Economics 210 Section Notes

Luke C.D. Stein

Autumn 2009[*]

Contents

These notes are based on (and at points copy without explicit citation from) notes by previous TAs Max Floetotto, Marcello Miccoli, Romans Pancs, and William "Gui" Woolston; Manuel Amador and Nir Jaimovich's slides and solution sets; Krueger, *Macroeconomic Theory* (manuscript), section 2.3; Ljungqvist and Sargent, *Recursive Macroeconomic Theory*, chapter 8; Stachurski, *Economic Dynamics: Theory and Computation*, chapters 3 and 6; Stokey, Lucas, and Prescott, *Recursive Methods in Economic Dynamics*, section 3.1; and the Wikipedia "Eigenspace" page.

[*]This version updated April 19, 2013.

1 Welcome

- email: lstein@stanford.edu

- homepage: http://www.lukestein.com

- Office Hours: Wednesdays, 3:15–5:05 in Landau Economics room 350

- Sections: Fridays, 12:15–2:05 in building 240 room 110

2 A taxonomy of economic models

1. **Static** Focus of Microeconomics core.

   Dynamic Focus of Macroeconomics core.

2. **Deterministic** e.g., $x_1 = 6, x_2 = 4$; $x_t = x_{t-1} + 1$ with $x_0 = 0$ (a recursive sequence defined by a difference equation; if the subscripts index time, we also call this a dynamical system). Nir's class.

   Stochastic e.g., $x_t \overset{\text{IID}}{\sim} \{H, T\}$ with probability $\frac{1}{2}$ each; $x_t = x_{t-1} + \varepsilon_t$ with $\varepsilon_t \overset{\text{IID}}{\sim} N(0, 1)$ and $x_0 = 0$ (a stochastic recursive sequence). Manuel's class.

3. **Finite horizon (discrete)** Use Kuhn-Tucker (i.e., Lagrangians).

   Infinite horizon (discrete) as Sequence Problem: Kuhn-Tucker; as Functional Equation: dynamic programming.

   Continuous time Use Hamiltonians.

4. **Discrete state/choice spaces** Good for computer.

   Continuous spaces Good for mathematical analysis.

5. **Steady state**

   Explicit dynamics Use linearization (Taylor approximation) and "diagonalization" (eigenvalue decomposition) or phase diagrams.

6. **One "agent"** Either a true single agent (partial equilibrium problems) or a social planner.

   Many agents interacting through markets; competitive equilibrium either in period-by-period (i.e., sequential) markets or in one big market at $t = 0$ (Arrow-Debreu).

3 Some thoughts about utility functions

Consider our canonical utility function (for problems with discrete time):

$$U\big(\{c_t\}_{t=0}^T\big) = \sum_{t=0}^T \beta^t u(c_t),$$

where T may be infinite. The function $u(\cdot)$ is called the **felicity** or **instantaneous utility** function. This utility function has several important properties, including:

Time separability The period utility at time t only depends on consumption at time t. For example, there is no habit persistence.

Exponential discounting β constant and $\beta < 1$ mean the agent values consumption today more than consumption tomorrow, with a constant "strength of preference" for consumption sooner vs. later.

Stationarity The felicity function is time invariant. Thus when $T = \infty$, the utility function evaluated over future consumption looks the same from any point in the future.

Most of the time we use one of a small number of felicity functions:

- $u(x) = x^\alpha$.

- $u(x) = \frac{x^{1-\sigma}-1}{1-\sigma}$; this represents the same preferences as the previous example.

- $u(x) = \log(x)$; the previous example approaches log utility as $\sigma \nearrow 1$.

These felicity functions give rise to additional useful properties of $U(\cdot)$, including:

Strict monotonicity $U(\cdot)$ is strictly increasing in c_t for all t.

Continuity Mathematically convenient.

Twice continuous differentiability Mathematically convenient.

Strict concavity $\partial^2 U / \partial c_t^2 < 0$ for all t, corresponding to decreasing marginal utility.

Inada conditions $\lim_{c_t \searrow 0} \partial U / \partial c_t = +\infty$ and $\lim_{c_t \nearrow +\infty} \partial U / \partial c_t = 0$ for all t, assuring that optimal $c_t \in (0, +\infty)$ for all t.[1]

Homotheticity Meaning that $\{c_t\}_t \succsim \{\tilde{c}_t\}_t$ if and only if $\{\lambda c_t\}_t \succsim \{\lambda \tilde{c}_t\}_t$, implying that the units in which consumption are measured don't matter (in many models).

Constant relative risk aversion (and elasticity of substitution). The Arrow-Pratt coefficient of relative risk aversion $-c_t u''(c_t)/u'(c_t)$, and the intertemporal elasticity of substitution $\varepsilon_{c_{t+1}/c_t, 1/R}$ do not depend on t, c_t, or c_{t+1}.

In particular, we will basically *always* assume the first four conditions, which help ensure that we can solve optimization problems using the Kuhn-Tucker (i.e., Lagrangian) algorithm.

4 Some math reminders

Including these here is a bit *ad hoc*, but the following are three mathematical concepts you need to be familiar with.

---

[1] In all, there are six Inada conditions. In addition to the two listed limit conditions, they are (1) continuous differentiability, (2) strict monotonicity, (3) strict concavity, and (4) $U(\vec{0}) = 0$.

4.1 Taylor approximation

Consider a differentiable real-valued function on (some subset of) Euclidean space, $f\colon \mathbb{R}^n \to \mathbb{R}$. The function can be approximated in the region around some arbitrary point $y \in \mathbb{R}^n$ by its tangent hyperplane.

If $f\colon \mathbb{R} \to \mathbb{R}$, this approximation takes the form

$$f(x) \approx f(y) + f'(y)(x - y).$$

If $f\colon \mathbb{R}^n \to \mathbb{R}$, this approximation takes the form

$$f(x) \approx f(y) + \sum_{i=1}^{n} f_i'(y)(x_i - y_i) = f(y) + \big[\nabla f(y)\big] \cdot (x - y),$$

where $\cdot$ is the vector dot product operator.

4.2 Concave functions

Gui Woolston has an excellent note on this subject, on which this section is based.

The following are all necessary and sufficient (i.e., equivalent) conditions for concavity of a twice-differentiable (real-to-real) function $f\colon \mathbb{R} \to \mathbb{R}$: for all $x, y \in \mathbb{R}$,

1. "Mixtures" give higher values than "extremes": for all $\alpha \in [0, 1]$,

$$f\big(\alpha x + (1 - \alpha)y\big) \geq \alpha f(x) + (1 - \alpha)f(y);$$

2. $f''(x) \leq 0$;

3. $x \geq y$ if and only if $f'(x) \leq f'(y)$, which can also be stated compactly as $\frac{f'(y) - f'(x)}{y - x} \leq 0$ or $(y - x)\big(f'(y) - f'(x)\big) \leq 0$; and

4. $f(\cdot)$ lies below its first-order Taylor approximation:

$$f(x) \leq \underbrace{f(y) + f'(y)(x - y)}_{\text{Taylor approx. at } x \text{ about } y}.$$

Also, for all concave functions (not just differentiable ones), a local maximum is a global maximum. And a concave function must be continuous on the interior of its domain (although it need not be continuous on the boundaries).

4.3 The envelope theorem

Envelope theorems relate the derivative of a value functions to the derivative of the objective function. Here is a simple envelope theorem for unconstrained optimization:

$$
\begin{aligned}
v(q) &= \max_{x} f(x, q) \\
&= f\big(x_*(q), q\big) \\
\frac{dv}{dq} &= \frac{\partial f}{\partial q}\big(x_*(q), q\big) + \underbrace{\frac{\partial f}{\partial x}\big(x_*(q), q\big)}_{=0 \text{ by FOC}} \cdot \frac{\partial x_*}{\partial q} \\
&= \frac{\partial f}{\partial q}\big(x_*(q), q\big).
\end{aligned}
$$

The fact that the derivative of the envelope equals the derivative of the objective function holding the choice variable fixed is illustrated for $v(z) \equiv \max_x -5(x - z)^2 - z(z - 1)$:
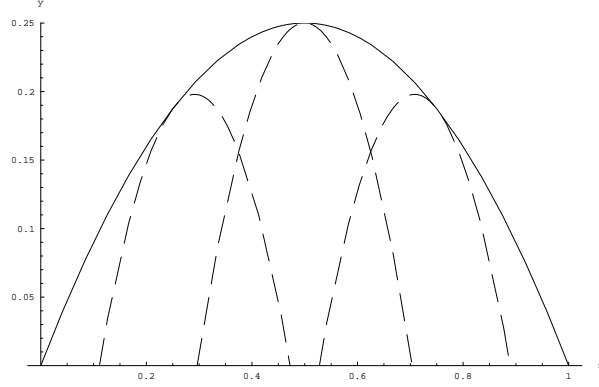
As an exercise, use the first-order and envelope conditions for the functional equation form of the NCGM (given in equation 8) to derive its intereuler. The derivation should really include an argument that the optimal k' is interior to the feasible set.

A more complete envelope theorem for constrained optimization is:

Theorem 1. *Consider a constrained optimization problem $v(\theta) = \max_x f(x,\theta)$ such that $g_1(x,\theta) \geq 0$, ..., $g_K(x,\theta) \geq 0$.*

Comparative statics on the value function are given by:

$$\frac{\partial v}{\partial \theta_i} = \left.\frac{\partial f}{\partial \theta_i}\right|_{x_*} + \sum_{k=1}^{K} \lambda_k \left.\frac{\partial g_k}{\partial \theta_i}\right|_{x_*} = \left.\frac{\partial \mathcal{L}}{\partial \theta_i}\right|_{x_*}$$

(for Lagrangian $\mathcal{L}(x,\theta,\lambda) \equiv f(x,\theta) + \sum_k \lambda_k g_k(x,\theta)$) for all θ such that the set of binding constraints does not change in an open neighborhood.

Roughly, this states that **the derivative of the value function is the derivative of the Lagrangian**.

Proof. The proof is given for a single constraint (but is similar for K constraints): $v(x,\theta) = \max_x f(x,\theta)$ such that $g(x,\theta) \geq 0$.

Lagrangian $\mathcal{L}(x,\theta) \equiv f(x,\theta) + \lambda g(x,\theta)$ gives FOC

$$\left.\frac{\partial f}{\partial x}\right|_* + \lambda \left.\frac{\partial g}{\partial x}\right|_* = \mathbf{0} \iff \left.\frac{\partial f}{\partial x}\right|_* = -\lambda \left.\frac{\partial g}{\partial x}\right|_* \tag{1}$$

where the notation $\cdot|_*$ means "evaluated at $(x_*(\theta),\theta)$ for some θ."

If $g(x_*(\theta),\theta) = 0$, take the derivative in θ of this equality condition to get

$$\left.\frac{\partial g}{\partial x}\right|_* \left.\frac{\partial x_*}{\partial \theta}\right|_\theta + \left.\frac{\partial g}{\partial \theta}\right|_* = \mathbf{0} \iff \left.\frac{\partial g}{\partial \theta}\right|_* = - \left.\frac{\partial g}{\partial x}\right|_* \left.\frac{\partial x_*}{\partial \theta}\right|_\theta. \tag{2}$$

Note that, $\frac{\partial \mathcal{L}}{\partial \theta} = \frac{\partial f}{\partial \theta} + \lambda \frac{\partial g}{\partial \theta}$. Evaluating at $(x_*(\theta),\theta)$ gives

$$\left.\frac{\partial \mathcal{L}}{\partial \theta}\right|_* = \left.\frac{\partial f}{\partial \theta}\right|_* + \lambda \left.\frac{\partial g}{\partial \theta}\right|_\theta.$$

If $\lambda = 0$, this gives that $\frac{\partial \mathcal{L}}{\partial \theta}|_* = \frac{\partial f}{\partial \theta}|_*$; if $\lambda > 0$, complementary slackness ensures $g(x_*(\theta),\theta) = 0$ so we can apply equation 2. In either case, we get that

$$= \frac{\partial f}{\partial \theta} - \lambda \left.\frac{\partial g}{\partial x}\right|_* \left.\frac{\partial x_*}{\partial \theta}\right|_\theta. \tag{3}$$

Applying the chain rule to $v(x, \theta) = f\big(x_*(\theta), \theta\big)$ and evaluating at $\big(x_*(\theta), \theta\big)$ gives

$$
\begin{aligned}
\frac{\partial v}{\partial \theta}\bigg|_* &= \frac{\partial f}{\partial x}\bigg|_* \frac{\partial x_*}{\partial \theta}\bigg|_\theta + \frac{\partial f}{\partial \theta}\bigg|_* \\
&= -\lambda \frac{\partial g}{\partial x}\bigg|_* \frac{\partial x_*}{\partial \theta}\bigg|_\theta + \frac{\partial f}{\partial \theta}\bigg|_* \\
&= \frac{\partial \mathcal{L}}{\partial \theta}\bigg|_*,
\end{aligned}
$$

where the last two equalities obtain by equations 1 and 3, respectively. □

5 Introduction to sequence problems

5.1 Two-period saving problem in partial equilibrium

In the first period, you can buy a risk free bonds which return Ra in the second period. Income y_1 and y_2 are available in the first and second periods, respectively. We get intereuler

$$
u'(\underbrace{y_1 - a}_{c_1}) = \beta R u'(\underbrace{y_2 + Ra}_{c_2}),
$$

which can be interpreted as follows: the left-hand side is the marginal benefit of consuming an extra unit today, the right-hand side is the marginal cost of consuming an extra unit today, comprising

1. R: Conversion from a unit of consumption today to units of consumption tomorrow,

2. $u'(c_2)$: Conversion from units of consumption tomorrow to felicits tomorrow,

3. β: Conversion from felicits tomorrow to utils.

 Things in other models can interfere with an intereuler like this holding; e.g.,

1. **Incomplete markets**, such as if R were stochastic;

2. **Non-time-separable utility**;

3. **Budget/borrowing constraints** including irreversible investment might prevent the agent from borrowing or saving all he would like to, resulting in $u'(c_1) \gtrless \beta R u'(c_2)$.

5.2 Neoclassical Growth Model

The canonical NCGM can be written as

$$
\begin{aligned}
\max_{\{(c_t, k_{t+1})\}_{t=0}^\infty} \sum_{t=0}^\infty \beta^t u(c_t) \\
\text{s.t. } \forall t, \quad c_t + k_{t+1} \leq \underbrace{F(k_t) + (1-\delta)k_t}_{\equiv f(k_t)}, \\
c_t, k_{t+1} \geq 0; \\
k_0 \text{ given.}
\end{aligned}
\tag{4}
$$

It can also be written many other ways; e.g., in terms of investment:

$$\max_{\{(c_t, i_t)\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t u(c_t)$$
$$\text{s.t. } \forall t, \quad c_t + i_t \leq F(k_t),$$
$$k_{t+1} = (1 - \delta)k_t + i_t,$$
$$c_t, k_{t+1} \geq 0;$$
$$k_0 \text{ given.} \tag{5}$$

There is an art to choosing the simplest formulation. We can turn some inequality constraints into equality constraints with simple arguments (e.g., monotone utility means no consumption will be "thrown away"), and Inada conditions (together with a no-free-lunch production function) can ensure that non-negativity constraints will never bind. Further, in deterministic problems, it is usually best to let the choice space be the state space; here, that means eliminating consumption and investment:

$$\max_{\{k_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t u\big(f(k_t) - k_{t+1}\big)$$
$$\text{s.t. } k_0 \text{ given.} \tag{6}$$

The resulting intereulers

$$u'\big(\underbrace{f(k_t) - k_{t+1}}_{c_t}\big) = \beta f'(k_{t+1}) u'\big(\underbrace{f(k_{t+1}) - k_{t+2}}_{c_{t+1}}\big)$$

have an interpretation almost identical to that of the two-period saving problem, with $f'(k_{t+1})$ playing the role of R: the marginal rate of transformation between c_{t+1} and c_t.

