ECON 605: Macroeconomic Theory I: Lecture Notes

February 11, 2026

1 Introduction

Detrending, Stationarity, and Filters

Why we detrend

The business cycle is about *fluctuations around a trend.* A convenient working decomposition is

$$y_t \;=\; \tau_t \;+\; x_t,$$

where y_t is the observed series (e.g., real GDP or employment), τ_t is a smooth "trend" (growth, demographics, technology, seasonality if unadjusted), and x_t is the cyclical/stationary component we want to study. The goal of detrending is to construct x_t so that its distribution does not drift over time, making correlation, impulse responses, and model fitting meaningful.

Intuition. Think of τ_t as the escalator and x_t as your steps on it. If you want to study how you walk (accelerations, pauses), you first subtract the escalator's steady motion.

1.1 Trends: deterministic vs. stochastic

Deterministic trends

A deterministic trend is a known (or well-approximated) function of time, e.g.,

$$\tau_t = a_0 + a_1 t \quad \text{(linear)}, \qquad \tau_t = a_0 + a_1 t + a_2 t^2 \quad \text{(quadratic)}.$$

Regressing y_t on time polynomials and using the residuals yields a detrended x_t.

Stochastic trends (unit roots)

A unit root process has a persistent, random drift:

$$y_t = y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \text{ i.i.d. (or weakly dependent)}.$$

First-differencing removes the unit root: $\Delta y_t = \varepsilon_t$ becomes stationary (up to short memory). This motivates using differences (or log-differences for growth rates).

Slide tie-in. Your slides explicitly write $y_t = \tau_t + x_t$ and list polynomial forms for τ_t on separate bullets; they also emphasize differencing when τ_t is stochastic and note unit-root testing (ADF) for stationarity diagnostics. [1]

1.2 Difference vs. log-difference

What each does

- **Difference** $\Delta y_t = y_t - y_{t-1}$ removes a unit root in levels.
- **Log-difference** $\Delta \ln y_t \approx \%$ growth. If y_t has a stochastic trend in logs, $\Delta \ln y_t$ is stationary under standard conditions.

Intuition. If the variable grows roughly exponentially (GDP, employment), taking logs makes the trend roughly linear; differencing logs then measures growth rates. For unit-root removal per se, both differences and log-differences purge the random walk part; choose logs when you want elasticities/growth interpretation.

1.3 Testing for stationarity in practice

Augmented Dickey–Fuller (ADF) tests and their variants assess whether a series has a unit root. In applied work:

1. Plot y_t and Δy_t (or $\Delta \ln y_t$).
2. Start with an ADF on y_t with a trend if plots suggest a trend.
3. If you fail to reject a unit root in levels, difference and re-test; usually you will then reject (i.e., differences are stationary).

Caution. Power is limited in small samples; combine tests with economic reasoning and visuals.

1.4 The Hodrick–Prescott (HP) filter

The HP trend $\{\tau_t\}_{t=1}^{T}$ solves

$$\min_{\{\tau_t\}_{t=1}^{T}} \sum_{t=1}^{T} (y_t - \tau_t)^2 \; + \; \lambda \sum_{t=2}^{T-1} \left(\Delta^2 \tau_t\right)^2, \qquad \Delta^2 \tau_t \equiv (\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1}),$$

i.e., it trades off fit against the curvature of the trend (penalizing changes in the *slope*). The cycle is $x_t = y_t - \tau_t$; the smoothing parameter λ controls how smooth the trend is (larger $\lambda \Rightarrow$ smoother trend).

Intuition. HP says: "Fit the data closely, but prefer a trend with nearly constant growth." Think of bending a thin metal ruler to trace y_t: the bending cost corresponds to the second-difference penalty. Endpoints are fragile (the *endpoint problem*); be cautious interpreting the last few observations.

Practice tips. Common conventions are: quarterly data $\lambda = 1600$, annual $\lambda \approx 6.25$, monthly $\lambda \approx$ 129,600. Treat these as useful defaults, not laws; always check robustness.

[1] *Notes from slides:* $y_t = \tau_t + x_t$ and polynomial τ_t variants are presented; unit-root tests ("Dickey–Fuller tests") are mentioned as tools to assess stationarity; the rule of thumb that differencing targets the stochastic trend appears alongside examples for GDP and employment.

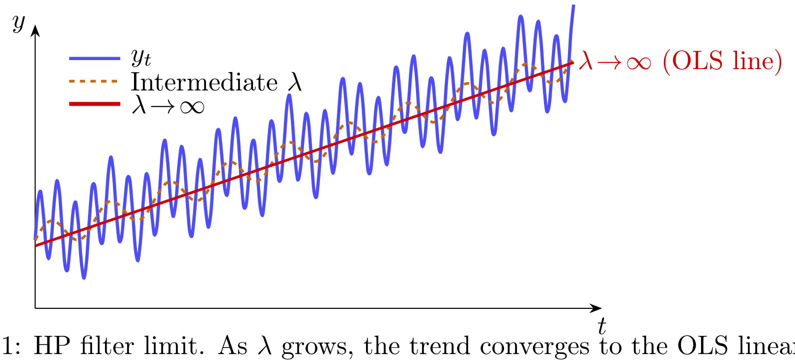


Figure 1: HP filter limit. As λ grows, the trend converges to the OLS linear trend.

Limit cases.

- $\lambda \to 0$: the penalty vanishes and $\tau_t \to y_t$ (virtually no smoothing).
- $\lambda \to \infty$: the curvature penalty dominates and the trend becomes *affine in* t, i.e. a straight line $\tau_t = a + bt$—specifically the OLS linear trend.

The case where $\lambda \to \infty$ is interesting to understand the logic. The HP objective is a tradeoff: fit the data (small $\sum(y_t - \tau_t)^2$) but avoid bending (small $\sum(\Delta^2\tau_t)^2$). When λ is huge, bending becomes infinitely costly, so we must have

$$\Delta^2\tau_t = \tau_{t+1} - 2\tau_t + \tau_{t-1} = 0 \quad \text{for all interior } t,$$

which implies a constant slope and therefore a straight line:

$$\tau_t = a + bt.$$

Proof. Set $d_t := \tau_t - \tau_{t-1}$. From $\Delta^2\tau_t = \tau_{t+1} - 2\tau_t + \tau_{t-1} = 0$ we get $d_{t+1} - d_t = 0$, hence

$$d_{t+1} = d_t = b\,\forall t.$$

Summing, $\tau_t - \tau_1 = \sum_{k=2}^{t} d_k = (t-1)b$, so

$$\tau_t = \tau_1 + (t-1)b = (\tau_1 - b) + bt := a + bt.$$

To pin down a, b, choose the least-squares fit to $\{(t, y_t)\}$:

$$(a, b) = \arg\min_{a,b} \sum_{t=1}^{T} (y_t - (a + bt))^2,$$

whose solution is the usual OLS line with $b = \frac{\sum(t-\bar{t})(y_t-\bar{y})}{\sum(t-\bar{t})^2}$, $a = \bar{y} - b\,\bar{t}$. □

1.5 Linear filters: time and frequency

Definition (time-domain view)

A *linear time-invariant* (LTI) filter transforms a series $\{y_t\}$ as

$$\hat{y}_t \;=\; \sum_{k=-K}^{K} a_k\, y_{t-k},$$

where $\{a_k\}$ are the *weights* (the impulse response) and K is the window half-width.

- **Two-sided (non-causal):** uses lags and leads ($k < 0$ and $k > 0$); best for retrospective analysis.
- **One-sided (causal / "real time"):** uses only lags ($k \geq 0$); avoids looking ahead but typically adds delay and ripple.
- **Level preservation:** if $\sum_k a_k = 1$, a constant mean passes through undistorted (unit gain at zero frequency).

This operation is a *convolution*: $\hat{y} = a * y$. The filter's response to a unit impulse is exactly $\{a_k\}$.

Definition (frequency-domain view)

Represent the series as a superposition of angular frequencies $\omega \in [-\pi, \pi]$ ($\omega = 2\pi/P$ for period P). Any LTI filter has a *frequency response*

$$\rho(\omega) \;=\; \sum_{k=-K}^{K} a_k \, e^{-i\omega k},$$

whose *magnitude* $|\rho(\omega)|$ is the **gain** (how much each frequency passes) and whose *argument* $\arg \rho(\omega)$ is the **phase** (horizontal shift). If x_t has spectral density $f_x(\omega)$, then

$$\operatorname{var}(\hat{x}_t) \;=\; \frac{1}{2\pi} \int_{-\pi}^{\pi} |\rho(\omega)|^2 \, f_x(\omega) \, d\omega \quad \text{(standard convention).}$$

Mental rule: set $|\rho(\omega)| \approx 1$ where you care (the "pass band"), and near 0 where you want attenuation (very low-frequency trend or very high-frequency noise).

Symmetry and phase. If the filter is *symmetric* ($a_{-k} = a_k$), then $\rho(\omega)$ is real and the phase is 0 or π (*zero phase*): the filter does not shift timing; it only rescales frequencies. Causal filters typically induce nonzero phase (delay).

Core examples

(i) M-term moving average (MA).

$$a_k = \begin{cases} 1/M, & k = 0, \ldots, M-1 \quad \text{(one-sided)} \\ 1/M, & k = -(M-1)/2, \ldots, (M-1)/2 \text{ (two-sided, } M \text{ odd)} \end{cases}$$

$$\rho(\omega) = \frac{1}{M} \sum_{k=0}^{M-1} e^{-i\omega k} = e^{-i\omega(M-1)/2} \, \frac{\sin\big(M\omega/2\big)}{M \sin(\omega/2)}.$$

Quick read: a **low-pass** smoother. It attenuates rapid wiggles (high frequencies) and preserves slow drift. The symmetric version avoids phase delay.

(ii) First difference (high-pass).

$$\hat{y}_t = y_t - y_{t-1}, \qquad a_0 = 1, \; a_1 = -1.$$

$$\rho(\omega) = 1 - e^{-i\omega} = e^{-i\omega/2} \cdot 2i \sin(\omega/2), \quad |\rho(\omega)| = 2 \, \big|\sin(\omega/2)\big|.$$

Quick read: **kills** zero frequency (the level/trend) and passes cycle; within $[0, \pi]$, higher ω gets larger gain.

(iii) Band-pass filter (ideal vs. approximations). The *ideal* keeps $|\rho(\omega)| = 1$ for $\omega \in [\omega_L, \omega_H]$ and 0 outside; it is not finitely implementable. Two popular approximations:

- **Baxter–King (BK):** two-sided, symmetric, finite window (removes frequencies outside a chosen range; sacrifices endpoints).
- **Christiano–Fitzgerald (CF):** band-pass approximation that can be implemented *almost* one-sided (better for real time, at the cost of phase and border behavior).

Business-cycle convention (quarterly): pass periods $P \in [6, 32]$ quarters, i.e. $\omega_H = 2\pi/6$ and $\omega_L = 2\pi/32$.

Useful connections

- **HP as a (near) linear filter:** HP acts like a *low-pass* LTI filter in the interior of the sample; it attenuates very low frequencies (trend) and passes cycle. At the endpoints it is no longer strictly LTI (the *endpoint* issue).
- **Sum of weights:** $\sum_k a_k = 1$ preserves the mean; $\sum_k a_k = 0$ (e.g., the difference) removes it.
- **Effect on second moments:** filtering x_t by $\{a_k\}$ multiplies its spectrum by $|\rho(\omega)|^2$. For ARMA data, you can combine the filter polynomial with the process polynomial.

Design and diagnostics

How to choose a filter (quick checklist).

1. **Goal:** growth (log-differences), cycle (band-pass), or smooth trend (low-pass)?
2. **Causality:** for *real time*, use one-sided; for historical analysis, use symmetric (zero phase).
3. **Frequency band:** set $[\omega_L, \omega_H]$ using periods of interest (e.g., 6–32 quarters).
4. **Robustness:** compare at least two methods (e.g., log-diffs vs. HP/BK/CF).

Common pitfalls and fixes.

- **Leakage:** finite filters don't separate bands perfectly; avoid claims driven by razor-thin peaks; inspect $|\rho(\omega)|$.
- **Gibbs ripples:** sharper band edges create more ringing; smoother windows reduce ripples at the cost of gentler transitions.
- **Endpoints:** two-sided filters lose data at the borders or become biased; document how many points you drop or use extensions/nowcasts.
- **Phase delay:** causal filters shift timing; report and, if needed, *correct* the delay when comparing peaks/troughs.

Mini-propositions (handy in class)

- **Symmetry $\Rightarrow$ zero phase:** if $a_{-k} = a_k$, then $\rho(\omega) \in \mathbb{R}$ and the filter does not shift time (it only rescales each frequency).
- **Difference removes the mean:** with $a_0 = 1$, $a_1 = -1$, we have $\sum_k a_k = 0$ and $|\rho(0)| = 0$; hence it eliminates the level trend.

- **Moving average preserves the mean:** $\sum_k a_k = 1$ and $|\rho(0)| = 1$; thus it does not remove a perfectly constant level.

Definition (VAR)**.** Suppose we have n stationary time series variables and we want to track their relationships over time. An order-p *vector autoregression* (VAR(p)) is a statistical model of the form

$$\boldsymbol{y}_t = \boldsymbol{c} + \sum_{s=1}^{p} \boldsymbol{\Phi}_s \boldsymbol{y}_{t-s} + \boldsymbol{\epsilon}_t \tag{1.1}$$

where

- $\boldsymbol{y}_t \in \mathbb{R}^{n\times 1}$ is a vector giving the observations of our n variables at time t;
- $\boldsymbol{c} \in \mathbb{R}^{n\times 1}$ is a constant;
- $\boldsymbol{\Phi}_s \in \mathbb{R}^{n\times n}$ captures the relationships between our time series variables at different lags, in the sense that $\boldsymbol{\Phi}_s^{(i,j)}$ captures the impact of the variable j from s periods ago on variable i today; and
- $\boldsymbol{\epsilon}_t \in \mathbb{R}^{n\times 1}$ is vector white noise: $\mathbb{E}[\boldsymbol{\epsilon}_t] = \mathbf{0}$, $\mathbb{E}\left[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^T\right] = \boldsymbol{\Omega} \in \mathbb{R}^{n\times n}$ for all t, where $\boldsymbol{\Omega} = [\sigma_{i,j}]_{j,i=1}^n$ is the variance-covariance matrix for $\boldsymbol{\epsilon}_t$, and $\mathbb{E}[\boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_j] = \mathbf{0}_{n\times n}$ for all $i \neq j$.

Remark (Estimating a VAR using OLS)**.** If a set of n variables follow a VAR(p), and we have data $\{\boldsymbol{y}_{-p+1}, \boldsymbol{y}_{-p+2}, ..., \boldsymbol{y}_0, \boldsymbol{y}_1, ..., \boldsymbol{y}_T\}$, then we can estimate Equation (1.1) using equation-by-equation OLS. In particular, for each $i = 1, ..., n$, row i of Equation (1.1) gives the value of variable i at time t as

$$y_{t,i} = c_i + \sum_{s=1}^{p} \sum_{j=1}^{n} \Phi_s^{(i,j)} y_{t-s,j} + \epsilon_{t,i} \tag{1.2}$$

with $\mathbb{E}[\epsilon_{t,i}] = 0$ and $\mathbb{E}\left[\epsilon_{t,i}^2\right] = \sigma_{i,i}^2$. We can use the sample $\{y_{1,i}, y_{2,i}, ..., y_{T,i}\}$ to estimate this equation, obtaining OLS estimates $\hat{c}_i$, and $\hat{\Phi}_s^{(i,j)}$ for each $s = 1, ..., p$ and $j = 1, ..., n$:

$$y_{t,i} = \hat{c}_i + \sum_{s=1}^{p} \sum_{j=1}^{n} \hat{\Phi}_s^{(i,j)} y_{t-s,j} + \hat{\epsilon}_{t,i} \tag{1.3}$$

After doing this for each variable, we can concatenate the intercept estimates to get $\hat{\boldsymbol{c}}$, and the coefficient estimates to get matrices $\hat{\Phi}_s$ for each $s = 1, ..., p$. We can estimate $\hat{\boldsymbol{\Omega}} = [\hat{\sigma}_{i,j}]_{i,j=1}^n$ by setting $\hat{\sigma}_{i,i} = \frac{1}{T}\sum_{t=1}^{T} \hat{\epsilon}_{t,i}^2$ for each $i = 1, ..., n$, and $\hat{\sigma}_{i,j} = \frac{1}{T}\sum_{t=1}^{T} \hat{\epsilon}_{t,i}\hat{\epsilon}_{t,j}$ for $i \neq j$.

Remark (Presenting a VAR without a constant)**.** Any VAR can be presented without a constant term by demeaning. Thus, without loss of generality, we can present any VAR without the constant term $\boldsymbol{c}$. To see this, suppose we have a VAR(p) as in Equation (1.1):

$$\boldsymbol{y}_t = \boldsymbol{c} + \sum_{s=1}^{p} \boldsymbol{\Phi}_s \boldsymbol{y}_{t-s} + \boldsymbol{\epsilon}_t.$$

Since each of the n underlying variables are assumed stationary, we can define $\boldsymbol{\mu} := \mathbb{E}[\boldsymbol{y}_t]$, which does not depend on time. Taking the expectation of both sides Equation (1.1), we then get

$$\boldsymbol{\mu} = \boldsymbol{c} + \sum_{s=1}^{p} \boldsymbol{\Phi}_s \boldsymbol{\mu}. \tag{1.4}$$

Thus,

$$\boldsymbol{c} = \boldsymbol{\mu} - \sum_{s=1}^{p} \boldsymbol{\Phi}_s \boldsymbol{\mu} \qquad\qquad = \left(\mathbf{I}_n - \sum_{s=1}^{p} \boldsymbol{\Phi}_s \right) \boldsymbol{\mu}$$

and so

$$\boldsymbol{\mu} = \left(\mathbf{I}_n - \sum_{s=1}^{p} \boldsymbol{\Phi}_s \right)^{-1} \boldsymbol{c}.$$

Now subtract $\boldsymbol{\mu}$ from both sides of Equation 1.1, and add and subtract $\left(\sum_{s=1}^{p} \boldsymbol{\Phi}_s\right) \boldsymbol{\mu}$ from the right-hand side:

$$\begin{aligned}
\boldsymbol{y}_t - \boldsymbol{\mu} &= \boldsymbol{c} - \boldsymbol{\mu} + \left(\sum_{s=1}^{p} \boldsymbol{\Phi_s} \right) \boldsymbol{\mu} + \sum_{s=1}^{p} \boldsymbol{\Phi}_s \boldsymbol{y}_{t-s} - \left(\sum_{s=1}^{p} \boldsymbol{\Phi_s} \right) \boldsymbol{\mu} + \boldsymbol{\epsilon}_t \\
&= \boldsymbol{c} - \left(\mathbf{I}_n - \sum_{s=1}^{p} \boldsymbol{\Phi}_s \right) \boldsymbol{\mu} + \sum_{s=1}^{p} \boldsymbol{\Phi}_s \left(\boldsymbol{y}_{t-s} - \boldsymbol{\mu} \right) + \boldsymbol{\epsilon}_t \\
&= \sum_{s=1}^{p} \boldsymbol{\Phi}_s \left(\boldsymbol{y}_{t-s} - \boldsymbol{\mu} \right) + \boldsymbol{\epsilon}_t.
\end{aligned}$$

Thus, defining $\boldsymbol{x}_t \equiv \boldsymbol{y}_t - \boldsymbol{\mu}$, we have

$$\boldsymbol{x}_t = \sum_{s=1}^{p} \boldsymbol{\Phi}_s \boldsymbol{x}_{t-s} + \boldsymbol{\epsilon}_t, \tag{1.5}$$

which is a VAR(p) with no constant term that has the same innovations and coefficient matrices as the VAR(p) we began with.

Remark (Representing VAR(p) as VAR(1))**.** Any VAR(p) (for finite p) can be represented as a VAR(1). Consider the VAR(p) on n variables (sin constante):

$$\boldsymbol{y}_t = \sum_{s=1}^{p} \boldsymbol{\Phi}_s \boldsymbol{y}_{t-s} + \boldsymbol{\epsilon}_t. \tag{1.6}$$

Para cada t, definí

$$\boldsymbol{\xi}_t = \begin{bmatrix} \boldsymbol{y}_t \\ \boldsymbol{y}_{t-1} \\ \vdots \\ \boldsymbol{y}_{t-p+1} \end{bmatrix} \in \mathbb{R}^{np\times 1}, \qquad \boldsymbol{v}_t = \begin{bmatrix} \boldsymbol{\epsilon}_t \\ \mathbf{0}_{n\times 1} \\ \vdots \\ \mathbf{0}_{n\times 1} \end{bmatrix} \in \mathbb{R}^{np\times 1},$$

y definí la matriz compañera:

$$\boldsymbol{F} = \begin{bmatrix} \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_2 & \cdots & \boldsymbol{\Phi}_{p-1} & \boldsymbol{\Phi}_p \\ \boldsymbol{I}_n & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_n & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{I}_n & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{np\times np}.$$

Notá que $\boldsymbol{v}_t$ es ruido blanco vectorial: para todo t,

$$\mathbb{E}[\boldsymbol{v}_t] = \mathbf{0}_{np\times 1}, \qquad \mathbb{E}\left[\boldsymbol{v}_t \boldsymbol{v}_t^\top\right] = \begin{bmatrix} \boldsymbol{\Omega} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix},$$

y para $t \neq \tau$,

$$\mathbb{E}\left[\boldsymbol{v}_t \boldsymbol{v}_\tau^\top\right] = \mathbf{0}_{np\times np}.$$

Entonces, el VAR(1) en np variables

$$\boldsymbol{\xi}_t = \boldsymbol{F}\,\boldsymbol{\xi}_{t-1} + \boldsymbol{v}_t$$

representa esencialmente el mismo proceso que el VAR(p) original.

