Dynamic Programming

Chapter 6: Stochastic Discounting

Thomas J. Sargent and John Stachurski

2023

Topics

- State-dependent discounting: motivation
- State-dependent discounting: valuation
- Asset pricing
- MDPs with state-dependent discounting

Motivation

One limitation of MDP model: discount rate is constant

This assumption can be problematic

Example. Cannot handle time preference shocks

- Krusell-Smith 1998
- Woodford 2011
- Christiano et al. 2011, 2014
- Schorfheide et al. 2018
- etc, etc.

Also, firms discount future profits using interest rates — which are stochastic and time-varying

Example.

- nominal rates for safe assets like US T-bills
- real interest rates
- rental cost of capital

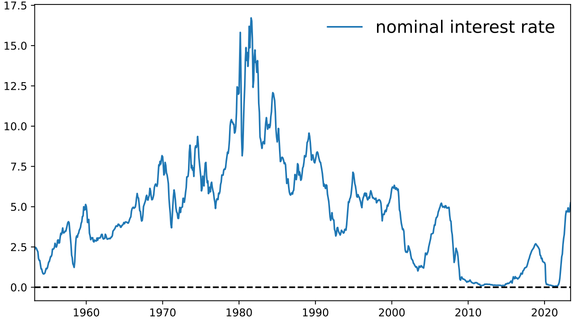

Figure: Nominal US interest rates

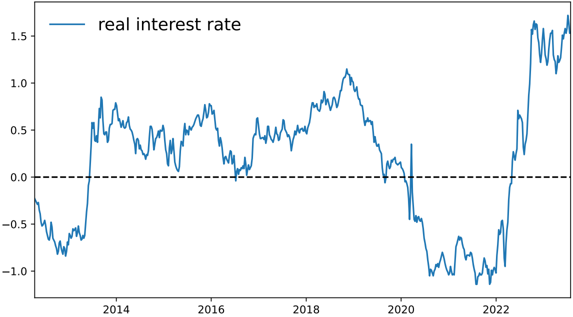

Figure: Real US interest rates

In this chapter we extend the MDP model to handle state-dependent discounting

Learn how to

1. compute lifetime values under state-dependent discounting
2. maximize these values via dynamic programming

We start with step 1...

Valuation under state-dependent discounting

Example. Consider a firm valuation problem where the interest rate follows stochastic process $(r_t)_{t \geqslant 0}$

The time zero expected present value of time t profit π_t is

$$\mathbb{E}\left\{\beta_1 \cdots \beta_t \cdot \pi_t\right\} \quad \text{where} \quad \beta_t := \frac{1}{1 + r_t}$$

The expected present value of the firm is

$$V_0 = \mathbb{E} \sum_{t=0}^{\infty} \left[\prod_{i=0}^{t} \beta_i \right] \pi_t \quad \text{where} \quad \beta_0 := 1$$

Questions:

- When is this valuation finite?
- How can we compute it?
- Are there any general results?

The next section answers these questions

Generalized geometric sums

Suppose

- X is finite and $P \in \mathcal{M}(\mathbb{R}^{\mathsf{X}})$
- $h \in \mathbb{R}^{\mathsf{X}}$
- b is a map from X × X to $(0, \infty)$

Let $(X_t)_{t \geqslant 0}$ be P-Markov and let

$$\beta_t := b(X_{t-1}, X_t) \text{ for } t \in \mathbb{N} \quad \text{with} \quad \beta_0 := 1$$

Let A be the **discount operator** defined by

$$A(x, x') := b(x, x') P(x, x')$$

Theorem. If $\rho(A) < 1$, then, for all $x \in \mathsf{X}$,

$$v(x) := \mathbb{E}_x \sum_{t=0}^{\infty} \left[\prod_{i=0}^{t} \beta_i \right] h(X_t) < \infty$$

Moreover, this function v satisfies

$$v = (I - A)^{-1} h = \sum_{t \geqslant 0} A^t h$$

Remark: $I - A$ is bijective by the Neumann series lemma (NSL)

Proof: Let all the primitives be as stated with $\rho(A) < 1$

As a first step, note that, for all $t \in \mathbb{N}$, $h \in \mathbb{R}^{\mathsf{X}}$ and $x \in \mathsf{X}$,

$$\mathbb{E}_x \left[\prod_{i=0}^{t} \beta_i \right] h(X_t) = (A^t h)(x)$$

Proof for $t = 1$:

$$\mathbb{E}_x \, \beta_1 \, h(X_1) = \sum_{x'} h(x') b(x, x') P(x, x') = (Ah)(x)$$

Ex. Extend this argument to general t

- Hint: use induction and law of iterated expectations
- A solution can be found in the book (Ch. 6)

Interpretation:

$$(A^t h)(x) = \text{ time zero present value of } h(X_t) \text{ given } X_0 = x$$

Now, fixing $x \in \mathsf{X}$, we have

$$v(x) = \mathbb{E}_x \sum_{t=0}^{\infty} \left[\prod_{i=0}^{t} \beta_i \right] h(X_t)$$

$$= \sum_{t=0}^{\infty} \mathbb{E}_x \left[\prod_{i=0}^{t} \beta_i \right] h(X_t)$$

$$= \sum_{t=0}^{\infty} (A^t h)(x)$$

Pointwise, this is $v = \sum_{t \geqslant 0} A^t h$

By the NSL and $\rho(A) < 1$, we have $\sum_{t \geqslant 0} A^t h = (I - A)^{-1} h$

