, given any $V \in \mathcal{V}_M$, for all $x \in X$
one has $M_* \leq (1 - \beta)\,[T^* V](x) \leq M^*$ for all $x \in X$.

Therefore both functionals $V \mapsto T^{\mathbf{u}} V$ and $V \mapsto T^* V$ map the set $\mathcal{V}_M$ of bounded functions into itself.

A First Contraction Mapping

The definition $[T^{\mathbf{u}} V](x) := f(x, u(x)) + \beta\, \mathbb{E}\, [V(\tilde{x})|x, u(x)]$ implies that for any two functions $V_1, V_2 \in \mathcal{V}_M$, one has

$$[T^{\mathbf{u}} V_1](x) - [T^{\mathbf{u}} V_2](x) = \beta\, \mathbb{E}\, [V_1(\tilde{x}) - V_2(\tilde{x})|x, u(x)]$$

The definition of the sup norm therefore implies that

$$\begin{aligned} \| T^{\mathbf{u}} V_1 - T^{\mathbf{u}} V_2 \| &= \sup_{x \in X} |[T^{\mathbf{u}} V_1](x) - [T^{\mathbf{u}} V_2](x)| \\ &= \sup_{x \in X} |\beta\, \mathbb{E}\, [V_1(\tilde{x}) - V_2(\tilde{x})|x, u(x)]\,| \\ &= \beta\, \sup_{x \in X} |\mathbb{E}\, [V_1(\tilde{x}) - V_2(\tilde{x})|x, u(x)]\,| \\ &\leq \beta\, \sup_{\tilde{x} \in X} |V_1(\tilde{x}) - V_2(\tilde{x})| \\ &= \beta\, \| V_1 - V_2 \| \end{aligned}$$

Hence $\mathcal{V}_M \ni V \mapsto T^{\mathbf{u}} V \in \mathcal{V}_M$
is a contraction mapping with factor $\beta < 1$.

Applying the Contraction Mapping Theorem, I

For each fixed policy $\mathbf{u}$, the contraction mapping

$$\mathcal{V}_M \ni V \mapsto T^{\mathbf{u}} V \in \mathcal{V}_M$$

has a unique fixed point in the form of a function $V^{\mathbf{u}} \in \mathcal{V}_M$.

Furthermore, given any initial function $V \in \mathcal{V}_M$,
consider the infinite sequence of mappings $[T^{\mathbf{u}}]^k V$ $(k \in \mathbb{N})$
that result from applying the operator $T^{\mathbf{u}}$ iteratively k times.

The contraction mapping property of $T^{\mathbf{u}}$
implies that $\|[T^{\mathbf{u}}]^k V - V^{\mathbf{u}}\| \to 0$ as $k \to \infty$.

Characterizing the Fixed Point, I

Starting from $V_0 = 0$ and given any initial state $x \in X$, note that

$$\begin{aligned} [T^{\mathbf{u}}]^k V_0(x) &= [T^{\mathbf{u}}]\left([T^{\mathbf{u}}]^{k-1} V_0\right)(x) \\ &= f(x, u(x)) + \beta\, \mathbb{E}\left[\left([T^{\mathbf{u}}]^{k-1} V_0\right)(\tilde{x}) | x, u(x)\right] \end{aligned}$$

It follows by induction on k that $[T^{\mathbf{u}}]^k V_0(\bar{x})$ equals
the expected discounted total payoff $\mathbb{E} \sum_{t=1}^{k} \beta^{t-1} f(x_t, u_t)$
of starting from $x_0 = \bar{x}$
and then following the policy $x \mapsto u(x)$ for k subsequent periods.

Taking the limit as $k \to \infty$, it follows that for any state $\bar{x} \in X$,
the value $V^{\mathbf{u}}(\bar{x})$ of the fixed point in $\mathcal{V}_M$
is the expected discounted total payoff

$$\mathbb{E} \sum_{t=1}^{\infty} \beta^{t-1} f(x_t, u_t)$$

of starting from $x_0 = \bar{x}$
and then following the policy $x \mapsto u(x)$ for ever thereafter.