So do these intereulers give us a solution to the original problem? **No**. They give a *second*-order difference equation (in k_t) with only *one* additional condition: the initial condition that k_0 is given. It turns out we need another condition, called the **transversality condition** (TVC), which is sufficient, together with the intereulers, for a solution (see SL Theorem 4.15):

$$\lim_{t \to \infty} \beta^t u'(c_t) f'(k_t) k_t = 0.$$

The intuition for the TVC is that if you consume too little and save too much, k_t and $u'(c_t)$ will grow so fast as to overwhelm the shrinking β^t and $f'(k_t)$. The intuition for the terms is:

1. k_t: Amount of capital,

2. $f'(k_t)$: Conversion from amount of capital to amount of consumption (at marginal—not average—product rate),

3. $u'(c_t)$: Conversion from amount of consumption to felicits at time t,

4. β^t: Conversion from felicits at time t to utils.

A final observation about the steady state of the NCGM. At a steady state $c_t = c_{t+1} = c_*$ and $k_{t+1} = k_*$, so the intereuler reduces to $f'(k_*) = 1/\beta$.

6 Balanced growth practice question

This question comes from the Economics 211 midterm examination in 2006. Its content is straightforward, but successfully completing it requires diligent care—it is very easy to get bogged down.

Question

Consider an economy with a measure one number of identical households. Preferences are given by:

$$\sum_0^\infty \frac{C_t^{1-\sigma}}{1-\sigma}.$$

Households have L units of labor, which they supply inelastically to firms that produce a consumption good using technology:

$$C_t = (u_t K_t)^\alpha L^{1-\alpha},$$

where K_t is physical capital and $u_t \in [0,1]$ is the fraction of the capital stock used to produce consumption goods.

Investment goods are produced according to:

$$I_t = A(1 - u_t)K_t,$$

where $1 - u_t$ represents the share of the capital stock used to produce capital goods.

The stock of capital evolves according to:

$$K_{t+1} = (1-\delta)K_t + I_t.$$

1. Write the problem of a Social Planner and find the Euler equation.

2. For the balanced growth path of this economy, find expressions for the growth rate of K_t, C_t and I_t in terms of the balanced growth path level of $u_t = u^*$.

3. Find the level of u_t along the balanced growth path.

Solution

1. *Write the problem of a Social Planner and find the Euler equation.*

   As we have discussed, there are several ways to write the same problem. Although it is tempting to substitute to get a problem written only in terms of capital, it turns out that here (and in many other problems with lots of choice variables), the notation can wind up getting very nasty. I started down that route, but returned to including more Lagrange multipliers instead. The social planner maximizes

   $$\max_{\{k_{t+1}, c_t, u_t\}_t} \sum_{t=0}^\infty \beta^t \frac{c_t^{1-\sigma}}{1-\sigma} \quad \text{s.t.}$$
   $$c_t = u_t^\alpha k_t^\alpha L^{1-\alpha}, \ \forall t;$$
   $$k_{t+1} = (1 - \delta + A - Au_t)k_t, \ \forall t;$$
   $$u_t \in [0,1], \ \forall t.$$

   With an appeal to an Inada condition on $u(\cdot)$ and the fact that $u_t = 0 \implies c_t = 0$, we will not worry about $u_t \geq 0$ binding. Without argument, we will also ignore $u_t \leq 1$; this could actually be a problem.

   The Lagrangian is

   $$\mathcal{L} = \sum_{t=0}^\infty \left[\beta^t \frac{c_t^{1-\sigma}}{1-\sigma} + \lambda_t (u_t^\alpha k_t^\alpha L^{1-\alpha} - c_t) + \mu_t \big((1 - \delta + A - Au_t)k_t - k_{t+1} \big) \right]$$

   Taking first-order conditions, we get

   $$\lambda_t = \beta^t c_t^{-\sigma};$$

$$\lambda_t \alpha \underbrace{u_t^{\alpha-1} k_t^\alpha L^{1-\alpha}}_{=c_t/u_t} = \mu_t A k_t \implies \mu_t = \frac{\beta^t \alpha c_t^{1-\sigma}}{u_t A k_t}; \text{ and}$$

$$\mu_t = \lambda_{t+1} \alpha \underbrace{u_{t+1}^{\alpha-1} k_{t+1}^\alpha L^{1-\alpha}}_{=c_{t+1}/u_t} + \mu_{t+1}(1 - \delta + A - A u_t) \implies$$

$$\frac{\cancel{\beta^t} \alpha c_t^{1-\sigma}}{u_t A k_t} = \frac{\beta^{t+1} \alpha c_{t+1}^{1-\sigma}}{k_{t+1}} + \frac{\beta^{t+1} \alpha c_{t+1}^{1-\sigma}}{u_{t+1} A k_{t+1}}(1 - \delta + A - A u_{t+1})$$

$$\frac{c_t^{1-\sigma}}{u_t k_t} = \frac{\beta A u_{t+1} c_{t+1}^{1-\sigma}}{u_{t+1} k_{t+1}} + \frac{\beta c_{t+1}^{1-\sigma}}{u_{t+1} k_{t+1}}(1 - \delta + A - A u_{t+1})$$

$$\left(\frac{c_t}{c_{t+1}}\right)^{1-\sigma} = \beta(1 - \delta + A) \frac{u_t k_t}{u_{t+1} k_{t+1}}.$$

Note there are other possible simplifications of this intereuler, based on the fact that $c_t/c_{t+1} = \left(\frac{u_t k_t}{u_{t+1} k_{t+1}}\right)^\alpha$.

2. *For the balanced growth path of this economy, find expressions for the growth rate of K_t, C_t and I_t in terms of the balanced growth path level of $u_t = u^*$.*

   Note that the production technology for the investment good is $i_t \propto A k_t$, which violates Inada conditions:

   $$\lim_{k \to \infty} \frac{\partial i_t}{\partial k_t} = A(1 - u_t) \neq 0.$$

   This can give us endogenous growth, since decreasing marginal returns never kick in.

   Along the balanced growth path, the law of motion for capital gives the growth rate of capital:

   $$\gamma_k \equiv k_{t+1}/k_t = 1 - \delta + A - A u_*.$$

   The investment goods production technology gives us the growth rate of investment:

   $$\gamma_i \equiv \frac{i_{t+1}}{i_t} = \frac{A(1 - u_*)k_{t+1}}{A(1 - u_*)k_t} = \gamma_k.$$

   The consumption goods production technology gives us the growth rate of consumption:

   $$\gamma_c \equiv \frac{c_{t+1}}{c_t} = \frac{(u_* k_{t+1})^\alpha L^{1-\alpha}}{(u_* k_t)^\alpha L^{1-\alpha}} = \left(\frac{k_{t+1}}{k_t}\right)^\alpha = \gamma_k^\alpha.$$

3. *Find the level of u_t along the balanced growth path.*

   Plugging the growth rates we just found, along with $u_t = u_{t+1} = u_*$ into our intereuler gives

   $$\gamma_c^{\sigma-1} = \beta(1 - \delta + A)\gamma_k^{-1}$$

   $$(1 - \delta + A - A u_*)^{1-\alpha(1-\sigma)} = \beta(1 - \delta + A)$$

   $$u_* = 1 + \frac{1 - \delta - [\beta(1 - \delta + A)]^{\frac{1}{1-\alpha+\alpha\sigma}}}{A}.$$

7 Introduction to dynamic programming

We consider "transforming" a sequence problem of the form given in equations 4–6 into a functional equation

$$V(k) = \max_{c,k'} \left[u(c) + \beta V(k') \right]$$
$$\text{s.t. } c + k' \leq f(k), \tag{7}$$
$$c, k' \geq 0.$$

As in the SP, we can recast the choice space to be the state space (also arguing that the inequality constraints do not bind):

$$V(k) = \max_{k' \in (0, f(k))} \left[u\big(f(k) - k'\big) + \beta V(k') \right]. \tag{8}$$

While the solution to the SP (as written in equation 4) is a joint sequence $\{(c_t, k_{t+1})\}_{t=0}^{\infty}$, the solution to the FE is a value function $v(k)$ and a policy function $y^* = g(k)$. Therefore in order to find a solution to the FE problem we need to understand a few things, namely:

- In what space do functions "live"?

- How do we define distance in this space? Convergence?

Once we know how to solve FEs, we have a few remaining questions.

- Is the solution unique? If so, we expect the solution to also solve the SP (this is ensured by the Principle of Optimality, but is not the direction we are interested in.)

- If there are multiple solutions to the FE, which one(s) solve(s) the SP?

Why do we want to solve the FE? We are shifting the problem from studying an infinite sequence to a function. Is this really easier? Analytically the answer is not clear; there are properties of the solution that are proved more easily using the SP formulation than the FE one. However, there is at least one clear advantage to a dynamic programming approach: many problems of interest to macroeconomists cannot be solved analytically. When we need to find numerical solutions, the FE formulation makes things easier.[2]

7.1 Where do functions live? A Metric space!

Definition 2. *A **real vector space** is a set X, with elements $x \in X$, together with two operations, addition and scalar multiplication,[3] satisfying the following axioms for any $x, y, z \in X$ and $\alpha, \beta \in \mathbb{R}$:*

    Axioms for addition

1. *$x + y \in X$ (closure under addition),*

2. *$x + y = y + x$ (commutativity),*

3. *$(x + y) + z = x + (y + z)$ (associativity),*

4. *there is a $\vec{0} \in X$ such that $x + \vec{0} = x$ (identity existence), and*

5. *there is $-x \in X$ such that $x + (-x) = \vec{0}$ (inverse existence).*

    Axioms for scalar multiplication

---

[2]This may not be intuitive. The solution to the SP is a vector in $\mathbb{R}^{\infty}$, while the solution to the latter problem is a function $V : \mathbb{R} \to \mathbb{R}$. The dimensionality of the former object is lower (countably infinite vs. uncountably so) but the function can much more easily be approximated. We do this by somehow restricting its domain, and then determining its value on a discrete grid of points.

[3]Actually there are two operations: left scalar multiplication and right scalar multiplication (with an additional requirement that these two give the same result).

1. $\alpha \cdot x \in X$ *(closure under scalar multiplication)*,

2. $\alpha \cdot x = x \cdot \alpha$ *(commutativity)*,

3. $(\alpha \cdot \beta) \cdot x = \alpha \cdot (\beta \cdot x)$ *(associativity)*,

4. $1 \cdot x = x$ *(identity existence)*, and

5. $0 \cdot x = \vec{0}$.

Distributive laws

1. $\alpha \cdot (x + y) = \alpha \cdot x + \alpha \cdot y$, *and*

2. $(\alpha + \beta) \cdot x = a \cdot x + \beta \cdot x$.

Examples of real vector spaces include:

- Euclidean spaces ($\mathbb{R}^n$),

- The set of real-valued functions $f\colon [a, b] \to \mathbb{R}$, and

- The set of continuous (real-to-real) functions $f\colon [a, b] \to \mathbb{R}$;

- The set of continuous, bounded (real-to-real) functions $f\colon [a, b] \to \mathbb{R}$; we refer to this set as $C[a, b]$.

Definition 3. *A **normed vector space** is a vector space S, together with a norm $\|\cdot\|\colon S \to \mathbb{R}$, such that, for all $x, y \in S$ and $\alpha \in \mathbb{R}$:*

1. $\|x\| = 0$ *if and only if $x = \vec{0}$,*

2. $\|\alpha \cdot x\| = |\alpha| \cdot \|x\|$, *and*

3. $\|x + y\| \le \|x\| + \|y\|$.

As an exercise, you can prove that these assumptions ensure that $\|x\| \ge 0$.
Examples of normed vector spaces include:

- On $\mathbb{R}$, absolute values $\|x\| = |x|$;

- On $\mathbb{R}^n$, the Euclidean norm $\|x\| = \sqrt{\sum_{i=i}^{n} x_i^2}$;

- On $\mathbb{R}^n$, the Manhattan norm $\|x\| = \sum_{i=i}^{n} |x_i|$;

- On $\mathbb{R}^n$, the sup norm $\|x\| = \sup_{i \in \{1, \dots, n\}} |x_i|$ (a proof that this is a norm is given in the appendix to this section); and

- On $C[a, b]$, the sup norm $\|f\| = \sup_{t \in [a, b]} |f(t)|$.

If we have a normed vector space, we can also define a **metric space**, which has a "metric" (also known as a "distance") function $\rho\colon S \times S \to \mathbb{R}$ defined by $\rho(x, y) = \|x - y\|$.

It may be useful for each of the example norms given above to sketch a ball: the locus of points in S less than distance r away from some element $x \in S$

Although every normed vector space can give rise to a metric space, not every metric space can be generated in this way. The general definition of a metric space follows; it is not hard to prove that the properties of a norm imply that $\rho(x, y) = \|x - y\|$ satisfies the properties required of a metric. (A proof is given in the appendix to this section.)

Definition 4. *A **metric space** is a set S, together with a metric $\rho : S \times S \to \mathbb{R}$, such that, for all $x, y, z \in S$:*

1. $\rho(x, y) = 0$ *if and only if $x = y$ (which implies, with the later axioms, that the distance between two distinct elements is strictly positive),*

2. $\rho(x, y) = \rho(y, x)$ *(symmetry), and*

3. $\rho(x, z) \le \rho(x, y) + \rho(y, z)$ *(triangle inequality).*

7.2 Convergence in metric spaces

Definition 5. *Let (S, ρ) be a metric space and $\{x_n\}_{n=0}^{\infty}$ be a sequence in S. We say that $\{x_n\}$ **converges to** $x \in S$, or that the sequence **has limit** x if*

$$\forall \varepsilon > 0, \exists N(\varepsilon) \text{ such that } n > N(\varepsilon) \implies \rho(x_n, x) < \varepsilon.$$

That is, a sequence is convergent if its terms get closer and closer to some point x, to the extent that, given an arbitrary small number $\varepsilon > 0$, we can always find some positive integer $N(\varepsilon)$ such that all terms of the sequence beyond $N(\varepsilon)$ will be no further than ε from x.

Checking directly whether a sequence convergences requires knowing its limit. Most of the time, we don't know the limit. The definition of Cauchy sequences will help us with that.

Definition 6. *A sequence $\{x_n\}_{n=0}^{\infty}$ in a metric space (S, ρ) is a **Cauchy sequence** if*

$$\forall \varepsilon > 0, \exists N(\varepsilon) \text{ such that } \forall m, n > N(\varepsilon), \rho(x_m, x_n) < \varepsilon.$$

It's clear that any converging sequence is also a Cauchy sequence, but the converse is not true: not every Cauchy sequence converges.[4] *However* there is a particular class of metric spaces, called complete metric spaces in which this converse *does* hold.

Definition 7. *A metric space (S, ρ) is **complete** if every Cauchy sequence contained in S converges to some point in S.*

Checking completeness is very difficult. We will take it as given that the real line $\mathbb{R}$ with the metric $\rho(x, y) = |x - y|$ is a complete metric space.

Theorem 8. *Let $X \subseteq \mathbb{R}^n$, and let $C(X)$ be the set of bounded, continuous functions $f \colon X \to \mathbb{R}$ with the sup norm, $\|f\| = \sup_{x \in X} |f(x)|$. Then $C(X)$ is a complete metric space.*

A proof is given in the appendix to this section.

7.3 How will convergence help us?

Consider a metric space S, ρ and a function $f \colon S \to S$. Define a sequence as follows, for some given $s_0 \in S$:

$$s_0, f(s_0), f\big(f(s_0)\big), \dots .$$

Clearly, this sequence can also be written as a recursive sequence $\{s_n\}_{n=0}^{\infty}$ defined by the difference equation $s_{t+1} = f(s_t)$ and the initial condition.

Does this sequence converge? It depends on $f(\cdot)$ and s_0. For example, consider the sequence defined as above for

- $S = \mathbb{R}$, $\rho(x, y) = |x - y|$, and $f(s) = s + 1$: the sequence diverges for any s_0;

- $S = \mathbb{R}$, $\rho(x, y) = |x - y|$, and $f(s) = s^2$: the sequence converges to 1 for $s_0 = \pm 1$, converges to 0 for $s_0 \in (-1, 1)$, and diverges otherwise;

- $S = \mathbb{R}_+$, $\rho(x, y) = |x - y|$, and $f(s) = \sqrt{s}$: the sequence converges for any s_0 (to 0 if $s_0 = 0$ and to 1 otherwise).

---

[4]Consider the metric space $S = (0, 1]$ with the metric derived from the absolute value norm $\rho(x, y) = |x - y|$. (This *is* a metric space, even though S does *not* form a vector space with regular addition and multiplication.) The Cauchy sequence $1, \frac{1}{2}, \frac{1}{3}, \dots$ lies in S but does not converge in S.