Remark (Representing a VAR(p) as a MA(∞)**.** Under certain conditions, we can give a VAR(p) a representation as an MA(∞) process. Without loss of generality, we consider a VAR(1) on n variables with no constant term:

$$\boldsymbol{y_t} = \boldsymbol{\Phi}\boldsymbol{y}_{t-1} + \boldsymbol{\epsilon}_t. \tag{1.7}$$

We can make recursive substitutions in Equation (1.7) to get

$$\begin{aligned}
\boldsymbol{y}_t &= \boldsymbol{\Phi}\left(\boldsymbol{\Phi}\boldsymbol{y}_{t-2} + \boldsymbol{\epsilon}_{t-1}\right) + \boldsymbol{\epsilon}_t \\
&= \boldsymbol{\Phi}^2\boldsymbol{y}_{t-2} + \boldsymbol{\Phi}\boldsymbol{\epsilon}_{t-1} + \boldsymbol{\epsilon}_t \\
&= \boldsymbol{\Phi}^2\left(\boldsymbol{\Phi}\boldsymbol{y}_{t-3} + \boldsymbol{\epsilon}_{t-2}\right) + \boldsymbol{\Phi}\boldsymbol{\epsilon}_{t-1} + \boldsymbol{\epsilon}_t \\
&= \boldsymbol{\Phi}^3\boldsymbol{y}_{t-3} + \boldsymbol{\Phi}^2\boldsymbol{\epsilon}_{t-2} + \boldsymbol{\Phi}\boldsymbol{\epsilon}_{t-1} + \boldsymbol{\epsilon}_t \\
&\vdots \\
&= \boldsymbol{\Phi}^{k+1}\boldsymbol{y}_{t-k-1} + \boldsymbol{\epsilon}_t + \sum_{j=1}^{k}\boldsymbol{\Phi}^j\boldsymbol{\epsilon}_{t-j}.
\end{aligned}$$

If all the eigenvalues of $\boldsymbol{\Phi}$ lie strictly within the unit circle, then $\lim_{k\to\infty}\boldsymbol{\Phi}^{k+1} = \mathbf{0}_{n\times n}$, the process is stationary, and we can take the limit of the recursion to get the following representation as an MA(∞):

$$\boldsymbol{y}_t = \boldsymbol{\epsilon}_t + \sum_{j=1}^{\infty}\boldsymbol{\Phi}^j\boldsymbol{\epsilon}_{t-j}. \tag{1.8}$$

This expresses today's observation as an infinite sum of all past innovations, with the influence of past innovations decreasing as we go further and further back in the past).

Remark (Impulse response functions using the MA(∞) representation)**.** Suppose we have the following VAR(p) on n variables (w.l.o.g. no constant):

$$\boldsymbol{y}_t = \sum_{s=1}^{p}\boldsymbol{\Phi}_s\boldsymbol{y}_{t-s} + \boldsymbol{\epsilon}_t. \tag{1.9}$$

Assume it admits an MA(∞) representation

$$\boldsymbol{y}_t = \boldsymbol{\epsilon}_t + \sum_{j=1}^{\infty}\boldsymbol{\Psi}_j\boldsymbol{\epsilon}_{t-j}. \tag{1.10}$$

Define the matrix polynomial $\boldsymbol{\Psi}(x) := \boldsymbol{I}_n + \sum_{j=1}^{\infty} \boldsymbol{\Psi}_j x^j$. Then (1.10) is

$$\boldsymbol{y}_t = \boldsymbol{\Psi}(L)\,\boldsymbol{\epsilon}_t, \tag{1.11}$$

where L is the lag operator, $L\boldsymbol{x}_t \equiv \boldsymbol{x}_{t-1}$. Likewise, defining $\boldsymbol{\Phi}(x) := \boldsymbol{I}_n - \sum_{s=1}^{p} \boldsymbol{\Phi}_s x^s$, we can rewrite (1.9) as

$$\boldsymbol{\epsilon}_t = \boldsymbol{\Phi}(L)\,\boldsymbol{y}_t. \tag{1.12}$$

Combining (1.11) and (1.12),

$$\begin{aligned}\boldsymbol{y}_t &= \boldsymbol{\Psi}(L)\,\boldsymbol{\epsilon}_t\\ &= \boldsymbol{\Psi}(L)\,\boldsymbol{\Phi}(L)\,\boldsymbol{y}_t,\end{aligned}$$

hence $\boldsymbol{\Psi}(L)\,\boldsymbol{\Phi}(L) = \boldsymbol{I}_n$. Expanding,

$$\begin{aligned}\boldsymbol{\Psi}(L)\,\boldsymbol{\Phi}(L) &= \Big(\boldsymbol{I}_n + \sum_{j=1}^{\infty} \boldsymbol{\Psi}_j L^j\Big)\Big(\boldsymbol{I}_n - \sum_{s=1}^{p} \boldsymbol{\Phi}_s L^s\Big)\\ &= \boldsymbol{I}_n - \sum_{s=1}^{p} \boldsymbol{\Phi}_s L^s + \sum_{j=1}^{\infty} \boldsymbol{\Psi}_j L^j - \sum_{j=1}^{\infty}\sum_{s=1}^{p} \boldsymbol{\Psi}_j \boldsymbol{\Phi}_s L^{j+s}\\ &= \boldsymbol{I}_n + \sum_{j=1}^{\infty} \Big(\boldsymbol{\Psi}_j - \sum_{i=1}^{p} \boldsymbol{\Psi}_{j-i}\boldsymbol{\Phi}_i\Big) L^j,\end{aligned}$$

where we set $\boldsymbol{\Psi}_0 = \boldsymbol{I}_n$ and $\boldsymbol{\Psi}_k = \mathbf{0}_{n\times n}$ for $k < 0$. Since the product equals $\boldsymbol{I}_n$, the coefficient of each power of L^j $(j \geq 1)$ must be $\mathbf{0}_{n\times n}$, yielding the recursion

$$\boldsymbol{\Psi}_j = \sum_{i=1}^{p} \boldsymbol{\Psi}_{j-i}\,\boldsymbol{\Phi}_i, \qquad j = 1, 2, \ldots$$

(equivalent a escribir la suma hasta $i \leq \min\{p, j\}$).

Why is this useful? From $\boldsymbol{y}_t = \boldsymbol{\epsilon}_t + \sum_{j=1}^{\infty} \boldsymbol{\Psi}_j \boldsymbol{\epsilon}_{t-j}$, the moving–average matrices $\{\boldsymbol{\Psi}_j\}$ describen cómo shocks reducidos se transmiten en el tiempo. En particular, el efecto contemporáneo en la variable i de un shock unitario en la componente j ocurrido s períodos atrás es la entrada (i, j) de $\boldsymbol{\Psi}_s$:

$$\frac{\partial y_{i,t+s}}{\partial \epsilon_{j,t}} = \big(\boldsymbol{\Psi}_s\big)_{ij}.$$

Graficar estos coeficientes en distintos leads/lags da la *impulse response function* (IRF). La información de $\{\boldsymbol{\Phi}_s\}$ y la de $\{\boldsymbol{\Psi}_j\}$ es equivalente (vía la relación anterior). **Importante:** la IRF de forma reducida *no* tiene interpretación causal sin una identificación estructural adicional (p.ej., restricciones contemporáneas, long-run, signos, etc.).

Definition. (Granger causality) Consider the following bivariate VAR(p):

$$\Big[y_{t,1} y_{t,2}\Big] = \sum_{s=1}^{p} \Big[\phi_s^{(1,1)} \quad \phi_s^{(1,2)} \phi_s^{(2,1)} \quad \phi_s^{(2,2)}\Big] \Big[y_{t-s,1} y_{t-s,2}\Big] + \Big[\epsilon_{t,1} \epsilon_{t,2}\Big]. \tag{1.13}$$

We say that y_2 does not *Granger cause* y_1 if each matrix ϕ_s is lower triangular; that is, $\phi_s^{(1,2)} = 0$ for all $s = 1, ..., p$, implying that past values of y_2 play no role in determining today's value of y_1. We test for Granger causility as follows. First, we estimate the following *unrestricted equation* using OLS (a direct reading off of the first line of Equation (1.13)):

$$y_{t,1} = \sum_{s=1}^{p} \phi_s^{(1,1)} y_{t-s,1} + \sum_{s=1}^{p} \phi_s^{(1,2)} y_{t-s,2} + \epsilon_{t,1}. \tag{1.14}$$

Save the sum of squares of the residuals from estimating Equation (1.14) as SSR_1. Next, estimate with OLS the following "restricted equation" postulating that past values of y_2 have no impact on contemporary y_1:

$$y_{t,1} = \sum_{s=1}^{p} \phi_s^{(1,1)} y_{t-s,1} + \epsilon_{t,1}. \tag{1.15}$$

Save the sum of squares of the residuals from estimate Equation (1.15) as SSR_0. It will always be the case that $SSR_1 \geq SSR_0$, since are adding more covariates to otherwise the same equation. The question is whether the difference is significant enough for us to reject lack of Granger causality. We conduct statistical tests on the difference $SSR_1 - SSR_0$ (either p or F tests).

Remark. (Limitations of reduced-form VARs) Without a minimum amount of theory, we cannot make causal statements using VARs. The innovations ϵ_t cannot be interpreted as shocks with economic meaning, because they are all correlated with each other. As an example, consider the following structural model:

$$\begin{aligned} y_{1,t} &= b_1 y_{2,t} + k_1 + b_{1,1} y_{1,t-1} + b_{1,2} y_{2,t-1} + u_{1,t}, \\ y_{2,t} &= b_2 y_{1,t} + k_2 + b_{2,1} y_{1,t-1} + b_{2,2} y_{2,t-1} + u_{2,t}, \end{aligned}$$

where $\begin{bmatrix} u_{1,2} & u_{2,t} \end{bmatrix}^T$ is vector white noise. (This is not yet a VAR, because the variables are in terms of each others' contemporaneous values). Concretely, thing of y_1 as GDP and y_2 as governmetn spending. If $1 - b_1 b_2 \neq 0$, we can use substitution to obtain the following representation of the model as a VAR (let's take it on faith):

$$\begin{aligned} y_{1,t} &= \frac{k_1 + b_1 k_2}{1 - b_1 b_2} + \frac{b_{11} + b_1 b_{21}}{1 - b_1 b_2} y_{1,t-1} + \frac{b_{12} + b_1 b_{22}}{1 - b_1 b_2} y_{2,t-1} + \frac{u_{1,t} + b_1 u_{2,t}}{1 - b_1 b_2}, \\ y_{2,t} &= \frac{k_2 + b_2 k_1}{1 - b_1 b_2} + \frac{b_{21} + b_2 b_{21}}{1 - b_1 b_2} y_{1,t-1} + \frac{b_{22} + b_2 b_{12}}{1 - b_1 b_2} y_{2,t-1} + \frac{u_{2,t} + b_2 u_{1,t}}{1 - b_1 b_2}. \end{aligned}$$

or

$$\boldsymbol{y}_t = \boldsymbol{c} + \boldsymbol{\Phi} \boldsymbol{y}_{t-1} + \boldsymbol{\epsilon}, \tag{1.16}$$

where

$$\boldsymbol{c} = \begin{bmatrix} \frac{k_1 + b_1 k_2}{1 - b_1 b_2} & \frac{k_2 + b_2 k_1}{1 - b_1 b_2} \end{bmatrix},$$

$$\boldsymbol{\Phi} = \begin{bmatrix} \frac{b_{11} + b_1 b_{21}}{1 - b_1 b_2} & \frac{b_{12} + b_1 b_{22}}{1 - b_1 b_2} & \frac{b_{21} + b_2 b_{21}}{1 - b_1 b_2} & \frac{b_{22} + b_2 b_{12}}{1 - b_1 b_2} \end{bmatrix},$$

and

$$\boldsymbol{\epsilon}_t = \begin{bmatrix} \frac{u_{1,t} + b_1 u_{2,t}}{1 - b_1 b_2} & \frac{u_{2,5} + b_2 u_{1,t}}{1 - b_1 b_2} \end{bmatrix}.$$

We can estimate these VAR parameters using OLS. But we are not able to recover the structural parameters b_1 and b_2. Note that if we knew b_1 and b_2, we could use our OLS estimates to estimate the other structural parameters by just solving some simple linear systems of equations. What if we don't know b_1 and b_2? We know that $\text{Var}(\epsilon_{1,t}) = \frac{\sigma_1^2 + b_1^2 \sigma_2^2}{(1 - b_1 b_2)^2}$, $\text{Var}(\epsilon_{2,t}) = \frac{\sigma_2^2 + b_2^2 \sigma_1^2}{(1 - b_1 b_2)^2}$, and $\text{Cov}(\epsilon_{1,t}, \epsilon_{2,t}) = \frac{b_2}{(1 - b_1 b_2)^2} \sigma_1^2 + \frac{b_1}{(1 - b_1 b_2)^2} \sigma_2^2$. This is three equations in four unknowns (σ_1, σ_2, b_1, and b_2). Without some more assumptions, we're screwed in terms of recovering structural parameters. If we assume that today's GDP has no contemporaneous effect on government spending (so $b_1 = 0$), we can make progress on estimating the other parameters. Indeed, our equations reduce to

$$\operatorname{Var}\left(\epsilon_{1,t}\right) = \sigma_1^2, \operatorname{Cov}\left(\epsilon_{1,t}, \epsilon_{2,t}\right) = b_2\sigma_1^2, \operatorname{Var}\left(\epsilon_{2,t}\right) = \sigma_2^2 + b_2^2\sigma_1^2$$

from we can estimate structural parameters, utilizing our OLS estimates:

$$\begin{aligned}
\hat{\sigma}_1^2 &= \operatorname{Var}\left(\hat{\epsilon}_{1,t}\right), \\
\hat{b}_2 &= \frac{\operatorname{Cov}\left(\hat{\epsilon}_{1,t}, \hat{\epsilon}_{2,t}\right)}{\hat{\sigma}_1^2} = \frac{\operatorname{Cov}\left(\hat{\epsilon}_{1,t}, \hat{\epsilon}_{2,t}\right)}{\operatorname{Var}\left(\hat{\epsilon}_{1,t}\right)}, \\
\hat{\sigma}_2^2 &= \operatorname{Var}\left(\hat{\epsilon}_{2,t}\right) - \hat{b}_2^2\hat{\sigma}_1^2 = \operatorname{Var}\left(\hat{\epsilon}_{1,t}\right) - \frac{\left(\operatorname{Cov}\left(\hat{\epsilon}_{1,t}, \hat{\epsilon}_{2,t}\right)\right)^2}{\operatorname{Var}\left(\hat{\epsilon}_{1,t}\right)}.
\end{aligned}$$

Remark. Let's more formally exhibit the identification problem inherent in structural VARs. Suppose the true, underlying linear structural model is given by

$$\boldsymbol{B}_0\boldsymbol{y}_t = \sum_{s=1}^{p} \boldsymbol{B}_s\boldsymbol{y}_{t-s} + \boldsymbol{u}_t. \tag{1.17}$$

where

- $\boldsymbol{y}_t \in \mathbb{R}^{n\times 1}$ is our collection of n variables at time t,
- $\boldsymbol{u}_t \in \mathbb{R}^{n\times 1}$ is the vector of *structural shocks* at time t. This is vector white noise, and by definition its components are uncorrelated. By rescaling, we may assume $\operatorname{Var}(u_{i,t}) = 1$ for each i. Thus, for all t,
$$\mathbb{E}[\boldsymbol{u}_t] = \mathbf{0}, \qquad \mathbb{E}[\boldsymbol{u}_t\boldsymbol{u}_t^\top] = \boldsymbol{I}_n,$$
and for all $i \neq j$,
$$\mathbb{E}[\boldsymbol{u}_i\boldsymbol{u}_j^\top] = \mathbf{0}_{n\times n}.$$
- $\boldsymbol{B}_s \in \mathbb{R}^{n\times n}$, $s = 0, 1, \ldots, p$, are coefficient matrices. For $s \geq 1$, the (i, j) entry of $\boldsymbol{B}_s$ gives the impact of variable j lagged s periods on variable i today. $\boldsymbol{B}_0$ is the *contemporaneous effects matrix.* (Note: after rescaling $\boldsymbol{u}_t$, the diagonal of $\boldsymbol{B}_0$ need not be 1.)

If $\boldsymbol{B}_0$ is not invertible, the model cannot be inverted. Suppose instead $\boldsymbol{B}_0$ is invertible. Then multiplying (1.17) by $\boldsymbol{B}_0^{-1}$ yields the VAR(p) representation:

$$\begin{aligned}
\boldsymbol{y}_t &= \sum_{s=1}^{p} \boldsymbol{B}_0^{-1}\boldsymbol{B}_s\boldsymbol{y}_{t-s} + \boldsymbol{B}_0^{-1}\boldsymbol{u}_t \\
&= \sum_{s=1}^{p} \boldsymbol{\Phi}_s\boldsymbol{y}_{t-s} + \boldsymbol{\epsilon}_t,
\end{aligned}$$

where $\boldsymbol{\Phi}_s := \boldsymbol{B}_0^{-1}\boldsymbol{B}_s$ and $\boldsymbol{\epsilon}_t := \boldsymbol{B}_0^{-1}\boldsymbol{u}_t$ are the reduced-form VAR innovations.

Note that for all t,

$$\begin{aligned}\mathbb{E}[\boldsymbol{\epsilon}_t] &= \mathbb{E}[\boldsymbol{B}_0^{-1}\boldsymbol{u}_t] \\ &= \boldsymbol{B}_0^{-1}\mathbb{E}[\boldsymbol{u}_t] \\ &= \boldsymbol{B}_0^{-1}\mathbf{0}_{n\times 1} \\ &= \mathbf{0}_{n\times 1},\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top] &= \mathbb{E}\left[(\boldsymbol{B}_0^{-1}\boldsymbol{u}_t)(\boldsymbol{B}_0^{-1}\boldsymbol{u}_t)^\top\right] \\ &= \boldsymbol{B}_0^{-1}\,\mathbb{E}[\boldsymbol{u}_t\boldsymbol{u}_t^\top]\,\boldsymbol{B}_0^{-T} \\ &= \boldsymbol{B}_0^{-1}\,\boldsymbol{I}_n\,\boldsymbol{B}_0^{-T} \\ &= \boldsymbol{B}_0^{-1}\boldsymbol{B}_0^{-T} \\ &:= \boldsymbol{\Omega}.\end{aligned}$$

For $i \neq j$,

$$\begin{aligned}\mathbb{E}[\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_j^\top] &= \mathbb{E}[(\boldsymbol{B}_0^{-1}\boldsymbol{u}_i)(\boldsymbol{B}_0^{-1}\boldsymbol{u}_j)^\top] \\ &= \boldsymbol{B}_0^{-1}\,\mathbb{E}[\boldsymbol{u}_i\boldsymbol{u}_j^\top]\,\boldsymbol{B}_0^{-T} \\ &= \boldsymbol{B}_0^{-1}\,\mathbf{0}_{n\times n}\,\boldsymbol{B}_0^{-T} \\ &= \mathbf{0}_{n\times n}.\end{aligned}$$

So from $\{\boldsymbol{y}_t\}$ we can estimate $\widehat{\boldsymbol{\Phi}}_s$, $\widehat{\boldsymbol{\epsilon}}_t$, and $\widehat{\boldsymbol{\Omega}}$. Ideally we would like to recover $\boldsymbol{u}_t$ and $\boldsymbol{B}_0$, but absent additional assumptions, this is impossible: $\boldsymbol{\Omega} = \boldsymbol{B}_0^{-1}\boldsymbol{B}_0^{-T}$ does not uniquely pin down $\boldsymbol{B}_0$. In fact, infinitely many $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ satisfy $\boldsymbol{\Omega} = \boldsymbol{A}^{-1}\boldsymbol{A}^{-T}$.

Indeed, let $\boldsymbol{R}$ be any orthogonal matrix ($\boldsymbol{R}\boldsymbol{R}^\top = \boldsymbol{I}_n$). Define fake shocks $\widetilde{\boldsymbol{u}}_t := \boldsymbol{R}\boldsymbol{u}_t$. Then

$$\begin{aligned}\mathbb{E}[\widetilde{\boldsymbol{u}}_t\widetilde{\boldsymbol{u}}_t^\top] &= \mathbb{E}[(\boldsymbol{R}\boldsymbol{u}_t)(\boldsymbol{R}\boldsymbol{u}_t)^\top] \\ &= \mathbb{E}[\boldsymbol{R}\boldsymbol{u}_t\boldsymbol{u}_t^\top\boldsymbol{R}^\top] \\ &= \boldsymbol{R}\,\mathbb{E}[\boldsymbol{u}_t\boldsymbol{u}_t^\top]\,\boldsymbol{R}^\top \\ &= \boldsymbol{R}\,\boldsymbol{I}_n\,\boldsymbol{R}^\top \\ &= \boldsymbol{I}_n.\end{aligned}$$

Thus,

$$\boldsymbol{\Omega} = (\boldsymbol{R}\boldsymbol{B}_0)^{-1}(\boldsymbol{R}\boldsymbol{B}_0)^{-T}.$$

So $\boldsymbol{R}\boldsymbol{B}_0$ is observationally equivalent to $\boldsymbol{B}_0$, and $\widetilde{\boldsymbol{u}}_t$ indistinguishable from the true structural shocks.

Remark (Recursive Cholesky decomposition)**.** Pick up where we left off in the last remark. Recall that a Cholesky decomposition matrix of a real matrix $\boldsymbol{A}$ is an equation $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^T$ where $\boldsymbol{L}$ is a real, lower-triangular matrix with nonnegative diagonal. It is a fact that every positive semidefinite matrix has a unique Cholesky decomposition. Also, we should know that the inverse of a lower triangular matrix is also a lower triangular matrix. Returning to $\boldsymbol{\Omega} = \boldsymbol{B}_0^{-1}\boldsymbol{B}_0^{-T}$. Note that $\boldsymbol{\Omega}$ and $\widehat{\boldsymbol{\Omega}}$ are positive semidefinite. If we assume the assumption that $\boldsymbol{B}_0$ (and hence $\boldsymbol{B}_0^{-1}$) is lower diagonal, then $\boldsymbol{\Omega} = \boldsymbol{B}_0^{-1}\boldsymbol{B}_{\mathbf{0}}{}^{-T}$ is the unique Cholesky decomposition of Ω. Hence we can uniquely identify $\boldsymbol{B}_0$ under the assumption it is lower diagonal (amounting to an economic reordering of variables). With $\boldsymbol{B}_0$ in hand, we can compute impulse response functions to structural shocks. Practically, given our data, we estimate $\widehat{\boldsymbol{\Omega}}$. Using MATLAB, find the Cholesky decomposition of $\widehat{\boldsymbol{\Omega}}$, which will turn out to be $\widehat{\boldsymbol{B}_0}^{-1}$. Then, if we want to simulate the response to a structural shock to $\boldsymbol{\epsilon}_t$, we can use our VAR via the innovation $\boldsymbol{u}_t = \widehat{\boldsymbol{B}_0}^{-1} \boldsymbol{\epsilon}_t$.