A Second Contraction Mapping

For each state $x \in X$, recall the definition

$$[T^* V](x) := \max_{u \in F(x)} \{ f(x,u) + \beta \mathbb{E} [V(\tilde{x}) | x, u] \}$$

Given any state $x \in X$ and any two functions $V_1, V_2 \in \mathcal{V}_M$, define $u_1, u_2 \in F(x)$ so that for $k = 1, 2$ one has

$$[T^* V_k](x) = f(x, u_k) + \beta \mathbb{E} [V_k(\tilde{x}) | x, u_k] \}$$

Note that $[T^* V_2](x) \geq f(x, u_1) + \beta \mathbb{E} [V_2(\tilde{x}) | x, u_1] \}$ implying that

$$\begin{aligned} [T^* V_1](x) - [T^* V_2](x) &\leq \beta \mathbb{E} [V_1(\tilde{x}) - V_2(\tilde{x}) | x, u_1] \} \\ &\leq \beta \| V_1 - V_2 \| \end{aligned}$$

Similarly, interchanging 1 and 2 in the above argument gives $[T^* V_2](x) - [T^* V_1](x) \leq \beta \| V_1 - V_2 \|$.

Hence $\| T^* V_1 - T^* V_2 \| \leq \beta \| V_1 - V_2 \|$, so T^* is also a contraction.

Applying the Contraction Mapping Theorem, II

Similarly the contraction mapping $V \mapsto T^* V$
has a unique fixed point in the form of a function $V^* \in \mathcal{V}_M$
such that $V^*(\bar{x})$ is the maximized expected discounted total payoff
of starting in state $x_0 = \bar{x}$
and following an optimal policy for ever thereafter.

Moreover, $V^* = T^* V^* = T^{\mathbf{u}^*} V^*$.

This implies that V^* is also the value
of following the policy $x \mapsto u^*(x)$ throughout,
which must therefore be an optimal policy.

Characterizing the Fixed Point, II

Starting from $V_0 = 0$ and given any initial state $x \in X$, note that

$$\begin{aligned} [T^*]^k V_0(x) &= [T^*] \left([T^*]^{k-1} V_0\right)(x) \\ &= \max_{u \in F(x)} \{ f(x,u) + \beta \, \mathbb{E}\left[\left([T^*]^{k-1} V_0\right)(\tilde{x}) | x, u\right] \} \end{aligned}$$

It follows by induction on k
that $[T^*]^k V_0(\bar{x})$ equals the maximum possible
expected discounted total payoff $\mathbb{E} \sum_{t=1}^{k} \beta^{t-1} f(x_t, u_t)$
of starting from $x_1 = \bar{x}$
and then following the "backward" sequence
of optimal policies $(u_k^*, u_{k-1}^*, u_{k-2}^*, \ldots, u_2^*, u_1^*)$,
where for each k the policy $x \mapsto u_k^*(x)$
is optimal when k periods remain.

Outline

Method of Successive Approximation

The method of successive approximation starts with an arbitrary function $V_0 \in \mathcal{V}_M$.

For $k = 1, 2, \ldots,$, it then repeatedly solves the pair of equations $V_k = T^* V_{k-1} = T^{u_k^*} V_{k-1}$ to construct sequences of:

1. state valuation functions $X \ni x \mapsto V_k(x) \in \mathbb{R}$;
2. policies $X \ni x \mapsto u_k^*(x) \in F(x)$ that are optimal given that one applies the preceding state valuation function $X \ni \tilde{x} \mapsto V_{k-1}(\tilde{x}) \in \mathbb{R}$ to each immediately succeeding state $\tilde{x}$.