Drawing pictures is useful here: graph $f(\cdot)$ and the 45-degree line on the same axes.[5]

Here's the important thing: suppose that for some s_0 the sequence converges to s. Then as long as f is continuous, s **is a fixed point of** f**; i.e.,** $f(s) = s$. Believe it or not, this will help us solve functional equations.

Recall the FE of equation 8:

$$V(k) = \max_{k' \in (0, f(k))} \left[u\big(f(k) - k'\big) + \beta V(k') \right] \quad \text{for all } k.$$

Just as our functions $f(\cdot)$ above mapped real numbers to real numbers, we consider an operator $T(\cdot)$ that maps functions to functions. In particular, let

$$T(v) \equiv \max_{k' \in (0, f(k))} \left[u\big(f(k) - k'\big) + \beta v(k') \right]$$

for any function v. This may take a bit to get your head around, but note that $T \colon \mathbb{R} \to \mathbb{R}$ and $T(v) \colon \mathbb{R} \to \mathbb{R}$. Given this definition of T, we can write our functional equation compactly as $V(k) = [T(V)](k)$ for all k, or even more simply as $V = T(V)$.

That means that V, the solution to our function equation, is a fixed point of the operator T. How can we find such a thing? Our earlier intuition suggests the following strategy:

1. Show the operator T is continuous (we will use this later);

2. Pick some initial function v_0;

3. Consider the sequence $\{v_n\}_{n=0}^{\infty}$ defined by $v_{n+1} = T(v_n)$ and the initial condition;

4. Show that this sequence converges under some distance over function spaces (we use that induced by the sup norm);

5. Find the thing V to which the sequence converges;

6. Note that continuity of T implies V is a fixed point of T and hence solves the functional equation.

How do we do this? If we can show that T is a contraction (perhaps using Blackwell's Theorem) then we are assured continuity, that the resulting sequence is Cauchy (and hence converges by the completeness of the metric space), and that it converges to the unique fixed point of T from any starting v_0. We will often have to use numerical methods to actually find the convergence point V, however.

7.4   Appendix: Additional proofs

Proofs of several results covered in this section follow.

Theorem 9. *If S and $\|\cdot\|$ are a normed vector space, then S and $\rho(x, y) \equiv \|x - y\|$ are a metric space.*

Proof. We must show that only identical elements have zero distance, that the distance function is symmetric, and that it satisfies the triangle inequality.

1. Let S be the vector space, for $x, y \in S$ let $z = (x - y) \in S$ (we can derive this from the axioms on vector spaces). Then $\rho(x, y) = \|x - y\| = \|z\| \geq 0$ by property (1) of the norm. Notice also that we can say that $z = \theta$ if and only if $x = y$, which implies that $\rho(x, y) = 0$ if and only if $x = y$.

2. Again let $z = (x - y) \in S$ and $-z = -(x - y) = (y - x) \in S$. Then $\|x - y\| = \|z\|$ and $\|y - x\| = \|-z\| = |-1|\|z\| = \|z\|$. Hence $\|x - y\| = \|y - x\|$.

---

[5]Another good exercise is to consider over what subset of $\mathbb{R}$ (if any) these functions are contractions. You can check for a set $S \subseteq \mathbb{R}$ (make sure to confirm that $f \colon S \to S$) using intuition and/or the definition of a contractions.

3. We want to show that $\rho(x, z) \leq \rho(x, y) + \rho(y, z)$ is true with our metric, i.e. that $\|x - z\| \leq \|x - y\| + \|y - z\|$.

   Again let $t = x - y \in S$ and $w = y - z \in S$. Then $\|x - z\| = \|x - y + y - z\| = \|t + w\| \leq \|t\| + \|w\| \leq \|x - y\| + \|y - z\|$. $\qquad \square$

Theorem 10. *The sup norm $\|x\| \equiv \sup_{i \in \{1, \ldots, n\}} |x_i|$ over $\mathbb{R}^n$ is a norm.*

Proof. We must show that only the zero element has zero norm, that scalar multiplication can be taken outside the norm, and that it satisfies a triangle inequality.

1. $\|x\| = \sup_i \{|x_i|, i = 1, \ldots, n\} = |x_j| \geq 0$ for some $1 \leq j \leq n$.

2. $\|\alpha \cdot x\| = \sup_i \{|\alpha \cdot x_i|, i = 1, \ldots, n\} = \sup_i \{|\alpha| \cdot |x_i|, i = 1, \ldots, n\} = |\alpha| \cdot \sup_i \{|x_i|, i = 1, \ldots, n\} = |\alpha| \cdot \|x\|$.

3. $\|x + y\| = \sup_i \{|x_i + y_i|, i = 1, \ldots, n\}$ Call the maximal elements x_* and y_*. Then: $\sup_i \{|x_i + y_i|, i = 1, \ldots, n\} = |x_* + y_*| \leq |x_*| + |y_*| \leq \sup_i \{|x_i|, i = 1, \ldots, n\} + \sup_i \{|y_i|, i = 1, \ldots, n\} = \|x\| + \|y\|$. $\qquad \square$

Theorem 11. *Let $X \subseteq \mathbb{R}^n$, and let $C(X)$ be the set of bounded continuous functions $f : X \to \mathbb{R}$ with the sup norm, $\|f\| = \sup_{x \in X} |f(x)|$. Then $C(X)$ is a complete metric space.*

Proof. We take as given that $C(X)$ with the sup norm metric is a metric space. Hence we must show that if $\{f_n\}$ is a Cauchy sequence, there exists $f \in C(X)$ such that

$$\text{for any } \varepsilon \geq 0, \text{ there exists } N(\varepsilon) \text{ such that } \|f_n - f\| \leq \varepsilon \text{ for all } n \geq N(\varepsilon).$$

There are three steps:

1. find f;

2. show that $\{f_n\}$ converges to f in the sup norm;

3. show that $f \in C(X)$, i.e., f is continuous and bounded.

So let's start:

1. We want to find the candidate function f. In what follows, we will indicate a general element of X as x and a particular element as x_0.

   Consider a Cauchy sequence $\{f_n\}$. Fix $x_0 \in X$, then $\{f_n(x_0)\}$ defines a sequence of real numbers. Let's focus on this sequence of real numbers (notice that now we are talking about something different than the Cauchy sequence of functions $\{f_n\}$): by the definition of sup and of sup norm we can say that:

   $$\left| f_n(x_0) - f_m(x_0) \right| \leq \sup_{x \in X} \left| f_n(x) - f_m(x) \right| = \|f_n - f_m\|.$$

   But then we know that the sequence $\{f_n\}$ is Cauchy by hypothesis (this time the sequence of functions), hence $\|f_n - f_m\| \leq \varepsilon$. But then:

   $$\left| f_n(x_0) - f_m(x_0) \right| \leq \sup_{x \in X} \left| f_n(x) - f_m(x) \right| = \|f_n - f_m\| \leq \varepsilon.$$

   Thus the sequence of real numbers $\{f_n(x_0)\}$ is also a Cauchy sequence, and since $\mathbb{R}$ is a complete metric space, it will converge to some limit point $f(x_0)$. Therefore we now have our candidate function $f : X \to \mathbb{R}$.

2. We want to show that the sequence of functions $\{f_n\}$ converges to f in the sup norm, i.e. that $\|f_n - f\| \to 0$ as $n \to \infty$.

   Fix an arbitrary $\varepsilon > 0$ and choose $N(\varepsilon)$ so that, $n, m \geq N(\varepsilon)$ implies $\|f_n - f_m\| \leq \varepsilon/2$ (we know that we can do this since $\{f_n\}$ is Cauchy).

   Now, for any fixed arbitrary $x_0 \in X$ and all $m \geq n \geq N(\varepsilon)$,

$$
\begin{aligned}
\left|f_n(x_0) - f(x_0)\right| &\leq \left|f_n(x_0) - f_m(x_0)\right| + \left|f_m(x_0) - f(x_0)\right| && \text{(by triangle inequality)} \\
&\leq \|f_n - f_m\| + \left|f_m(x_0) - f(x_0)\right| && \text{(by def of sup and sup norm)} \\
&\leq \varepsilon/2 + \left|f_m(x_0) - f(x_0)\right| && \text{(by } \{f_n\} \text{ being Cauchy).}
\end{aligned}
$$

   Since $\{f_m(x_0)\}$ converges to $f(x_0)$ (this is how we constructed $f(x_0)$), then we can choose m for each fixed $x_0 \in X$ so that $|f_m(x_0) - f(x_0)| \leq \varepsilon/2$. Hence we have:

$$
\left|f_n(x_0) - f(x_0)\right| \leq \varepsilon.
$$

   But since the choice of x_0 was arbitrary, this will hold for all $x \in X$, in particular for $\sup_{x \in X}|f_n(x) - f(x)| = \|f_n - f\|$. And since also the choice of ε was arbitrary, then we have obtained that for any $\varepsilon > 0$, $\|f_n - f\| \leq \varepsilon$, for all $n \geq N(\varepsilon)$.

3. Now we want to show that $f \in C(X)$, i.e. that f is bounded and countinuous.

   f is bounded by construction. To prove that f is continuous, we need to show that for every $\varepsilon \geq 0$ and every $x \in X$, there exists $\delta \geq 0$ such that $|f(x) - f(y)| \leq \varepsilon$ if $\|x - y\|_E < \delta$, where $\|x - y\|_E = \sqrt[n]{\sum_{i=1}^{n}|x_i - y_i|^n}$ is the Euclidean norm on $\mathbb{R}^n$.

   Fix arbitrary $\varepsilon > 0$ and $x_0 \in X$. Then choose k such that $\|f - f_k\| < \varepsilon/3$ (we can do this since $\{f_n\}$ converges to f in the sup norm). Then notice that f_k is continuous (by hypothesis the sequence is in $C(X)$, hence it continuous. Therefore there exists δ such that:

$$
\|x_0 - y\|_E < \delta \text{ implies } \left|f_k(x_0) - f_k(y)\right| < \varepsilon/3.
$$

   Then, almost as we did before:

$$
\begin{aligned}
\left|f(x_0) - f(y)\right| &\leq \left|f(x_0) - f_k(y_0)\right| + \left|f_k(x_0) - f_k(y)\right| + \left|f_k(y) - f(y)\right| && \text{(by triangle inequality)} \\
&\leq 2 \cdot \|f - f_k\| + \left|f(x_0) - f_k(y_0)\right| && \text{(by def of sup and sup norm)} \\
&< \varepsilon && \text{(by continuity of } f_k\text{).}
\end{aligned}
$$

   But since the choice of ε and x_0 was arbitrary, this will hold generally, so we have proved our statement. $\qquad\square$

8   Algorithms for solving dynamic programming problems

There are a several different algorithms for solving dynamic programming problems like the canonical

$$
V(k) = \max_{k' \in \Gamma(k)} \left[u(k, k') + \beta V(k')\right].
$$

The two algorithms you are expected to "know" for Economics 210 are "guess and verify" and value (function) iteration. They are discussed below, along with a third algorithm: policy (function) iteration.

8.1 "Guess and verify"

This is an analytical algorithm; the two following algorithms will be numerical.

1. Develop an initial "guess" for the value function, $V_0(\cdot)$, with an appropriately parameterized functional form.

2. Iterate over V_i using

$$V_{i+1}(k) = \max_{k' \in \Gamma(k)} \left[u(k, k') + \beta V_i(k') \right]$$

until V_i and V_{i+1} have the same functional form.

3. Equate the parameters of V_i and V_{i+1} so that the two functions are equal; this is a solution to the FE.

This will only work for some u and Γ, and perhaps only with a very good guess for V_0. If we are lucky, $V_0(k) = 0$ will work.

8.2 Value iteration

1. Develop an initial "guess" for the value function, $V_0(\cdot)$. If the space is finite and discrete, this guess can be represented by a vector of V_0 evaluated at all the values k can take on.

2. Iterate over V_i using

$$V_{i+1}(k) = \max_{k' \in \Gamma(k)} \left[u(k, k') + \beta V_i(k') \right]$$

until V_i and V_{i+1} are sufficiently close to each other (typically measured by their sup norm distance).

3. Conclude that V_{i+1} is an approximate solution to the FE.

For some u and Γ, we may be able to use the contraction mapping theorem to ensure that this process converges for any initial guess, V_0.

8.3 Policy iteration

This algorithm works explicitly with the optimal policy function

$$g(k) = \operatorname*{argmax}_{k' \in \Gamma(k)} \left[u(k, k') + \beta V(k') \right],$$

where $V(\cdot)$ is the solution to the FE.

1. Develop an initial "guess" for the policy function, $g_0(\cdot)$. If the space is finite and discrete, this guess can be represented by a vector of g_0 evaluated at all the values k can take on.

2. Iterate over g_i as follows:

   - Form V_i from g_i by

   $$V_i(k) = u\big(k, g_i(k)\big) + \beta u\big(g_i(k), g_i(g_i(k))\big) + \beta^2 u\big(g_i(g_i(k)), g_i(g_i(g_i(k)))\big) + \cdots$$

   $$= \sum_{t=0}^{\infty} \beta^t u\big(g_i^t(k), g_i^{t+1}(k)\big).$$

   One way to implement this is to approximate V_i using the first T terms of this sum for some large T.[6]

---

[6] The policy function can be expressed as a transition matrix G_i (containing all zeros, except for a single one in each row), in which case $V_i = \sum_t \beta^t G_i^t \mathbf{u}$ for an appropriately formed vector $\mathbf{u}$.

- Form g_{i+1} from V_i by
$$g_{i+1}(k) = \operatorname*{argmax}_{k' \in \Gamma(k)} \big[u(k, k') + \beta V_i(k') \big].$$

Continue iterating until g_i and g_{i+1} are sufficiently close to each other (typically measured by their sup norm distance).

3. Conclude that V_{i+1} is an approximate solution to the FE.

As with value iteration, for some u and Γ we may be able to use the contraction mapping theorem to ensure that this process converges for any initial guess, g_0.

The contraction mapping theorem can be used to ensure that this process converges for any initial guess. This may seem like a more complex algorithm than value iteration, but can in fact be easier to implement.

8.4 Continuous spaces

Solving models with continuous spaces—like all the models we have seen—with numerical methods will typically rely on "discretizing" the state space in some way; that is, approximating the model with one in which the state and choice variables can only take on a finite number of values.

The first important thing is to be smart about setting up this "grid" of discrete values. It often makes sense to choose values whose spacing is logarithmic rather than linear; this allows the grid to be finer in areas where the value or policy function has more curvature.

Another thing to consider is *where* to discretize. For example, let K be the grid of values that we restrict k to taking on. A naïve approach (which you will see taken in the solutions for your first problem set) might conduct value iteration by finding

$$V_{i+1}(k) = \max_{k' \in \Gamma(k) \cap K} \big[u(k, k') + \beta V_i(k') \big]$$

for each $k \in K$ by exhaustively considering all the $k' \in \Gamma(k) \cap K$ to find the maximizer. A more sophisticated approach still only considers $k \in K$, but would allow k' to take on any value. This suggests attempting to iterate with

$$V_{i+1}(k) \overset{?}{=} \max_{k' \in \Gamma(k)} \big[u(k, k') + \beta V_i(k') \big].$$

This proposed approach raises two questions. First, exhaustive search of a continuous choice set is not feasible; how do we solve the maximization when k' can take on a continuum of values? Fortunately, numerical computing tools (e.g., Matlab, Scilab, SciPy) offer many built-in optimization algorithms. Secondly, how can we even evaluate the maximand for $k' \notin K$ when V_i was only defined for values in the grid? There are a number of ways of doing this; the simplest (which also has attractive theoretical properties) is simply to linearly interpolate between adjacent elements of K. Thus the algorithm, which is called *fitted value iteration*, actually iterates using

$$V_{i+1}(k) = \max_{k' \in \Gamma(k)} \big[u(k, k') + \beta \bar{V}_i(k') \big],$$

where $\bar{V}_i$ means the function that linearly interpolates V_i between grid points. There is also a related algorithm called fitted policy iteration.

9 From the steady state towards explicit dynamics

Once we have "solved" a deterministic infinite-horizon dynamic model, typically the first thing we will do is look to characterize its steady state. This is not hard: just find a recursive formulation of the solution (i.e., one or more difference equations pinning it down), and then substitute, for each time-varying value x,

$$x_* \equiv \cdots = x_{t-1} = x_t = x_{t+1} = \cdots.$$

Typically, the requisite recursive formulation will come from the intereuler(s).