Remark (Other canonical identification ideas)**.** We present an alternative identification assumption to assuming that $\boldsymbol{B}_0$ is lower triangular. Recall the MA(∞) representation of the reduced-form VAR:

$$\boldsymbol{y}_t = \boldsymbol{\epsilon}_t + \sum_{s=1}^{\infty} \boldsymbol{\Psi}_s \boldsymbol{\epsilon}_{t-s} = \boldsymbol{\Psi}\left(\boldsymbol{L}\right) \boldsymbol{\epsilon}_t,$$

where $\boldsymbol{\Psi}\left(\boldsymbol{L}\right)$ is the infinite polynomial $\boldsymbol{\Psi}\left(\boldsymbol{x}\right) = \boldsymbol{I}_n + \sum_{s=1}^{\infty} \boldsymbol{\Psi}_s \boldsymbol{x}^s$. Plugging in that $\boldsymbol{\epsilon}_t = \boldsymbol{B}_0^{-1}\boldsymbol{u}_t$, we get that

$$\boldsymbol{y}_t = \boldsymbol{B_0}^{-1}\boldsymbol{u}_t + \sum_{s=1}^{\infty} \boldsymbol{\Psi}_s \boldsymbol{B}_0^{-1} \boldsymbol{u}_{t-s} \qquad = \boldsymbol{\Psi}\left(\boldsymbol{L}\right)\left(\boldsymbol{B}_0^{-1}\boldsymbol{u}_t\right) = \boldsymbol{A}\left(\boldsymbol{L}\right)\boldsymbol{u}_t,$$

where $\boldsymbol{A}\left(\boldsymbol{L}\right)$ is the infinite polynomial $\boldsymbol{A}\left(\boldsymbol{x}\right) = \boldsymbol{B}_0^{-1} + \sum_{s=1}^{\infty} \boldsymbol{\Psi}_s \boldsymbol{B}_0^{-1} \boldsymbol{x}^s$. Note that $\boldsymbol{A}\left(\boldsymbol{0}\right) = \boldsymbol{B_0}^{-1}$. So, the Cholesky identification strategy amounts to placing restrictions on $\boldsymbol{A}\left(\boldsymbol{0}\right)$. Consider instead $\boldsymbol{A}\left(\boldsymbol{1}\right)$:

$$\boldsymbol{A}\left(\boldsymbol{1}\right) = \boldsymbol{B}_0^{-1} + \sum_{s=1}^{\infty} \boldsymbol{\Psi}_s \boldsymbol{B}_0^{-1} \qquad = \left(\boldsymbol{I}_n + \sum_{s=1}^{\infty} \boldsymbol{\Psi}_s\right)\boldsymbol{B}_0^{-1} = \boldsymbol{\Psi}\left(\boldsymbol{1}\right)\boldsymbol{B}_0^{-1}.$$

Note that $\boldsymbol{A}\left(\boldsymbol{1}\right)$ gives the cumulative sum of all responses of VAR variables to a structural shock $\boldsymbol{u}_t$. Hence if our variables are in differences to begin with, then $\boldsymbol{A}\left(\boldsymbol{1}\right)$ gives the sum of the total response of differences to the shock–ie, the long-run change in the level. Therefore, by making assumptions on $\boldsymbol{A}\left(\boldsymbol{1}\right)$, such as assuming certain long-run relationships between variables hold, or saying we don't want certain structural shocks to effect certain variables, etc, we can aid identification. Eg, assume $\boldsymbol{A}\left(\boldsymbol{1}\right)$ is lower triangular, etc.

In practice, we estimate the variance-covariance matrix $\widehat{\boldsymbol{\Omega}}$ from our data for the reduced-form VAR. We estimate a candidate $\widehat{\boldsymbol{B}_0}$ as the inverse of the Cholesky decomposition $\widehat{\boldsymbol{\Lambda}}$ of $\widehat{\boldsymbol{\Omega}}$. Then, we draw orthogonal matrices $\boldsymbol{R}$ and consider a new contemporaneous effects matrix $\boldsymbol{R}\boldsymbol{B}_0$. We continue searching until the matrix $\boldsymbol{A}\left(\boldsymbol{1}\right) = \boldsymbol{\Psi}\left(\boldsymbol{1}\right)\left(\boldsymbol{R}\boldsymbol{B}_0\right)^{-1} = \boldsymbol{\Psi}\left(\boldsymbol{1}\right)\left(\boldsymbol{\Lambda}\boldsymbol{R}^{-1}\right)$ is zero in the elements for which we want to assume zero long-run effects.

2 Consumption

Remark. (Simple Model of Consumption Smoothing) We consider a simple model of consumption smoothing.

- Time is discrete, indexed by t. We begin at time $t = 1$.
- Each period t, agents choose consumption $C_t \geq 0$ and saving/borrowing.. The agent's current stock of wealth is denoted by A_t.
- Agents possess perfect foresight and certainty about:
  - End of life: finite and known. The agent dies at time T.
  - Labor income/endowment stream $(Y_1, Y_2, ..., Y_T)$ received each period, with $Y_t \geq 0$ for all $t = 1, ..., T$.
  - Interest rates are constant across time, and zero: $r_t = r = 0$ for all $t = 1, ..., T$. This is a "pure storage technology."
- No government.
- Perfect capital markets (same interest rate for borrowing and lending). The resources placed in the capital markets are non-perishable.
- Default is forbidden: the agent must end life with positive wealth. $A_T \geq 0$.

The agent derives present utility from consumption through a per-period *felicity* (or flow-utility) function, $u : \mathbb{R}_+ \to \mathbb{R}$. We assume u to be:

- Strictly increasing.
- Strictly concave.
- We assume u to be (once or twice) differentiable as needed. If it is once differentiable, then the assumption that u is strictly increasing implies $u'(C) > 0$ for all $C \in \mathbb{R}_+$. If it is twice differentiable, then the strict concavity assumption implies $u''(C) < 0$ for all $C \in \mathbb{R}_+$.
- Also, we assume that the felicity functions are such that zero consumption will never be chosen in any period. If it is differentiable, this is accomplished by requiring $\lim\limits_{C \to 0^+} u'(C) = +\infty$.

The consumption good C_t is non-durable–it disappears after giving flow utility to the consumer.

Intertemporal (or *lifetime*) utility is given by a discounted sum of per-period flow utilities. Specifically, for a *discount factor* $\beta \in \mathbb{R}_+$, lifetime utility is given by the function $U : \mathbb{R}_+^T \to \mathbb{R}$ defined by $U(C_1, C_2, ..., C_T) = \sum_{t=1}^T \beta^{t-1} u(C_t)$. Note that this form of intertemporal utility implies that more or less consumption yesterday or tomorrow has no impact on the marginal utility of more consumption today: utility is *additively separable* in consumption from different periods. From the corresponding assumptions on u, we see that U is strictly increasing in each argument and strictly concave. The *discount rate* γ is derived implicitly from the discount factor β via $\beta := \frac{1}{1+\gamma}$. For now, assume $\beta = 1$. Then, $\gamma = 0$–the agent discounts the future subjectively at the same rate as the capital market does (interest rate of zero). (Note that in infinite horizon models, we need $\beta < 1$ so that intertemporal utility is finite). The agent is born with resources A_0. Each period, she enters with wealth A_{t-1}, obtains income A_t and chooses her consumption C_t, implicitly defining saving/borrowing A_t as the difference $A_t = A_{t-1} + Y_t - C_t$. Recall that agents must die with nonnegative networth: $A_T \geq 0$. Moreover, we assume there is no bequest motive, so agents will never want to die with nonnegative wealth, as they can get flow utility from consuming whatever is left with no future consequences: $A_T \leq 0$. Together, these conditions imply $A_T = 0$. The agent's choices for $(C_t)_{t=1}^T$ must satisfy the *intertemporal budget constraint* (IBC):

$$\sum_{t=1}^{T} C_t \leq A_0 + \sum_{t=1}^{T} Y_t.$$

In words, her lifetime consumption cannot exceed her lifetime wealth. (We assume that the exogenous variables A_0 and Y_t are such that $A_0 + \sum_{t=1}^T Y_t \geq 0$, or else the problem is not sensible). Since intertemporal utility U is strictly increasing, it is clear that the IBC will always hold with equality. Where does the IBC come from? In fact, the IBC is derived from the institutional per-period *flow-budget constraints* (FBC), which say that each period's consumption, minus borrowing or plus saving, cannot exceed the resources available to the consumer at the beginning of the period:

$$C_1 + A_1 \leq A_0 + Y_1, C_2 + A_2 \qquad \leq A_1 + Y_2, \vdots C_t + A_t \leq A_{t-1} + Y_t, \vdots C_T \qquad \leq A_{T-1} + Y_T.$$

(For the period-T FBC, recall that the agent always chooses $A_T = 0$). Rearrange each flow budget constraint to get a bound on each period's consumption:

$$C_1 \leq A_0 - A_1 + Y_1, C_2 \qquad \leq A_1 - A_2 + Y_2, \vdots C_t \leq A_{t-1} - A_t + Y_t, \vdots C_T \qquad \leq A_{T-1} + Y_T.$$

These constrainst state that each period's consumption cannot exceed that period's income less the change in wealth. Summing these T equations gives the IBC:

$$\begin{aligned}\sum_{t=1}^{T} C_t &\leq \sum_{t=1}^{T-1}((A_{t-1} - A_t) + Y_t) + A_{T-1} + Y_T \\ &= A_0 + \sum_{t=1}^{T} Y_t.\end{aligned}$$

Alternatively, we could start with the IBC and derive the flow-budget constraints: TODOTDO

We proceed to formally state the agent's decision problem: given initial endowment $A_0 \in \mathbb{R}$ and a sequence of labor endowments $(Y_1, Y_2, ..., Y_T) \in \mathbb{R}_+^T$ satisfying $A_0 + \sum_{t=1}^T Y_t$,

$$\max_{(C_1,...,C_T)\in\mathbb{R}_+^T} U(C_1, ..., C_t) = \sum_{t=1}^{T} C_t \text{ s.t. } \sum_{t=1}^{T} C_t \leq A_0 + \sum_{t=1}^{T} Y_t.$$

We see that the seemingly dynamic problem of choosing consumption each period is mathematically equivalent to a static decision. We solve the problem. Recall that zero consumption will never be chosen in any period. Moreover, recall that the IBC will always be satisfied with equality. Thus we can write define the Lagrangian $\mathcal{L} : \mathbb{R}_{++}^T \times \mathbb{R}_+ \to \mathbb{R}$ via

$$\mathcal{L}(C_1, C_2, ..., C_T, \lambda) = \sum_{t=1}^{T} u(C_t) - \lambda[\sum_{t=1}^{T} C_t - A_0 - \sum_{t=1}^{T} Y_t],$$

and solve the unconstrained maximization problem

$$\max_{(C_1,C_2,...,C_T)\in\mathbb{R}_+^T,\lambda\in\mathbb{R}_+} \mathcal{L}(C_1, C_2, ..., C_T, \lambda).$$

Since U is concave, any critical points of $\mathcal{L}$ will solve this problem. For each $t = 1, ..., T$, the first-order condition yields

$$u'(C_t^*) - \lambda^* = 0 \Rightarrow u'(C_t^*) = \lambda^*.$$

In particular, this tells us that $u'(C_1^*) = u'(C_2^*) = ... = u'(C_T^*)$. Since $u''(C) < 0$ for all C, then u' is strictly monotone, and hence it is invertible. Therefore, we have that $C_1^* = C_2^* = ... = C_T^*$. Denote by $\overline{C}$ the value C_t^* for all $t = 1, ..., T$. The first-order conditions have told us about the growth path of C_t (namely, that it stays constant), but we still do not know the level. To find this, we take the first-order condition for λ:

$$\sum_{t=1}^{T} C_t^* - A_0 - \sum_{t=1}^{T} Y_t = 0 \Rightarrow \qquad \sum_{t=1}^{T} \overline{C} = A_0 + \sum_{t=1}^{T} Y_t \Rightarrow \overline{C} = \frac{1}{T}[A_0 + \sum_{t=1}^{T} Y_t].$$

Thus, we see that the agent chooses the same consumption every period, equal to the average of total lifetime income. The marginal propensity to consume out of current income is the same each period, at $\frac{1}{T}$. What does this result imply about saving? Defining per-period saving as $S_t = Y_t - C_t =$ for $t = 1, ..., T$, we find

$$S_t = Y_t - C_t \qquad = Y_t - \frac{1}{T}[A_0 + \sum_{t=1}^{T} Y_t].$$

As an alternative frame for saving, note

$$\begin{aligned} A_t - A_{t-1} &= (A_{t-1} + Y_t - C_t) - A_{t-1} \\ &= Y_t - C_t \\ &= S_t \end{aligned}$$

so saving/dissaving is the transitory change in net asset position. So, $S_t > 0$ precisely when the period's transitory income is greater than average lifetime income, and $S_t < 0$ when transitory income is less than average lifetime income. Punchline: even under total certainty and no risk, the agent will save and borrow to smooth consumption!

Cookie–Cutter: FBC → IBC (telescoping trick)

Consider the flow budget constraint (no labor income):

$$A_{t+1} = (A_t - C_t)(1 + r) = (1 + r)A_t - (1 + r)C_t.$$

Divide by $(1 + r)^{t+1}$ to obtain the *discounted-assets form*:

$$\frac{A_{t+1}}{(1 + r)^{t+1}} = \frac{A_t}{(1 + r)^t} - \frac{C_t}{(1 + r)^t}. \tag{$\star$}$$

Sum $(\star)$ from $t = 0$ to T; the A-terms telescope:

$$\frac{A_{T+1}}{(1 + r)^{T+1}} - \frac{A_0}{1} = -\sum_{t=0}^{T} \frac{C_t}{(1 + r)^t}.$$

Rearranging,

$$\sum_{t=0}^{T} \frac{C_t}{(1 + r)^t} = A_0 - \frac{A_{T+1}}{(1 + r)^{T+1}}.$$

Impose the *No-Ponzi / transversality condition* $\lim_{T\to\infty} \frac{A_{T+1}}{(1+r)^{T+1}} = 0$ to get the **intertemporal budget constraint (IBC)**:

$$\boxed{\sum_{t=0}^{\infty} \frac{C_t}{(1+r)^t} = A_0.}$$

Mnemonic for exams: **D–S–S–L** = *Discount* by $(1+r)^{t+1}$, *Shift* indices, *Sum* (telescopes), take the *Limit.* [a]

[a]Useful variants:

- With income Y_t: $A_{t+1} = (A_t + Y_t - C_t)(1+r) \Rightarrow \sum_{t\geq 0} \frac{C_t}{(1+r)^t} = A_0 + \sum_{t\geq 0} \frac{Y_t}{(1+r)^t}$.
- Time-varying safe rate r_t: discount by $R_{0,t} \equiv \prod_{j=0}^{t-1}(1+r_j)$; IBC becomes $\sum_{t\geq 0} \frac{C_t}{R_{0,t}} = A_0 + \sum_{t\geq 0} \frac{Y_t}{R_{0,t}}$ with $\lim_{T\to\infty} \frac{A_{T+1}}{R_{0,T+1}} = 0$.
- Using $\beta \equiv \frac{1}{1+r}$: $A_{t+1}\beta^{t+1} = A_t\beta^t - C_t\beta^t \Rightarrow \sum_{t\geq 0} \beta^t C_t = A_0$ (and similarly with Y_t).
- Risky/uncertain returns: use the stochastic discount factor M_{t+1}; discounted assets are $M_{0,t}A_t$ and the IBC is $\mathbb{E}_0\left[\sum_{t\geq 0} M_{0,t}C_t\right] = A_0 + \mathbb{E}_0\left[\sum_{t\geq 0} M_{0,t}Y_t\right]$ with $\lim_{T\to\infty} \mathbb{E}_0[M_{0,T+1}A_{T+1}] = 0$.
- Gross return notation $R=1+r$: replace $(1+r)^t$ by R^t throughout.
- Growing measurement units (inflation or population growth g): work in real/per-capita terms or discount additionally by the growth factor so the transversality condition holds.
- Finite horizon T: no limit; instead $\sum_{t=0}^{T} \frac{C_t}{R_{0,t}} = A_0 + \sum_{t=0}^{T} \frac{Y_t}{R_{0,t}} - \frac{A_{T+1}}{R_{0,T+1}}$.

Remark (Two timing conventions yield the same intertemporal budget)**.** *Notation:* $R = 1+r$, $q = 1/R$.

(A) **Pay–tomorrow (payoff units).**

$$A_{t+1} = R\,(A_t + Y_t - C_t) \qquad \Longleftrightarrow \qquad C_t + qA_{t+1} = A_t + Y_t.$$

Divide by R^{t+1}, sum, and impose No–Ponzi:

$$\sum_{t\geq 0} \frac{C_t}{R^t} = A_0 + \sum_{t\geq 0} \frac{Y_t}{R^t}.$$

(B) **Buy–today (present–value units).**

$$C_t + \frac{W_{t+1}}{R} = W_t + Y_t \qquad \Longleftrightarrow \qquad W_{t+1} = R\,(W_t + Y_t - C_t).$$

Divide by R^t, sum, and impose No–Ponzi:

$$\sum_{t\geq 0} \frac{C_t}{R^t} = W_0 + \sum_{t\geq 0} \frac{Y_t}{R^t}.$$

Dictionary. At each date t, $W_t = A_t$ (same "today–pesos"). The difference is whether R multiplies savings ($A_{t+1} = R(\cdot)$) or appears in the purchase price of assets (W_{t+1}/R).

Remark. (Simple model of retirement savings) Let us make Model into a very basic life cycle model. Let T_j be the (known) *retirement date*, and set $Y_t = \overline{Y}$ for $t = 1, ..., T_j$ for fixed $\overline{Y} \geq 0$ and $Y_t = 0$ for $t \geq T_j + 1$. Assume $A_0 = 0$ for simplicity. Application of the formula $C_t = \overline{C} = \frac{1}{T}\left[A_0 + \sum_{t=1}^{T} Y_t\right]$ for all $t \in \{1, ..., T\}$ to our specific circumstances yields

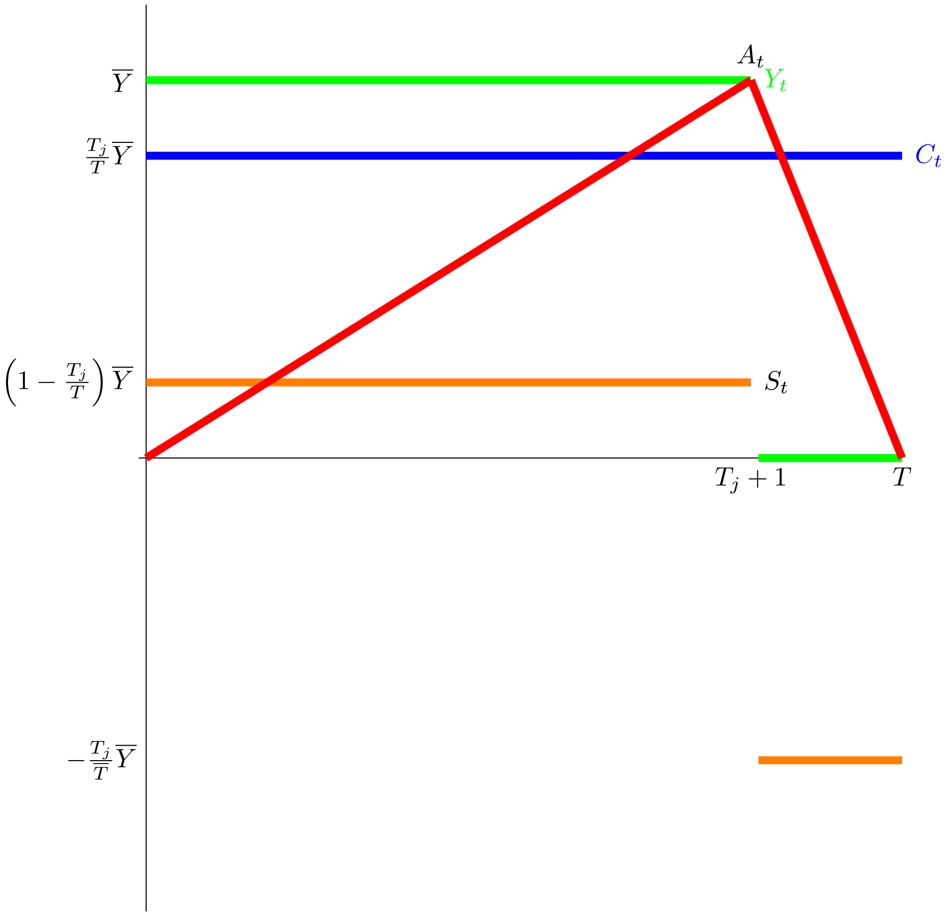
$$C_t = \overline{C} = \frac{T_j}{T}\,\overline{Y}.$$

Thus, savings $S_t = Y_t - C_t$ is given by

$$S_t = \begin{cases} \left(1 - \frac{T_j}{T}\right)\overline{Y}, & t = 1, \dots, T_j, \\ \overline{Y}, & t = T_j + 1, \dots, T. \end{cases}$$

So during the working life, $S_t > 0$, and during retirement, $S_t < 0$. So, the lifecycle paths of the variables are as follows:

- Consumption stays constant every period at $\overline{C} = \frac{T_j}{T}\overline{Y}$.
- Income: constant at $\overline{Y}$ for $t = 1, ..., T_j$, then constant at zero for $t = T_j + 1, ..., T$.
- Savings: the difference between income and consumption.
- Assets: triangular shape: increase until T_j, then declining to zero until T.