Because the operator $V \mapsto T^* V$ on $\mathcal{V}_M$ is a contraction mapping, the method produces a convergent sequence $(V_k)_{k=1}^\infty$ of state valuation functions whose limit satisfies $V^* = T^* V^* = T^{u^*} V^*$ for a suitable policy $X \ni x \mapsto u^*(x) \in F(x)$.

Monotonicity

For all functions $V \in \mathcal{V}_M$, policies $\mathbf{u}$ in the form of functions $X \ni x \mapsto u(x) \in F(x)$, and states $x \in X$, we have defined

$$\begin{aligned} [T^{\mathbf{u}} V](x) &:= f(x, u(x)) + \beta\, \mathbb{E}\, [V(\tilde{x}) | x, u(x)] \\ \text{and} \quad [T^* V](x) &:= \max_{u \in F(x)} \{ f(x, u) + \beta\, \mathbb{E}\, [V(\tilde{x}) | x, u] \} \end{aligned}$$

Notation

Given any pair $V_1, V_2 \in \mathcal{V}_M$, we write $V_1 \geqq V_2$ to indicate that the inequality $V_1(x) \geq V_2(x)$ holds for all $x \in X$.

Definition

An operator $\mathcal{V}_M \ni V \mapsto TV \in \mathcal{V}_M$ is monotone just in case whenever $V_1, V_2 \in \mathcal{V}_M$ satisfy $V_1 \geqq V_2$, one has $TV_1 \geqq TV_2$.

Theorem

The following operators on $\mathcal{V}_M$ are monotone:

1. $V \mapsto T^{\mathbf{u}} V$ *for all policies* $\mathbf{u}$*;*
2. $V \mapsto T^* V$ *for the optimal policy.*

The proofs will occupy the next two slides.

Proof that the Operator $T^{\mathbf{u}}$ is Monotone

Given any state $x \in X$ and any two functions $V_1, V_2 \in \mathcal{V}_M$, the definition of $T^{\mathbf{u}}$ implies that

$$\begin{array}{rrcl} & [T^{\mathbf{u}} V_1](x) & := & f(x, u(x)) + \beta \, \mathbb{E} \, [V_1(\tilde{x}) | x, u(x)] \\ \text{and} & [T^{\mathbf{u}} V_2](x) & := & f(x, u(x)) + \beta \, \mathbb{E} \, [V_2(\tilde{x}) | x, u(x)] \end{array}$$

Subtracting the second equation from the first implies that

$$[T^{\mathbf{u}} V_1](x) - [T^{\mathbf{u}} V_2](x) = \beta \, \mathbb{E} \, [V_1(\tilde{x}) - V_2(\tilde{x}) | x, u(x)]$$

If $V_1 \geqq V_2$ and so the inequality $V_1(\tilde{x}) \geq V_2(\tilde{x})$ holds for all $\tilde{x} \in X$, it follows that $[T^{\mathbf{u}} V_1](x) \geq [T^{\mathbf{u}} V_2](x)$.

Since this holds for all $x \in X$,
we have proved that $T^{\mathbf{u}} V_1 \geqq T^{\mathbf{u}} V_2$. □

Proof that the Operator T^* is Monotone

Given any state $x \in X$ and any two functions $V_1, V_2 \in \mathcal{V}_M$, define the two feasible policies $\mathbf{u}_1, \mathbf{u}_2$ so that for $k = 1, 2$ one has

$$\begin{aligned} [T^* V_k](x) &= \max_{u \in F(x)} \{ f(x, u) + \beta \, \mathbb{E} \, [V_k(\tilde{x}) | x, u] \} \\ = \quad [T^{\mathbf{u}_k} V_k](x) &= f(x, u_k) + \beta \, \mathbb{E} \, [V_k(\tilde{x}) | x, u_k] \end{aligned}$$

Because $\mathbf{u}_2$ may be suboptimal given V_1, it follows that

$$\begin{aligned} [T^* V_1](x) &\geq f(x, u_2) + \beta \, \mathbb{E} \, [V_1(\tilde{x}) | x, u_2] \\ \text{whereas} \quad [T^* V_2](x) &= f(x, u_2) + \beta \, \mathbb{E} \, [V_2(\tilde{x}) | x, u_2] \end{aligned}$$