Once we have characterized a steady state, we may

- Evaluate comparative statics of the steady state with regard to exogenous parameters,

- Draw conclusions about dynamics following small deviations from the steady state, or

- Explicitly characterize the system's dynamics.

How does this last activity relate to identifying a steady state? In a sense, it does not. In fact, we have already been investigating transition dynamics without solving for a steady state (for example, solving for policy functions by "guessing and verifying" or using numerical algorithms). A different approach recognizes that a major challenge in analytically investigating dynamics comes from the fact that they are nonlinear—consider for example the second-order difference equation

$$u'\big(f(k_t) - k_{t+1}\big) = \beta f'(k_{t+1})u'\big(f(k_{t+1}) - k_{t+2}\big)$$

for arbitrary u and f. One way of making this system more tractable is to consider instead a linear approximation of this difference equation instead:

$$k_{t+2} \approx \alpha_0 + \alpha_1 k_{t+1} + \alpha_2 k_t$$

for some appropriately chosen $\alpha_0, \alpha_1, \alpha_2 \in \mathbb{R}$. We generally conduct such approximations about the steady state (which is usually the only point on $g(\cdot)$ we can find analytically).

There will be more talk of approximating dynamic systems soon. When we get there, several mathematical tools will be essential.

9.1 Taylor approximation (a.k.a. "linearization")

Consider a differentiable real-valued function on (some subset of) Euclidean space, $g \colon \mathbb{R}^n \to \mathbb{R}$. The function can be approximated in the region around some arbitrary point $x_* \in \mathbb{R}^n$ by its tangent hyperplane.

If $g \colon \mathbb{R} \to \mathbb{R}$, this approximation takes the form[7]

$$g(x) \approx g(x_*) + g'(x_*)(x - x_*).$$

If a system evolves according to $x_t = g(x_{t-1})$, we get a "linearized" system that evolves according to $x_t \approx g(x_*) + g'(x_*)(x_{t-1} - x_*)$. If x_* is the system's steady state, $x_* = g(x_*)$; at this point, the approximation is perfect.

The slope $g'(x_*)$ tells us about the speed of convergence of the linearized system: if $g'(x_*) = 0$, then the approximation tells us that $x_t \approx g(x_*) = x_*$ for all x_{t-1}; convergence is instantaneous. In contrast, if $g'(x_*) = 1$, then the approximation tells us that $x_t \approx g(x_*) + x_{t-1} - x_* = x_{t-1}$ for all x_{t-1}, and there is no convergence towards x_* whatsoever.

9.2 Eigenvectors and eigenvalues

An eigenvector $\mathbf{p}$ ($\neq \mathbf{0}$) of a square matrix W has an associated eigenvalue λ if $W\mathbf{p} = \lambda\mathbf{p}$. If you think of left-multiplication by the matrix W as representing some linear transformation in Euclidean space (rotation, reflection, stretching/compression, shear, or any combination of these), W's eigenvectors point in directions that are unchanged by the transformation, while eigenvalues tell us how much these vectors' magnitudes change.

---

[7]If $g \colon \mathbb{R}^n \to \mathbb{R}$, this approximation takes the form $g(x) \approx g(x_*) + \sum_{i=1}^n g_i'(x_*)(x_i - x_{*\,i}) = g(x_*) + \big[\nabla g(x_*)\big] \cdot (x - x_*)$, where $\cdot$ is the vector dot product operator.

How do we find eigenvectors and eigenvalues? Eigenvector $\mathbf{p}$ is associated with eigenvalue λ if

$$W\mathbf{p} = \lambda\mathbf{p}$$
$$W\mathbf{p} - \lambda\mathbf{p} = \mathbf{0}$$
$$(W - \lambda I)\mathbf{p} = \mathbf{0} \tag{9}$$

where I is the identity matrix. To solve for the eigenvalues, note that this equation can be satisfied if and only if $W - \lambda I$ is singular (i.e., non-invertible), or

$$\det[W - \lambda I] = 0.$$

This is called the *characteristic equation*; the left-hand side is a polynomial in λ whose order is the dimension of W. Unfortunately, for polynomials of degree exceeding four, there is no general solution in radicals (per the Abel-Ruffini Theorem). Fortunately, there are other ways to calculate eigenvalues and you are unlikely to need to calculate eigenvalues by hand for a matrix larger than 2×2.[8] If you don't remember the quadratic formula or how to find the determinant of a 2×2 matrix, now would be a good time to remind yourself.

After solving for the eigenvalues, you can find the eigenvectors using $W\mathbf{p} = \lambda\mathbf{p}$. Of course, $\mathbf{p}$ will not be pinned down entirely: if $\mathbf{p}$ is an eigenvector, then $\alpha\mathbf{p}$ is also an eigenvector for any $\alpha \neq 0$. In practice, people sometimes set the first element of each eigenvector equal to 1 (as long as eigenvector does not have a 0 in that entry) and solve for the rest of the vector.[9]

Theorem 12 (Eigen Decomposition Theorem). *Consider a square matrix W; denote its k distinct eigenvectors $\mathbf{p}_1, \ldots, \mathbf{p}_k$ and the associated eigenvalues $\lambda_1, \ldots, \lambda_k$. Let P be the matrix containing the eigenvectors as columns and Λ be the diagonal matrix with the eigenvalues on the diagonal:*

$$P \equiv \begin{bmatrix} | & | & & | \\ \mathbf{p}_1 & \mathbf{p}_2 & \cdots & \mathbf{p}_k \\ | & | & & | \end{bmatrix} \qquad \Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix}.$$

If P is a square matrix, W can be decomposed into

$$W = P\Lambda P^{-1}.$$

Proof.

$$P\Lambda = \begin{bmatrix} | & & | \\ \lambda_1\mathbf{p}_1 & \cdots & \lambda_k\mathbf{p}_k \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ W\mathbf{p}_1 & \cdots & W\mathbf{p}_k \\ | & & | \end{bmatrix} = WP,$$

where the first and third equalities follow from the matrix multiplication algorithm, and the second from the definition of and eigenvector and its associated eigenvalue. Postmultiplying by P^{-1} completes the proof. $\square$

We will use this result extensively to analyze dynamic systems characterized by linear difference equations (where the linearity typically arises through linear approximation around the steady state). Here's how: suppose that a dynamic system evolves according to

$$\underbrace{\begin{bmatrix} k_{t+2} \\ k_{t+1} \end{bmatrix}}_{\equiv \mathbf{x}_{t+1}} = W \underbrace{\begin{bmatrix} k_{t+1} \\ k_t \end{bmatrix}}_{\equiv \mathbf{x}_t}.$$

---

[8]In the 2×2 case, the characteristic equation is

$$\det \begin{bmatrix} w_{11} - \lambda & w_{12} \\ w_{21} & w_{22} - \lambda \end{bmatrix} = 0,$$

or $(w_{11} - \lambda)(w_{22} - \lambda) - w_{21}w_{12} = 0$. Note that there may be zero, one, or two real solutions to this equation.

[9]A complete specification of this normalization is that the first *non-zero* element of each eigenvector must be 1. Other options include requiring that $\sum_i p_i = 1$, or that $\|\mathbf{p}\| \equiv \left(\sum_i p_i^2\right)^{1/2} = 1$.

We therefore have that $\mathbf{x}_t = W^t \mathbf{x}_0$. Off the bat, we might not have much to say about the matrix W raised to a high power. However, using the eigen decomposition of W gives

$$\mathbf{x}_t = (P \Lambda P^{-1})^t \mathbf{x}_0 = P \Lambda \cancel{P^{-1}} \cancel{P} \Lambda \cancel{P^{-1}} \cancel{P} \Lambda \cdots \Lambda P^{-1} \mathbf{x}_0 = P \Lambda^t P^{-1} \mathbf{x}_0,$$

where the structure of Λ makes it very easy to calculate Λ^t:

$$\Lambda^t = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_k \end{bmatrix}^t = \begin{bmatrix} \lambda_1^t & & \\ & \ddots & \\ & & \lambda_k^t \end{bmatrix}.$$

If all the eigenvalues of W have magnitude less than one, then the $\mathbf{x}_t = W^t \mathbf{x}_0 = P \Lambda^t P^{-1} \mathbf{x}_0$ form a convergent process no matter what $\mathbf{x}_0 = [k_1 \quad k_2]'$ we start from—systems like this are sometimes called "sinks." *However*, if W has an eigenvalue with magnitude greater than one (an "exploding" eigenvalue), the system only converges if $\mathbf{x}_0$ is such that the element of $P^{-1} \mathbf{x}_0$ corresponding to the exploding eigenvalue is zero. Such a system is said to be a "source" or have a "saddle path," depending on whether some or all eigenvalues are explosive. More on all this is still to come.

10 Introducing competitive equilibrium

A single agent—typically the "social planner"—has gotten us thus far, but most economic models have multiple agents. For multiple agents' presence to have any real significance, these agents will need to interact with each other; in macroeconomic models, this interaction usually takes place in a market where

- Pricing is linear (e.g., there are no quantity discounts),

- Pricing is anonymous (i.e., the price paid by one buyer is paid by all buyers and the price received by one seller is received by all sellers),

- All agents are price-takers/-makers,

- The price paid is the price received, and

- The quantity sold is the quantity bought.

In investigating a multi-agent model, we will look for *competitive equilibrium*. Since the meaning of a competitive equilibrium can depend on the economic environment, typically the first thing we do is to *define* one. A canonical definition is that a competitive equilibrium comprises **prices** and **quantities** (or allocations) such that

1. All **agents maximize** given the prices (and any other parameters they cannot control) they face,

2. **Markets clear** (i.e., prices paid are prices received and quantities sold are quantities bought), and

3. Physical **resource constraints are satisfied**.

11 Comparing planners' outcomes with competitive equilibria

Consider a set of economies, each of which has a single perishable good, and a mass of agents with utility function $U\left(\{c_t\}_{t=0}^{\infty}\right) = \sum_t \beta^t u(c_t)$, each of whom receives a constant endowment each period equal to y.

11.1 Asset-free economy

This economy has no assets. It therefore does not matter whether we consider a social planner problem or a competitive equilibrium: there are no prices, and the economy's resource constraint is equivalent to the representative agent's budget constraint. The problem of the economy is therefore to

$$\max_{\{c_t\}_t} \sum_t \beta^t u(c_t) \quad \text{s.t. } c_t = y, \ \forall t.$$

The solution is trivial: $c_t^* = y$ for all t.

We also know how to write this problem as something like a functional equation, although the value "function" has no arguments (since there are no state variables):

$$V = u(y) + \beta V.$$

11.2 Partial equilibrium asset market

In this economy, you can borrow or save money in a bank that exists outside the economy and pays/charges gross interest rate R_t (known in the previous period) . The only price in the economy is exogenous, and the economy has no aggregate resource constraint. Thus it again does not matter whether we consider a social planner problem or a competitive equilibrium. The problem of the economy is to

$$\max_{\{c_t, a_{t+1}\}_t} \sum_t \beta^t u(c_t) \quad \text{s.t.}$$

$$c_t + a_{t+1} = y + R_t a_t, \ \forall t;$$

$$a_0 = 0; \text{ and}$$

$$\text{an appropriate TVC.}$$

To solve the model, we consider set up a Lagrangian for the representative agent's problem

$$\mathcal{L} = \sum_t \beta^t u(\underbrace{y + R_t a_t - a_{t+1}}_{c_t})$$

and use the FOCs to generate the intereuler: $u'(c_t) = \beta R_{t+1} u'(c_{t+1})$ for all t.

We can also form a functional equation as long as the interest rates are constant ($R_t = R$ for all t):

$$V(a) = \max_{a'} u(y + Ra - a') + \beta V(a').$$

11.3 General equilibrium asset markets: social planner

In this economy, you can borrow or save money, but *only with other agents inside the economy.* That means that one agent can only borrow when another lends her his savings. Let the gross interest rate (known in the previous period) be R_t, which now arises endogenously to clear the financial market.

We first consider a social planner in this economy. Market clearing means that $a_t = 0$ for all t, and the economy-wide resource constraint means $c_t = y$ for all t. We are effectively back in the asset-free economy.

11.4 General equilibrium sequential asset markets: competitive equilibrium

This economy is almost the same as the last one, but each period t, a competitive market for financial claims determines the gross interest rate R_{t+1} between the current and subsequent periods. A competitive equilibrium comprises allocations $\{c_t, a_{t+1}\}_t$ and prices $\{R_{t+1}\}_t$ such that

- Agents maximize: $\{c_t, a_{t+1}\}_t$ solve

$$\max_{\{c_t, a_{t+1}\}_t} \sum_t \beta^t u(c_t) \quad \text{s.t.}$$

$$c_t + a_{t+1} = y + R_t a_t, \ \forall t;$$

$$a_0 = 0; \text{ and}$$

an appropriate TVC.

This gives the same intereuler as in the partial equilbrium assets market: $u'(c_t) = \beta R_{t+1} u'(c_{t+1})$ for all t. Note that this will *always* be the case; individual agents are price-takers/-makers, so there is no difference (yet) between GE and PE.

- Markets clear: $a_t = 0$ for all t.

- The resource constraint holds: $c_t = y$ for all t.

By either market clearing and the agents' budget constraint, or by the economy's resource constraint, we have that $u'(c_{t+1}) = u'(c_t)$ for all t. Thus the intereuler pins down the equilibrium interest rate $R_t = \beta^{-1}$ for all $t \geq 1$.

Note that the allocations are the same as the social planner's solution under general equilibrium asset markets. This is a direct implication of the First Welfare Theorem.

11.5 General equilibrium Arrow-Debreu asset market: competitive equilibrium

This economy is almost the same as the last one, but instead of trading financial claims each period (one period ahead of the contracted delivery), all trading takes place before the economy starts. In other words, at time -1, all the agents buy and sell promises to deliver units of the consumption good at each future period. Let the price of one consumption unit in period t be p_t. We could think of measuring the number of contracts each agent buys and sells for each period, but it is easier just to measure her consumption, which pins down her contract position (given y).

A competitive equilibrium thus comprises allocations $\{c_t\}_t$ and prices $\{p_t\}_t$ such that

- Agents maximize: $\{c_t\}_t$ solve

$$\max_{\{c_t\}_t} \sum_t \beta^t u(c_t) \quad \text{s.t.}$$

$$\sum_t p_t c_t = \sum_t p_t y; \text{ and}$$

an appropriate TVC.

The Lagrangian

$$\mathcal{L} = \sum_t \left[\beta^t u(c_t) + \lambda(y - c_t) \right]$$

gives FOCs $\beta^t u'(c_t) = \lambda p_t$ for all t.

- Markets clear: $y - c_t = 0$ for all t.

- The resource constraint holds: $c_t = y$ for all t.

By either market clearing or by the economy's resource constraint, we have that $u'(c_t) = u'(y)$ for all t. Thus the intereuler implies that $p_t \equiv \beta^t \rho$ for all t, where $\rho \equiv u'(y)/\lambda$. (This just gives our one allowable normalization in prices.)

What are equilibrium interest rates? They are $R_t = p_t/p_{t-1} = \beta$, just as we would expect. Again, the allocations are the same as the social planner's solution under general equilibrium asset markets.

12 Pareto Optimality

Definition 13. *An allocation is **pareto optimal** if it is:*

1. *Feasible: the sum of consumptions is less than or equal to the total endowment; and*

2. *Pareto: it is not possible to make any person better off without making at least one other person worse off.*

The sections below consider two ways of modelling an economy: as a "Pareto problem," and as a social planner problem. We argue that the solutions are the same, and further note that the First Welfare Theorem ensures that any competitive equilibrium can also be found as the solution to a Pareto or social planner problem. We will not actually have too much more to say in this class about the Fundamental Welfare Theorems,[10] but there is further treatment in the general equilibrium section of Economics 202/202N.

Keep in mind that these results (as usual) rely on the concavity of the utility function.