Example. Consider the simple (constant discount case)

$$b \equiv \beta \in (0, 1)$$

Then

- $\prod_{i=0}^{t} \beta_i = \beta^t$
- $A = \beta P$
- $\rho(A) = \rho(\beta P) = \beta < 1$
- $v = (I - A)^{-1} h = (I - \beta P)^{-1} h$

Example. Consider again the firm valuation problem with

$$v(x) = \mathbb{E}_x \sum_{t=0}^{\infty} \left[\prod_{i=0}^{t} \beta_i \right] \pi_t \quad \text{where} \quad \beta_0 := 1$$

Suppose

- (X_t) is P-Markov on X
- $\beta_t = \beta(X_t)$ for some fixed $\beta \in \mathbb{R}^{\mathsf{X}}$
- $\pi_t = \pi(X_t)$ for some fixed $\pi \in \mathbb{R}^{\mathsf{X}}$

Let

$$A(x, x') := \beta(x) P(x, x')$$

Ex. Show the following: If $\rho(A) < 1$, then v is finite and satisfies

$$v(x) = \mathbb{E}_x \sum_{t=0}^{\infty} \left[\prod_{i=0}^{t} \beta_i \right] \pi_t$$

Moreover,

$$v = (I - A)^{-1} \pi$$

Proof: This is immediate from the result on slide 11 with

$$b(X_{t-1}, X_t) = \beta(X_t) \text{ and } h = \pi$$

Ex. Show the following: If $\rho(A) < 1$, then v is finite and satisfies

$$v(x) = \mathbb{E}_x \sum_{t=0}^{\infty} \left[\prod_{i=0}^{t} \beta_i \right] \pi_t$$

Moreover,

$$v = (I - A)^{-1} \pi$$

Proof: This is immediate from the result on slide 11 with

$$b(X_{t-1}, X_t) = \beta(X_t) \text{ and } h = \pi$$

Introduction to asset pricing

As an application of state-dependent discouting we now discuss asset pricing

In standard neoclassical asset pricing, state-dependent discounting arises naturally from basic assumptions

- prices are determined by market equilibria
- no arbitrage

Readers who prefer to move on to dynamic programming can jump to slide 41

Risk-neutral pricing

Consider an asset with payoff G_{t+1} next period

What current price Π_t should we assign?

Risk neutral pricing says

$$\Pi_t = \mathbb{E}_t \, \beta \, G_{t+1}$$

for some $\beta \in (0, 1)$

If the payoff is in k periods, then the price is $\mathbb{E}_t \, \beta^k \, G_{t+k}$

Example. The time t risk-neutral price of a **European call option** is

$$\Pi_t = \mathbb{E}_t \, \beta^k \max\{S_{t+k} - K, 0\}$$

where

- S_t is the price of the underlying asset (e.g., stock)
- K is the strike price
- k is the duration
- $\beta = 1/(1+r)$ where r is the discount rate

But assuming risk neutrality for all investors is **not consistent with the data**

Example. Consider the rate of return $r_{t+1} := (G_{t+1} - \Pi_t)/\Pi_t$

From $\Pi_t = \mathbb{E}_t \, \beta \, G_{t+1}$ we get

$$\mathbb{E}_t \, \beta \, \frac{G_{t+1}}{\Pi_t} = 1 \quad \iff \quad \mathbb{E}_t \, \beta (1 + r_{t+1}) = 1$$

Hence

$$\mathbb{E}_t \, r_{t+1} = \frac{1 - \beta}{\beta}$$

Thus, risk neutrality implies that all assets have the same expected rate of return

In fact riskier assets usually have higher average rates of return

- incentivize investors to bear risk

Example. The **risk premium** $:=$ expected rate of return minus the rate of return on a risk-free asset

Risk-neutrality $\implies$ risk premium is zero for all assets

But calculations based on post-war US data show that

$$\text{average risk premium for equities} \approx 8\% \text{ per annum}$$

These facts motivate a more general theory. . .

Fundamental theorem of asset pricing

Here is an informal statement from standard neoclassical finance — Stephen Ross, LP Hansen, David Kreps, etc.

There exists a positive random variable M_{t+1} such that the price Π_t of any payoff G_{t+1} obeys

$$\Pi_t = \mathbb{E}_t \, M_{t+1} \, G_{t+1}$$

- assumptions $\approx$ representative agent, no arbitrage
- M_{t+1} is called the **stochastic discount factor** (SDF)
- can handle risk aversion, different rates of return
- key assertion: same SDF can price <u>any</u> asset

Special case: two period Lucas tree

How should M_{t+1} be constructed?