Remark. (Relative risk aversion) For a twice differentiable, strictly increasing, strictly concave felicity function $u : \mathbb{R}_+ \to \mathbb{R}$, we define the *relative risk aversion* at $C \in \mathbb{R}_{++}$ by $-\frac{u''(C)C_t}{u'(C)}$. (Recall $u''(C) > 0$ for all C, so relative risk aversion is positive). Intuitively, $-u''(C)$ governs the degree of concavity in the consumer's preferences, thereby governing their risk aversion–how much they prefer a sure bet in the middle versus a risky bet with the same expected utility. To get this into a form that is invariant across affine transformations of a utility function (which do not affect preferences), divide by $u'(C)$. (Indeed, if $v : \mathbb{R}_+ \to \mathbb{R}$ is another utility function given by $v(C) = au(C) + b$, then $-\frac{v''(C)}{v'(C)} = -\frac{au''(C)}{au'(C)} = -\frac{u''(C)}{u'(C)}$.) Finally, note that the expression $-\frac{u''(C)}{u'(C)}$ is currently in units of inverse consumption; to make it unitless, we multiply by C to obtain $-\frac{u''(C)C}{u'(C)}$, the coefficient of relative risk aversion. In general, the coefficient of relative risk aversion varies depending on C. But a special class of utility functions have a constant relative risk aversion at all levels of consumption. For any $\sigma \geq 0$, we define the CRRA (constant relative risk aversion) utility function $u : \mathbb{R}_{++} \to \mathbb{R}$ by

$$u(C) = \frac{C^{1-\sigma} - 1}{1-\sigma}, \ \sigma \neq 1,$$
$$u(C) = \ln(C), \ \sigma = 1$$

which has constant relative risk aversion of σ at each level of consumption (stay tuned). In fact, this is the *only* class of utility functions with constant relative risk aversion (prove). (Often, the minus 1 in the numerator is omitted because it does not change preferences, but it is useful theoretically). Note that for $\sigma = 0$, we simply have $u(C) = C - 1$, so utility is simply linear. We will justify below that $\ln(C)$ is the limiting case of $u(C)$ as $\sigma \to 1$. For $1 \neq \sigma \geq 0$, we show that the CRRA utility function with parameter σ has constant coefficient of relative risk aversion σ. Indeed, for all $C \in \mathbb{R}_{++}$,

$$u'(C) = C^{-\sigma}, u''(C) = -\sigma C^{-\sigma-1}$$

and so

$$\text{Relative risk aversion at } C = -\frac{u''(C)C}{u'(C)} = -\frac{-\sigma C^{-\sigma-1}C}{C^{-\sigma}} = \sigma$$

so relative risk aversion is the same at all consumption levels, and is the parameter σ. We show that $u(C) \to \ln(C)$ as $\sigma \to 1$: using L'Hôpital's rule,

$$\lim_{\sigma \to 1} \frac{C^{1-\sigma} - 1}{1-\sigma} = \lim_{\sigma \to 1} \frac{-\ln(C)\, C^{1-\sigma}}{-1} = \ln(C).$$

And the utility function $u(C) = \ln(C)$ has constant relative risk aversion of $\sigma = 1$:

$$u'(C) = \frac{1}{C}, u''(C) = -\frac{1}{C^2}$$

and so

$$\text{Relative risk aversion at } C = -\frac{u''(C)C}{u'(C)} = -\frac{-\frac{1}{C^2}C}{\frac{1}{C}} = 1.$$

Remark. (General model of consumption and saving under perfect foresight) We generalize Model as follows:

1. Discount factor $\beta < 1$, implying a discount rate $\gamma = \frac{1-\beta}{\beta} > 0$. The agent (subjectively) discounts the future.
2. Strictly positive interest rate $r > 0$. Still constant over time and same for borrowing and saving.
3. Government levying proportional, time-varying consumption taxes τ_t^c and τ_t^y, $t = 1, ..., T$. Government revenues are thrown in the ocean.

Otherwise, we keep the same general setting. The intertemporal utility function is now

$$U(C_1, C_2, ..., C_T) = \sum_{t=1}^{T} \beta^{t-1} u(C_t).$$

The flow budget-constraints:

$$
\begin{aligned}
C_1(1+\tau_1^c) + A_1 &= Y_1(1-\tau_1^y) + A_0(1+r), \\
C_2(1+\tau_2^c) + A_2 &= Y_2(1-\tau_2^y) + A_1(1+r), \\
&\vdots \\
C_t(1+\tau_t^c) + A_t &= Y_2(1-\tau_t^y) + A_{t-1}(1+r), \\
&\vdots \\
C_T(1+\tau_T^c) &= Y_T(1-\tau_T^y) + A_{T-1}(1+r).
\end{aligned}
$$

(Recall that we assume $A_T = 0$–no bequest motive and no default). To derive the intertemporal budget constraint, we multiply the period-t flow budget constraint by $\frac{1}{(1+r)^{t-1}}$ for $t = 1, ..., T$, to obtain

$$
\frac{1}{(1+r)^{t-1}} \left[C_t\left(1+\tau_t^c\right) + A_t\right] = \frac{1}{\left(1+r\right)^{t-1}} \left[Y_2\left(1-\tau_2^y\right)\right] + \frac{1}{\left(1+r\right)^{t-2}} A_{t-1},\ t = 1, ..., T.
$$

Rearrange to

$$
\frac{C_t\left(1+\tau_t^c\right)}{\left(1+r\right)^{t-1}} = \frac{Y_t\left(1-\tau_2^y\right)}{\left(1+r\right)^{t-1}} + \frac{A_{t-1}}{\left(1+r\right)^{t-2}} - \frac{A_t}{\left(1+r\right)^{t-1}},\ t = 1, ..., T.
$$

Finally, sum up these T equations to get the intertemporal budget constraint:

$$
\begin{aligned}
\sum_{t=1}^{T} \frac{C_t\left(1+\tau_t^c\right)}{\left(1+r\right)^{t-1}} &= \sum_{t=1}^{T} \frac{Y_t\left(1-\tau_2^y\right)}{\left(1+r\right)^{t-1}} + \sum_{t=1}^{T} \frac{A_{t-1}}{\left(1+r\right)^{t-2}} - \sum_{t=1}^{T} \frac{A_t}{\left(1+r\right)^{t-1}} \\
&= \sum_{t=1}^{T} \frac{Y_t\left(1-\tau_2^y\right)}{\left(1+r\right)^{t-1}} + \sum_{t=0}^{T-1} \frac{A_t}{\left(1+r\right)^{t-1}} - \sum_{t=1}^{T} \frac{A_t}{\left(1+r\right)^{t-1}} \\
&= \sum_{t=1}^{T} \frac{Y_t\left(1-\tau_2^y\right)}{\left(1+r\right)^{t-1}} + \frac{A_0}{\left(1+r\right)^{0-1}} + \frac{A_T}{\left(1+r\right)^{T-1}} \\
&= \sum_{t=1}^{T} \frac{Y_t\left(1-\tau_2^y\right)}{\left(1+r\right)^{t-1}} + \left(1+r\right) A_0.
\end{aligned}
$$

That is,

$$
\sum_{t=1}^{T} \frac{C_t\left(1+\tau_t^c\right)}{\left(1+r\right)^{t-1}} = \sum_{t=1}^{T} \frac{Y_t\left(1-\tau_2^y\right)}{\left(1+r\right)^{t-1}} + \left(1+r\right) A_0,
$$

that is, the present value of (consumption tax-inclusive) lifetime consumption equals the present value of lifetime income. We now formally state the problem:

$$
\begin{aligned}
&\max_{C_1, C_2, ..., C_T \in \mathbb{R}_+^T} U(C_1, C_2, ..., C_T) = \sum_{t=1}^{T} \beta^{t-1} u(C_t), \\
&\text{s.t.} \ \sum_{t=1}^{T} \frac{C_t\left(1+\tau_t^c\right)}{\left(1+r\right)^{t-1}} \leq \sum_{t=1}^{T} \frac{Y_t\left(1-\tau_t^y\right)}{\left(1+r\right)^{t-1}} + \left(1+r\right) A_0
\end{aligned}
$$

with $A_0 \in \mathbb{R}$ given, $A_T = 0$, $(Y_1, ..., Y_T) \in \mathbb{R}_+^T$ given, $\beta < 1$ given, $r > 0$ given, and $(\tau_t^c, \tau_t^y)_{t=1}^T$ given, with $\sum_{t=1}^{T} \frac{Y_t\left(1-\tau_2^y\right)}{(1+r)^{t-1}} + (1+r)A_0 \geq 0$. As before, our assumptions on felicity imply that each C_t chosen will be strictly positive, and the intertemporal budget constraint will hold with equality. Hence our decision is equivalent to:

$$\max_{(C_1,...,C_T)\in\mathbb{R}^T_{++},\lambda\in\mathbb{R}_{++}} \mathcal{L}(C_1,...,C_T,\lambda)$$
$$=\sum_{t=1}^{T}\beta^{t-1}u(C_t)+\lambda\left[\sum_{t=1}^{T}\frac{Y_t\left(1-\tau_t^y\right)}{\left(1+r\right)^{t-1}}+\left(1+r\right)A_0-\sum_{t=1}^{T}\frac{C_t\left(1+\tau_t^c\right)}{\left(1+r\right)^{t-1}}\right].$$

This problem is concave, so it suffices to find the critical points. For each $t=1,...T-1$, the FOC for C_t is

$$\beta^{t-1}u'(C_t^*)=\lambda^*\frac{1+\tau_t^c}{\left(1+r\right)^{t-1}}$$

and the FOC for C_{t+1} is

$$\beta^t u'\left(C_{t+1}^*\right)=\lambda^*\frac{1+\tau_{t+1}^c}{\left(1+r\right)^t}.$$

Dividing these two conditions gives the *consumption Euler equation*:

$$\beta\frac{u'\left(C_{t+1}^*\right)}{u'\left(C_t^*\right)}=\frac{1+\tau_{t+1}^c}{\left(1+r\right)\left(1+\tau_t^c\right)},\ t=1,...,T-1.$$

Notice that income Y_t and labor income taxes τ_t^y do not factor into this equation at all. Do a first-order Taylor approximation of u' about C_t^*, evaluated at C_{t+1}^*:

$$u'\left(C_{t+1}^*\right)=u'\left(C_t^*\right)+u''\left(C_t^*\right)\left(C_{t+1}^*-C_t^*\right)+O\left(\left(C_{t+1}^*-C_t^*\right)^2\right),\ t=1,...,T-1.$$

Divide each of these equations through by $u'\left(C_t^*\right)$ to get

$$\frac{u'\left(C_{t+1}^*\right)}{u'\left(C_t^*\right)}=1+\frac{u''\left(C_t^*\right)\cdot C_t^*}{u'\left(C_t^*\right)}\cdot\frac{C_{t+1}^*-C_t^*}{C_t^*}+O\left(\left(C_{t+1}^*-C_t^*\right)^2\right).$$

(Drop stars from now on). We then see that to a first-order,

$$\frac{u'\left(C_{t+1}\right)}{u'\left(C_t\right)}=1-\text{ coefficient of relative risk aversion}*\text{growth rate of consumption}$$

so $\frac{u'(C_{t+1})}{u'(C_t)}$ is inversely related to the growth rate of consumption.
Recall that the *elasticity of intertemporal substitution* at time t is given by

$$-\frac{d\left(\frac{C_{t+1}}{C_t}\right)}{d\left(\frac{u'(C_{t+1})}{u'(C_t)}\right)}\cdot\frac{\frac{u'(C_{t+1})}{u'(C_t)}}{\frac{C_{t+1}}{C_t}}$$

This captures the degree to which the growth rate of consumption is sensitive to changes in the marginal rate of substitution $(u'(C_{t+1})/u'(C_t))$. If it is high, then a small increase in the intertemporal rate of substitution is associated with a large decrease in the growth rate of consumption. Note we can write the elasticity of intertemporal substitution as

$$-\left(\frac{d\left(\ln\left(\frac{C_{t+1}}{C_t}\right)\right)}{d\left(\ln\left(\frac{u'(C_{t+1})}{u'(C_t)}\right)\right)}\right).$$

If our utility function is CRRA with parameter $-\sigma$, then this further reduces to

$$-\left(\frac{d\left(\ln\left(\frac{C_{t+1}}{C_t}\right)\right)}{d\left(\ln\left(\left(\frac{C_{t+1}}{C_t}\right)^{-\sigma}\right)\right)}\right) = \qquad \frac{1}{\sigma}\left(\frac{d\left(\ln\left(\frac{C_{t+1}}{C_t}\right)\right)}{d\left(\ln\left(\frac{C_{t+1}}{C_t}\right)\right)}\right) = \frac{1}{\sigma}.$$

So under CRRA utility, the elasticity of intertemporal substitution is equal to the inverse of the coefficient of relative risk aversion. That is weird! Back to the model. Denote by $\sigma_t = -\frac{u''(C_t)C_t}{u'(C_t)}$ the coefficient of relative risk aversion at time t. (If utility is CRRA, then $\sigma_t = \sigma$ for all t). Then, we found that for $t = 1, ..., T_1$,

$$\beta\frac{u'\left(C_{t+1}\right)}{u'\left(C_t\right)} = \frac{1+\tau_{t+1}^c}{\left(1+r\right)\left(1+\tau_t^c\right)}$$

and up to the first order,

$$\frac{u'\left(C_{t+1}\right)}{u'\left(C_t\right)} = 1-\sigma_t\cdot\frac{\Delta C_{t+1}}{C_t}.$$

Equating these expressions gives that, up to a first order, for each $t = 1, ..., T-1$, we have the following approximation for the growth rate of consumption:

$$\frac{u'\left(C_{t+1}\right)}{u'\left(C_t\right)} = \frac{1+\tau_{t+1}^c}{\beta\left(1+r\right)\left(1+\tau_t^c\right)} = 1-\sigma_t\cdot\frac{\Delta C_{t+1}}{C_t} \Rightarrow \qquad \frac{\Delta C_{t+1}}{C_t} = \frac{1}{\sigma_t}\left[1-\frac{1+\tau_{t+1}^c}{\beta\left(1+r\right)\left(1+\tau_t^c\right)}\right].$$

This is called a *Keynes-Ramsey* formula: the Euler equation has been transformed from an expression involving marginal utilities into an expression about the growth rate of consumption. Some remarks on the above Keynes-Ramsey formula:

- Income Y_t, income taxes τ_t^y, and initial wealth A_0 are all irrelevant for consumption growth.
- Suppose consumption taxes will icnrease tomorrow: $\tau_t^c < \tau_{t+1}^c$. This will cause a decrease in the growth rate of consumption (and if $\beta\left(1+r\right) = 1$, it will certainly cause negative consumption growth rate). The level of consumption today depends on the balance between the intertemporal substitution effect and the intertemporal wealth effect from future tax increases. Could potentially cause a consumption boom this period depending.

From now on, we assume all consumption taxes are constant: $\tau_t^c = \tau$ for all t. We made our point about unconventional fiscal policy. Then, the Keynes-Ramsey formula reduces to

$$\begin{aligned}\frac{\Delta C_{t+1}}{C_t} &= \frac{1}{\sigma_t}\left[1-\frac{1}{\beta\left(1+r\right)}\right]\\ &= \frac{1}{\sigma_t}\left[\frac{\beta\left(1+r\right)-1}{\beta\left(1+r\right)}\right]\\ &= \frac{1}{\sigma_t}\left[\frac{\beta\left(1+r\right)-\beta\left(1+\gamma\right)}{\beta\left(1+r\right)}\right]\\ &= \frac{1}{\sigma_t}\left[\frac{\beta\left(\left(1+r\right)-\left(1+\gamma\right)\right)}{\beta\left(1+r\right)}\right]\end{aligned}$$

where, recall that γ is the discount rate defined such that $\beta = \frac{1}{1+\gamma}$. Finally, we see that, again up to a first order, and assuming σ_t is roughly constant at some σ (literally true if felicity is CRRA),

$$\frac{\Delta C_{t+1}}{C_t} = \frac{r-\gamma}{\sigma\left(1+r\right)}.$$

From this new Keynes-Ramsey formula, we get some more insights:

1. If relative risk aversion σ is high, so that $\frac{1}{\sigma}$ is low, then any change in interest or discount rates translates into less change on the consumption growth rate than would a lower σ.
2. If $r = \gamma$, implying $r = \frac{1-\beta}{\beta} \Rightarrow \beta r = 1 - \beta \Rightarrow \beta\left(1+r\right) = 1$, then the market interest rate equals the discount rate, and consumption growth is zero. Intuitively, the market compensates the agent just enough for her impatience that she does not tilt consumption.
3. If $r > \gamma$, then the market discounts the future more than the agent does. The agent is patient and defers consumption–consumption growth is positive.
4. If $r < \gamma$, then the market discounts the future less than does the agent. The agent is too impatient to reap the rewards of waiting, and consumption growth is negative.

We can do some comparitive statics on what happens if expogenous parameters change:

1. Suppose that the discount rate γ rises, that is, β increases. We see that consumption growth will fall. Intuitively, the agent discounts the future more, is more impatient, and shifts relatively more ocnsumption to today.
2. The derivative of the Keynes-Ramsey formula with respect to r is

   $$\frac{\sigma\left(1+r\right) - \left(r-\gamma\right)\sigma}{\sigma^2\left(1+r\right)^2} = \frac{1+\gamma}{\sigma^2\left(1+r\right)^2} > 0$$

   implying that an increase in the interest rate will cause consumption growth to increase.
3. Again, none of this has to do with levels.

Punchline: up to a first order, we have that consumption will be constant over time (as in Model) if consumption taxes are constant over time and the consumers are as patient as the market. In general, however, consumption will not be constant, but grows and shrinks over time. Even in this perfect foresight world, there is *consumption tilting* when consumption taxes are changing or consumers have differing patience from the market.

In continuous time (need $\beta < 1$), the annoying $1+r$ in the denominator of the Keynes-Ramsey formula drops out. Assuming consumption taxes are constant over time, and that discounting is proportional to time. We can write (some of this math is a little iffy)

$$\beta\frac{u'\left(C_{t+1}\right)}{u'\left(C_t\right)} = \frac{1}{1+r} \Rightarrow \qquad \frac{u'\left(C_{t+1}\right)}{u'\left(C_t\right)} = \frac{1+\gamma}{1+r}$$

so for small increments of time Δt,

$$\frac{u'\left(C_{t+\Delta t}\right)}{u'\left(C_t\right)} = \frac{1+\gamma\Delta t}{1+r\Delta t}.$$

We can approximate $C_{t+\Delta t}$ as $C_t + \Delta t\dot{C}_t$, where $\dot{C}_t = \frac{\partial C}{\partial t}$ is the instantaneous rate of change of C. Then we use a Taylor expansion of u' about C_t to get the approximation $u'\left(C_t + \Delta t\dot{C}_t\right) = u'\left(C_t\right) +$

$u''(C_t)\,\Delta t\dot{C}_t$. Thus,

$$\begin{aligned}
&\frac{u'(C_t)+u''(C_t)\,\Delta t\dot{C}_t}{u'(C_t)}=\frac{1+\gamma\Delta t}{1+r\Delta t}\\
\Rightarrow\ &1+\frac{u''(C_t)\,\Delta t\dot{C}_t}{u'(C_t)}=\frac{1+\gamma\Delta t}{1+r\Delta t}\\
\Rightarrow\ &\frac{u''(C_t)\,\Delta t\dot{C}_t}{u'(C_t)}=\frac{1+\gamma\Delta t-1-r\Delta t}{1+r\Delta t}=\frac{\Delta t\,(\gamma-r)}{1+r\Delta t}\\
\Rightarrow\ &\frac{u''(C_t)\,\dot{C}_t}{u'(C_t)}=\frac{\gamma-r}{1+r\Delta t}.
\end{aligned}$$

Multiply the left-hand side by $\frac{C_t}{C_t}$. Assume the elasticity of intertemporal substitution is is roughly constant at $\frac{1}{\sigma}=-\frac{u'(C)}{u''(C)C}$. Then,

$$\frac{u''(C_t)\,C_t}{u'(C_t)}\cdot\frac{\dot{C}_t}{C_t}=\frac{\gamma-r}{1+r\Delta t}\Rightarrow\qquad -\sigma\cdot\frac{\dot{C}_t}{C_t}=\frac{\gamma-r}{1+r\Delta t}\Rightarrow\ \frac{\dot{C}_t}{C_t}=\frac{1}{\sigma}\cdot\frac{r-\gamma}{1+r\Delta t}.$$

Taking $\Delta t\to 0$:

$$\frac{\dot{C}_t}{C_t}=\frac{r-\gamma}{\sigma}$$

saying that the rate of change of consumption is the difference between the market and subjective interest rates, modulated by the intertemporal elasticity of substitution. Return now to the case of discrete time. Assume that consumption taxes are constant. Then, the Euler equation reads

$$\beta\,(1+r)\,u'(C_{t+1})=u'(C_t)\,.$$

We can interpret this intuitively with a marginal benefit-marginal cost analysis. If the consumer is behaving optimally, then any infinitesimal change should not result in a change in utility. So suppose she saves an infinitesimal amount dC more today and consumes it tomorrow. How does that impact her utility? She loses approximately $u'(C_t)\,dC$ utils from the loss of consumption today, but gains $\beta\,(1+r)\,u'(C_{t+1})\,dC$ tomorrow. Thus these quantities should be the same if she is at the optimum. We can solve the model in closed-form in the case of $\beta\,(1+r)=1$, that is, $\gamma=r$. Assume also that consumption taxes are zero. In all periods. Then the Euler equation reads

$$u'(C_{t+1})=u'(C_t)$$

so since u' is invertible, we have $C_1=C_2=...=C_T=\overline{C}$. Denote lifetime income by $W=A_0\,(1+r)+$

$\sum_{t=1}^{T}\frac{1-\tau_t^y}{(1+r)^{t-1}}Y_t$. Plug these into the intertemporal budget constraint to get

$$\begin{aligned}
\sum_{t=1}^{T}\frac{C_t}{(1+r)^{t-1}} &= W \\
\Rightarrow \overline{C}\sum_{t=1}^{T}\frac{1}{(1+r)^{t-1}} &= W \\
\Rightarrow \overline{C}\frac{\left(\frac{1}{1+r}\right)^T-1}{\left(\frac{1}{1+r}\right)-1} &= W \\
\Rightarrow \overline{C}\frac{(1+r)\left(\left(\frac{1}{1+r}\right)^T-1\right)}{-r} &= W \\
\Rightarrow \overline{C} = \frac{r}{(1+r)\left(1-\left(\frac{1}{1+r}\right)^T\right)}W.
\end{aligned}$$

As $T\to\infty$, then each period's consumption approaches the "annuity value" of lifetime wealth (with the pesky $1+r$ in the denominator, an artifact of discrete time):

$$\overline{C} = \frac{r}{1+r}W.$$

Note that the marginal propensity to consume out of current income is $r\frac{1}{1+r} = \left(\frac{1-\beta}{\beta}\right)\beta = 1-\beta$. Since β is thought of as close to one, this means the marginal propensity to consume is small. In the general case of $\beta(1+r)\neq 1$ (that is, $r\neq\gamma$), we solve numerically using a shooting algorithm. We only need to solve for one thing, C_1, since perfect foresight and the Euler equations imply everything else. We guess a value for C_1, use the Euler equations to compute the implied $C_2, C_3, ..., C_T$, and check how close the present value of lifetime consumption is to W. Iterate until you are arbitrarily close to equality in the IBC. This type of algorithm can accommodate all sorts of time-varying stuff (interest rates, discount factors, taxes, etc.) so long as it is deterministic and known.