Subtracting the second equation from the first inequality gives

$$[T^* V_1](x) - [T^* V_2](x) \geq \beta \, \mathbb{E} \, [V_1(\tilde{x}) - V_2(\tilde{x}) | x, u_2]$$

If $V_1 \geqq V_2$ and so the inequality $V_1(\tilde{x}) \geq V_2(\tilde{x})$ holds for all $\tilde{x} \in X$, it follows that $[T^* V_1](x) \geq [T^* V_2](x)$.

Since this inequality holds for all $x \in X$, we have proved that $T^* V_1 \geqq T^* V_2$. □

The Method of Policy Improvement

The method of policy improvement starts
with any fixed policy $\mathbf{u}_0$ or $X \ni x \mapsto u_0(x) \in F(x)$,
along with the value $V^{\mathbf{u}_0} \in \mathcal{V}_M$ of following that policy for ever.

Note that, among the domain $\mathcal{V}_M$ of suitably bounded functions,
the value $V^{\mathbf{u}_0}$ is the unique fixed point satisfying $V^{\mathbf{u}_0} = T^{\mathbf{u}_0} V^{\mathbf{u}_0}$.

At each step $k = 1, 2, \ldots$, given the previous policy $\mathbf{u}_{k-1}$
and associated value $V^{\mathbf{u}_{k-1}}$ satisfying $V^{\mathbf{u}_{k-1}} = T^{\mathbf{u}_{k-1}} V^{\mathbf{u}_{k-1}}$,
choose:

1. the policy $\mathbf{u}_k$ or $X \ni x \mapsto u_k(x) \in F(x)$ optimally,
   so that $T^* V^{\mathbf{u}_{k-1}} = T^{\mathbf{u}_k} V^{\mathbf{u}_{k-1}}$;
2. the state valuation function $x \mapsto V_k(x)$
   as the unique fixed point in $\mathcal{V}_M$ of the operator $T^{\mathbf{u}_k}$.

Policy Improvement Theorem

Theorem

The infinite sequence $(\mathbf{u}_k, V^{\mathbf{u}_k})_{k \in \mathbb{N}}$
consisting of pairs of policies $\mathbf{u}_k$
with their associated valuation functions $V^{\mathbf{u}_k} \in \mathcal{V}_M$ *satisfies*

1. $V^{\mathbf{u}_k} \geqq V^{\mathbf{u}_{k-1}}$ *for all* $k \in \mathbb{N}$ *(policy improvement);*
2. $\|V^{\mathbf{u}_k} - V^*\| \to 0$ *as* $k \to \infty$*,*
   where V^* *is the infinite-horizon optimal*
   state valuation function in $\mathcal{V}_M$ *that satisfies* $T^* V^* = V^*$*.*

The proof will occupy the next three slides.

Proof of Policy Improvement

By definition of the optimality operator T^*,
one has $T^* V \geqq T^{\mathbf{u}} V$ for all functions $V \in \mathcal{V}_M$ and all policies $\mathbf{u}$.

So at each step k of the policy improvement routine, one has

$$T^{\mathbf{u}_k} V^{\mathbf{u}_{k-1}} = T^* V^{\mathbf{u}_{k-1}} \geqq T^{\mathbf{u}_{k-1}} V^{\mathbf{u}_{k-1}} = V^{\mathbf{u}_{k-1}}$$

In particular, $V^{\mathbf{u}_{k-1}} \leqq T^{\mathbf{u}_k} V^{\mathbf{u}_{k-1}}$.