To answer this we need a model

Let's look at a simple example

- a two-period model
- CRRA utility
- one asset and one agent

Agent takes Π_t as given and solves

$$\max_{0 \leqslant \alpha \leqslant 1} \{u(C_t) + \beta \mathbb{E}_t u(C_{t+1})\}$$

$$\text{subject to} \quad C_t = E_t - \Pi_t \alpha \quad \text{and} \quad C_{t+1} = E_{t+1} + \alpha G_{t+1}$$

Here

- u is a flow utility function and β measures impatience
- G_{t+1} is the payoff of the asset and Π_t is the time-t price
- E_t and E_{t+1} are endowments and
- α is the share of the asset purchased by the agent

Rewrite as

$$\max_{\alpha}\{u(E_t - \Pi_t \alpha) + \beta \mathbb{E}_t u(E_{t+1} + \alpha G_{t+1})\}$$

Differentiating w.r.t. α leads to first order condition

$$u'(E_t - \Pi_t \alpha)\Pi_t = \beta \mathbb{E}_t u'(E_{t+1} + \alpha G_{t+1}) G_{t+1}$$

Rearranging gives us

$$\Pi_t = \mathbb{E}_t M_{t+1} G_{t+1} \quad \text{where} \quad M_{t+1} := \beta \frac{u'(C_{t+1})}{u'(C_t)}$$

Example. In the CRRA case $u(c) = c^{1-\gamma}/(1-\gamma)$ we get

$$M_{t+1} = \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma}$$

Alternatively,

$$M_{t+1} = \beta \exp(-\gamma g_{t+1}) \quad \text{where} \quad g_{t+1} := \ln(C_{t+1}/C_t)$$

Applies

- heavier discounting in states of the world where consumption growth is high
- lower discounting in states of the world where consumption growth is low

Favors assets that hedge against the risk of low consumption states

Question: How well does this model work when confronted with data?

Answer: badly — search "equity premium puzzle" for an introduction

1. Shiller (1982), Mehra and Prescott (1985), etc.

Recent quantitative models build more sophisticated SDFs to try to get closer to the data

- Epstein–Zin preferences
- ambiguity
- long-run risk models, etc.

Question: How well does this model work when confronted with data?

Answer: badly — search “equity premium puzzle” for an introduction

1. Shiller (1982), Mehra and Prescott (1985), etc.

Recent quantitative models build more sophisticated SDFs to try to get closer to the data

- Epstein–Zin preferences
- ambiguity
- long-run risk models, etc.

General SDF, Markov state

Let's go back to the general case

$$\Pi_t = \mathbb{E}_t \, M_{t+1} \, G_{t+1}$$

How can we compute this?

Suppose $(X_t)_{t \geqslant 0}$ is P-Markov on X,

$$M_{t+1} = m(X_t, X_{t+1}) \quad \text{and} \quad G_{t+1} = g(X_t, X_{t+1})$$

With $\pi(x) := \mathbb{E}_x M_{t+1} \, G_{t+1}$, we have

$$\pi(x) = \sum_{x' \in \mathsf{X}} m(x, x') g(x, x') P(x, x')$$

Pricing a dividend stream

Consider the price of a claim on dividend stream $(D_t)_{t \geqslant 0}$

Let the price at time t be Π_t

Buying at t and selling at $t+1$ pays $\Pi_{t+1} + D_{t+1}$

Hence the price sequence $(\Pi_t)_{t \geqslant 0}$ must obey

$$\Pi_t = \mathbb{E}_t \, M_{t+1} (\Pi_{t+1} + D_{t+1})$$

Current price depends on future price — how can we solve it?

Recall the key equation

$$\Pi_t = \mathbb{E}_t \, M_{t+1}(\Pi_{t+1} + D_{t+1})$$

Let

- $D_t = d(X_t)$ where $(X_t)_{t \geqslant 0}$ is P-Markov
- $\pi(x) =$ current price given $X_t = x$

We get

$$\pi(x) = \sum_{x'} m(x, x')(\pi(x') + d(x'))P(x, x') \qquad (x \in \mathsf{X})$$

Rewrite the last expression as

$$\pi = A\pi + Ad$$

where

$$A(x, x') := m(x, x')P(x, x')$$

Neumann series lemma: $\rho(A) < 1 \implies$ the unique solution is

$$\pi^* = (I - A)^{-1} Ad$$

- π^* is called an **equilibrium price function**
- A is called the **Arrow–Debreu discount operator**

Nonstationary Dividends

A more realistic model is one where dividends grow over time

A standard model of dividend growth is

$$\ln \frac{D_{t+1}}{D_t} = \kappa(X_t, \eta_{t+1}) \qquad t = 0, 1, \dots,$$

Here

- κ is a fixed function
- (X_t) is P-Markov on finite set X
- (η_t) is $\textsc{iid}$ with density φ
- $M_{t+1} = m(X_t, X_{t+1})$ for some positive function m

Growing dividends $\implies$ growing prices

- no π such that $\Pi_t = \pi(X_t)$ for all t

Instead we try to solve for the **price-dividend ratio** $V_t := \Pi_t / D_t$

Ex. Show that $\Pi_t = \mathbb{E}_t \left[M_{t+1}(D_{t+1} + \Pi_{t+1}) \right]$ implies

$$V_t = \mathbb{E}_t \, \left[M_{t+1} \exp(\kappa(X_t, \eta_{t+1})) \, (1 + V_{t+1}) \right]$$