Remark. (Model of consumption and savings under income risk, with quadratic felicity) We now present a model of consumption and savings where we drop the perfect foresight assumption. Specifically, we allow per-period income to be random. We utilize the *quadratic felicity function*

$$u'(C_t) = C_t - \frac{a}{2}C_t^2$$

for a constant $a>0$. Note that marginal utility is linear: $u'(C_t) = 1-aC_t$. This is convenient analytically, but quadratic felicity has some drawbacks:

1. Since $u'(0)=1$, the condition that $\lim_{C\to 0^+} u'(C)=\infty$ does not hold. Thus we could get stuck at a boundary solution where someone finds it optimal to starve to death.

2. Note that marginal utility is negative for $C>\frac{1}{a}$. This means that quadratic felicity violates the assumption of strictly increasing marginal utility. In practice, we can often get around this issue by setting a to be as small as we need.

For simplicity, assume that all taxes are zero: $\tau_t^c=\tau_t^y=0$ for all t. Assume as well that $\gamma=r=0$, so $\beta=1$. None of these assumptions are fundamental, but just for expositional clarity so we can abstract away from consumption tilting. The income stream $\{Y_t\}_{t=1}^T$ is a stochastic process. At $t=1$, the consumer only knows Y_1, and can only immediately control C_1. The consumer's problem is to choose decision rules $C_1(Y_1)$, $C_2(Y_1,Y_2)$, ..., $C_T(Y_1,...,Y_T)$, where $C_t(Y_1,...,Y_t)$ is a mapping from the set of all possible histories $Y_1,...,Y_t$ into a nonnegative consumption level, subject to the intertemporal budget constraint. Formally, the problem is

$$\max_{C_1\geq 0,\ldots,C_T\geq 0}\mathbb{E}_1\left[\sum_{t=1}^{T}\left(C_t-\frac{a}{2}C_t^2\right)\right],\qquad\qquad \text{s.t.}\ \sum_{t=1}^{T}C_t\leq A_0+\sum_{t=1}^{T}Y_t$$

where again, $C_t\geq 0$ is not a single number, but rather a mapping from the space of histories $(Y_1,...,Y_t)$ into $\mathbb{R}_+$. The budget constraint must hold with in *every* state of the world, not just in expectation. We require that $A_0+\sum_{t=1}^{T}Y_t\geq 0$ in every state of the world, so that $A_T<0$ does not happen from now fault of the consumer. We solve for now using a variational argument. We assume equality in the intertemporal budget constraint (even though inequality could occur with quadratic felicity, since agents don't always want to consume everything.). Suppose we are at a solution, and that we defer an infinitesimal amount of consumption from period 1 to period t. Note that the utility cost of giving up an infinitesimal amount of consumption today is $u'(C_1)\,dC=(1-aC_1)\,dC$. The expected utility gain from the future consumption is $\mathbb{E}_1[u'(C_t)\,dC]=\mathbb{E}_1[(1-aC_t)\,dC]=(1-a\mathbb{E}_1(C_t))\,dC$. Equating the benefit and cost gives $(1-aC_1)\,dC=(1-a\mathbb{E}_1(C_t))\,dC$, which yields the Euler equation

$$C_1=\mathbb{E}_1[C_t]\ \text{ for all } t.$$

So consumption is constant in expectation. Hence we will smooth consumption in expectation. Now, assuming equality in the IBC, we take the expectation of both sides:

$$\begin{aligned}
&\sum_{t=1}^{T}C_t=A_0+\sum_{t=1}^{T}Y_t\\
&\Rightarrow\mathbb{E}_1\left[\sum_{t=1}^{T}C_t\right]=\mathbb{E}_1\left[A_0+\sum_{t=1}^{T}Y_t\right]\\
&\Rightarrow C_1+\sum_{t=2}^{T}\mathbb{E}_1[C_t]=A_0+\sum_{t=1}^{T}\mathbb{E}_1[Y_t]\\
&\Rightarrow TC_1=A_0+\sum_{t=1}^{T}\mathbb{E}_1[Y_t]\\
&\Rightarrow C_1=\frac{1}{T}\left(A_0+\sum_{t=1}^{T}\mathbb{E}_1[Y_t]\right)
\end{aligned}$$

so period-1 consumption is the time average of *expected* lifetime income, given the knowledge we have at period 1. This behavior is called *certainty equivalence*: the consumer behaves almost as under certainty, just with the known income stream being replaced by the expected income stream. There is no precautionary savings present. It turns out that certainty equivalence is unique to the quadratic felicity function. The result ultimately stems from the fact that for quadratic felicity, $\mathbb{E}[u'(C_t)]=u'(\mathbb{E}[C_t])$, which does not hold in general due to Jensen's inequality. Under quadratic felicity, the derivation of $\mathbb{E}_1[C_t]=C_1$ for all t can clearly be generalized: $\mathbb{E}_t[C_{t+s}]=C_t$ for all $t<s\leq T$, using an identical variational argument. In particular, we have that *random walk result*:

$$\mathbb{E}_t[C_{t+1}]=C_t\ \text{for all } t.$$

This says that individual consumption follows a martingale: today's consumption is the best predictor of tomorrow's consumption. We reformulate the random-walk result in terms of changes in consumption.

By the definition of the expectation operator, we can write

$$C_t = \mathbb{E}_{t-1}[C_t] + e_t,$$

where e_t is a random variable satisfying $\mathbb{E}[E_t] = 0$. Using the random walk result, we can rewrite this as

$$C_t = C_{t-1} + e_t \Rightarrow \qquad \Delta C_t = e_t.$$

Therefore, the random walk result implies that changes in individual consumption are unpredictable from information in the previous period. We discuss two tests of the random walk result in the literature. Using aggregate consumption data, Hall (1978) regressed aggregate consumption change ΔC_t on variables known at time $t-1$, such as C_{t-1}, Y_{t-1}, and a stock market index. Aside from the stock market index, none of them were predictive. Some issues: how do we know we are capturing the whole information set at $t-1$? There could be other relevant variables there that we don't observe. Moreover, since we are using aggregate data but the result was derived at an individual level, Hall is implicitly invoking a representative agent assumption. Campbell and Mankiw (1989) tested the random walk result against a specific alternative. Namely, the assume that a fraction $1-\lambda$ of the population follows the random walk in their consumption, while a fraction λ is hand-to-mouth: $\Delta C_t = \Delta Y_t$. Then, aggregate consumption is $\Delta C_t = \lambda \Delta Y_t + (1-\lambda) e_t$. Since $(1-\lambda) e_t$ is a regression error term, they estimated this equation to estimate λ. One issue with OLS is that ΔY_t and e_t are likely correlated, since they represent news about current income and news about changes in permanent income. As an IV for ΔY_t, they use lagged ΔY_t and lagged ΔC_t, and estimate $\lambda \approx 0.5$, a significant departure from the random walk result, which implies $\lambda = 0$. Still, there is the issue that they are using aggregate consumption data, and things are in partial equilibrium. Newer papers try to microfound things more rigorously. Back to the model. Recall that

$$C_1 = \frac{1}{T}\left(A_0 + \sum_{t=1}^{T} \mathbb{E}_1[Y_t]\right).$$

Let us find period 2 consumption. Taking the expectation of the IBC in period 2, we have

$$\begin{aligned}
\mathbb{E}_2\left[\sum_{t=1}^{T} C_t\right] &= \mathbb{E}_2\left[A_0 + \sum_{t=1}^{T} Y_t\right] \\
\Rightarrow C_1 + C_2 + \sum_{t=3}^{T} \mathbb{E}_2[C_t] &= A_0 + Y_1 + \sum_{t=2}^{T} \mathbb{E}_2[Y_t] \\
\Rightarrow C_1 + (T-1) C_2 &= A_0 + Y_1 + \sum_{t=2}^{T} \mathbb{E}_2[Y_t] \\
\Rightarrow (T-1) C_2 &= (A_0 + Y_1 - C_1) + \sum_{t=2}^{T} \mathbb{E}_2[Y_t] \\
\Rightarrow C_2 &= \frac{1}{T-1}\left(A_1 + \sum_{t=2}^{T} \mathbb{E}_2[Y_t]\right).
\end{aligned}$$

So C_2 is the average of the expected remaining lifetime resources. We can further massage this expression to

$$
\begin{aligned}
C_2 &= \frac{1}{T-1}\left(A_0 + Y_1 - C_1 + \sum_{t=2}^{T}\mathbb{E}_1\left[Y_t\right] + \sum_{t=2}^{T}\mathbb{E}_2\left[Y_t\right] - \sum_{t=2}^{T}\mathbb{E}_1\left[Y_t\right]\right) \\
&= \frac{1}{T-1}\left(A_0 - C_1 + \sum_{t=1}^{T}\mathbb{E}_1\left[Y_t\right] + \sum_{t=2}^{T}\mathbb{E}_2\left[Y_t\right] - \sum_{t=2}^{T}\mathbb{E}_1\left[Y_t\right]\right) \\
&= \frac{1}{T-1}\left(TC_1 - C_1 + \sum_{t=2}^{T}\mathbb{E}_2\left[Y_t\right] - \sum_{t=2}^{T}\mathbb{E}_1\left[Y_t\right]\right) \\
&= C_1 + \frac{1}{T-1}\left(\sum_{t=2}^{T}\mathbb{E}_2\left[Y_t\right] - \sum_{t=2}^{T}\mathbb{E}_1\left[Y_t\right]\right).
\end{aligned}
$$

Therefore, the change in consumption is

$$
\begin{aligned}
\Delta C_2 &= C_2 - C_1 \\
&= \frac{1}{T-1}\left(\sum_{t=2}^{T}\mathbb{E}_2\left[Y_t\right] - \sum_{t=2}^{T}\mathbb{E}_1\left[Y_t\right]\right);
\end{aligned}
$$

that is, the change in consumption is the change in expected lifetime income that occurs between period 1 and period 2. So consumption will only change in response to previously unexpected changes in income.

Remark. (Model with income risk, generalized) We consider the same model of income risk with quadratic felicity as in Model , but we drop the assumption of $r = \gamma = 0$. Specifically, we still impose that $r = \gamma$, but we allow it to take values greater than zero: $r = \gamma > 0$, so $\beta < 1$. The new optimization problem is

$$
\begin{aligned}
&\max_{C_1,\ldots,C_T} \mathbb{E}_1\left[\sum_{t=1}^{T}\beta\left(C_t - \frac{a}{2}C_t^2\right)\right] \\
&\text{s.t. } \sum_{t=1}^{T}\frac{C_t}{\left(1+r\right)^{t-1}} \leq A_0\left(1+r\right) + \sum_{t=1}^{T}\frac{Y_t}{\left(1+r\right)^{t-1}},
\end{aligned}
$$

where as before, we are not choosing single values for C_t, but rather contingency plans (mappings from the realized history up to period t into C_t), and the Y_t's are random variables, so the budget constraint holds in every possible history. We use a variational argument to show that consumption follows a martingale. The utility cost of giving up an infinitesimal amount of period-1 consumption is $u'\left(C_1\right) = 1 - aC_1$. If it is invested until period t and then consumed, the expected utility gain as of period 1 is $\mathbb{E}_1\left[\beta^{t-1}\left(1+r\right)^{t-1}u'\left(C_t\right)\right] = 1 - a\mathbb{E}_1\left[C_t\right]$ (recall that $r = \gamma$, so $\beta\left(1+r\right) = \beta\left(1+\gamma\right) = 1$. Equating the benefits and costs gives

$$
\mathbb{E}_1\left[C_t\right] = C_1.
$$

Take the period-1 expectation of the IBC to obtain

$$\begin{aligned}
\mathbb{E}_1\left[\sum_{t=1}^{T}\frac{C_t}{(1+r)^{t-1}}\right] &= \mathbb{E}_1\left[(1+r)A_0+\sum_{t=1}^{T}\frac{Y_t}{(1+r)^{t-1}}\right] \\
\Rightarrow \sum_{t=1}^{T}\frac{\mathbb{E}_1[C_t]}{(1+r)^{t-1}} &= (1+r)A_0+\sum_{t=1}^{T}\frac{\mathbb{E}_1[Y_t]}{(1+r)^{t-1}} \\
\Rightarrow TC_1\sum_{t=1}^{T}\frac{1}{(1+r)^{t-1}} &= (1+r)A_0+\sum_{t=1}^{T}\frac{\mathbb{E}_1[Y_t]}{(1+r)^{t-1}} \\
\Rightarrow C_1 &= \frac{r}{(1+r)\left(1-\left(\frac{1}{1+r}\right)^T\right)}\left[(1+r)A_0+\sum_{t=1}^{T}\frac{\mathbb{E}_1[Y_t]}{(1+r)^{t-1}}\right].
\end{aligned}$$

As $T\to\infty$, this expression approaches

$$C_1=\frac{r}{1+r}\left[(1+r)A_0+\sum_{t=1}^{\infty}\frac{\mathbb{E}_1[Y_t]}{(1+r)^{t-1}}\right].$$

This shows that we have certainty equivalence once again, where we behave almost as in the perfect foresight case, but just replacing actual lifetime income with expected lifetime income. In general, we have that as $T\to\infty$

$$C_t=\frac{r}{1+r}\left[A_{t-1}(1+r)+\sum_{s=0}^{\infty}\frac{\mathbb{E}_t[Y_{t+s}]}{(1+r)^{s}}\right].$$

Let's derive that. Fix t. We have the martingale result that $\mathbb{E}_t[Y_s]=Y_t$ for $s>t$, by a variational argument: $u'(C_t)=\beta^{t-s}(1+r)^{t-s}\,\mathbb{E}_t[u'(C_s)]$. With this in hand, take the period-t expectation of the IBC to obtain

$$\begin{aligned}
\mathbb{E}_t\left[\sum_{s=1}^{T}\frac{C_s}{(1+r)^{s-1}}\right] &= \mathbb{E}_t\left[(1+r)A_0+\sum_{s=1}^{T}\frac{Y_s}{(1+r)^{s-1}}\right] \\
\Rightarrow \sum_{s=1}^{t-1}\frac{C_s}{(1+r)^{s-1}}+\frac{C_t}{(1+r)^{t-1}}+\sum_{s=t+1}^{T}\frac{\mathbb{E}_t[C_s]}{(1+r)^{s-1}} &= (1+r)A_0+\sum_{s=1}^{t-1}\frac{Y_s}{(1+r)^{s-1}}+\sum_{s=t}^{T}\frac{\mathbb{E}_t[Y_s]}{(1+r)^{s-1}} \\
\Rightarrow C_t\sum_{s=t}^{T}\frac{1}{(1+r)^{s-1}} &= (1+r)A_0+\sum_{s=1}^{t-1}\frac{Y_s-C_s}{(1+r)^{s-1}}+\sum_{s=t}^{T}\frac{\mathbb{E}_t[Y_s]}{(1+r)^{s-1}} \\
\Rightarrow C_t\sum_{s=0}^{T-s}\frac{1}{(1+r)^{s+t-1}} &= (1+r)A_0+\sum_{s=1}^{t-1}\frac{A_s-(1+r)A_{s-1}}{(1+r)^{s-1}}+\sum_{s=0}^{T-s}\frac{\mathbb{E}_t[Y_{t+s}]}{(1+r)^{s+t-1}}.
\end{aligned}$$

Now, note that

$$\begin{aligned}
(1+r)A_0+\sum_{s=1}^{t-1}\frac{A_s-(1+r)A_{s-1}}{(1+r)^{s-1}} &= (1+r)A_0+\sum_{s=1}^{t-1}\frac{A_s}{(1+r)^{s-1}}-\sum_{s=1}^{t-1}\frac{A_{s-1}}{(1+r)^{s-2}} \\
&= \sum_{s=0}^{t-1}\frac{A_s}{(1+r)^{s-1}}-\sum_{s=0}^{t-2}\frac{A_s}{(1+r)^{s-1}} \\
&= \frac{A_{t-1}}{(1+r)^{t-2}}.
\end{aligned}$$

Plugging this back into into our expression, we get

$$\begin{aligned}
C_t \sum_{s=0}^{T-t} \frac{1}{(1+r)^{s+t-1}} &= \frac{A_{t-1}}{(1+r)^{t-2}} + \sum_{s=0}^{T-t} \frac{\mathbb{E}_t\left[Y_{t+s}\right]}{(1+r)^{s+t-1}} \\
\Rightarrow C_t \sum_{s=0}^{T-t} \frac{1}{(1+r)^{s}} &= (1+r)\, A_{t-1} + \sum_{s=0}^{T-t} \frac{\mathbb{E}_t\left[Y_{t+s}\right]}{(1+r)^{s}} \\
\Rightarrow C_t &= \frac{r}{(1+r)\left(1-\left(\frac{1}{1+r}\right)^{T+t-1}\right)} \left[(1+r)\, A_{t-1} + \sum_{s=0}^{T-t} \frac{\mathbb{E}_t\left[Y_{s+t}\right]}{(1+r)^{s}}\right].
\end{aligned}$$

Taking $T \to \infty$ gives the desired result:

$$C_t = \frac{r}{1+r}\left[A_{t-1}(1+r) + \sum_{s\geq 0} \frac{1}{(1+r)^s}\mathbb{E}_t\left[Y_{t+s}\right]\right].$$

It can be then derived (todo on PS2) that the change in consumption depends only on updates in news. This holds under certainty equivalence and without motives for consumption tilting ($r = \gamma$). Indeed, we have

$$\begin{aligned}
C_t &= \frac{r}{1+r}\left(A_{t-1}(1+r) + \sum_{s=0}^{\infty} \frac{1}{(1+r)^s} E_t[Y_{t+s}]\right) \\
&= \frac{r}{1+r}\left[(1+r)\left(Y_{t-1} + (1+r)\, A_{t-2} - C_{t-1}\right) + \sum_{s=0}^{\infty} \frac{E_{t-1}[Y_{t+s}]}{(1+r)^s} + \sum_{s=0}^{\infty} \frac{E_t[Y_{t+s}]}{(1+r)^s} - \sum_{s=0}^{\infty} \frac{E_{t-1}[Y_{t+s}]}{(1+r)^s}\right] \\
&= \frac{r}{1+r}\left[(1+r)^2 A_{t-2} + (1+r)\, Y_{t-1} - (1+r)\, C_{t-1} + (1+r)\sum_{s=1}^{\infty} \frac{E_t[Y_{t-1+s}]}{(1+r)^s} + \sum_{s=0}^{\infty} \frac{E_t[Y_{t+s}]}{(1+r)^s} - \sum_{s=0}^{\infty} \frac{E_{t-1}[Y_{t+s}]}{(1+r)^s}\right] \\
&= (1+r)\frac{r}{1+r}\left[(1+r)\, A_{t-2} + \sum_{s=0}^{\infty} \frac{E_t[Y_{t-1+s}]}{(1+r)^s}\right] - rC_{t-1} + \frac{r}{1+r}\left[\sum_{s=0}^{\infty} \frac{E_t[Y_{t+s}]}{(1+r)^s} - \sum_{s=0}^{\infty} \frac{E_{t-1}[Y_{t+s}]}{(1+r)^s}\right] \\
&= C_{t-1} + \frac{r}{1+r}\left[\sum_{s=0}^{\infty} \frac{E_t[Y_{t+s}]}{(1+r)^s} - \sum_{s=0}^{\infty} \frac{E_{t-1}[Y_{t+s}]}{(1+r)^s}\right]
\end{aligned}$$

and so

$$\Delta C_t = \frac{r}{1+r}\left[\sum_{s=0}^{\infty} \frac{E_t[Y_{t+s}]}{(1+r)^s} - \sum_{s=0}^{\infty} \frac{E_{t-1}[Y_{t+s}]}{(1+r)^s}\right].$$

If we make assumptions on the process generating Y_t, we can get nicer results for how consumption changes in response to shocks. For instance, first suppose that Y_t follows a process $Y_t = \mu + e_t$, where e_t is iid white noise. Then,

$$\begin{aligned}
C_t - C_{t-1} &= \frac{r}{1+r}\left[\sum_{s=0}^{\infty} \frac{E_t[Y_{t+s}]}{(1+r)^s} - \sum_{s=0}^{\infty} \frac{E_{t-1}[Y_{t+s}]}{(1+r)^s}\right] \\
&= \frac{r}{1+r}\left[E_t\left[\mu + e_t\right] - E_{t-1}\left[\mu + e_t\right] + \sum_{s=1}^{\infty} \frac{\mathbb{E}_t\left[\mu + e_{t+s}\right] - E_{t-1}\left[\mu + e_{t+s}\right]}{(1+r)^s}\right] \\
&= \frac{r}{1+r}\left[\mu + e_t - \mu - 0 + \sum_{s=1}^{\infty} \frac{\mu + 0 - \mu - 0}{(1+r)^s}\right] \\
&= \frac{r}{1+r} e_t.
\end{aligned}$$

Therefore, we see that consumption barely changes in response to a transitory income shock e_t (since

$\frac{r}{1+r}$ is likely much less than one).