Because the operator $T^{\mathbf{u}_k}$ is monotonic, applying it iteratively implies that

$$V^{\mathbf{u}_{k-1}} \leqq T^{\mathbf{u}_k} V^{\mathbf{u}_{k-1}} \leqq [T^{\mathbf{u}_k}]^2 V^{\mathbf{u}_{k-1}} \leqq \dots$$
$$\dots \leqq [T^{\mathbf{u}_k}]^r V^{\mathbf{u}_{k-1}} \leqq [T^{\mathbf{u}_k}]^{r+1} V^{\mathbf{u}_{k-1}} \leqq \dots$$

But the definition of $V^{\mathbf{u}_k}$ implies that for all $V \in \mathcal{V}_M$,
including $V = V^{\mathbf{u}_{k-1}}$, one has $\|[T^{\mathbf{u}_k}]^r V - V^{\mathbf{u}_k}\| \to 0$ as $r \to \infty$.

Hence $V^{\mathbf{u}_k} = \sup_r [T^{\mathbf{u}_k}]^r V^{\mathbf{u}_{k-1}} \geqq V^{\mathbf{u}_{k-1}}$,
thus confirming that the policy $\mathbf{u}_k$ does improve $\mathbf{u}_{k-1}$. □

Proof of Convergence, I

Recall that at each step $k \in \mathbb{N}$ of the policy improvement routine, one has $T^{\mathbf{u}_k} V^{\mathbf{u}_{k-1}} = T^* V^{\mathbf{u}_{k-1}}$ and also $T^{\mathbf{u}_k} V^{\mathbf{u}_k} = V^{\mathbf{u}_k}$.

Now, for each state $x \in X$, define $\hat{V}(x) := \sup_{k \in \mathbb{N}} V^{\mathbf{u}_k}(x)$.

First, because $V^{\mathbf{u}_k} \geqq V^{\mathbf{u}_{k-1}}$ and $T^{\mathbf{u}_k}$ is monotonic, one has $V^{\mathbf{u}_k} = T^{\mathbf{u}_k} V^{\mathbf{u}_k} \geqq T^{\mathbf{u}_k} V^{\mathbf{u}_{k-1}} = T^* V^{\mathbf{u}_{k-1}}$.

Then, because $V^{\mathbf{u}_k} \geqq V^{\mathbf{u}_{k-1}}$ and T^* is monotonic, it follows that

$$\hat{V} = \sup_k V^{\mathbf{u}_k} \geqq \sup_k T^* V^{\mathbf{u}_{k-1}} = T^*(\sup_k V^{\mathbf{u}_{k-1}}) = T^* \hat{V}$$

Second, the definition of T^* implies that $V^{\mathbf{u}_k} = T^{\mathbf{u}_k} V^{\mathbf{u}_k}$.

Similarly $V^{\mathbf{u}_k} = T^{\mathbf{u}_k} V^{\mathbf{u}_k} \leqq T^* V^{\mathbf{u}_k}$ for each $k \in \mathbb{N}$.

Because T^* is monotonic, it follows that

$$\hat{V} = \sup_k V^{\mathbf{u}_k} = \sup_k T^{\mathbf{u}_k} V^{\mathbf{u}_k} \leqq \sup_k T^* V^{\mathbf{u}_k} = T^*(\sup_k V^{\mathbf{u}_k}) = T^* \hat{V}$$

So we have proved that $\hat{V} \geqq T^* \hat{V}$ and that $\hat{V} \leqq T^* \hat{V}$, which together imply that $\hat{V} = T^* \hat{V}$.

Proof of Convergence, II

We have proved that $\hat{V} = T^* \hat{V}$ where $\hat{V} = \sup_k V^{\mathbf{u}_k}$.

But earlier we defined V^* as the unique fixed point of the contraction mapping T^*, satisfying $V^* = T^* V^* = T^{\mathbf{u}^*} V^*$.

It follows that $V^* = \hat{V} = \sup_k V^{\mathbf{u}_k}$.

But by construction the sequence $V^{\mathbf{u}_k}(x)$ is non-decreasing.

It follows that $V^{\mathbf{u}_k}(x) \to V^*(x)$ for each $x \in X$. $\square$

Lecture Outline

Unbounded Utility

In economics the boundedness condition $M_* \leq f(x, u) \leq M^*$ is rarely satisfied!