Conditioning on $X_t = x$,

$$v(x) = \sum_{x' \in \mathsf{X}} m(x, x') \int \exp(\kappa(x, \eta)) \varphi(\mathrm{d}\eta) \left[1 + v(x') \right] P(x, x')$$

Let

$$A(x, x') := m(x, x') \int \exp(\kappa(x, \eta)) \varphi(\mathrm{d}\eta) P(x, x')$$

Now we see a v that solves

$$v(x) = \sum_{x' \in \mathsf{X}} \left[1 + v(x')\right] A(x, x')$$

Equivalent:

$$v = A\mathbb{1} + Av$$

If $\rho(A) < 1$, then the unique solution is

$$v^* = (I - A)^{-1} A\mathbb{1}$$

Example. Dividend growth is

$$\kappa(X_t, \eta_{d,t+1}) = \mu_d + X_t + \sigma_d \, \eta_{d,t+1} \quad \text{where} \quad (\eta_{d,t})_{t \geqslant 0} \overset{\text{IID}}{\sim} N(0,1)$$

Consumption growth is given by

$$\ln \frac{C_{t+1}}{C_t} = \mu_c + X_t + \sigma_c \, \eta_{c,t+1} \quad \text{where} \quad (\eta_{c,t})_{t \geqslant 0} \overset{\text{IID}}{\sim} N(0,1)$$

We use the Lucas CRRA SDF, implying that

$$M_{t+1} = \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} = \beta \exp(-\gamma(\mu_c + X_t + \sigma_c \eta_{c,t+1}))$$

```julia
using QuantEcon, LinearAlgebra

"Creates an instance of the asset pricing model with Markov state."
function create_asset_pricing_model(;
        n=200,              # state grid size
        ρ=0.9, ν=0.2,       # state persistence and volatility
        β=0.99, γ=2.5,      # discount and preference parameter
        μ_c=0.01, σ_c=0.02, # consumption growth mean and volatility
        μ_d=0.02, σ_d=0.1)  # dividend growth mean and volatility
    mc = tauchen(n, ρ, ν)
    x_vals, P = exp.(mc.state_values), mc.p
    return (; x_vals, P, β, γ, μ_c, σ_c, μ_d, σ_d)
end
```

```julia
" Build the discount matrix A. "
function build_discount_matrix(model)
    (; x_vals, P, β, γ, μ_c, σ_c, μ_d, σ_d) = model
    e = exp.(μ_d - γ*μ_c + (γ^2*σ_c^2 + σ_d^2)/2 .+ (1-γ)*x_vals)
    return β * e .* P
end

"Compute the price-dividend ratio associated with the model."
function pd_ratio(model)
    (; x_vals, P, β, γ, μ_c, σ_c, μ_d, σ_d) = model
    A = build_discount_matrix(model)
    @assert maximum(abs.(eigvals(A))) < 1 "Requires r(A) < 1."
    n = length(x_vals)
    return (I - A) \ (A * ones(n))
end
```

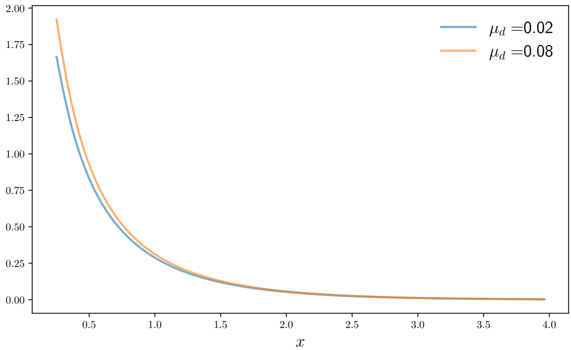

Figure: Price-dividend ratio as a function of x

Back to dynamic programming

Soon we turn to dynamic programming with stochastic discounting

We will use

- some stability conditions related to spectral radii
- an extension of Banach's fixed point theorem

We discuss this fixed point results first

Eventual contractions

Fix

1. $U \subset \mathbb{R}^{\mathsf{X}}$
2. norm $\| \cdot \|$ on U
3. self-map T on U

We call T **eventually contracting** on U if $\exists$ a $k \in \mathbb{N}$ such that T^k is contracting on U under $\| \cdot \|$

Theorem. If U is closed and T is eventually contracting, then T is globally stable on U

Ex. Prove the theorem [Hint: Use Banach's theorem]

Example. If $Su = Au + b$ for $b \in \mathbb{R}^{\mathsf{X}}$ and $A \in \mathcal{L}(\mathbb{R}^{\mathsf{X}})$, then

$$\begin{aligned} \|S^k u - S^k v\| &= \|A^k u - A^k v\| \\ &= \|A^k (u - v)\| \leqslant \|A^k\| \|u - v\| \end{aligned}$$

If $\rho(A) < 1$, we can choose $k \in \mathbb{N}$ such that $\|A^k\| < 1$

$\therefore \quad S$ is eventually contracting and hence globally stable

The unique fixed point satisfies $u^* = Au^* + b$

By the NSL, $I - A$ is bijective, so

$$u^* = (I - A)^{-1} b$$

The last example shows the connection to the Neumann series lemma

- adds global stability

But eventual contractions have much wider scope than the Neumann series lemma

- can also be applied in nonlinear settings

A Spectral Radius Condition

Let T be a self-map on $U \subset \mathbb{R}^{\mathsf{X}}$.