12.1 A simple Pareto problem

There is a single period in which a quantity Y of a consumption good must be shared between $I > 2$ agents, each with a utility function that satisfies the Inada conditions. The Pareto problem can be specified as:

$$\max_{\{c_i\}_{i=1}^I} u_1(c_1) \quad \text{s.t.}$$
$$u_i(c_i) \geq u_i^*, \; \forall i \geq 2;$$
$$c_i \geq 0, \; \forall i \geq 1;$$
$$\sum_{i=1}^I c_i \leq Y.$$

The first set of constraints can be thought of as "promise-keeping" constraints: the Pareto optimizer seeks to maximize the utility of agent one conditional on promises that she has made to deliver utility of at least u_i^* to each other agent $i \geq 2$.

At the optimum, we will clearly have

$$c_i^* \equiv c_i^*(u_i^*) = \inf_x \{x | u_i(x) \geq u_i^*\}$$

for $i \geq 2$; that is, c_i^* is the minimum level of consumption that i needs to achieve utility u_i^*. By varying $\{c_i^*\}_{i \geq 2}$ and solving the above optimization, we can identify the full set of Pareto Optimal outcomes. Note that under these assumptions, since utility is increasing, it must always be the case that $\sum_{i=1}^I c_i^* = Y$, so that knowing $\{c_i^*\}_{i \geq 2}$ allows us to back out the optimal $c_1^* = Y - \sum_{i \geq 2} c_i^*$.

Setting up a Lagrangian (and ignoring non-negativity constraints as usual) gives

$$\mathcal{L} \equiv u_1(c_1) + \left[\sum_{i=2}^I \theta_i \big[u_i(c_i) - u_i^*\big]\right] + \omega\left[Y - \sum_{i=1}^I c_i\right].$$

Note that the problem is well behaved: the choice set is convex, and the objective function is concave and differentiable. Thus the FOCs characterize the solution. Taking FOCs,

$$u_1'(c_1^*) = \omega,$$
$$\theta_i u_i'(c_i^*) = \omega, \; \forall i \geq 2.$$

---

[10]The Second Welfare Theorem ensures that any solution to a Pareto or social planner problem can be "supported" as a competitive equilibrium.

Thus

$$\frac{u_1'(c_1^*)}{u_i'(c_i^*)} = \theta_i$$

for all $i \geq 2$ (or for all i, if we define an extra parameter $\theta_1 \equiv 1$).

12.2 A simple social planner problem

As above, there is a single period in which a quantity Y of a consumption good must be shared between $I > 2$ agents, each with a utility function that satisfies the Inada conditions. The social planner problem is to maximize the weighted sum of the utility of all agents. That is, the social planner wants to

$$\max_{\{c_i\}_{i=1}^I} \sum_{i=1}^I \lambda_i u_i(c_i) \quad \text{s.t.}$$
$$c_i \geq 0, \forall i;$$
$$\sum_{i=1}^I c_i \leq Y,$$

where the λs are the weights—exogenous to the problem—that the social planner places on each agent. Note that only the ratios of the λs matters (we could double the value of each and leave the problem unchanged), so without loss of generality we can normalize $\lambda_1 = 1$.

Setting up a Lagrangian (and ignoring non-negativity constraints as usual) gives

$$\mathcal{L} \equiv \left[\sum_{i=1}^I \lambda_i u_i(c_i) \right] + \gamma \left[Y - \sum_{i=1}^I c_i \right].$$

As in the Pareto problem, this is well behaved (the choice set is convex, and the objective function is concave and differentiable) so the FOCs characterize the solution:

$$\lambda_i u_i'(c_i^*) = \gamma, \forall i.$$

These imply that

$$\frac{u_1'(c_1^*)}{u_i'(c_i^*)} = 1/\lambda_i$$

for all i.

Thus we can achieve any Pareto optimal allocation in the social planner problem by choosing appropriate weights: $\lambda_i = 1/\theta_i$. There is an equivalence between the two problems.

13 Competitive equilibrium practice question

The question that follows was the longest question on the Fall, 2006 midterm. Per the question's point value, you might have expected to spend about 36 minutes on it. It did not test any dynamic programing or tricky math, but it was long and required that you correctly answer the first part of the question before going on the second parts. This problem tests your ability to take FOCs, interpret the budget constraint, and do some algebraic manipulation. It also tests something you have already been asked to do on your problem set and in interpreting Nir's notes: read between the lines to fully specify a model, some details of which may have been left out.

Question

Consider the neoclassical growth model with endogenous hours and government spending. That is, the representative agent maximizes

$$\sum_{t=0}^{\infty} \beta^t U(C_t, N_t)$$

$$\text{s.t.} \quad C_t + K_{t+1} - (1-\delta)K_t + g_t = W_t N_t + R_t K_t$$

and the representative firm produces output according to

$$Y_t = F(K_t, N_t).$$

The functions U and F are strictly increasing in each argument, strictly concave, differentiable and they satisfy the Inada conditions. We also assume that the corresponding TVC holds.

1. Assume that

$$U(C_t, N_t) = \log(C_t) - \frac{N_t^{1+\chi}}{1+\chi}$$

   and that the production function is Cobb-Douglas

$$F(K_t, N_t) = K_t^\alpha N_t^{1-\alpha}.$$

   (a) Derive the FOCs of the hh and the firm and explain each of the four equations you get (two for the hh and two for the firm).

   (b) Define a competitive equilibrium.

   (c) Is this allocation Pareto-optimal? Give a short intuitive argument.

   (d) Describe the steady state of the economy. Hint: you will not be able to find closed form solutions for the endogenous variables. Use the intereuler condition to pin down the capital-labor ratio, which you can define as X. Use X to simplify the intraeuler condition and the resource constraint of the economy.

   (e) Consider now a permanent increase in g_t. In the new steady state (i.e., ignoring transition dynamics), what is the effect of this change on the allocations in the economy (N, K, C, and Y)?

2. Now, assume that the momentary utility function is given by

$$U(C_t, N_t) = \log\left(C_t - \frac{N_t^{1+\chi}}{1+\chi}\right).$$

   (a) Derive the FOCs for this economy (two equations for the hh; the firm problem did not change).

   (b) Consider again a permanent increase in g_t with this new utility function. In the new steady state (i.e., ignoring transition dynamics), what is the effect of this change on the allocations in the economy (N, K, C, and Y)?

   (c) Compare the effect of changing g for the two cases with different utility functions. Provide intuition for your result.

Solution

The first thing to do is figure out what's going on in this economy. Common questions include,

1. Who owns capital?

2. What are consumers' sources and uses of income?

3. What are firms' expenses?

4. What happens to firms' profit (i.e., who owns firms)?

5. In what units are prices measured?

6. What is the physical resource constraint of the economy?

Answers are not always given explicitly in the model, so you may need to use your judgment. In this model, the answers to these questions are:

1. Households.

2. Sources: wages, rental income of capital, (depreciated) capital stock. Uses: consumption, future capital stock, tax.

3. Wages and capital rental.

4. It's not clear who owns firms—perhaps someone outside the economy? However, it will turn out that the form of the production function ensures there are no profits. (This is not a coincidence; we will often see this.)

5. Consumption goods. An equivalent way to state this is that the price of consumption is normalized to one.

6. $C_t + g_t + K_{t+1} = Y_t + (1-\delta)K_t$.

We now proceed with a solution. Note that the discussion below goes into significantly more depth of explanation than would be expected on an examination. If you were short on time, you could still could do much of this problem quickly. You should be able to take the FOCs (1a), define a competitive equilibrium (1b), have a quick discussion of Pareto Optimally (1c), define a steady state and evaluate the FOCs at a steady state (1d), take the new FOCs (2a), and mention/define income effects (2c).

1. (a) You should take the FOCs correctly. The rest of the problem relies on them, so you need to get them correct. As Nir wrote in his solution set,

   "Make sure not to lose points on the FOCs! This problem was certainly not easy in the later parts, but very standard in its general setup. You should not lose any points and/or time on questions like part 1a, 1b and 2a. Those alone gave you half the points for this question."

   As for "explaining" them, most of the explanations are not interesting ("you need to balance marginal this against marginal that, discounted by the interest rate and β"). The intuition should be straightforward for each FOC, and you might consider spending most of your time on the math rather than writing out long explanations of the FOCs. In fact, on Nir's solutions, he didn't explain the FOCs at all. Do write something (something correct!), but you need not write more than a sentence or two here.

   Firm Problem   We have to set up the firms' problem, which is not given explicitly in the problem; this is the first of several things that we will just have to use our judgment about. Looking at the households' budget constraints suggests that

   - Households own capital, which they rent to firms at a price of R_t,
   - Firms hire workers at a wage of W_t, and
   - The "units" in which these prices are measured are consumption units.

27

Thus a firm's revenue is the quantity it produces, Y_t (since the sale price is one), and its cost is $R_t K_t + W_t N_t$. Thus it solves

$$\max_{K_t, N_t} K_t^\alpha N_t^{1-\alpha} - R_t K_t - W_t N_t.$$

We have not included any non-negativity constraints, justified with an appeal to F's satisfaction of Inada conditions. The FOCs are below:

$$[\text{firm, labor}] \ W_t = (1-\alpha)K_t^\alpha N_t^{-\alpha}.$$

The firm should be willing to hire more workers if the wage is less than the worker's marginal product $(W_t < (1-\alpha)K_t^\alpha N_t^{-\alpha})$ and should want to fire workers otherwise. Hence in equilibrium, $W_t = (1-\alpha)K_t^\alpha N_t^{-\alpha}$.

$$[\text{firm, capital}] \ R_t = \alpha K_t^{\alpha-1} N_t^{1-\alpha}.$$

The firm should be able to rent more capital if the rental price is less than the marginal product of capital and should want to get rid of capital ("rent it to the market") otherwise. Hence in equilibrium, $R_t = \alpha K_t^{\alpha-1} N_t^{1-\alpha}$.

Household problem Each household seeks to

$$\max_{\{K_{t+1}, N_t\}_t} \sum_{t=0}^\infty \beta^t \left[\log\Big(\underbrace{W_t N_t + R_t K_t - K_{t+1} - g_t + (1-\delta)K_t}_{C_t} \Big) - \frac{N_t^{1+\chi}}{1+\chi} \right]$$

given K_0 (and again ignoring non-negativity) or, expanding around a given point,

$$\cdots + \beta^t \left[\log\Big(\underbrace{W_t N_t + R_t K_t - K_{t+1} - g_t + (1-\delta)K_t}_{C_t} \Big) - \frac{N_t^{1+\chi}}{1+\chi} \right] +$$

$$\beta^{t+1} \left[\log\Big(\underbrace{W_{t+1} N_{t+1} + R_{t+1} K_{t+1} - K_{t+2} - g_{t+1} + (1-\delta)K_{t+1}}_{C_{t+1}} \Big) - \frac{N_{t+1}^{1+\chi}}{1+\chi} \right] + \cdots.$$

Taking FOCs, we get that

$$[\text{hh, capital}] \ \frac{1}{C_t} = \beta \frac{1}{C_{t+1}} \big(R_{t+1} + (1-\delta) \big).$$

This is the standard intertemporal Euler equation (intereuler). The LHS is the marginal utility of consumption today; the RHS is the marginal utility of consumption tomorrow, discounted by β and multiplied by the effective interest rate, $R_{t+1} + (1-\delta)$, which can be thought of as the marginal rate of transformation between consumption today and consumption tomorrow.

$$[\text{hh, labor}] \ N_t^\chi C_t = W_t$$

$$N_t^\chi = W_t \frac{1}{C_t}$$

The is the standard within-period Euler equation (intraeuler). Working more gets you W_t which, multiplied by the marginal utility of consumption, $\frac{1}{C_t}$, gives you the utility gain from working a bit more. The household needs to balance this against the marginal disutilty from working, N_t^χ.

(b) A competitive equilibrium is prices $\{W_t, R_t\}_{t=0}^\infty$ and allocations $\{C_t, N_t^s, K_t^s, N_t^d, K_t^d, Y_t\}_{t=0}^\infty$ such that

   i. Given prices $\{W_t, R_t\}_{t=0}^{\infty}$, allocations $\{C_t, N_t^s, K_{t+1}^s\}_{t=0}^{\infty}$ solve the household problem;

   ii. Given prices $\{W_t, R_t\}_{t=0}^{\infty}$, allocations $\{N_t^d, K_t^d\}_{t=0}^{\infty}$ solve the firm problem;

   iii. Markets clear: $N_t^s = N_t^d \equiv N_t$ and $K_t^s = K_t^d \equiv K_t$ for all t; and

   iv. The resource constraint is satisfied: $C_t + g_t + K_{t+1} = Y_t + (1 - \delta)K_t$ for all t.

(c) There are no externalities and no market failures, So a CE is Pareto Optimal by the First Fundamental Welfare Theorem. We could also formally show that the FOC of the social planner is the same, but since Nir asked for a "short intuitive" answer, this is beyond the scope of the question. Note that reducing government spending is *not* a feasible Pareto Improvement, as g is an exogenous parameter here. You lost points if you said that reducing g was a Pareto Improvement. That argument, that reducing g makes everyone better off, would be analogous to saying that we can have a Pareto Improvement by lowering the depreciation rate δ.

(d) Recall the intereuler

$$\frac{1}{C_t} = \beta \frac{1}{C_{t+1}} \left(R_{t+1} + (1 - \delta) \right).$$

In the steady state, $C_t = C_{t+1}$, so

$$\frac{1}{\beta} - (1 - \delta) = R_* = \alpha \left(\frac{K_*}{N_*} \right)^{\alpha-1}$$

(where the second equality comes from the firm's FOC with respect to capital), or

$$X_* \equiv \frac{K_*}{N_*} = \left(\frac{1 - \beta(1 - \delta)}{\alpha\beta} \right)^{\frac{1}{\alpha-1}}.$$

Thus the capital/labor ratio X is pinned down by β, δ, and α; this pins down $R_* = \alpha X_*^{\alpha-1}$. From the intraeuler and the firm's FOC with respect to labor, we have that

$$C_* N_*^{\chi} = W_* = (1 - \alpha)X_*^{\alpha}.$$

From the economy's resource constraint, at steady state

$$\begin{aligned} C_* + g_* + \ K_* &= K_*^{\alpha} N_*^{1-\alpha} + (1 - \delta)K_* \\ C_* + g_* + \delta K_* &= K_*^{\alpha} N_*^{1-\alpha} \\ &= X_*^{\alpha} N_*. \end{aligned}$$

Substituting in $K_* = X_* N_*$ and $C_* = (1 - \alpha)X_*^{\alpha} N_*^{-\chi}$ gives that

$$(1 - \alpha)X_*^{\alpha} N_*^{-\chi} + g_* + \delta X_* N_* = X_*^{\alpha} N_*.$$

This pins down N_*, from which we can also solve K_* and C_*.

(e) Suppose now that g_* increases. Note that the intereuler does not change; the capital/labor ratio X_* remains fixed. Recall from the last part that

$$g_* = (X_*^{\alpha} - \delta X_*)N_* - (1 - \alpha)X_*^{\alpha} N_*^{-\chi}.$$

If g_* increases,

- Labor N_* increases as well since the right-hand side of the equation above is increasing in N_* (check this![11]),

---

[11] There are two terms on the right-hand side. Notice the second term $(-(1 - \alpha)X_*^{\alpha} N_*^{-\chi})$ is increasing in N_*, and is negative. Since the left-hand side of the equation is positive, the right-hand side must be as well, so its first term $((X_*^{\alpha} - \delta X_*)N_*)$ must be positive, and therefore increasing in N_*.

- Capital $K_* = X_* N_*$ increases (by the same proportion as labor),
- Output $Y_* = K_*^\alpha N_*^{1-\alpha}$ increases (by the same proportion as labor and capital), and
- Consumption $C_* \propto N_*^{-\chi}$ decreases.

We can think of this as the (long-run) effect of an ongoing lump-sum tax.