If Y_t is a stationary AR(p) $Y_t = \sum_{i=1}^{p} \rho_i Y_{t-i} + e_t$, with $\sum_{i=1}^{p} \rho_i \leq 1$, then (prove)

$$C_t - C_{t-1} = \frac{r}{1+r} \cdot \frac{1}{1 - \sum_{i=1}^{p} \frac{\rho_i}{(1+r)^i}} \cdot e_t.$$

In the case of an AR(1), this simplifies to (prove)

$$C_t - C_{t-1} = \frac{r}{1+r-\rho} e_t.$$

So as $\rho \to 1$, C_t increasingly approaches following a random walk, where consumption changes one-for-one with a transitory income shock:

$$C_t - C_{t-1} = e_t.$$

Remark (Income risk with precautionary savings)**.** We turn to the case of income risk where we move away from quadratic felicity, so that certainty equivalence does not hold. Under CARA utility and a normal income process, we will show that consumption now follows a random walk with drift. For simplicity, assume that $r = \gamma = 0$, so $\beta = 1$. We make use of the *constant absolute risk aversion (CARA)* utility function

$$u(C) = -\frac{1}{\alpha} e^{-\alpha C}.$$

The CARA function, fittingly enough, has constant absolute risk aversion $-\frac{u''(C)}{u'(C)} = \alpha$. We assume labor income follows a process

$$Y_t = Y_{t-1} + e_t,$$

with $e_t \sim N(0, \sigma^2)$ iid. As before, we state the consumer's problem:

$$\max_{C_1,\ldots,C_T} \sum_{t=1}^{T} u(C_t) \qquad \text{s.t.} \quad \sum_{t=1}^{T} C_t \leq A_0 + \sum_{t=1}^{T} Y_t.$$

We derive the FOC using a variational argument. If we give up one unit of consumption today, the utility cost is $u'(C_t) = e^{-\alpha C_t}$. If we save this amount to tomorrow and consume it then, the utility benefit is $\beta(1+r) E_t[u'(C_{t+1})] = E_t[e^{-\alpha C_{t+1}}]$. Hence we have the first-order condition

$$e^{-\alpha C_t} = E_t\left[e^{-\alpha C_{t+1}}\right].$$

We conjecture that consumption obeys the following random walk with drift, and then confirm that it satisfies that first-order condition:

$$C_{t+1} = C_t + \mu + e_{t+1}.$$

Note that the shocks to the consumption process are the same as the shocks to the income process. It is a fact that for a normal random variable X, we have $E\left[e^{\alpha X}\right] = e^{\alpha E[X]} \cdot e^{\frac{\alpha^2 \operatorname{Var}(x)}{2}}$ (prove). Plugging this into the right-hand side of the first-order condition gives

$$\begin{aligned} E_t[e^{-\alpha C_{t+1}}] &= E_t\left[e^{-\alpha(C_t+\mu+e_{t+1})}\right] \\ &= E_t\left[e^{-\alpha C_t} \cdot e^{-\alpha\mu} \cdot e^{-\alpha e_{t+1}}\right] \\ &= e^{-\alpha C_t} \cdot e^{-\alpha\mu} \cdot E_t\left[e^{-\alpha e_{t+1}}\right] \\ &= e^{-\alpha C_t} \cdot e^{-\alpha\mu} \cdot e^{\frac{\alpha^2\sigma^2}{2}}. \end{aligned}$$

Equating this to the left-hand side of the first-order condition yields

$$e^{-\alpha C_t} = e^{-\alpha C_t} \cdot e^{-\alpha\mu} \cdot e^{\frac{\alpha^2\sigma^2}{2}} \Rightarrow \qquad e^{\alpha\mu} = e^{\frac{\alpha^2\sigma^2}{2}} \Rightarrow \alpha\mu = \frac{\alpha^2\sigma^2}{2} \Rightarrow \qquad \mu = \frac{\alpha\sigma^2}{2}.$$

Hence, the solution is

$$C_{t+1} = C_t + \frac{\alpha\sigma^2}{2} + e_{t+1}.$$

Therefore, consumption is a random walk with drift, where the drift is increasing in absolute risk aversion and the variance of the income process.

Note that we can write

$$C_2 = C_1 + \frac{\alpha\sigma^2}{2} + e_2 = C_1 + \frac{\alpha\sigma^2}{2} + (Y_2 - Y_1)$$

and

$$C_3 = C_2 + \frac{\alpha\sigma^2}{2} + (Y_3 - Y_2)$$

and so

$$C_3 = C_1 + 2 \cdot \frac{\alpha\sigma^2}{2} + (Y_3 - Y_1).$$

Iterating this forward, we get that for any t,

$$C_t = C_1 + (t-1)\frac{\alpha\sigma^2}{2} + (Y_t - Y_1).$$

Plug this into the IBC:

$$\begin{aligned}
&\sum_{t=1}^{T}\left(C_1 + (t-1)\frac{\alpha\sigma^2}{2} + Y_t - Y_1\right) = A_0 + \sum_{t=1}^{T} Y_t \\
\Rightarrow\ & TC_1 + \frac{\alpha\sigma^2}{2}\sum_{t=1}^{T-1} t + \sum_{t=1}^{T} Y_t - TY_1 = A_0 + \sum_{t=1}^{T} Y_t \\
\Rightarrow\ & TC_1 = A_0 + TY_1 - \frac{\alpha\sigma^2 \cdot T(T-1)}{4} \\
\Rightarrow\ & C_1 = \frac{A_0}{T} + Y_1 - \frac{\alpha\sigma^2(T-1)}{4}.
\end{aligned}$$

Recall that in the certainty-equivalence case, with quadratic felicity and $\gamma = r = 0$, we had $C_1' = \frac{A_0}{T} + \frac{1}{T}\sum_{t=1}^{T} E_1[Y_t]$. If we made the random-walk hypothesis on labor income, then we would have $C_1' = \frac{A_0}{T} + Y_1$. Hence, under CARA-normal assumptions, we have that $C_1 = C_1' - \frac{\alpha\sigma^2(T-1)}{4}$. Consumption is lower because of *precautionary saving*. Note that precautionary savings is increasing in absolute risk aversion α, the volatility of the income process σ, and lifespan T. In general for any t, we have that for $s \geq t$ (show),

$$C_s = C_t + (s-t)\frac{\alpha\sigma^2}{2} + Y_s - Y_t.$$

Therefore, manipulating the budget constraint, we find that

$$
\begin{aligned}
&\sum_{s=1}^{T} C_s = A_0 + \sum_{s=1}^{T} Y_s \\
&\Rightarrow \sum_{s=1}^{t-1} C_s + \sum_{s=t}^{T} \left(C_t + (s-t)\frac{\alpha\sigma^2}{2} + Y_s - Y_t \right) = A_0 + \sum_{s=1}^{T} Y_s \\
&\Rightarrow (T-t+1)\, C_t + \frac{\alpha\sigma^2}{2} \sum_{s=0}^{T-s} s + \sum_{s=t}^{T} Y_s - (T-t+1)\, Y_t = A_0 + \sum_{s=1}^{t-1} (Y_s - C_s) + \sum_{s=t}^{T} Y_s \\
&\Rightarrow (T-t+1)\, C_t + \frac{\alpha\sigma^2\,(T-s)\,(T-t+1)}{4} - (T-t+1)\, Y_t = A_0 + \sum_{s=1}^{t-1} (A_s - A_{s-1}) = A_{t-1} \\
&\Rightarrow C_t = \frac{A_{t-1}}{T-t+1} + Y_t - \frac{\alpha\sigma^2\,(T-t)}{4}.
\end{aligned}
$$

Note that this formula gives $C_T = A_{T-1} + Y_T$, as it should be. Finally, we must show that the process does indeed follow the random walk with drift. First, note that for any t,

$$
\begin{aligned}
A_{t-1} &= A_{t-2} + Y_{t-1} - C_{t-1} \\
&= A_{t-2} + Y_{t-1} - \frac{A_{t-2}}{T-s+2} - Y_{t-1} + \frac{\alpha\sigma^2\,(T-s+1)}{4} \\
&= A_{t-2}\left(1 - \frac{1}{T-s+2}\right) + \frac{\alpha\sigma^2\,(T-t+1)}{4} \\
&= A_{t-2}\left(\frac{T-t+1}{T-t+2}\right) + \frac{\alpha\sigma^2\,(T-t+1)}{4}.
\end{aligned}
$$

Then, we compute the change in consumption as

$$
\begin{aligned}
C_t - C_{t-1} &= \frac{A_{t-1}}{T-t+1} + Y_t - \frac{\alpha\sigma^2\,(T-t)}{4} - \frac{A_{t-2}}{T-t+2} - Y_{t-1} + \frac{\alpha\sigma^2\,(T-t+1)}{4} \\
&= \frac{A_{t-2}}{T-t+2} + \frac{\alpha\sigma^2}{4} + \frac{\alpha\sigma^2\,((T-t+1)-(T-t))}{4} - \frac{A_{t-2}}{T-t+2} + Y_t - Y_{t-1} \\
&= \frac{\alpha\sigma^2}{2} + \underbrace{Y_t - Y_{t-1}}_{=e_t} \\
&= \frac{\alpha\sigma^2}{2} + e_t.
\end{aligned}
$$

Prudence, mean-preserving spreads, and precautionary saving

We have shown that, for a particular combination of a felicity function (CARA) and income process (Gaussian random walk), we can derive a *particular* consumption process (random walk with drift, with innovations the same as in the income process) as the outcome. When directly compared to the certainty-equivalent case (where the consumption outcome is a random walk), we obtain *precautionary saving* (captured by the drift of the random walk).

In what follows, we show what causes precautionary saving *more generally*, independently of the particular function forms assumed above: the condition $u'''(\cdot) > 0$, i.e., strict convexity of marginal felicity.

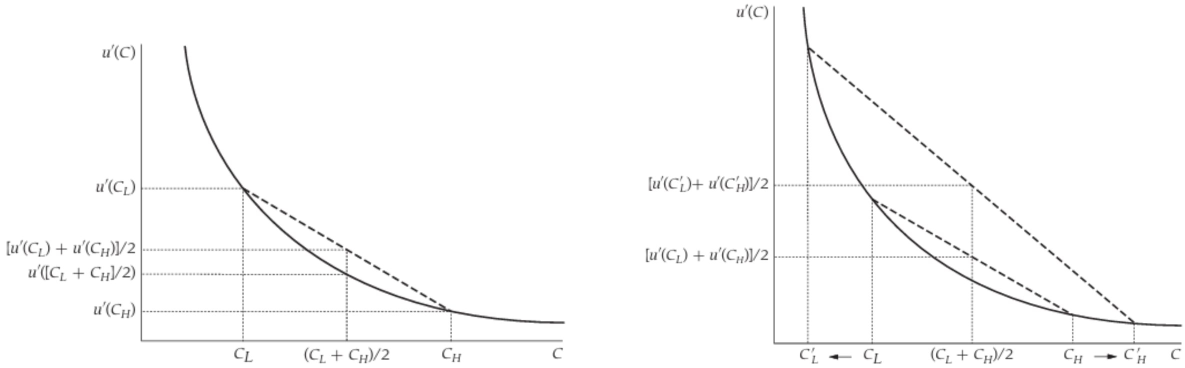
The role of $u'''(\cdot) > 0$ under CARA felicity

For CARA,

$$u(C_t) \;=\; -\frac{1}{\alpha}e^{-\alpha C_t}, \qquad u'(C_t) \;=\; e^{-\alpha C_t} \;>\; 0 \text{ (strict monotonicity)},$$

$$u''(C_t) \;=\; -\alpha e^{-\alpha C_t} \;<\; 0 \text{ (strict concavity)}, \qquad u'''(C_t) \;=\; \alpha^2 e^{-\alpha C_t} \;>\; 0 \text{ (strict convexity of } u'(C_t))$$

Looking at the third derivative we know that the marginal felicity $u'(\cdot)$ in the CARA case is *strictly convex*.



(a) CARA: $u'(C)$ is decreasing and strictly convex. (Romer, 2018, p. 401)

(b) Mean-preserving spread on C. (Romer, 2018, p. 401)

To make the argument cleanest, consider $r = \gamma$ (so that $\beta(1 + r) = 1$). Consider a *mean-preserving spread* (MPS) applied to C_{t+1}, which by definition preserves the conditional mean:

$$\underbrace{\mathbb{E}_t\left[C_{t+1}^{\text{pre MPS}}\right] = \mathbb{E}_t\left[C_{t+1}^{\text{post MPS}}\right]}_{\text{mean preservation}}$$

Using the Euler equation (with $r = \gamma$): I do not totally understand the math here.

$$\underbrace{u'(C_t) \;=\; \mathbb{E}_t\left[u'\left(C_{t+1}^{\text{post MPS}}\right)\right]}_{\text{Euler}} \;>\; \underbrace{\mathbb{E}_t\left[u'\left(C_{t+1}^{\text{pre MPS}}\right)\right]}_{\text{MPS } + \, u'''(\cdot) > 0} \;>\; \underbrace{u'\left(\mathbb{E}_t\left[C_{t+1}^{\text{pre MPS}}\right]\right)}_{\text{Jensen } + \, u'''(\cdot) > 0}.$$

From $u'(C_t) > u'\left(\mathbb{E}_t[C_{t+1}^{\text{pre MPS}}]\right)$ and since $u''(\cdot) < 0$ (i.e., $u'(\cdot)$ is strictly decreasing), it follows that

$$C_t \;<\; \mathbb{E}_t\left[C_{t+1}^{\text{pre MPS}}\right].$$

By the definition of a mean-preserving spread,

$$C_t \;<\; \mathbb{E}_t\left[C_{t+1}^{\text{post MPS}}\right].$$

The same chain of arguments for the certainty-equivalent (quadratic felicity, i.e. linear $u'(\cdot)$) case yields

$$C_t^{\text{c.e.}} \;=\; \mathbb{E}_t\left[C_{t+1}^{\text{c.e., post MPS}}\right].$$

Choosing the *same* MPS across felicity specifications (so that the comparison is sensible) implies

$$\mathbb{E}_t\left[C_{t+1}^{\text{post MPS}}\right] \;=\; \mathbb{E}_t\left[C_{t+1}^{\text{c.e., post MPS}}\right].$$

Taken together, we conclude:

$$C_t^{\text{c.e.}} \;>\; C_t, \qquad \text{or equivalently} \qquad S_t^{\text{precautionary}} \;\equiv\; C_t^{\text{c.e.}} - C_t \;>\; 0.$$

Intuition. Consider two people, A with *linear* marginal felicity (quadratic utility; $u''' = 0$) and B with *strictly convex* marginal felicity ($u''' > 0$). Start both with a consumption lottery for tomorrow with the same expectation (and, for ease of comparison, the same variance). Expose both to the same mean-preserving spread (riskier lottery while the expected consumption tomorrow remains the same). Person A will not change her consumption today. Person B will reduce her consumption today and save more (precautionary saving).

Income risk is a case where the *third* derivative matters in Economics. Economists call the property $u'''(\cdot) > 0$ *prudence*. As we have seen, prudence (a property of the felicity function—a primitive) and precautionary saving (an *outcome*) are tightly connected: $u'''(\cdot) > 0 \Rightarrow$ marginal felicity is strictly convex $\Rightarrow$ mean-preserving spreads raise $\mathbb{E}[u'(\cdot)]$ and induce lower C_t today.

Addendum. Risk aversion ($u''(C) < 0$) and prudence ($u'''(C) > 0$) play different roles. Risk aversion governs dislike of risk *in levels*; prudence governs the *precautionary* response of current consumption to future risk. Under $\beta(1 + r) = 1$ and a mean-preserving spread (MPS) in C_{t+1}:

- If $u''(C) < 0$ and $u'''(C) = 0$ (e.g., quadratic utility), the MPS leaves C_t *unchanged* $\Rightarrow$ certainty equivalence.
- If $u''(C) < 0$ and $u'''(C) > 0$ (e.g., CARA/CRRA), the MPS raises $\mathbb{E}_t[u'(C_{t+1})]$, so by Euler $u'(C_t)$ must rise $\Rightarrow C_t$ falls $\Rightarrow$ higher saving (precautionary saving).
- If $u''(C) < 0$ and $u'''(C) < 0$ ("imprudence"), the MPS lowers $\mathbb{E}_t[u'(C_{t+1})] \Rightarrow C_t$ rises $\Rightarrow$ lower saving.

Dynamic Programming

An Initial Remark This subsection is methodological. We will first apply dynamic programming to the consumption problem, but the techniques are broadly applicable beyond that context. The discussion is slightly more abstract than the substance-oriented material elsewhere in the lecture; the goal is to equip you with a general tool.

Statement of the Problem

We work in discrete time, indexed by $t = 0, 1, 2, \ldots$ (starting at $t = 0$ is a convention). Intertemporal aggregation is *additively separable*: lifetime value is the discounted sum of per-period flow payoffs.[2] The horizon is infinite for most of what follows, and—for now—we assume perfect foresight/certainty.

Given an initial state $x_0 \in \mathbb{R}^n$, the planner's value function is

$$V(x_0) \;\equiv\; \max_{\{u_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \, r(x_t, u_t)$$

subject to the law of motion

$$x_{t+1} = g(x_t, u_t), \qquad t = 0, 1, 2, \ldots$$

For ease of notation we suppress boldface; it should be clear that x_t and u_t can be vectors:

- $x_t \in \mathbb{R}^n$ for all t (the *state*).
- $u_t \in \mathbb{R}^k$ for all t (the *control*).
- The discount factor is $\beta \in (0, 1)$ in the infinite-horizon case (so that the objective is well-defined).
- The per-period flow payoff is a real-valued mapping

$$r : \; \mathbb{R}^n \times \mathbb{R}^k \longrightarrow \mathbb{R}, \qquad (x, u) \mapsto r(x, u).$$

- $V(x_0)$ is measured in the same "units" as r and represents the maximal discounted value starting at x_0.

What does $r(x_t, u_t)$ capture? In general r can depend on both states and controls. In the baseline consumption application we will take

$$r(x_t, u_t) = u(C_t),$$

i.e., felicity depends only on the choice C_t (do *not* confuse the function $u(\cdot)$ here with the control vector notation). One could also allow r to depend on components of the state (e.g., preference shocks), in which case $r = r(x_t, u_t)$ truly uses both arguments.

Example (time-varying CRRA). A simple specification with time-varying risk aversion σ_t is

$$u_t(C_t) = \frac{C_t^{1-\sigma_t}}{1 - \sigma_t},$$

where the subscript on u_t reminds us that felicity parameters may change over time.

[2] Additive separability means the objective contains no cross-time interaction terms such as $h(x_t, x_{t+1})$; instead it is $\sum_{t=0}^{\infty} \beta^t r(x_t, u_t)$.

States vs. Controls: Formulation Dependence

What one labels as a *state* or a *control* is formulation-dependent. In the baseline consumption formulation, today's consumption C_t is the control and beginning-of-period assets are a component of the state, e.g. $x_t = (A_{t-1}, Y_t, \dots)$. An equivalent formulation replaces C_t with a flow budget constraint and takes next period's assets A_t as the control; then A_t becomes part of tomorrow's state x_{t+1}. Thus, today's control can become tomorrow's state via the transition g.

Minimal example. Budget constraint with zero interest for simplicity:

$$A_t \;=\; A_{t-1} + Y_t - C_t, \qquad C_t \geq 0,$$

and per-period payoff $r(x_t, u_t) = u(C_t)$. The stochastic object for next period is Y_{t+1}.

Formulation A (choose consumption). State today: $x_t = (A_{t-1}, Y_t)$. Control today: $u_t = C_t \in [0,\, A_{t-1} + Y_t]$. Transition (law of motion):

$$x_{t+1} = \big(A_t,\, Y_{t+1}\big) \quad \text{with} \quad A_t \equiv A_{t-1} + Y_t - C_t.$$

Bellman equation:

$$V(A_{t-1}, Y_t) = \max_{0 \leq C \leq A_{t-1}+Y_t} \Big\{ u(C) + \beta\, \mathbb{E}\big[V(A_{t-1} + Y_t - C,\, Y_{t+1}) \mid Y_t\big] \Big\}.$$

Formulation B (choose next period's assets). Same state today: $x_t = (A_{t-1}, Y_t)$. Control today: $u_t = A_t \in [\underline{A},\, A_{t-1} + Y_t]$ (an occasionally imposed borrowing limit $\underline{A}$). Consumption is now *implied* by the choice of A_t:

$$C_t \equiv A_{t-1} + Y_t - A_t.$$

Transition:

$$x_{t+1} = \big(A_t,\, Y_{t+1}\big).$$

Bellman equation:

$$V(A_{t-1}, Y_t) = \max_{\underline{A} \leq A' \leq A_{t-1}+Y_t} \Big\{ u\big(A_{t-1} + Y_t - A'\big) + \beta\, \mathbb{E}\big[V(A',\, Y_{t+1}) \mid Y_t\big] \Big\}.$$

Why this is the same problem. Formulation A optimizes directly over C and then *induces* $A' = A_{t-1} + Y_t - C$. Formulation B optimizes directly over A' and then *induces* $C = A_{t-1} + Y_t - A'$. Substituting $A' \leftrightarrow C$ turns one Bellman equation into the other. Hence, "today's control becomes part of tomorrow's state" simply because the control feeds the transition map g:

$$g(A_{t-1}, Y_t;\, C_t) = \big(A_{t-1} + Y_t - C_t,\, Y_{t+1}\big) \quad \text{or} \quad g(A_{t-1}, Y_t;\, A_t) = \big(A_t,\, Y_{t+1}\big).$$

Mnemonic. *State* = what you need to carry into the Bellman problem to evaluate feasibility and continuation value next period. *Control* = what you pick today; once picked, it is recorded in the next state via $x_{t+1} = g(x_t, u_t)$.