Proposition. If $\exists$ a positive $L \in \mathcal{L}(\mathbb{R}^{\mathsf{X}})$ with $\rho(L) < 1$ and

$$|Tv - Tw| \leqslant L|v - w| \qquad \text{for all } v, w \in U$$

then T is an eventual contraction on U

Proof: Fixing $v, w \in U$ and $k \in \mathbb{N}$, we have

$$|T^k v - T^k w| \leqslant L|T^{k-1} v - T^{k-1} w|$$

or

$$e_k \leqslant L e_{k-1} \quad \text{where} \quad e_k := |T^k v - T^k w|$$

Ex. Show that $e_k \leqslant L^k e_0$ for all $k \in \mathbb{N}$

Proof: We know that

$$e_k \leqslant L e_{k-1} \quad \text{for all } k \in \mathbb{N} \tag{1}$$

We prove by induction

The claim is true at $k = 1$ by (1)

Now suppose that it is true at k, so $e_k \leqslant L^k e_0$

Since L is order-preserving on U and (1) holds, we have

$$e_{k+1} \leqslant L e_k \leqslant L L^k e_0 = L^{k+1} e_0$$

Hence the claim is also true at $k + 1$ — QED

Ex. Show that $e_k \leqslant L^k e_0$ for all $k \in \mathbb{N}$

Proof: We know that

$$e_k \leqslant L e_{k-1} \quad \text{for all } k \in \mathbb{N} \tag{1}$$

We prove by induction

The claim is true at $k = 1$ by (1)

Now suppose that it is true at k, so $e_k \leqslant L^k e_0$

Since L is order-preserving on U and (1) holds, we have

$$e_{k+1} \leqslant L e_k \leqslant L L^k e_0 = L^{k+1} e_0$$

Hence the claim is also true at $k + 1$ — QED

We have verified $e_k \leqslant L^k e_0$, or

$$|T^k v - T^k w| \leqslant L^k |v - w|$$

Let $\| \cdot \|$ be the Euclidean norm

Since $0 \leqslant a \leqslant b$ implies $\|a\| \leqslant \|b\|$, we get

$$\|T^k v - T^k w\| \leqslant \|L^k |v - w|\| \leqslant \|L^k\| \|v - w\|$$

Since $\rho(L) < 1$, we have $\|L^k\| \to 0$ as $k \to \infty$

Hence T is eventually contracting on U

Blackwell for eventual contractions

In Ch. 2 we studied Blackwell's condition for a contraction

Here we provide an analogous result for eventual contractions

Let $U \subset \mathbb{R}^{\mathsf{X}}$ be such that $v, c \in U$ and $c \geqslant 0$ implies $v + c \in U$

Proposition. Let T be an order-preserving self-map on U

If $\exists$ a positive $L \in \mathcal{L}(\mathbb{R}^{\mathsf{X}})$ such that $\rho(L) < 1$ and

$$T(v + c) \leqslant Tv + Lc \quad \text{for all } c, v \in \mathbb{R}^{\mathsf{X}} \text{ with } c \geqslant 0$$

then T is eventually contracting on U

Proof: Let U, T, L be as in the statement of the proposition

Fix $v, w \in U$

By the assumed properties on T, we have

$$Tv = T(v + w - w) \leqslant T(w + |v - w|) \leqslant Tw + L|v - w|$$

Rearranging gives $Tv - Tw \leqslant L|v - w|$

Reversing the roles of v and w yields $|Tv - Tw| \leqslant L|v - w|$

The claim now follows from the proposition on slide 44

MDPs with State-Dependent Discounting

We begin with an MDP $\mathscr{M} = (\Gamma, \beta, r, P)$ with state space X, action space A and feasible state-action pairs G

We replace the constant β with a function β from G $\times$ X to $\mathbb{R}_+$

The **Bellman operator** takes the form

$$(Tv)(x) = \max_{a \in \Gamma(x)} \left\{ r(x, a) + \sum_{x'} v(x') \beta(x, a, x') P(x, a, x') \right\}$$

where $x \in$ X and $v \in \mathbb{R}^{\mathsf{X}}$

Let Σ be the set of **feasible policies** (as for regular MDPs)

The **policy operator** T_σ corresponding to $\sigma \in \Sigma$ is

$$(T_\sigma \, v)(x) = r(x, \sigma(x)) + \sum_{x'} v(x') \beta(x, \sigma(x), x') P(x, \sigma(x), x')$$

- T_σ maps $v \in \mathbb{R}^{\mathsf{X}}$ to $T_\sigma \, v \in \mathbb{R}^{\mathsf{X}}$

Given $v \in \mathbb{R}^{\mathsf{X}}$, a policy σ is called v-greedy if, for all x in X,

$$\sigma(x) \in \operatorname*{argmax}_{a \in \Gamma(x)} \left\{ r(x, a) + \sum_{x'} v(x') \beta(x, a, x') P(x, a, x') \right\}$$

Ex. Show that σ is v-greedy whenever $T_\sigma \, v = T v$

When does T_σ have a unique fixed point in $\mathbb{R}^{\mathsf{X}}$?