2. (a) Here, the household's problem is

$$\max_{\{K_{t+1}, N_t\}_t} \sum_{t=0}^{\infty} \beta^t \log\left(\underbrace{W_t N_t + R_t K_t - K_{t+1} - g_t + (1-\delta)K_t}_{C_t} - \frac{N_t^{1+\chi}}{1+\chi} \right)$$

or, expanding around a given point

$$\cdots + \beta^t \log\left(\underbrace{W_t N_t + R_t K_t - K_{t+1} - g_t + (1-\delta)K_t}_{C_t} - \frac{N_t^{1+\chi}}{1+\chi} \right) +$$

$$\beta^{t+1} \log\left(\underbrace{W_{t+1} N_{t+1} + r_{t+1} K_{t+1} - K_{t+2} - g_{t+1} + (1-\delta)K_{t+1}}_{C_{t+1}} - \frac{N_{t+1}^{1+\chi}}{1+\chi} \right) + \cdots .$$

With this utility function the FOC are

$$[\text{hh, capital}] \quad \left(C_t - \frac{N_t^{1+\chi}}{1+\chi} \right)^{-1} = \beta \left(C_{t+1} - \frac{N_{t+1}^{1+\chi}}{1+\chi} \right)^{-1} \left(R_{t+1} + (1-\delta) \right)$$

and

$$[\text{hh, labor}] \quad N_t^\chi = W_t.$$

(b) The intereuler here gives the same steady-state result as with the other utility function: $\beta^{-1} - (1-\delta) = R_*$. Since the firm's FOCs are the same, we get exactly the same capital/labor ratio:

$$X_* \equiv \frac{K_*}{N_*} = \left(\frac{1 - \beta(1-\delta)}{\alpha\beta} \right)^{\frac{1}{\alpha-1}}.$$

From the intraeuler and the firm's FOC with respect to labor, we have that

$$N_*^\chi = W_* = (1-\alpha)X_*^\alpha.$$

If g_* increases,
- Labor N_* remains fixed, since X_* does,
- Capital $K_* = X_* N_*$ remains fixed,
- Output $Y_* = K_*^\alpha N_*^{1-\alpha}$ remains fixed, and
- Consumption $C_* = K_*^\alpha N_*^{1-\alpha} - \delta K_* - g_*$ (from the economy's resource constraint) falls by the same amount that government expenditures increase.

(c) Can we say something about the fall in C_* in the two models? Yes! Since output increased in the economy with income effects (the first economy) but remained constant in the second, the same change in g_* will result in a smaller consumption decline in the first model.[12] In the model with income effects, the influence of the lump sum tax on consumption is partly offset by the increasing supply of capital and labor. Hence for every dollar the government taxes, consumption falls by less

---

[12]To be more precise, we should consider not output, but $Y_* - \delta K_*$. However, since Y_* and K_* increase by the same fraction, the difference must increase as long as it was positive to begin with.

than one dollar in the first example. In the second example, there are no income effects. Having the government tax a household one dollar causes consumption to fall by one dollar.

Note that the first model has income effects in the sense that a change in consumption changes the within-period problem, illustrated by the intraeuler $C_t N_t^\chi = W_t$. Increasing C_t must make N_t fall; hence hours depend on income and there is an income effect.

In the second model, in contrast, increasing consumption does not change the household's within-period problem, per the intraeuler $N_t^\chi = W_t$. Given a saving rate, the household must each period maximize

$$\log\left(C_t - \frac{N_t^{1+\chi}}{1+\chi}\right)$$

which is the same as simply maximizing $C_t - \frac{N_t^{1+\chi}}{1+\chi}$. Since the within-period decision is quasi-linear in consumption, there are no income effects.

14 Constant returns to scale production

In the question above, we noted that firms earned no profits. This was ensured by the Cobb-Douglas production function, and in particular the fact that it demonstrates constant returns to scale (CRS, or constant RTS).

Definition 14. *A production function f exhibits **constant returns to scale** if*

$$f(\alpha k, \alpha n) = \alpha f(k, n)$$

for all $\alpha > 0$

One reason that CRS technologies are attractive for economic models is that if firms exhibit increasing RTS ($f(\alpha k, \alpha n) > \alpha f(k, n)$ for $\alpha > 1$), multiple firms will always want to combine and if they exhibit decreasing RTS ($f(\alpha k, \alpha n) < \alpha f(k, n)$ for $\alpha > 1$), firms will always want to split into a multiplicity of smaller firms. It is also the case that at a competitive equilibrium, firms with CRS production functions will earn zero profits, ensuring there is no incentive for firms to enter or exit the market.

Claim 15. *If $f(\alpha k, \alpha n) = \alpha f(k, n)$ for all $\alpha > 0$, then*

$$\max_{k,n} f(k, n) - rk - wn = 0.$$

Proof. Consider the definition of CRS:

$$\alpha f(k, n) = f(\alpha k, \alpha n).$$

Differentiating with respect to α,

$$f(k, n) = \frac{\partial f}{\partial k}(\alpha k, \alpha n) \cdot k + \frac{\partial f}{\partial n}(\alpha k, \alpha n) \cdot n.$$

Since this must hold for all $\alpha > 0$, k, and n, we can plug in $\alpha = 1$, $k = k_*$, and $n = n_*$:

$$f(k_*, n_*) = \frac{\partial f}{\partial k}(k_*, n_*) \cdot k_* + \frac{\partial f}{\partial n}(k_*, n_*) \cdot n_*.$$

By the first-order conditions for the maximization problem, $\frac{\partial f}{\partial k}(k_*, n_*) = r$ and $\frac{\partial f}{\partial n}(k_*, n_*) = w$, so

$$f(k_*, n_*) = rk_* + wn_*$$

$$\underbrace{f(k_*, n_*) - rk_* + wn_*}_{=\max_{k,n} f(k,n) - rk - wn} = 0. \qquad \square$$

A production function demonstrates CRS if and only if it is homogeneous of degree one, where a function f is said to be homogeneous of degree k if

$$f(\alpha \mathbf{x}) = \alpha^k f(\mathbf{x})$$

where $\mathbf{x}$ may be multidimensional. Using techniques very similar to those above, we could prove

Theorem 16 (Euler's Law). *If f is homogeneous of degree k,*

$$\nabla f(\mathbf{x}) \cdot \mathbf{x} = k f(\mathbf{x})$$

for all $\mathbf{x}$.

The following result is often cited as a corollary of Euler's Law:

Theorem 17. *If f is homogeneous of degree one, then ∇f is homogeneous of degree zero.*

Proof. Homogeneity of degree one means that

$$f(\lambda p) = \lambda f(p).$$

Taking the derivative with respect to p,

$$\lambda \nabla f(\lambda p) = \lambda \nabla f(p)$$
$$\nabla f(\lambda p) = \nabla f(p). \qquad \square$$

15 Continuous-time optimization

Much of these notes comprise (what I hope is) a slightly clearer version of the lecture notes' treatment of the same subject.

The key notation we will use in the following is that "dotted" variables represent derivatives with respect to time ($\dot{x} \equiv \partial x / \partial t$).

15.1 Finite-horizon

Consider the following maximization problem:

$$V(0) = \max_{c\colon [0,T] \to \mathbb{R}} \int_0^T v\big(k(t), c(t), t\big)\, dt \quad \text{s.t.}$$
$$\dot{k} \le g\big(k(t), c(t), t\big), \forall t;$$
$$k(0) = k_0 > 0 \text{ given; and}$$
$$\text{a no-Ponzi condition.}$$

Note that capital is a state variable and consumption is a control. The first constraint is is the "transition equation" or "equation of motion" for the state variable k. The no-Ponzi condition ensures that capital at the end of time, $k(T)$, cannot be "too negative."[13]

First, imagine setting up a Lagrangian

$$\mathcal{L} = \int_0^T v\big(k(t), c(t), t\big)\, dt + \int_0^T \mu(t) \big[g\big(k(t), c(t), t\big) - \dot{k}\big]\, dt + \gamma k(T) e^{-T\bar{r}(T)}. \tag{10}$$

---

[13]In this problem, the no-Ponzi condition takes the form $k(T) \exp\big(-T\bar{r}(T)\big) \ge 0$, where $\bar{r}(t)$ is the average interest rate from time 0 to t. This implies that assets at time T are weakly positive.

When we have a countable number of constraints, we multiply each by a Lagrange multiplier and then add these; here the transition equation gives a continuum of constraints, each of which we multiply by a Lagrange multiplier $\mu(t)$ and then integrate across these.

If we try to solve this Lagrangian using typical methods, we face difficulty due to the presence of $\dot{k}$; thus we attempt to eliminate it. Note that

$$\int_0^T \dot{\mu}k(t) + \dot{k}\mu(t)\, dt = \int_0^T \frac{d}{dt}\big(\mu(t)k(t)\big)\, dt = k(T)\mu(T) - k(0)\mu(0),$$

where the first equality follows from the product rule ($\frac{d}{dt}(\mu k) = \dot{\mu}k + \dot{k}\mu$) and the second from the Fundamental Theorem of Calculus. Hence

$$\int_0^T \dot{k}\mu(t)\, dt = k(T)\mu(T) - k(0)\mu(0) - \int_0^T \dot{\mu}k(t)\, dt.$$

We can substitute this into equation 10, giving

$$\mathcal{L} = \int_0^T \underbrace{v\big(k(t), c(t), t\big) + \mu(t)g\big(k(t), c(t), t\big)}_{\equiv \mathcal{H}(k,c,t,\mu)}\, dt + \int_0^T \dot{\mu}k(t)\, dt + k(0)\mu(0) - k(T)\mu(T) + \gamma k(T)e^{-T\bar{r}(T)}.$$

With the Hamiltonian defined as above, this is

$$= \int_0^T \mathcal{H}(k, c, t, \mu) + \dot{\mu}k(t)\, dt + k(0)\mu(0) - k(T)\mu(T) + \gamma k(T)e^{-T\bar{r}(T)}.$$

Finding the optimal paths

In general, maximizing over the choice of functions is difficult. Fortunately, Pontryagin proposed a method that allows us to recast the problem in terms of a single variable, allowing us to use tools we already know to solve the optimization problem.

Consider a path $\overset{*}{c}(t)$ (which pins down $\overset{*}{k}(t)$ per $\dot{k} = g\big(k(t), c(t), t\big)$ given $k(0)$) that solves the maximization problem. That means that for any other feasible path, the objective function must be lower. We can express all other feasible consumption paths as the sum of the optimal path $\overset{*}{c}$ and some "perturbation function":

$$c(t) = \overset{*}{c}(t) + \varepsilon p_c(t).$$

Similarly, we can define $p_k(\cdot)$ satisfying

$$k(t) = \overset{*}{k}(t) + \varepsilon p_k(t).$$

Checking the optimality of $\overset{*}{c}$ is equivalent to confirming that for any perturbation function $p_c(\cdot)$, the objective function is (weakly) maximized at $\varepsilon = 0$. This requires that $d\mathcal{L}/d\varepsilon = 0$ for all functions $p_c(\cdot)$ and $p_k(\cdot)$. Recasting our Lagrangian in terms of ε gives

$$\mathcal{L}(\varepsilon, \dots) = \int_0^T \mathcal{H}\big(k(\cdot, \varepsilon), c(\cdot, \varepsilon), t, \mu\big) + \dot{\mu}k(t, \varepsilon)\, dt + k(0, \varepsilon)\mu(0) - k(T, \varepsilon)\mu(T) + \gamma k(T, \varepsilon)e^{-T\bar{r}(T)}$$

$$\frac{d\mathcal{L}}{d\varepsilon} = \int_0^T \left[\frac{d\mathcal{H}}{d\varepsilon} + \dot{\mu}\frac{\partial k}{\partial \varepsilon}\right] dt + \frac{\partial k(T, \varepsilon)}{\partial \varepsilon}\big[\gamma e^{-T\bar{r}(T)} - \mu(T)\big]$$

Now, notice that the chain rule and definition of the perturbation functions gives

$$\frac{d\mathcal{H}}{d\varepsilon} = \frac{\partial \mathcal{H}}{\partial c} \cdot \frac{\partial c}{\partial \varepsilon} + \frac{\partial \mathcal{H}}{\partial k} \cdot \frac{\partial k}{\partial \varepsilon} = \frac{\partial \mathcal{H}}{\partial c}p_c(t) + \frac{\partial \mathcal{H}}{\partial k}p_k(t)$$

33

and

$$\frac{\partial k(T, \varepsilon)}{\partial \varepsilon} = p_k(T).$$

Substituting in gives

$$\frac{d\mathcal{L}}{d\varepsilon} = \int_0^T \left[\underbrace{\frac{\partial \mathcal{H}}{\partial c} p_c(t)}_{\equiv A} + \underbrace{\left(\frac{\partial \mathcal{H}}{\partial k} + \dot{\mu} \right) p_k(t)}_{\equiv B} \right] dt + \underbrace{p_k(T) \left[\gamma e^{-T\bar{r}(T)} - \mu(T) \right]}_{\equiv C}.$$

Optimality requires that this be zero when evaluated at $\varepsilon = 0$ for all p_c and p_k. For this to hold, one can show that this must hold component-by-component. (In general, it is *not* the case that $A + B + C = 0$ implies A, B, and C are all zero; it is true here, however.) That is, we need

$$\frac{\partial \mathcal{H}}{\partial c} = 0$$
$$\frac{\partial \mathcal{H}}{\partial k} = -\dot{\mu}$$
$$\gamma e^{-T\bar{r}(T)} = \mu(T).$$

Connectively, these are know as "The Maximum Principle." The first-two play the role of first-order conditions, and the third is the no-Ponzi condition.[14]

15.2   Infinite time

When time is infinite, the problem becomes

$$V(0) = \max_{c:\, \mathbb{R}_+ \to \mathbb{R}} \int_{\mathbb{R}} v\big(k(t), c(t), t\big)\, dt \quad \text{s.t.}$$
$$\dot{k} \leq g\big(k(t), c(t), t\big), \forall t;$$
$$k(0) = k_0 > 0 \text{ given; and}$$
$$\text{a no-Ponzi condition.}$$

The no-Ponzi condition is now ensured by a TVC: $\lim_{t \to \infty} k(t) \exp\big(-tr(t)\big) \geq 0$. This means that assets can be negative everywhere, but that they cannot grow more quickly than the interest rate.

There is some discussion in Barro and Sala-I-Martin as to if this TVC is really needed. They report that in some circumstance, you may be able to solve the problem with a relaxed TVC. If you are interested, consult the book.

15.3   The Hamiltonian "cookbook"

15.3.1   One control, one state variable

1. **Construct the Hamiltonian**. Define

$$\mathcal{H}(k, c, t, \mu) \equiv \underbrace{v\big(k(t), c(t), t\big)}_{\text{felicity}} + \underbrace{\mu(t)}_{\text{multiplier}} \cdot \underbrace{g\big(k(t), c(t), t\big)}_{\text{RHS of transition eqn.}}.$$

---

[14]Note that $\mu(t)$ measures the shadow price of capital at time t (in utils). The condition that $\gamma \exp\big(-T\bar{r}(T)\big) = \mu(T)$ means that the (shadow) "cost of the no-Ponzi condition"—given by its multiplier γ discounted back to the present—is equal to the shadow cost of the capital stock available at the end of time. If, for some reason, the no-Ponzi condition did not bind, $\gamma = 0$. But then the agent could use a bit more capital in the last period and still not violate the no-Ponzi condition. For this to be optimal, the value of capital in the last period, $\mu(T)$, must be 0 as well. If the no-Ponzi condition binds, then capital must be costly.

2. **Take FOC for control variable**. Set $\frac{\mathcal{H}}{\partial c} = 0$:

$$\frac{\partial \mathcal{H}(k, c, t, \mu)}{\partial c(t)} = \frac{\partial v\big(k(t), c(t), t\big)}{\partial c(t)} + \mu(t)\frac{\partial g\big(k(t), c(t), t\big)}{\partial c(t)} = 0$$

for all t.

3. **Take FOC for state variable**. Set $\frac{\partial \mathcal{H}}{\partial k} = -\dot{\mu}$:

$$\frac{\partial \mathcal{H}(k, c, t, \mu)}{\partial k(t)} = \frac{\partial v\big(k(t), c(t), t\big)}{\partial k(t)} + \mu(t)\frac{\partial g\big(k(t), c(t), t\big)}{\partial k(t)} = -\dot{\mu}(t)$$

for all t.

4. **Identify TVC**. If $T < \infty$, then the TVC is

$$\mu(T)k(T) = 0;$$

that is, either capital equals 0 in the last period (the constraint binds) or the constraint does not bind and $\mu(T) = 0$.