Regularity: Existence, Boundedness, and Concavity TB Revised For existence of an optimal policy and well-behaved value functions it is standard to assume: (i) the feasible correspondence $U(x)$ is nonempty, compact-valued, and (jointly) continuous in x; (ii) $g(x, u)$ is continuous; (iii) $r(x, u)$ is continuous and bounded above (or $\beta < 1$ with growth conditions). Under these conditions standard fixed-point arguments (e.g., Blackwell's sufficient conditions) deliver a unique bounded solution to the Bellman equation.

If, in addition, the *feasible set is convex* (for each x, the set $\{(x', u) : x' = g(x, u),\ u \in U(x)\}$ is convex) and the flow payoff r is *concave* in (x, u), then the Bellman operator preserves concavity and the value function V is concave.

The Sequence Problem

Given $x_0 \in \mathbb{R}^n$, $\beta \in (0, 1)$, a flow payoff $r : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$, and a transition map $g : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n$, the decision maker solves

$$V(x_0) \equiv \max_{\{u_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t r(x_t, u_t) \quad \text{s.t.} \quad x_{t+1} = g(x_t, u_t),\ x_0 \text{ given.}$$

This is the *sequence problem.* The objective is *additively separable over time.*

Dynamic Programming and the Recursive Problem

The idea is to replace the infinite sequence $\{u_t\}_{t\geq 0}$ with a (time-invariant, in infinite horizon) *policy function* $u_t = h(x_t)$ so that the path generated by $x_{t+1} = g(x_t, h(x_t))$ solves the sequence problem. Many macro problems admit such *recursive solutions* (state $\to$ choice $\to$ next state); with finite memory one can pack lags into x_t.

Bellman's optimality principle. Whatever the current state and decision, the remaining decisions must be optimal for the state induced by today's choice.

Bellman equation. Dropping time subscripts and denoting next period with a prime,

$$V(x) = \max_{u \in U(x)} \Big\{ r(x, u) + \beta\, V\big(g(x, u)\big) \Big\}, \qquad x' = g(x, u).$$

Here $r(x, u)$ is the *flow value* and $V(g(x, u))$ the (discounted) *continuation value.* A maximizer defines $u = h(x)$ and then

$$V(x) = r\big(x, h(x)\big) + \beta\, V\big(g(x, h(x))\big).$$

Both displays are *functional equations* in the two unknowns $h(\cdot)$ and $V(\cdot)$.

Existence, Uniqueness, and Iteration Under standard regularity (nonempty compact feasible sets $U(x)$, continuity of r and g, and boundedness or $\beta < 1$ with growth control), the Bellman operator has a unique bounded fixed point V. With concave r and a convex feasible set, V is concave. Value iteration $V_{j+1} = \mathcal{T} V_j$ with

$$(\mathcal{T}V)(x) = \max_u \{r(x, u) + \beta V(g(x, u))\}$$

converges to V from any bounded continuous V_0.

First-Order and Envelope Conditions

Assume interior solutions and differentiability of r and g. A necessary condition for $u = h(x)$ is

$$\frac{\partial r(x, u)}{\partial u} \;+\; \beta\, \frac{\partial V(x')}{\partial x'}\, \frac{\partial g(x, u)}{\partial u} \;=\; 0, \qquad x' = g(x, u).$$

Differentiating the Bellman equation w.r.t. x at the optimum $u = h(x)$ yields

$$\frac{\partial V(x)}{\partial x} = \frac{\partial r(x,u)}{\partial x} + \frac{\partial r(x,u)}{\partial u}\frac{\partial h(x)}{\partial x} + \beta\frac{\partial V(x')}{\partial x'}\left[\frac{\partial g(x,u)}{\partial x} + \frac{\partial g(x,u)}{\partial u}\frac{\partial h(x)}{\partial x}\right].$$

If, in addition, g can be written as $g(u)$ (or more generally $\partial g/\partial x = 0$ after re-formulation), the *Benveniste–Scheinkman envelope condition* simplifies to

$$\boxed{\frac{\partial V(x)}{\partial x} = \frac{\partial r(x,u)}{\partial x} \quad \text{at } u = h(x).}$$

Substituting the envelope condition back into the FOC gives the standard *Euler equation* that contains $h(\cdot)$ but no $V(\cdot)$:

$$\frac{\partial r(x,u)}{\partial u} + \beta\,\frac{\partial r(x',u')}{\partial x'}\frac{\partial g(x,u)}{\partial u} = 0, \qquad x' = g(x,u),\;\; u' = h(x').$$

This functional equation in the policy h can be attacked with analytical or numerical methods.

Remark. Before an example:

- The FOC will almost always exist, as long as we have differentiability of the functions involved.
- However, the FOC in a useful way, both economically and for numerical purposes, i.e., as an Euler equation without the value function, does not always exist; only when the Benveniste-Scheinkman formula holds.
- If this is not the case, the solution has to come directly from the Bellman equation (which is also a function equation, but in $V(x)$).

Example I: Consumption–Saving

We revisit the standard consumption–saving problem under perfect foresight. Let assets be A, income Y, interest rate r, and period felicity $u(\cdot)$.

$$r(x,u) = u(C), \qquad g(x,u):\; A' = (1+r)A + Y - C,$$

with Y exogenous, control variable C, and state variables (A,Y). State the problem at $t = 0$ as

$$\max_{\{C_t, A_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t u(C_t) \quad \text{s.t.} \quad A_{t+1} = (1+r)A_t + Y_t - C_t,\;\; A_0 \text{ given.}$$

Equivalently, using forward substitution,

$$\sum_{t=0}^{\infty} \frac{1}{(1+r)^t} C_t = A_0(1+r) + \sum_{t=0}^{\infty} \frac{1}{(1+r)^t} Y_t.$$

With state (A,Y) and control C,

$$V(A,Y) = \max_{C,A'} \bigl\{u(C) + \beta V(A',Y')\bigr\} \quad \text{s.t.} \quad A' = (1+r)A + Y - C.$$

Plugging the constraint,

$$V(A,Y) = \max_{C}\; u(C) + \beta V\bigl((1+r)A + Y - C,\; Y'\bigr).$$

The FOC w.r.t. C is

$$u'(C) - \beta V_A(A', Y') = 0 \quad \Longleftrightarrow \quad u'(C) = \beta V_A(A', Y'),$$

which is *not yet* an Euler equation because it still contains $V_A(\cdot)$. Because $A' = (1+r)A + Y - C(A)$ depends on A, the Benveniste–Scheinkman envelope in its simple form does not apply directly. Differentiating the Bellman equation at the optimum $C = C(A, Y)$ gives

$$\frac{\partial V(A,Y)}{\partial A} = u'(C)\,\frac{\partial C}{\partial A} - \beta V_A(A', Y')\,\frac{\partial C}{\partial A} + \underbrace{\beta V_A(A', Y')}_{=u'(C)}(1+r) = u'(C)(1+r).$$

Using the FOC $u'(C) = \beta V_A(A', Y')$ to cancel the bracketed terms yields

$$\frac{\partial V(A,Y)}{\partial A} = \underbrace{\beta V_A(A', Y')}_{=u'(C)}(1+r) \quad \Longrightarrow \quad V_A(A', Y') = u'(C')(1+r) \text{ (holds tomorrow).}$$

Substituting back into the FOC produces the familiar Euler equation:

$$\boxed{u'(C) = \beta(1+r)\,u'(C')}.$$

Equivalent Formulation Making Envelope Immediate

We can instead take next period's assets as the control, $u = A'$, and write

$$V(A, Y) = \max_{A'} \; u\big(A(1+r) + Y - A'\big) + \beta V(A', Y').$$

Here $g(x, u)$ is the identity in u[3]
so the Benveniste–Scheinkman envelope gives

$$V_A(A, Y) = u'\big(C\big)\,(1+r).$$

The FOC $-u'(C) + \beta V_A(A', Y') = 0$ then implies

$$V_A(A', Y') = \frac{u'(C)}{\beta} \quad \Rightarrow \quad u'(C) = \beta(1+r)\,u'(C'),$$

the same Euler condition as above.

Remarks Three important points:

- Y is technically part of the state; under perfect foresight it can be suppressed in notation without loss.
- Both derivations highlight the "formulation dependence" of states/controls: taking C or A' as the control leads to the same optimality condition, but different algebra.

[3] "$g(x, u)$ is the identity at u" means that, when taking A' as the control, the transition is $x' = g(x, u) = (A', Y')$: the first component of the future state is exactly the control, so that $\partial A'/\partial u = 1$ and, crucially for the envelope with respect to A, $\partial x'/\partial A = (\partial A'/\partial A, \partial Y'/\partial A) = (0, 0)$ (because $A' = u$ and Y' is exogenous). Intuition: if I choose C as the control, the constraint $A' = (1+r)A + Y - C$ makes A' *dependent* on A; then when deriving V with respect to A, a "cross" term appears via A' (and must be canceled using the FOC). On the other hand, if I choose A' as the control variable, the future state no longer changes when I vary A in that derivative: there is no cross effect and the envelope condition is immediate, $V_A(A, Y) = u'(C)(1+r)$.

- Under the usual concavity/compactness conditions, the Bellman problem admits a unique time-invariant value function and policy.

Statement of the Problem II

Sequence problem

Given an initial state x_0, consider

$$V(x_0) \;\equiv\; \max_{\{u_t\}_{t\geq 0}} \; E_0 \sum_{t=0}^{\infty} \beta^t \, r(x_t, u_t)$$

subject to

$$x_{t+1} = g(x_t, u_t, \varepsilon_{t+1}), \qquad \{\varepsilon_t\} \text{ i.i.d. with distribution } F, \quad x_0 \in \mathbb{R}^n \text{ given.}$$

Recursive formulation (Bellman equation)

$$V(x) \;=\; \max_u \Big\{ r(x,u) \;+\; \beta \, \mathbb{E}\big[V\big(g(x,u,\varepsilon)\big) \,\big|\, x \big] \Big\},$$

where the conditional expectation is

$$\mathbb{E}\big[V\big(g(x,u,\varepsilon)\big) \,\big|\, x \big] \;\equiv\; \int V\big(g(x,u,\varepsilon)\big) \, dF(\varepsilon).$$

First-order condition For an interior optimum,

$$\frac{\partial r(x,u)}{\partial u} \;+\; \beta \, E\left[\frac{\partial V(x')}{\partial x'} \, \frac{\partial g(x,u,\varepsilon)}{\partial u} \,\Big|\, x \right] \;=\; 0.$$

Envelope (Benveniste–Scheinkman). If $\partial g(x,u,\varepsilon)/\partial x = 0$, the usual reasoning yields

$$\frac{\partial V(x)}{\partial x} \;=\; \frac{\partial r(x,u)}{\partial x}.$$

And you get

$$\frac{\partial r(x,u)}{\partial u} \;+\; \beta \, E\left[\frac{\partial r(x',u')}{\partial x'} \, \frac{\partial g(x,u,\varepsilon)}{\partial u} \,\Big|\, x \right] \;=\; 0.$$

Example II: Consumption–Savings with Stochastic Income

Let current assets A and current income Y be the state. The Bellman problem (taking next period's assets A' as the control) is

$$V(A,Y) \;=\; \max_{A'} \; u\big(A(1+r) + Y - A'\big) \;+\; \beta \, \mathbb{E}\big[V\big(A', Y'\big) \,\big|\, Y \big]. \tag{Bellman}$$

Why is Y a state? It carries information about Y' (e.g., if Y follows a Markov process), so it matters for expectations even though today it only enters the budget through resources.

Envelope (identity in the control). With A' as the control, the transition for the asset component is the identity: the choice A' *is* next period's asset state. Hence, when differentiating with respect to A, we hold A' fixed and obtain the envelope directly:

$$V_A(A,Y) \;=\; u_C(C)\,(1+r), \qquad C \equiv A(1+r) + Y - A'. \tag{Envelope}$$

Note: If instead C were the control, $A' = (1+r)A + Y - C$ depends on A and a cross term would appear; it cancels with the FOC below.

FOC.

$$-\,u_C(C) \;+\; \beta\,\mathbb{E}\left[V_A\big(A',Y'\big)\,\big|\,Y\right] \;=\; 0. \tag{FOC}$$

Using the envelope *one period ahead*, $V_A(A',Y') = (1+r)\,u_C(C')$, we obtain the familiar stochastic Euler equation:

$$u_C(C) \;=\; \beta(1+r)\,\mathbb{E}\left[u_C\big(C'\big)\,\big|\,Y\right]. \tag{Euler}$$

What not to confuse (derivative bookkeeping).

- $\dfrac{\partial}{\partial A'}\big\{u(A(1+r)+Y-A')\big\} \;=\; -\,u_C(C).$
- $\dfrac{\partial}{\partial A}\big\{u(A(1+r)+Y-A')\big\} \;=\; u_C(C)\,(1+r)$ is the derivative wrt the *current* state A, not wrt the *control* A'. Do not mix the role of A' as control with the role of A as (future) state variable.
- Hence $\dfrac{\partial r(x',u')}{\partial x'}$ in the slide's abstract notation corresponds here to $(1+r)\,u_C(C')$.
- If g denotes the law of motion for the asset state, then with A' as control we have $g(A',\cdot) = A'$ (identity), so $\partial g/\partial A' = 1$.

Combining the FOC with the one-step-ahead envelope gives

$$-\,u_C(C) \;+\; \beta(1+r)\,\mathbb{E}\left[u_C\big(C'\big)\,\big|\,Y\right] = 0 \quad\Longleftrightarrow\quad u_C(C) \;=\; \beta(1+r)\,\mathbb{E}\left[u_C\big(C'\big)\,\big|\,Y\right].$$

To avoid any confusion about what to plug into the general Euler equation, start from the FOC

$$-\,u_C(C) \;+\; \beta\,\mathbb{E}\left[V_A\big(A',Y'\big)\,\big|\,Y\right] = 0,$$

then apply Benveniste–Scheinkman to get $V_A(A,Y) = u_C(C)\,(1+r)$ and shift it forward to (A',Y') to obtain $V_A(A',Y') = (1+r)u_C(C')$. Substituting back delivers the Euler equation.

If Y is i.i.d., define *cash on hand*

$$Z \;\equiv\; A(1+r)+Y.$$

Now Y does not forecast Y' and can be dropped from the state. The Bellman equation collapses to a single state variable:

$$V(Z) \;=\; \max_{A'}\; u\big(Z-A'\big) \;+\; \beta\,\mathbb{E}\left[V\big(Z'\big)\right], \qquad Z' \;=\; A'(1+r)+Y'.$$

Economic takeaway. Y entered the state only because it predicted Y'. When that forecasting role disappears (i.i.d. case), current resources are fully summarized by Z. In general, look for transformations that compress forecasting information into fewer state variables (e.g., cash-on-hand with i.i.d. shocks; companion-form stacking when only finitely many lags matter).

Summary so far

So far, Dynamic Programming has delivered (via Bellman + Benveniste–Scheinkman) the Euler equation. The next step is more substantive: can we compute/characterize the value function $V(x)$ and the policy $h(x)$? To push in that direction, we will need a bit more mathematical background.

Contraction Mapping for the Bellman Operator

Bellman operator as a map on functions

Let X be the state space and U the set of feasible controls at each $x \in X$. Given:

- a bounded one–period return (flow utility) $r : X \times U \to \mathbb{R}$,
- a transition $g : X \times U \times E \to X$ with shock $\varepsilon \in E$,
- a discount factor $\beta \in (0, 1)$,
- and a conditional expectation operator $\mathbb{E}[\cdot \mid x]$ over ε,

define the *Bellman operator* $T : \mathcal{S} \to \mathcal{S}$ by

$$[T(V)](x) \;\equiv\; \max_{u \in U(x)} \Big\{ r(x, u) + \beta\, \mathbb{E}\left[V\big(g(x, u, \varepsilon)\big) \,\middle|\, x \right] \Big\}.$$

Here $\mathcal{S}$ is a set of real-valued functions on X (specified below).

Interpretation. T takes a candidate continuation value V and produces a new function: (i) forecast next period's value through g and expectation, (ii) discount, (iii) add current flow r, (iv) maximize over u. The Bellman equation is the fixed-point condition

$$T(V) = V.$$

Let $\mathcal{S}$ be the space of bounded real functions on X, and endow it with the sup norm

$$\|V\|_\infty \;\equiv\; \sup_{x \in X} |V(x)|, \qquad d(V, \widetilde{V}) = \|V - \widetilde{V}\|_\infty.$$

Then $(\mathcal{S}, d)$ is a complete metric space (the space of bounded functions is complete under the sup norm).

Contraction Mapping Theorem (Banach)

Theorem 1 (CMT)**.** *Suppose $T : \mathcal{S} \to \mathcal{S}$ is a contraction with modulus $\beta \in (0, 1)$ under the metric d, i.e.*

$$d\big(TV, T\widetilde{V}\big) \le \beta\, d(V, \widetilde{V}) \qquad \forall\, V, \widetilde{V} \in \mathcal{S}.$$

Then:

1. ***Existence and Uniqueness.*** *T has exactly one fixed point $V^\star \in \mathcal{S}$ with $T(V^\star) = V^\star$.*
2. ***Convergence of Value Iteration.*** *For any $V_0 \in \mathcal{S}$, the sequence $V_{n+1} = T(V_n)$ satisfies $V_n \to V^\star$.*
3. ***Geometric error bound.*** *For all $n \ge 0$,*

$$d\big(V_n, V^\star\big) \;\le\; \beta^n\, d\big(V_0, V^\star\big).$$

4. ***Operational bound (no $V^\star$).*** *For all $n \ge 0$,*

$$d\big(V_n, V^\star\big) \;\le\; \frac{\beta^n}{1 - \beta}\, d\big(V_n, V_{n+1}\big).$$

This bound is convenient because it only involves iterates you can compute.

Intuition A contraction shrinks distances by a fixed factor $\beta < 1$. Iterating T keeps shrinking the gap to the (unique) fixed point. The "geometry" is that T pulls every candidate value function closer to $V^\star$ by at least a factor β each step.

When is the Bellman operator a contraction?

In our setting with $\|\cdot\|_\infty$,

$$\big\|TV - T\widetilde{V}\big\|_\infty \;=\; \sup_x \Big|\max_u\{r(x,u) + \beta\,\mathbb{E}[V(x')|x]\} - \max_u\{r(x,u) + \beta\,\mathbb{E}[\widetilde{V}(x')|x]\}\Big| \;\le\; \beta\,\|V - \widetilde{V}\|_\infty,$$

because the r cancels and the max is 1-Lipschitz: $\big|\max_i a_i - \max_i b_i\big| \le \max_i |a_i - b_i|$. Moreover, conditional expectation is a contraction in sup norm: $\sup_x \big|\mathbb{E}[V(x') - \widetilde{V}(x') \mid x]\big| \le \|V - \widetilde{V}\|_\infty$. Hence with $\beta \in (0,1)$, T is a contraction on $(\mathcal{S}, \|\cdot\|_\infty)$.

Remark (Quick checklist)**.** To argue T is a contraction under $\|\cdot\|_\infty$ it suffices that: (i) r is bounded and $U(x)$ nonempty; (ii) $\mathbb{E}[\cdot \mid x]$ maps bounded functions to bounded functions and is 1-Lipschitz in sup norm; (iii) $\beta \in (0,1)$. In practice this is exactly the standard consumption–savings setup: bounded (or suitably normalized) returns, discounting, and well-behaved expectations.

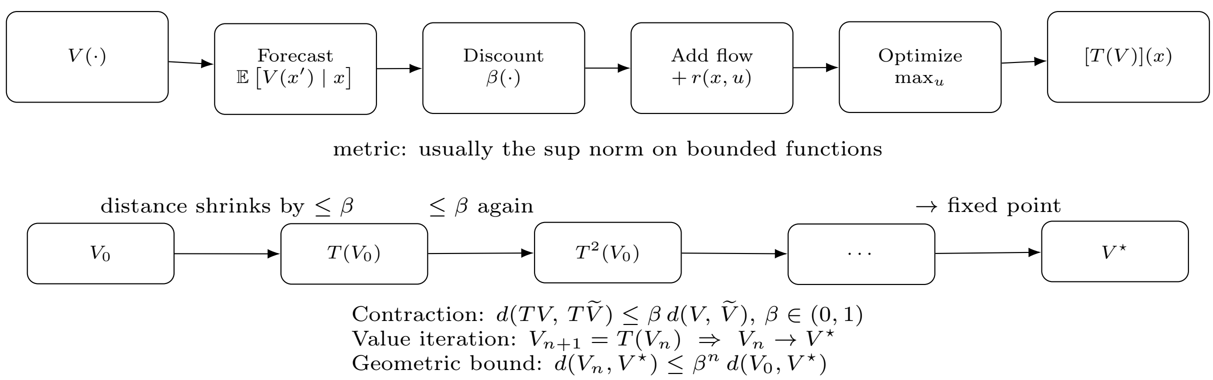
Why this is useful

If T is a contraction with modulus β:

- **Existence** of a solution to the Bellman equation.
- **Uniqueness** of that solution.
- **Computation**: value iteration $V_{n+1} = T(V_n)$ converges to $V^\star$ from any V_0, with a known rate (β) and practical error bounds.

In applications we often just iterate and see whether V_n stabilizes. Convergence suggests (and under contraction guarantees) existence and uniqueness. A simple robustness check is to try different V_0 and verify they converge to the same limit; monitor $d(V_n, V_{n+1})$ to certify accuracy via the operational bound.

Figure 3: Bellman operator as a pipeline (top) and as a contraction driving value iteration to the unique fixed point $V^\star$ (bottom).