Suppose $\exists$ a $b < 1$ such that

$$\beta(x, a, x') \leqslant b \text{ for all } (x, a, x') \in \mathsf{G} \times \mathsf{X}$$

Then the Bellman and policy operators are all contraction maps

In this setting it is easy to extend MDP optimality results

Unfortunately, this assumption is too strict for many applications

We return to this point below

When does T_σ have a unique fixed point in $\mathbb{R}^{\mathsf{X}}$?

Suppose $\exists$ a $b < 1$ such that

$$\beta(x, a, x') \leqslant b \text{ for all } (x, a, x') \in \mathsf{G} \times \mathsf{X}$$

Then the Bellman and policy operators are all contraction maps

In this setting it is easy to extend MDP optimality results

Unfortunately, this assumption is too strict for many applications

We return to this point below

So consider the following weaker condition

Assumption SD. $\exists$ an $L \in \mathcal{L}(\mathbb{R}^{\mathsf{X}})$ with $\rho(L) < 1$ and

$$\beta(x, a, x')P(x, a, x') \leqslant L(x, x') \quad \text{for all } (x, a) \in \mathsf{G} \text{ and } x' \in \mathsf{X}$$

To state the next result, we set $r_\sigma(x) = r(x, \sigma(x))$ and

$$L_\sigma(x, x') := \beta(x, \sigma(x), x')P(x, \sigma(x), x')$$

Ex. Show that, when Assumption SD holds, T_σ is globally stable on $\mathbb{R}^{\mathsf{X}}$ with unique fixed point

$$v_\sigma = (I - L_\sigma)^{-1} r_\sigma$$

Proof: Given the definition of L_σ, we can write T_σ as

$$T_\sigma \, v = r_\sigma + L_\sigma \, v$$

Since Assumption SD holds,

$$\beta(x, \sigma(x), x') P(x, \sigma(x), x') \leqslant L(x, x') \text{ for all } (x, x') \in \mathsf{X} \times \mathsf{X}$$

Therefore, $0 \leqslant L_\sigma \leqslant L$

But then $\rho(L_\sigma) \leqslant \rho(L)$

Hence $\rho(L_\sigma) < 1$

The claim now follows from the result on slide 42

The next exercise helps us interpret v_σ

Ex. Let

- $P_\sigma(x, x') = P(x, \sigma(x), x')$
- (X_t) be P_σ-Markov
- $\beta_t = \beta(X_t, \sigma(X_t), X_{t+1})$ for all $t \in \mathbb{N}$ with $\beta_0 = 1$

Prove that, under Assumption SD, the function v_σ obeys

$$v_\sigma(x) = \mathbb{E}_x \sum_{t=0}^{\infty} \left[\prod_{i=0}^{t} \beta_i \right] r_\sigma(X_t) \tag{2}$$

for all $x \in \mathsf{X}$

Proof: Recall that

$$L_\sigma(x, x') = \beta(x, \sigma(x), x') P(x, \sigma(x), x')$$

This is the "discount operator" associated with σ

Recall also that $\rho(L_\sigma) < 1$

It follows directly from the result on slide 11 that

$$v(x) := \mathbb{E}_x \sum_{t=0}^{\infty} \left[\prod_{i=0}^{t} \beta_i \right] r_\sigma(X_t)$$

is a well-defined element of $\mathbb{R}^{\mathsf{X}}$ and equals $(I - L_\sigma)^{-1} r_\sigma$

In other words, $v = v_\sigma$, as was to be shown

When Assumption SD holds, we can define the **value function** via

$$v^* := \bigvee_{\sigma \in \Sigma} v_\sigma$$

A policy σ is called **optimal** if $v_\sigma = v^*$

A function $v \in \mathbb{R}^{\mathsf{X}}$ is said to satisfy **the Bellman equation** if

$$v(x) = \max_{a \in \Gamma(x)} \left\{ r(x,a) + \sum_{x'} v(x') \beta(x,a,x') P(x,a,x') \right\}$$

for all $x \in \mathsf{X}$

Here is our main optimality result for MDPs with state-dependent discounting

Proposition. If Assumption SD holds, then

1. v^* is the unique fixed point of the Bellman operator,
2. a policy $\sigma \in \Sigma$ is optimal if and only if it is v^*-greedy, and
3. at least one optimal policy exists

We state and prove a more general result later

- in our discussion of recursive dynamic programs
- allows us to add recursive preferences, etc.