If time is infinite, the TVC is

$$\lim_{t \to \infty} \mu(t)k(t) = 0.$$

15.3.2 Multiple control or state variables

Now, let there be n control variables and m state variables. That is, we choose $c_1(\cdot), c_2(\cdot), \ldots, c_n(\cdot)$ to maximize

$$V(0) = \max_{\vec{c}(\cdot)} \int_0^T v \underbrace{\big(k_1(t), \ldots, k_m(t), c_1(t), \ldots, c_n(t), t\big)}_{\equiv (\vec{k}(t), \vec{c}(t), t)}\, dt$$

$$\text{s.t.}$$

$$\dot{k}_1(t) \leq g_1\big(k_1(t), \ldots, k_m(t), c_1(t), \ldots, c_n(t), t\big), \forall t;$$

$$\vdots$$

$$\dot{k}_m(t) \leq g_m\big(k_1(t), \ldots, k_m(t), c_1(t), \ldots, c_n(t), t\big), \forall t; \text{ and}$$

$$k_1(0), \ldots, k_m(0) > 0 \text{ given.}$$

1. **Construct the Hamiltonian**. Define

$$\mathcal{H}(\vec{k}, \vec{c}, t, \mu) \equiv v\big(\vec{k}(t), \vec{c}(t), t\big) + \sum_{j=1}^m \mu_j(t)g_j\big(\vec{k}(t), \vec{c}(t), t\big).$$

2. **Take FOCs for control variables**. Set $\frac{\mathcal{H}}{\partial c_i} = 0$:

$$\frac{\partial \mathcal{H}(\vec{k}, \vec{c}, t, \mu)}{\partial c_i(t)} = \frac{\partial v\big(\vec{k}(t), \vec{c}(t), t\big)}{\partial c_i(t)} + \sum_{j=1}^m \mu_j(t)\frac{\partial g_j\big(\vec{k}(t), \vec{c}(t), t\big)}{\partial c_i(t)} = 0$$

for $i \in \{1, \ldots, n\}$ and all t.

3. **Take FOCs for state variables**. Set $\frac{\partial \mathcal{H}}{\partial k_i} = -\dot{\mu}_i$:

$$\frac{\partial \mathcal{H}(\vec{k}, \vec{c}, t, \mu)}{\partial k_i(t)} = \frac{\partial v(\vec{k}(t), \vec{c}(t), t)}{\partial k_i(t)} + \sum_{j=1}^{m} \mu_j(t) \frac{\partial g_j(\vec{k}(t), \vec{c}(t), t)}{\partial k_i(t)} = -\dot{\mu}_i(t)$$

for $i \in \{1, \ldots, m\}$ and all t.

4. **Identify TVCs**. If $T < \infty$, then the TVCs are

$$\mu_i(T) k_i(T) = 0$$

for $i \in \{1, \ldots, m\}$. If time is infinite, then the TVCs take the form

$$\lim_{t \to \infty} \mu_i(t) k_i(t) = 0.$$

15.4   Current-value Hamiltonians

So far we have used what is known as the "present-value Hamiltonian." There is another formulation of the problem called the "current-value Hamiltonian," which is equivalent for the models that we consider in this class. I am not aware of a particular advantage of this approach, but you might want to know that it exists in case you come across a Hamiltonian that seems to have been created differently than you expect.

Consider a model with objectives of the form

$$\int_0^T \underbrace{e^{-\rho t} u\big(k(t), c(t)\big)}_{\equiv v(k(t), c(t), t)} \, dt.$$

The present-value Hamiltonian is

$$\begin{aligned}
\mathcal{H}(k, c, t, \mu) &= v\big(k(t), c(t), t\big) + \mu(t) g\big(k(t), c(t), t\big) \\
&= e^{-\rho t} u\big(k(t), c(t)\big) + \mu(t) g\big(k(t), c(t), t\big).
\end{aligned}$$

But, sometimes it is common instead to proceed as follows. Consider multiplying by $e^{\rho t}$ to get the current-value Hamiltonian:

$$\begin{aligned}
\widetilde{\mathcal{H}}(k, c, t, \mu) &\equiv e^{\rho t} \mathcal{H}(k, c, t, \mu) \\
&= u\big(k(t), c(t)\big) + \underbrace{e^{\rho t} \mu(t)}_{\equiv \lambda(t)} g\big(k(t), c(t), t\big).
\end{aligned}$$

$\lambda(t)$ is the current-value shadow price: it gives the value of a unit of capital at time t measured in time-t utils (i.e., felicits), rather than in time-0 utils.

The Maximum Principle tells us that at an optimum, we have an FOC in the choice variable

$$\frac{\partial \mathcal{H}}{\partial c(t)} = e^{-\rho t} \frac{\partial \widetilde{\mathcal{H}}}{\partial c(t)} = 0 \quad \Longleftrightarrow \quad \frac{\partial \widetilde{\mathcal{H}}}{\partial c(t)} = 0.$$

The FOC in the state variable takes the form

$$\frac{\partial \mathcal{H}}{\partial k(t)} = e^{-\rho t} \frac{\partial \widetilde{\mathcal{H}}}{\partial k(t)} = -\dot{\mu}(t),$$

36

which, since $\mu(t) = e^{-\rho t}\lambda(t)$,

$$= \rho\lambda(t)e^{-\rho t} - \dot{\lambda}(t)e^{-\rho t}$$

$$\frac{\partial\widetilde{\mathcal{H}}}{\partial k(t)} = \rho\lambda(t) - \dot{\lambda}(t).$$

Finally, the TVC is

$$\mu(T)k(T) = e^{-\rho T}\lambda(T)k(T) = 0$$

or

$$\lim_{t \to \infty} \mu(t)k(t) = \lim_{t \to \infty} e^{-\rho t}\lambda(t)k(t) = 0.$$

16   Log-linearization

Recall that we have thus far linearized systems like $g\colon \mathbb{R}^n \to \mathbb{R}$ using the first-order Taylor approximation about the steady state x^*:

$$g(x) \approx g(x^*) + g_1'(x^*)(x_1 - x_1^*) + g_2'(x^*)(x_2 - x_2^*) + \cdots + g_n'(x^*)(x_n - x_n^*)$$
$$= g(x^*) + \left[\nabla g(x^*)\right] \cdot (x - x^*).$$

Linearizing gives an approximation that is linear in $x - x^*$. That is, a one *unit* change in in x causes the approximated of $g(x)$ to increase by a constant $g'(x^*)$. Log-linearization instead gives an approximation for $g(\cdot)$ that is linear in $\hat{x} \equiv \frac{x-x^*}{x^*}$, or x's *percentage* deviation from steady state.

Why would we like to do this? Consider trying to describe the economies of Palo Alto and the United States using a single model. Given the difference in the economy's scales, it makes more sense to draw conclusions about how each would respond to, say, a 5% budget surplus—which might in some sense affect Palo Alto and the U.S. similarly—than to draw conclusions about how each would respond to a billion dollar budget surplus.

We sometimes think in terms of percent movements ("stocks went down 2%") rather than absolute movements ("the Dow dropped by 400 points"). One advantage of the former approach is that it allows us to express our conclusions in unitless measures; they are therefore robust to unit conversion. This advantage also accrues to log-linearization: if x is measured in Euros, then so is $x - x^*$, while $\hat{x} \equiv \frac{x-x^*}{x^*}$ is a unitless measure.

16.1   Why are approximations in terms of $\hat{x} \equiv \frac{x-x^*}{x^*}$ called "log-linear"?

Consider a first-order Taylor approximation of the (natural) log function, the most important Taylor approximation in economics:[15]

$$\log(x) \approx \log(x_*) + \frac{1}{x_*}(x - x_*)$$

$$\log(x) - \log(x_*) \approx \frac{x - x_*}{x_*} \equiv \hat{x}.$$

Thus just as a standard Taylor approximation is linear in $(x - x_*)$, the log-linearization is linear in $\hat{x} \approx \log(x) - \log(x_*)$; that is, it is linear in logarithms.

---

[15]You should memorize it and get used to recognizing it. In particular, you should get used to recognizing when people treat this approximation as if it holds exactly; typically this occurs when the log first-difference of a time series $(\log(y_t) - \log(y_{t-1}))$ is treated as equal to the series' growth rate $(y_t/y_{t-1} - 1)$.

16.2   Log-linearization: first approach

To get a log-linearization, there are many techniques. One is to start with the standard linearized (i.e., Taylor approximated) version and "build up" $\hat{x} \equiv \frac{x - x_*}{x_*}$ and $\widehat{f(x)} \equiv \frac{f(x) - f(x_*)}{f(x_*)}$.[16]

$$f(x) \approx f(x_*) + f'(x_*)(x - x_*)$$
$$f(x) - f(x_*) \approx f'(x_*)(x - x_*)$$
$$\frac{f(x) - f(x_*)}{f(x_*)} \approx \frac{f'(x_*)(x - x_*)}{f(x_*)}$$
$$\approx \frac{f'(x_*)x_*}{f(x_*)} \cdot \frac{(x - x_*)}{x_*}$$
$$\widehat{f(x)} \approx \frac{f'(x_*)x_*}{f(x_*)} \hat{x}$$

Here, we have an expression that describes what happens to the deviation of $f(x)$ from its steady state in percent terms as x deviates from its steady state in percent terms. A 1% increase in x from the steady state causes $f(x)$ to increase by about $\frac{f'(x_*)x_*}{f(x_*)}$ percent.

You might recognize that this last expression has another name: the elasticity of $f(x)$ with respect to x. Perhaps it is easier to see when written as

$$\frac{f'(x_*)x_*}{f(x_*)} = \left(\frac{\partial f}{\partial x} \cdot \frac{x}{f(x)} \right) \bigg|_{x_*}.$$

It is important to get some practice with quickly log-linearizing simple functions. Several examples follow:

- A Cobb-Douglas production function

$$y_t = k_t^\alpha n_t^{1-a}$$
$$y_t \approx y^* + \alpha \frac{k_*^\alpha n_*^{1-a}}{k_*}(k_t - k_*) + (1 - \alpha)\frac{k_*^\alpha n_*^{1-a}}{n_*}(n_t - n_*)$$
$$y_t - y^* \approx \alpha \frac{y^*}{k_*}(k_t - k_*) + (1 - \alpha)\frac{y^*}{n_*}(n_t - n_*)$$
$$\frac{y_t - y^*}{y^*} \approx \alpha \frac{(k_t - k_*)}{k_*} + (1 - \alpha)\frac{(n_t - n_*)}{n_*}$$
$$\hat{y}_t \approx \alpha \hat{k}_t + (1 - \alpha)\hat{n}_t.$$

- An intereuler

$$\frac{1}{c_t} = \beta \frac{1}{c_{t+1}} \alpha z_{t+1} k_{t+1}^{\alpha-1} n_{t+1}^{1-\alpha}$$

where z_{t+1} is fixed. We want to log linearize and find $\hat{c}_{t+1}$ as a function of $\hat{c}_t$, $\hat{k}_{t+1}$, and $\hat{n}_{t+1}$.

---

[16]Our notation here is that $\hat{y} \equiv \frac{y - y_*}{y_*} \approx \log(y) - \log(y_*)$ for any y. Sometimes we will instead use $\hat{y} \equiv \log(y) - \log(y_*) \approx \frac{y - y_*}{y_*}$. Both are standard in Nir's class, but not necessarily generally; in particular, some authors use capitalization to distinguish a percentage/log deviation from steady state (or trend, or whatever else is being linearized around).

Rearranging, we see that

$$c_{t+1} = \beta c_t \alpha z_{t+1} k_{t+1}^{\alpha-1} n_{t+1}^{1-\alpha}$$

$$\approx c_* + \left.\frac{\partial \left(c_t \alpha z_{t+1} k_{t+1}^{\alpha-1} n_{t+1}^{1-\alpha}\right)}{\partial c_t}\right|_* (c_t - c_*) + \left.\frac{\partial \left(c_t \alpha z_{t+1} k_{t+1}^{\alpha-1} n_{t+1}^{1-\alpha}\right)}{\partial k_{t+1}}\right|_* (k_{t+1} - k_*)$$

$$+ \left.\frac{\partial \left(c_t \alpha z_{t+1} k_{t+1}^{\alpha-1} n_{t+1}^{1-\alpha}\right)}{\partial n_{t+1}}\right|_* (n_{t+1} - n_*)$$

$$\approx c_* + \frac{c_*}{c_*}(c_t - c_*) + \frac{c_*}{k_*}(k_{t+1} - k_*) + \frac{c_*}{n_*}(n_{t+1} - n_*)$$

$$c_{t+1} - c_* = \frac{c_*}{c_*}(c_t - c_*) + \frac{c_*}{k_*}(\alpha-1)(k_{t+1} - k_*) + \frac{c_*}{n_*}(1-\alpha)(n_{t+1} - n_*)$$

$$\frac{c_{t+1} - c_*}{c_*} \approx \frac{(c_t - c_*)}{c_*} + (\alpha-1)\frac{(k_{t+1} - k_*)}{k_*} + (1-\alpha)\frac{(n_{t+1} - n_*)}{n_*}$$

$$\hat{c}_{t+1} \approx \hat{c}_t + (\alpha-1)\hat{k}_{t+1} + (1-\alpha)\hat{n}_{t+1}.$$

- The exponential

$$x_t^n \approx x_*^n + n x_*^{n-1}(x_t - x_*)$$

$$\approx x_*^n + x_* n x_*^{n-1}\frac{x_t - x_*}{x_*}$$

$$\approx x_*^n + n x_*^n \hat{x}_t.$$

- Another function

$$e^{x_t + y_t} \approx e^{x_* + y_*} + e^{x_* + y_*}(x_t - x_*) + e^{x_* + y_*}(y_t - y_*)$$

$$\approx e^{x_* + y_*} + x_* e^{x_* + y_*}\frac{x_t - x_*}{x_*} + y_* e^{x_* + y_*}\frac{y_t - y_*}{y_*}$$

$$\approx e^{x_* + y_*}(1 + x_* \cdot \hat{x}_t + y_* \cdot \hat{y}_t).$$

- A sum of functions

$$x_t + y_t \approx x_* + y_* + (x_t - x_*) + (y_t - y_*)$$

$$\approx x_* + y_* + x_* \frac{x_t - x_*}{x_*} + y_* \frac{y_t - y_*}{y_*}$$

$$\approx x_* + y_* + x_* \cdot \hat{x}_t + y_* \cdot \hat{y}_t.$$

16.3   Log-linearization: second approach

This way may or may not be faster; it depends on your preferences and the problem at hand. It is easiest to explain this method by way of example. The idea is that we will re-express the function in terms of logs of the variables, and then take a linear approximation in the logs; then we will have log-linearized.

For simplicity of notation, assume that for any variable y_t, we define $Y_t \equiv \log(y_t)$. Thus we also have $\hat{y}_t \approx Y_t - Y_*$.