Blackwell's sufficient conditions

Let $B(X)$ be the set of bounded, real-valued functions on a state space X and

$$d(V, \widetilde{V}) \;=\; \|V - \widetilde{V}\|_\infty \;\equiv\; \sup_{x \in X} |V(x) - \widetilde{V}(x)|$$

be the sup-norm metric. For $V, \widetilde{V} \in B(X)$, write $V \le \widetilde{V}$ to mean $V(x) \le \widetilde{V}(x)$ for all x; and for $a \in \mathbb{R}$, $(V + a)(x) \equiv V(x) + a$.

Theorem 2 (Blackwell)**.** *Let $T : B(X) \to B(X)$ satisfy:*

1. ***Monotonicity:*** *if $V \le \widetilde{V}$ then $TV \le T\widetilde{V}$.*
2. ***Discounting:*** *there exists $\beta \in (0, 1)$ such that for all $V \in B(X)$ and $a \ge 0$,*

$$T(V + a) \;\le\; T(V) + \beta a.$$

Then T is a contraction on $(B(X), \|\cdot\|_\infty)$ with modulus β:

$$d(TV,\, T\widetilde{V}) \;\le\; \beta\, d(V, \widetilde{V}) \qquad \forall\, V, \widetilde{V} \in B(X).$$

Consequently, T has a unique fixed point $V^\star$, and for any V_0 the iterates $V_{n+1} = T(V_n)$ converge to $V^\star$ with geometric rate β.

Proof (three short steps). Step 1. For any $V, \widetilde{V}$,

$$V(x) = \widetilde{V}(x) + \big(V(x) - \widetilde{V}(x)\big) \;\le\; \widetilde{V}(x) + \sup_{y \in X} |V(y) - \widetilde{V}(y)| \;=\; \widetilde{V}(x) + d(V, \widetilde{V}),$$

so $V \le \widetilde{V} + d(V, \widetilde{V})$.

Step 2. By monotonicity and discounting,

$$TV \;\le\; T\big(\widetilde{V} + d(V, \widetilde{V})\big) \;\le\; T\widetilde{V} + \beta\, d(V, \widetilde{V}).$$

Swapping $V, \widetilde{V}$ gives $T\widetilde{V} \le TV + \beta\, d(V, \widetilde{V})$.

Step 3. The two inequalities imply, for all x,

$$|\, TV(x) - T\widetilde{V}(x)\,| \;\le\; \beta\, d(V, \widetilde{V}),$$

hence $d(TV, T\widetilde{V}) \le \beta\, d(V, \widetilde{V})$. □

Intuition. Monotonicity: improving continuation values cannot hurt today's evaluation. Discounting: adding a constant a to continuation values moves T by at most βa. Together they bound T's "sensitivity" to changes in V by the factor $\beta < 1$, which is exactly the contraction property.

Garbage In—Garbage Out (GIGO) for Bellman fixed points

Let $(\mathcal{S}, d)$ be a complete metric space of functions (e.g. $B(X)$ with the sup norm). A subset $\mathcal{S}' \subset \mathcal{S}$ is *closed* if every convergent sequence in $\mathcal{S}'$ has its limit in $\mathcal{S}'$.

Closedness facts (with $\|\cdot\|_\infty$). (i) Bounded and continuous functions $X \to \mathbb{R}$ form a closed set in $B(X)$. (ii) On an interval $[a, b]$, bounded and increasing functions $[a, b] \to \mathbb{R}$ form a closed set. (iii)

On $[a,b]$, bounded and convex functions form a closed set. (Each follows since uniform limits preserve continuity, monotonicity, and convexity.)

Lemma (GIGO)**.** *Let $T : \mathcal{S} \to \mathcal{S}$ be a contraction with modulus $\beta \in (0,1)$ and let $V^\star$ be its unique fixed point. If $\emptyset \neq \mathcal{S}' \subset \mathcal{S}$ is closed and* invariant *under T (i.e. $T(\mathcal{S}') \subset \mathcal{S}'$), then $V^\star \in \mathcal{S}'$. More generally, if $T(\mathcal{S}') \subset \mathcal{S}'' \subset \mathcal{S}'$ with $\mathcal{S}''$ closed, then $V^\star \in \mathcal{S}''$.*

Proof. Pick any $V_0 \in \mathcal{S}'$. By invariance, $V_n \equiv T^n(V_0) \in \mathcal{S}'$ for all n. By contraction, $V_n \to V^\star$ in d. Closedness of $\mathcal{S}'$ gives $V^\star \in \mathcal{S}'$ (and similarly for $\mathcal{S}''$). □

Applications. Let T be the Bellman operator on $(B(X), \|\cdot\|_\infty)$.

- *Continuity:* If T maps bounded continuous functions into bounded continuous functions, then $V^\star$ is continuous.
- *Monotonicity:* On $[a,b]$, if T maps bounded increasing functions into increasing functions, then $V^\star$ is increasing.
- *Convexity/concavity:* On $[a,b]$, if T maps bounded convex (resp. concave) functions into convex (resp. concave) functions, then $V^\star$ is convex (resp. concave).

Strict versions. If, for every increasing V, the image $T(V)$ is *strictly* increasing, then starting from any increasing V_0 the sequence $T^n(V_0)$ is strictly increasing for all n and the limit $V^\star$ is strictly increasing. (Analogous statements hold for strict convexity/concavity.)

Intuition. Start with a "good" guess V_0 (continuous, increasing, convex, ...). If T never leaves that class (invariance) and value iteration converges, the limit inherits the good property by closedness. Garbage in $\Rightarrow$ garbage out; good in $\Rightarrow$ good out.

Equivalence of the sequence and the recursive problems

Let states evolve by $x_{t+1} = g(x_t, u_t, \varepsilon_{t+1})$ and per–period payoff be $r(x_t, u_t)$ with discount $\beta \in (0,1)$.

Finite horizon (perfect foresight for simplicity). Fix a terminal value $V_{T+1}(\cdot)$. Define recursively

$$V_t(x) \;=\; \max_u \left\{ r(x,u) + \beta\, V_{t+1}(g(x,u)) \right\}, \qquad t = T, \dots, 0.$$

By successive substitution,

$$V_0(x_0) \;=\; \max_{\{u_0,\dots,u_T\}} \sum_{t=0}^{T} \beta^t r(x_t,u_t) \;+\; \beta^{T+1} V_{T+1}(x_{T+1}),$$

subject to $x_{t+1} = g(x_t, u_t)$. Hence the recursive and sequence problems are the *same* object written differently; there is no fixed point (values are time–indexed).

Infinite horizon. Let V solve the Bellman equation

$$V(x) \;=\; \max_u \Big\{ r(x,u) + \beta\, \mathbb{E}\left[V(g(x,u,\varepsilon')) \mid x \right] \Big\}.$$

Iterating the recursion $T+1$ times along any feasible policy $\{u_t\}$ yields the telescoping representation

$$V(x_0) \;=\; \sup_{\{u_t\}} \; \mathbb{E}\left[\sum_{t=0}^{T} \beta^t r(x_t,u_t) \;+\; \beta^{T+1} V(x_{T+1}) \;\middle|\; x_0 \right].$$

If the *tail term* vanishes,

$$\lim_{T\to\infty} \beta^{T+1}\, \mathbb{E}[\,|V(x_{T+1})|\,|\,x_0] = 0,$$

then

$$V(x_0) \;=\; \sup_{\{u_t\}} \; \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t r(x_t,u_t) \;\middle|\; x_0 \right],$$

so the sequence and recursive formulations are equivalent.

Easy sufficient condition (practical). If $|r(x,u)| \le \bar{r} < \infty$ and $\beta \in (0,1)$, the fixed point satisfies

$$\|V\|_\infty \;\le\; \frac{\bar{r}}{1-\beta},$$

hence $\beta^{T+1}\,\mathbb{E}[|V(x_{T+1})|] \le \beta^{T+1}\|V\|_\infty \to 0$ and equivalence follows. (Variants: any bound of the form $|V(x)| \le K$ suffices; in growth models one uses no–Ponzi/borrowing constraints or weighted norms to ensure the tail goes to zero.)

Intuition. Backward induction (finite T) or repeated substitution (infinite horizon) "unwraps" the Bellman equation into the sequence objective plus a single terminal piece. Under discounting, the terminal piece dies out; what remains is exactly the sequence problem.

Numerical solution: value function iteration

Given the Bellman operator

$$[T(V)](x) \;\equiv\; \max_{u\in U(x)} \Big\{ r(x,u) + \beta\; \mathbb{E}\big[V\big(g(x,u,\varepsilon)\big) \mid x\big] \Big\},$$

value function iteration (VFI) computes V by fixed–point iteration.

Algorithm. Choose any V_0 (e.g., $V_0 \equiv 0$) and for $j = 0,1,2,\ldots$ update

$$V_{j+1}(x) \;=\; \max_{u\in U(x)} \Big\{ r(x,u) + \beta\; \mathbb{E}\big[V_j\big(g(x,u,\varepsilon)\big) \mid x\big] \Big\}.$$

Stop when $\|V_{j+1} - V_j\|_\infty <$ tol. Under the contraction conditions, $V_j \to V^\star$ and the *operational* error bound holds:

$$\|V^\star - V_j\|_\infty \;\le\; \frac{\beta}{1-\beta}\,\|V_j - V_{j-1}\|_\infty.$$

Practical remarks. (i) With a finite state space, the update is a finite–dimensional fixed point in $\mathbb{R}^n$; contraction carries over and convergence is guaranteed. (ii) With continuous states, one works on a grid and interpolates (e.g., linear, monotone cubic). Contraction applies to the *true* Bellman operator; the discretized/interpolated operator need not be a contraction—if convergence is slow/unstable, refine the grid or use structure–preserving interpolation (monotonicity/concavity). (iii) Expectations: use quadrature or simulation consistent with the shock process and the interpolation scheme.

Howard policy improvement / modified policy iteration. Let h_j be the maximizer (policy) associated with V_j. Rather than re–maximizing every step, *evaluate* the current policy several times before the next maximization:

$$V_{j,\ell+1}(x) \;=\; r\big(x, h_j(x)\big) + \beta\, \mathbb{E}\big[V_{j,\ell}\big(g(x, h_j(x), \varepsilon)\big) \mid x\big], \qquad \ell = 0, \ldots, L-1,$$

starting from $V_{j,0} = V_j$; then set $V_{j+1} = V_{j,L}$ and update h_{j+1}. This typically yields large speedups because costly maximizations are done only every L iterations and values improve monotonically.

Euler–equation methods (policy approximation)

When the first–order approach is valid, the optimal policy $u = h(x)$ satisfies the functional equation

$$R\big(x, h(x)\big) \;\equiv\; \frac{\partial r(x,u)}{\partial u}\Big|_{u=h(x)} \;+\; \beta\, \mathbb{E}\left[\frac{\partial r(x',u')}{\partial x'}\,\frac{\partial g(x,u,\varepsilon)}{\partial u}\Big|_{u=h(x),\, u'=h(x')}\;\middle|\; x\right] = 0.$$

Approximate $h(\cdot)$ in a parametric family $h(x;\theta)$ and choose θ so that $R(x, h(x;\theta)) \approx 0$:

- *Collocation:* enforce $R(\xi_k, h(\xi_k;\theta)) = 0$ at nodes $\{\xi_k\}$.
- *Residual minimization:* choose θ to minimize $\sum_k w_k\, R(\xi_k, h(\xi_k;\theta))^2$ (deterministic or simulation–based).

This targets the *policy* directly. It is fast but requires (a) valid FOCs (concavity, interior solutions), (b) careful treatment of constraints (e.g., occasionally binding), and (c) accurate expectations consistent with the parameterization.

Summary and guidance

- VFI solves for $V(\cdot)$; the policy follows from the argmax. It is robust (works with kinks/non–differentiabilities) but can be slow.
- Euler–equation methods solve for $h(\cdot)$; very fast when FOCs are valid and constraints are well handled.
- Once you have either V or h, the other is essentially implied.
- For speed with reliability: start with VFI on a coarse grid to get a good initial policy, then switch to policy improvement and/or an Euler residual refinement.

Closed-form solutions: when and why

Only a handful of dynamic programs admit paper-and-pencil solutions. Two canonical ones: (i) Brock–Mirman (optimal growth, Cobb–Douglas output, $\delta = 1$, log utility); (ii) consumption–saving with HARA preferences under special income processes. In both, the strategy is *guess and verify*: posit a parametric form for V (and often for the policy) and check optimality.

Why useful? (i) Solving by hand builds intuition for numerical algorithms (what objects the Bellman update is pushing). (ii) These are limit cases of richer models; their closed forms are excellent initial guesses for value/policy iteration.

Brock–Mirman (deterministic, capital only)

Planner chooses $\{C_t, K_{t+1}\}_{t\geq 0}$ to maximize

$$V(K_0) \equiv \max_{\{C_t, K_{t+1}\}} \sum_{t=0}^{\infty} \beta^t \log C_t \quad \text{s.t.} \quad K_{t+1} = K_t^{\alpha} - C_t, \;\; C_t \geq 0, \;\; K_{t+1} \geq 0,$$

with $\beta \in (0,1)$ and $\alpha \in (0,1)$. Resource feasibility implies $K_{t+1} \in [0, K_t^{\alpha}]$.

Recursive form.

$$V(K) \;=\; \max_{K' \in [0, K^{\alpha}]} \Big\{ \log\big(K^{\alpha} - K'\big) + \beta\, V(K') \Big\}.$$

Guess. Try $V(K) = A + B \log K$ with constants (A, B) to be determined.

Verification. FOC of the Bellman problem under the guess:

$$-\frac{1}{K^{\alpha} - K'} + \beta \frac{B}{K'} = 0 \quad \Longrightarrow \quad K' = \frac{\beta B}{1 + \beta B} K^{\alpha} \equiv s\, K^{\alpha}.$$

Thus the policy is *proportional savings.* Plug the Euler equation from the primal problem,

$$\frac{1}{C_t} = \beta \frac{1}{C_{t+1}} \alpha K_{t+1}^{\alpha - 1},$$

together with $C_t = (1-s)K_t^{\alpha}$, $K_{t+1} = sK_t^{\alpha}$, to obtain $s = \alpha\beta$. Equating the two expressions for s pins down B:

$$\frac{\beta B}{1 + \beta B} = \alpha\beta \quad \Longrightarrow \quad B = \frac{\alpha}{1 - \alpha\beta}.$$

Finally, identify A by substituting the optimal policy ($K' = \alpha\beta K^{\alpha}$, $C = (1-\alpha\beta)K^{\alpha}$) into the Bellman equation:

$$A = \frac{\log(1-\alpha\beta) + \beta B \log(\alpha\beta)}{1 - \beta} \qquad \text{with} \qquad B = \frac{\alpha}{1 - \alpha\beta}.$$

Closed forms.

$$\boxed{K_{t+1} = \alpha\beta\, K_t^{\alpha}, \qquad C_t = (1-\alpha\beta)\, K_t^{\alpha}}$$

$$\boxed{V(K) = A + B \log K, \quad B = \frac{\alpha}{1-\alpha\beta}, \quad A = \frac{\log(1-\alpha\beta) + \beta B \log(\alpha\beta)}{1-\beta}}$$

Feasibility requires only $\alpha\beta \in (0,1)$, which holds given $\alpha, \beta \in (0,1)$.

Notes. (i) The proportional-savings rule is the hallmark of log utility plus Cobb–Douglas with full depreciation. (ii) With i.i.d. multiplicative shocks $Y_t = Z_t K_t^{\alpha}$ and $\mathbb{E}[\log Z_t] = 0$, the same guess delivers $K_{t+1} = \alpha\beta\, Z_t K_t^{\alpha}$ and the value function gains an $\mathbb{E}[\log Z]$ term in A; the *share* $\alpha\beta$ is unchanged. (iii) These formulas are excellent warm starts for VFI or policy iteration in nearby calibrated models.

Cookie–Cutter: Guess and verify under a certain value function form

Setup. State x (e.g. assets A), control y (e.g. next assets A').
Bellman: $V(x) = \max_y \{\, u(x,y) + \beta V(f(x,y))\,\}$ with easy-to-eliminate y via the flow budget.
1) Guess. Pick a functional form closed under the operator, e.g.

$$V(x) = \alpha + \kappa\, g(x)^{\,1-\rho} \quad \text{or} \quad V(x) = \alpha + \kappa \log g(x).$$

2) Reduce. Eliminate consumption using the FBC (e.g. $C = A - A'/R$), so the Bellman is $V(A) = \max_{A'}\{u(C(A,A')) + \beta V(A')\}$.
3) FOC/Policy. Compute $\partial/\partial A'$: solve for $A' = h(A)$ (often affine). If messy, do a change of variables that linearizes marginal utility (e.g. HARA: $Z \equiv \frac{a}{\sigma}C + b$).
4) Euler pin-down. Use envelope $V'(A) = u'(C)$ to turn FOC into the Euler $u'(C_t) = \beta R\, u'(C_{t+1})$; this pins the slope/ratio of the policy (e.g. $Z_{t+1} = \psi Z_t$, $\psi = (\beta R)^{1/\sigma}$).
5) Verify (levels). Plug $h(A)$ and the guessed $V(\cdot)$ back into Bellman:

$$u(C(A,h(A))) + \beta V(h(A)) \overset{!}{=} V(A) \quad \forall A.$$

Match coefficients to solve unknown constants $(\alpha, \kappa, \dots)$.
6) Checks. (i) Domain (e.g. $Z > 0$), (ii) TVC, (iii) corner solutions, (iv) knife-edges ($\beta R = 1$, $\rho \to 1$), (v) monotonicity/concavity as needed.

Consumption–savings with $Y = 0$ (log utility)

We revisit the single-agent problem with no exogenous income, $Y = 0$. Preferences are $u(C) = \log C$ and assets evolve with the flow budget constraint

$$A' \;=\; A(1+r) - C.$$

Idea (interpretation). If human capital can be traded at the market rate of return, we can subsume labor income into asset income and set $Y = 0$. Perfect foresight is not essential for the main insight.

Bellman equation and guess.

$$V(A) \;=\; \max_{C>0} \left\{ \log C + \beta V(A') \right\}, \qquad A' = A(1+r) - C.$$

Guess a linear–affine transformation of felicity:

$$V(A) = E + F \log A.$$

Verification. Plugging the guess,

$$E + F\log A = \max_{C>0} \left\{ \log C + \beta E + \beta F \log\big((1+r)A - C\big) \right\}.$$

FOC:

$$\frac{1}{C} = \frac{\beta F}{(1+r)A - C} \quad\Longrightarrow\quad C = \frac{1+r}{1+\beta F}\, A.$$

Substitute back and compare coefficients on $\log A$:

$$F \;=\; 1 + \beta F \quad\Longrightarrow\quad F = \frac{1}{1-\beta}.$$

Therefore the (policy) consumption rule is

$$\boxed{C = (1-\beta)(1+r)\, A}.$$

Special case: if $\beta = \frac{1}{1+r}$, then $C = rA$. *Note on timing:* under the alternative timing $A' = (A-C)(1+r)$ the rule becomes $C = (1-\beta)A$.

Constants. The level E follows from the constant terms in the Bellman equation (uninteresting for policy and comparative statics).

Generalization: HARA felicities. The same logic extends to HARA utility; the value function is a linear–affine transform of felicity. HARA is

$$u(C) = \frac{\sigma}{1-\sigma}\left(\frac{aC}{\sigma} + b\right)^{1-\sigma},$$

with parameters (σ, a, b). Notable members:

- **Linear:** $\sigma = 0 \;\Rightarrow\; u(C) = aC$ (limit case).
- **Quadratic:** $\sigma = -1 \;\Rightarrow\; u(C) = -\frac{1}{2}(-aC + b)^2 \;=\; ab\,C - \frac{1}{2}b^2 - \frac{1}{2}a^2C^2$.
- **CRRA:** $a = \sigma,\; b = 0 \;\Rightarrow\; u(C) = \dfrac{\sigma}{1-\sigma}\,C^{1-\sigma}$ (the front factor is irrelevant).

In each HARA case (with $Y = 0$), $V(A) = E + F \cdot \text{felicity}(A)$ with F pinned down by the same coefficient-matching argument.

CARA as a limiting case of HARA. CARA utility can be derived as a limit of the HARA class. Set $b = 1$ and take $\sigma \to \infty$. Then:

$$\lim_{\sigma\to\infty} \frac{\sigma}{1-\sigma}\left(\frac{aC}{\sigma} + 1\right)^{1-\sigma} = \lim_{\sigma\to\infty} \frac{\sigma}{1-\sigma} \cdot \lim_{\sigma\to\infty}\left(\frac{aC}{\sigma} + 1\right)^{1-\sigma}.$$

Clearly,

$$\lim_{\sigma\to\infty} \frac{\sigma}{1-\sigma} = -1,$$

so we can focus on the second limit.

Inspection shows this is an indeterminate form of type $1^{-\infty}$. To apply L'Hôpital's rule, it is convenient to work with logarithms:

$$\lim_{\sigma\to\infty} (1-\sigma)\,\log\!\left(\frac{aC}{\sigma} + 1\right),$$

which is of the type $(-\infty)\cdot 0$. Rewriting gives a $0/0$ form:

$$\lim_{\sigma\to\infty} \frac{\log\!\left(\frac{aC}{\sigma} + 1\right)}{\frac{1}{1-\sigma}}.$$

Applying L'Hôpital's rule:

$$\lim_{\sigma\to\infty} \frac{\left(-\frac{1}{\frac{aC}{\sigma}+1}\right)\frac{aC}{\sigma^2}}{\left(\frac{1}{1-\sigma}\right)^2} = -\,aC \lim_{\sigma\to\infty} \frac{1 - 2\sigma + \sigma^2}{aC\sigma + \sigma^2} = -aC.$$

Since $-aC$ is the limit of the logarithm, the original expression's limit is its exponential:

$$\lim_{\sigma\to\infty} \frac{\sigma}{1-\sigma}\left(\frac{aC}{\sigma} + 1\right)^{1-\sigma} = -\,e^{-aC}.$$

Hence CARA (constant absolute risk aversion) is obtained as

$$u(C) = -\,\frac{1}{a}\,e^{-aC},$$

which is a special case of the HARA family of felicity functions.