For now let's

- Show that T is globally stable
- Consider algorithms
- Look at a special case, where (β_t) is purely exogenous
- Look at some applications

Ex. Prove T is globally stable on $\mathbb{R}^{\mathsf{X}}$ under Assumption SD

Proof: Fixing $v, c \in \mathbb{R}^{\mathsf{X}}$ with $c \geqslant 0$, we have

$$
\begin{aligned}
&(T(v+c))(x) \\
&\qquad = \max_{a \in \Gamma(x)} \left\{ r(x,a) + \sum_{x'} [v(x') + c(x')] \beta(x,a,x') P(x,a,x') \right\} \\
&\qquad \leqslant (Tv)(x) + \max_{a \in \Gamma(x)} \sum_{x'} c(x') \beta(x,a,x') P(x,a,x')
\end{aligned}
$$

Hence $T(v+c) \leqslant Tv + Lc$

$\therefore$ T is eventually contracting (slide 47) and hence globally stable (slide 41)

Ex. Prove T is globally stable on $\mathbb{R}^{\mathsf{X}}$ under Assumption SD

Proof: Fixing $v, c \in \mathbb{R}^{\mathsf{X}}$ with $c \geqslant 0$, we have

$$(T(v+c))(x)$$

$$= \max_{a \in \Gamma(x)} \left\{ r(x,a) + \sum_{x'} [v(x') + c(x')] \beta(x,a,x') P(x,a,x') \right\}$$

$$\leqslant (Tv)(x) + \max_{a \in \Gamma(x)} \sum_{x'} c(x') \beta(x,a,x') P(x,a,x')$$

Hence $T(v+c) \leqslant Tv + Lc$

$\therefore$ T is eventually contracting (slide 47) and hence globally stable (slide 41)

Algorithms for state-dependent discounting MDPs

For MDPs we studied

- value function iteration (VFI)
- Howard policy iteration (HPI)
- optimistic policy iteration (OPI)

Do these algorithms still converge?

Under what conditions?

The statement of the VFI and OPI algorithms are identical

- of course, use the new definitions for T, T_σ

The statement of HPI is as follows

Algorithm 1: HPI for MDPs with state-dependent discounting

```
input $\sigma_0 \in \Sigma$, an initial guess of $\sigma^*$
$k \leftarrow 0$
$\varepsilon \leftarrow 1$
while $\varepsilon > 0$ do
    $v_k \leftarrow (I - L_{\sigma_k})^{-1} r_{\sigma_k}$
    $\sigma_{k+1} \leftarrow$ a $v_k$-greedy policy
    $\varepsilon \leftarrow \mathbb{1}\{\sigma_k \neq \sigma_{k+1}\}$
    $k \leftarrow k + 1$
end
return $\sigma_k$
```

Theorem. Under Assumption SD (slide 57),

1. the sequence of values (v_k) generated by OPI and VFI converge to v^*
2. HPI returns an optimal policy in finitely many steps

Exogenous discounting

Let's specialize to a setting where the discount process is purely exogenous

We consider a model (Γ, β, r, Q, R) where

1. Γ is a nonempty correspondence from $\mathsf{Y} \to \mathsf{A}$; set

$$\mathsf{G} := \{(y, a) \in \mathsf{Y} \times \mathsf{A} : a \in \Gamma(y)\}$$

2. β is a function from Z to $\mathbb{R}_+$
3. r is a function from G to $\mathbb{R}$
4. Q is a stochastic matrix on Z
5. R is a stochastic kernel from G to Y

A summary of dynamics:

- (Z_t) is Q-Markov
- The **discount factor process** $(\beta_t)_{t \geqslant 0}$ obeys $\beta_t := \beta(Z_t)$
- Given $Y_t = y$ and current action a, current reward is $r(y, a)$
- Y_{t+1} is drawn from distribution $R(y, a, \cdot)$
- Y_{t+1} and Z_{t+1} are updated independently given (y, z, a)

The Bellman equation becomes

$$v(y,z) = \max_{a \in \Gamma(y)} \left\{ r(y,a) + \beta(z) \sum_{z',\, y'} v(y',z') Q(z,z') R(y,a,y') \right\}$$

for all $(y,z) \in \mathsf{X}$

Given $\sigma \in \Sigma$, the **policy operator** is

$$(T_\sigma\, v)(y,z) = r(y, \sigma(y,z)) + \\ \beta(z) \sum_{z',\, y'} v(y',z') Q(z,z') R(y, \sigma(y,z), y')$$

Proposition Let

$$L(z, z') := \beta(z) Q(z, z')$$

If $\rho(L) < 1$, then all of the optimality results for MDPs with state-dependent discounting on slide 57 apply

Ex. Prove it

Proof: To formulate this problem as an MDP with state-dependent discounting we set

- $\mathsf{X} = \mathsf{Y} \times \mathsf{Z}$ and $x = (y, z)$
- $\beta(x, a, x') = \beta(x) = \beta(y, z) = \beta(z)$
- $P(x, a, x') = P((y, z), a, (y', z')) = Q(z, z')R(y, a, y')$

Then

$$\begin{aligned} \beta(x, a, x')P(x, a, x') &= \beta(z)Q(z, z')R(y, a, y') \\ &\leqslant \beta(z)Q(z, z') \\ &=: L(z, z') \end{aligned}$$

Since $\rho(L) < 1$, we are done

How tight is the condition $\rho(L) < 1$?

Ex. Show that $\rho(L) < 1$ when

$$\exists\, b < 1 \text{ such that } \beta(z) \leqslant b \text{ for all } z \in \mathsf{Z}$$

But this condition is too strict

Remember $\beta < 1$ and $\beta = 1/(1+r)$ implies $r > 0$

How tight is the condition $\rho(L) < 1$?