Consider for example, for each parameter value $\epsilon > 0$, the isoelastic utility function $\mathbb{R}_{++} \ni c \mapsto u(c; \epsilon) \in \mathbb{R}$ defined by

$$u(c; \epsilon) = \begin{cases} \dfrac{c^{1-\epsilon}}{1-\epsilon} & \text{if } \epsilon > 0 \text{ and } \epsilon \neq 1 \\ \ln c & \text{if } \epsilon = 1 \end{cases}$$

This function is obviously:

1. bounded below but unbounded above in case $0 < \epsilon < 1$;
2. unbounded both above and below in case $\epsilon = 1$;
3. bounded above but unbounded below in case $\epsilon > 1$.

Also commonly used is the negative exponential utility function defined by $u(c) = -e^{-\alpha c}$
where α is the constant absolute rate of risk aversion (CARA).

This function is bounded above and, provided that $c \geq 0$, also bounded below.

Warning Example: The Optimization Problem

The following example shows that there can be irrelevant unbounded solutions to the Bellman equation.

Example

Consider the problem of maximizing $\sum_{t=0}^{\infty} \beta^t (1 - u_t)$

where $u_t \in [0, 1]$, $0 < \beta < 1$, and $x_{t+1} = \frac{1}{\beta}(x_t + u_t)$, with $x_0 > 0$.

Notice that $x_{t+1} \geq \frac{1}{\beta} x_t$ implying that $x_t \geq \beta^{-t} x_0 \to \infty$ as $t \to \infty$.

Of course the return function $[0, 1] \ni u \mapsto f(x, u) = 1 - u \in [0, 1]$ is uniformly bounded.

Warning Example: Unbounded Spurious Solution

The Bellman equation is

$$\left.\begin{array}{rcl} J(x) & = & \\ u^*(x) & = & \arg \end{array}\right\} \max_{u\in[0,1]} \left\{ 1 - u + \beta J\left(\frac{1}{\beta}(x+u)\right)\right\}$$

Even though the return function is uniformly bounded,
this Bellman equation has an unbounded spurious solution.

Indeed, we find a spurious solution
with $J(x) \equiv \gamma + x$ for a suitable constant γ.

The condition for this to solve the Bellman equation is that

$$\begin{aligned} \gamma + x &= \max_{u\in[0,1]} \left\{ 1 - u + \beta \left[\gamma + \frac{1}{\beta}(x+u)\right]\right\} \\ &= \max_{u\in[0,1]} \{1 + \beta\gamma + x\} = 1 + \beta\gamma + x \end{aligned}$$

This is true if and only if $\gamma = 1 + \beta\gamma$, implying that $\gamma = (1-\beta)^{-1}$.

Warning Example: True Solution

The problem is to maximize $\sum_{t=0}^{\infty} \beta^t (1 - u_t)$

where $u_t \in [0, 1]$, $0 < \beta < 1$, and $x_{t+1} = \frac{1}{\beta}(x_t + u_t)$, with $x_0 > 0$.

The obvious optimal policy is to choose $u_t = 0$ for all t,
giving the maximized value $J(x) = \sum_{t=0}^{\infty} \beta^t = (1 - \beta)^{-1}$.

Indeed, consider the bounded function $J(x) = (1 - \beta)^{-1}$,
together with $u^* = 0$, both independent of x.

These do indeed solve the Bellman equation because

$$\begin{aligned} J(x) &= \max_{u \in [0,1]} \left\{ 1 - u + \beta J\left(\frac{1}{\beta}(x + u)\right) \right\} \\ &= \max_{u \in [0,1]} \left\{ 1 - u + \beta (1 - \beta)^{-1} \right\} \\ &= 1 + \frac{\beta}{1 - \beta} = \frac{1}{1 - \beta} \text{ when } u = 0 \quad \square \end{aligned}$$