Ex. Show that $\rho(L) < 1$ when

$$\exists\, b < 1 \text{ such that } \beta(z) \leqslant b \text{ for all } z \in \mathsf{Z}$$

But this condition is too strict

Remember $\beta < 1$ and $\beta = 1/(1 + r)$ implies $r > 0$

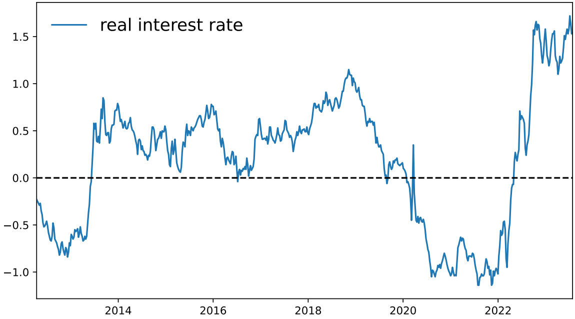

Figure: Real US interest rates are sometimes negative

Also, household preferences are sometimes assumed to have occasionally negative discount rates

- implies β_t sometimes > 1

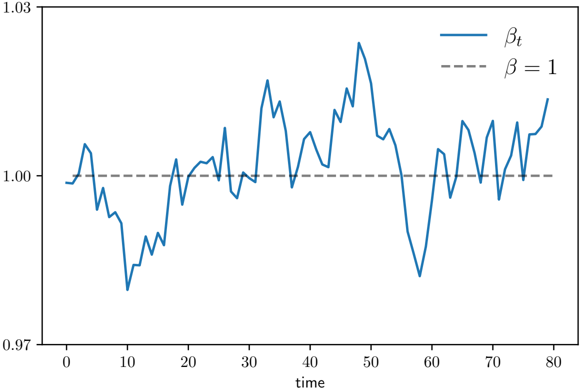

Figure: $(\beta)_{t \geqslant 0}$ process in Hills, Nakata and Schmidt (2019)

The process in Hills, Nakata and Schmidt (2019) is AR(1)

Following them, we discretize via Tauchen approximation

```
mc = tauchen(n, ρ, σ, 1 - ρ, m)
```

Parameters are as in Hills et al. (2019)

We find $\rho(L) = 0.9996$, so optimality results apply

In summary,

- $\rho(L) < 1$ allows the discount factor to exceed one at times
- Reasonable for economic applications
- Yet strong enough for optimality

Application: Inventory Management

Recall the inventory management model with Bellman equation

$$v(x) = \max_{a \in \Gamma(x)} \left\{ r(x, a) + \beta \sum_{d \geqslant 0} v(f(x, a, d)) \varphi(d) \right\}$$

- $x \in \mathsf{X} := \{0, \ldots, K\}$ is the current inventory level
- a is the current inventory order
- $r(x, a)$ is current profits
- $f(x, a, d) := (x - d) \vee 0 + a$
- d is an IID demand shock with distribution φ

We now replace β with $\beta_t = \beta(Z_t)$

- $(Z_t)_{t \geqslant 0}$ is Q-Markov on Z

This is an MDP with state-dependent discounting

The Bellman equation becomes

$$v(y, z) = \max_{a \in \Gamma(y)} \left\{ r(y, a) + \beta(z) \sum_{d, z'} v(f(y, a, d), z') \varphi(d) Q(z, z') \right\}$$

All optimality results hold when

- $L(z, z') := \beta(z) Q(z, z')$ and
- $\rho(L) < 1$

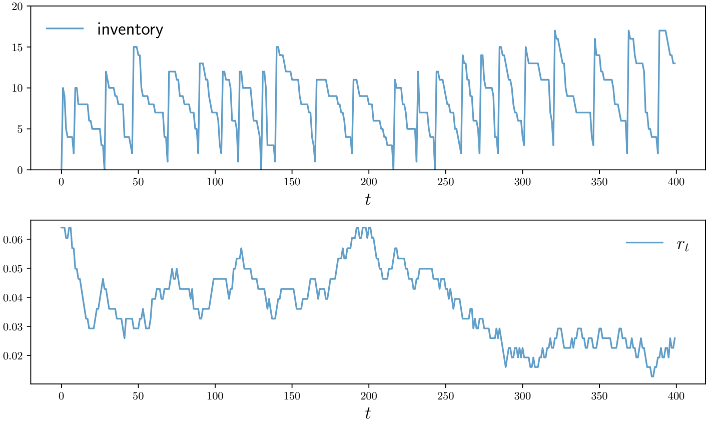

Figure: Inventory dynamics with time-varying interest rates

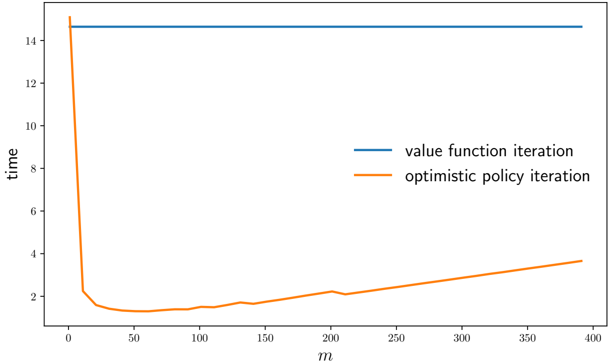

Figure: OPI vs VFI timings for the inventory model