Several examples of this technique follow:

- A Cobb-Douglas production function, $y_t = k_t^\alpha n_t^{1-a}$. Let us start with the left-hand side:

$$
\begin{aligned}
y_t &= e^{\log(y_t)} \\
&= e^{Y_t} \\
&\approx e^{Y_*} + \left.\frac{\partial\left(e^{Y_t}\right)}{\partial Y_t}\right|_{Y_t=Y_*} (Y_t - Y_*) \\
&\approx e^{\log(y_*)} + e^{\log(y_*)}\left(Y_t - Y_*\right) \\
&\approx y_* + y_* \cdot \hat{y}_t.
\end{aligned}
$$

We now proceed with the right-hand side

$$
\begin{aligned}
k_t^\alpha n_t^{1-a} &= e^{\alpha \log(k_t) + (1-\alpha)\log(n_t)} \\
&= e^{\alpha K_t + (1-\alpha)N_t} \\
&\approx e^{\alpha K_* + (1-\alpha)N*} + \left.\frac{\partial\left(e^{\alpha K_t + (1-\alpha)N_t}\right)}{\partial K_t}\right|_{K_t=K_*} (K_t - K_*) \\
&\quad + \left.\frac{\partial\left(e^{\alpha K_t + (1-\alpha)N_t}\right)}{\partial N_t}\right|_{N_t=N_*} (N_t - N_*) \\
&\approx y_* + \alpha y_* \cdot \hat{k}_t + (1-\alpha)y_* \cdot \hat{n}_t
\end{aligned}
$$

equating the two sides, we see that

$$
\begin{aligned}
y_* + y_* \cdot \hat{y}_t &\approx y_* + \alpha y_* \cdot \hat{k}_t + (1-\alpha)y_* \cdot \hat{n}_t \\
\hat{y}_t &\approx \alpha \hat{k}_t + (1-\alpha)\hat{n}_t
\end{aligned}
$$

- An intereuler

$$
\frac{1}{c_t} = \beta \frac{1}{c_{t+1}} \alpha z_{t+1} k_{t+1}^{\alpha-1} n_{t+1}^{1-\alpha}
$$

where z_{t+1} and k_{t+1} are fixed. We want to find $\hat{c}_{t+1}$ as a function of $\hat{c}_t$, $\hat{k}_{t+1}$, and $\hat{n}_{t+1}$. Rearranging yields that

$$
c_{t+1} = \beta c_t \alpha z_{t+1} k_{t+1}^{\alpha-1} n_{t+1}^{1-\alpha}
$$

Log-linearizing both the left- and right-hand sides gives

$$
\begin{aligned}
c_* + c_* \cdot \hat{c}_{t+1} &\approx \beta \alpha e^{C_{t+1}+Z_{t+1}+(\alpha-1)K_{t+1}+(1-\alpha)N_{t+1}} \\
&\approx \beta\alpha e^{Z_*+(\alpha-1)K_*+(1-\alpha)N_*} + \beta\alpha e^{C_{t+1}+Z_{t+1}+(\alpha-1)K_{t+1}+(1-\alpha)N_{t+1}} \cdot [C_t - C_*] + \\
&\quad \beta\alpha e^{Z_*+(\alpha-1)K_*+(1-\alpha)N_*} \cdot (\alpha-1)\left[K_{t+1} - K_*\right] + \\
&\quad \beta\alpha e^{Z_*+(\alpha-1)K_*+(1-\alpha)N_*} \cdot (1-\alpha)\left[N_t - N_*\right] \\
&\approx c_* + c_*\hat{c}_t + c_*(\alpha-1)\hat{k}_{t+1} + c_*(1-\alpha)\hat{n}_t \\
\hat{c}_{t+1} &\approx \hat{c}_t + (\alpha-1)\hat{k}_{t+1} + (1-\alpha)\hat{n}_t
\end{aligned}
$$

17 Log-linearizing the NCGM in continuous time

We seek to maximize

$$
\int_0^\infty e^{-\rho t} U(c_t)\, dt
$$

such that

$$\dot{k}_t = w_t n_t + r_t k_t - \delta k_t - c_t. \tag{11}$$

We set up the Hamiltonian as

$$\mathcal{H}_t \equiv e^{-\rho t} U(c_t) + \mu_t (w_t n_t + r_t k_t - \delta k_t - c_t).$$

Note that there is no disutility associated with working, so they will be chosen to be $n = 1$ (there is no FOC in n_t). Per the Maximum Principle, the correct FOCs are:

$$\frac{\partial \mathcal{H}}{\partial c_t} = 0 \qquad \Longleftrightarrow \qquad e^{-\rho t} U'(c_t) = \mu_t; \text{ and} \tag{12}$$

$$\frac{\partial \mathcal{H}}{\partial k_t} = -\frac{d}{dt}\mu_t \qquad \Longleftrightarrow \qquad \mu_t(r_t - \delta_t) = -\dot{\mu}_t; \tag{13}$$

(along with a TVC). Equation 11 is a dynamic condition on capital, and equation 13 is a dynamic condition on the shadow price of capital, μ. Combining equation 12 (which implies $\dot{\mu}_t = e^{-\rho t} \left[U''(c_t)\dot{c}_t - \rho U'(c_t) \right]$) with equation 13 yields an analogous condition on consumption:

$$\dot{c}_t = \frac{U'(c_t)}{U''(c_t)}(\rho + \delta - r_t). \tag{14}$$

This is the analogue of the intereulers we get in discrete-time models.

At this point we impose CRRA utility, hence

$$U(c) = \frac{c^{1-\sigma}}{1-\sigma} \implies U'(c) = c^{-\sigma} \text{ and } U''(c) = -\sigma c^{-\sigma-1},$$

and Cobb-Douglas production (recall that $n = 1$), hence

$$F(k, n) = k^\alpha n^{1-\alpha} \implies r = \alpha k^{\alpha-1} \text{ and } w = (1-\alpha)k^\alpha.$$

Thus equation 11 becomes

$$\dot{k}_t = (1-\alpha)k_t^\alpha + \alpha k_t^{\alpha-1} k_t - \delta k_t - c_t$$
$$= k_t^\alpha - \delta k_t - c_t$$
$$\frac{\dot{k}_t}{k_t} = k_t^{\alpha-1} - \delta - \frac{c_t}{k_t}.$$

From now on, we will use the notation that "hatted" variables represent the relative deviation from steady state ($\hat{x} \equiv \log(x) - \log(x_*) \approx (x - x_*)/x_*$). Note that $\dot{k}_t/k_t = \frac{d}{dt}(\log k_t) = \frac{d}{dt}(\log k_t - \log k_*) \equiv \dot{\hat{k}}_t$. This gives

$$\dot{\hat{k}}_t = k_t^{\alpha-1} - \delta - \frac{c_t}{k_t}. \tag{15}$$

Similarly, equation 14 becomes

$$\dot{c}_t = \tfrac{1}{\sigma} c_t (\alpha k_t^{\alpha-1} - \rho - \delta)$$
$$\dot{\hat{c}}_t = \tfrac{1}{\sigma}(\alpha k_t^{\alpha-1} - \rho - \delta). \tag{16}$$

We know that at steady state, equation 16 equals zero, so

$$k_*^{\alpha-1} = \frac{\rho+\delta}{\alpha} \tag{17}$$

and similarly for equation 15,

$$\frac{c_*}{k_*} = k_*^{\alpha-1} - \delta$$

$$= \frac{\rho + \delta}{\alpha} - \delta = \frac{\rho + \delta(1 - \alpha)}{\alpha}. \tag{18}$$

Log-linearizing equation 15 (noting that $\dot{\hat{k}}_* = 0$) gives

$$\dot{\hat{k}}_t \approx (\alpha - 1)k_*^{\alpha-1}\hat{k}_t - \frac{c_*}{k_*}\hat{c}_t + \frac{c_*}{k_*}\hat{k}_t$$

Substituting in using the steady state results from equations 17 and 18,

$$= \left[(\alpha - 1)\frac{\rho + \delta}{\alpha} + \frac{\rho + \delta(1 - \alpha)}{\alpha}\right]\hat{k}_t - \frac{\rho + \delta(1 - \alpha)}{\alpha}\hat{c}_t$$

$$= \rho\hat{k}_t - \frac{\rho + \delta(1 - \alpha)}{\alpha}\hat{c}_t \tag{19}$$

Log-linearizing equation 16 (noting that $\dot{\hat{c}}_* = 0$) gives

$$\dot{\hat{c}}_t \approx \frac{1}{\sigma}\alpha(\alpha - 1)k_*^{\alpha-1}\hat{k}_t.$$

Substituting in using the steady state results from equation 17,

$$= \frac{(\alpha - 1)(\rho + \delta)}{\sigma}\hat{k}_t \tag{20}$$

Combining equations 19 and 20 in matrix notation, we have the dynamic system

$$\begin{bmatrix} \dot{\hat{k}}_t \\ \dot{\hat{c}}_t \end{bmatrix} = \begin{bmatrix} \rho & -\frac{\rho + \delta(1 - \alpha)}{\alpha} \\ \frac{(\alpha-1)(\rho+\delta)}{\sigma} & 0 \end{bmatrix} \begin{bmatrix} \hat{k}_t \\ \hat{c}_t \end{bmatrix}.$$

If we consider this system as $\dot{x}_t = Ax_t$, we could "decouple" the system using the eigen decomposition $A = P\Lambda P^{-1}$. Thus the system becomes $\frac{d}{dt}P^{-1}x_t = \Lambda P^{-1}x_t$. The solution is $x_t = Pe^{\Lambda t}P^{-1}x_0$.[17]

18   Optimal taxation

18.1   The Ramsey model

In our models so far, we have only had two types of agents: households and firms. When we have considered government spending, we have always specified it entirely in terms of an exogenous stream $\{g_t\}_t$ that is taken from households. Because the revenue requirement was exogenous, and because the government only had a single instrument by which to collect it—lump-sum, or "head"—taxes, there was no flexibility in the government's behavior and no reason to treat it as an agent in the model.

If we relax either or both of these restrictions—allowing the government to choose the level of revenue collected each period and/or the means of collecting it—we move to a class of models called "optimal taxation" or "Ramsey" models. How does the government make its taxation choices in these models? We will generally

---

[17]To see this, first note that the system is $\frac{d}{dt}\tilde{x}(t) = \Lambda\tilde{x}(t)$, where we define $\tilde{x}(t) \equiv P^{-1}x(t)$. This gives separable differential equations in each element $i \in \{1, 2\}$ of $\tilde{x}$ of the form $d\tilde{x}_i(t)/\tilde{x}_i(t) = \Lambda_i \, dt \implies \int d\tilde{x}_i(t)/\tilde{x}_i(t) = \int \Lambda_i \, dt \implies \log(\tilde{x}_i(t)) = t\Lambda_i + \log(\tilde{x}_i(0))$, where the constant of integration is pinned down by the initial condition. Finally, this implies $\tilde{x}_i(t) = \exp(t\Lambda_i) \cdot \tilde{x}_i(0)$, or $P^{-1}x_i(t) = \exp(t\Lambda_i)P^{-1}x_i(0) \implies x_i(t) = P\exp(t\Lambda_i)P^{-1}x_i(0)$.

assume that the government's objective is to maximize household welfare, noting that the government knows households will maximize their own welfare subject to tax policy. A stylized way of describing this is as follows:

$$\max_{\tau \text{ s.t. } \ldots} \left[\max_c U(c, \tau) \right] \qquad (21)$$

where τ are the government's tax policies, c are the representative household's choices, and there is some constraint on the outer maximization (i.e., the government's) of the form $g(\tau, c_*(\tau)) \geq 0$, designed to capture the fact that the government must raise enough money to achieve some exogenous goal.[18] A solution gives a "Ramsey equilbrium": allocations, prices, and taxes such that

1. Households maximize utility subject to their budget constraints, taking prices and taxes as given,

2. Government maximizes households' utility while financing government expenditures (i.e., meeting its budget constraint or constraints),

3. Markets clear, and

4. The economy's resource constraint is satisfied.

Consider an example: an economy with a representative household and the government. In this economy there is no capital; households can produce a perishable consumption good with technology $f(n) = n$, and they can invest in government bonds. Further suppose that the government must raise (exogenous) g_t each period, which it can do through an income tax or through government debt.

We can write the household problem as

$$\max_{\{c_t, n_t, b_{t+1}\}} \sum_{t=0}^{\infty} \beta^t u(c_t, n_t)$$
$$\text{s.t.} \quad c_t + q_t b_{t+1} = n_t(1 - \tau_t) + b_t, \quad \forall t;$$
$$b_0 \text{ given.}$$

The household's resource uses are consumption and purchase of bonds; its sources are production $(f(n_t) = n_t)$ net of taxes $(\tau_t n_t)$ and bond coupons. The government's resource sources are bond sales and tax revenue, and its uses are bond repayment and government spending; thus the government budget constraints are

$$q_t b_{t+1} + \tau_t n_t = b_t + g_t, \quad \forall t. \qquad (22)$$

Finally, it is useful to write the economy's resource constraint (which must hold by the combination of household and government budget constraints):

$$c_t + g_t = n_t. \qquad (23)$$

So what does the government do? Following the approach suggested by equation 21, it could

1. Find the household optimum as a function of tax rates $\vec{\tau} \equiv \{\tau_t\}_t$ and bond prices $\vec{q} \equiv \{q\}_t$. Setting up and solving the household problem gives the following intra- and intereulers: for all t,

$$(1 - \tau_t) \cdot u_c(c_t, n_t) = -u_n(c_t, n_t), \text{ and} \qquad (24)$$
$$q_t \cdot u_c(c_t, n_t) = \beta u_c(c_{t+1}, n_{t+1}). \qquad (25)$$

These help pin down the optimal $\left\{ c_t^*(\vec{\tau}, \vec{q}), n_t^*(\vec{\tau}, \vec{q}) \right\}_t$; the household budget constraints then allow us to find the bond holdings $\left\{ b_{t+1}^*(\vec{\tau}, \vec{q}) \right\}_t$.

---

[18]Note that τ captures all tax policies (in each period, for each instrument available to the government), c captures all household choices (e.g., consumption, hours, assets, and/or capital), and the government's budget constraint could potentially be imposing period-by-period revenue requirements.

2. Choose $\vec{\tau}$ and $\vec{q}$ to maximize

$$U^*(\vec{\tau}, \vec{q}) \equiv U\left(\left\{c_t^*(\vec{\tau}, \vec{q}), n_t^*(\vec{\tau}, \vec{q})\right\}_t\right),$$

subject to the government's budget constraints,

$$q_t \cdot b_{t+1}^*(\vec{\tau}, \vec{q}) + \tau_t \cdot n_t^*(\vec{\tau}, \vec{q}) = b_t^*(\vec{\tau}, \vec{q}) + g_t.$$

18.2   The Primal approach

The solution procedure described above turns out typically to form a *very* difficult problem. An alternative technique, called the "primal approach," is usually much easier to solve. In the primal approach, the government optimizes not over taxes, but rather over allocations, subject to a constraint that the allocations it chooses are optimal for the household under *some* tax regime. Roughly, this is equivalent to

$$\max_{c \text{ s.t. } \ldots} U(c),$$

where the government is constrained to choose a c that are the household's optimal choice ($c = c_*(\tau)$) for some τ satisfying the government's budget constraint $g(\tau, c_*(\tau)) \geq 0$. This is called the "implementability constraint."

Proceeding with our example from above, which allocations $\{c_t, n_t, b_{t+1}\}_t$ are consistent with household optimization for some tax rate and bond prices satisfying the government's budget constraint? The intereuler (equation 25) allows us to pin down bond prices in terms of allocations:

$$q_t = \beta \frac{u_c(c_{t+1}, n_{t+1})}{u_c(c_t, n_t)}.$$

With the government budget constraint (equation 22), this lets us pin down the tax rates in terms of allocations.

$$\begin{aligned}
\tau_t &= \frac{b_t + g_t - q_t b_{t+1}}{n_t} \\
&= \frac{b_t + g_t - \beta \frac{u_c(c_{t+1}, n_{t+1})}{u_c(c_t, n_t)} b_{t+1}}{n_t}.
\end{aligned} \tag{26}$$

Plugging this into the intraeuler (equation 24) gives

$$1 - \frac{b_t + g_t - \beta \frac{u_c(c_{t+1}, n_{t+1})}{u_c(c_t, n_t)} b_{t+1}}{n_t} = -\frac{u_n(c_t, n_t)}{u_c(c_t, n_t)},$$

where either c or n (here c) can be eliminated using the economy's resource constraint (equation 23):

$$1 - \frac{b_t + g_t - \beta \frac{u_c(n_{t+1} - g_{t+1}, n_{t+1})}{u_c(n_t - g_t, n_t)} b_{t+1}}{n_t} = -\frac{u_n(n_t - g_t, n_t)}{u_c(n_t - g_t, n_t)}.$$

Thus we can write the government's optimal taxation problem as

$$\max_{\{n_t, b_{t+1}\}} \sum_{t=0}^{\infty} \beta^t u(n_t - g_t, n_t)$$

$$\text{s.t.} \quad \underbrace{1 - \frac{b_t + g_t - \beta \frac{u_c(n_{t+1} - g_{t+1}, n_{t+1})}{u_c(n_t - g_t, n_t)} b_{t+1}}{n_t} + \frac{u_n(n_t - g_t, n_t)}{u_c(n_t - g_t, n_t)}}_{\equiv \eta(n_t, g_t, b_t, n_{t+1}, g_{t+1}, b_{t+1})} = 0.$$

44

Although it looks ugly, this "primal" problem is actually straightforward to solve![19]

In summary, the primal approach uses the "solution" to the household problem, the government's budget constraint, and the economy's resource constraint to eliminate taxes (and prices) from the government's problem. Instead, the government faces an implementability constraint (here, $\eta(n_t, g_t, b_t, n_{t+1}, g_{t+1}, b_{t+1}) = 0$ for all t). Maximizing over allocations should then be relatively easy, and the implementability constraint ensures that the allocations are consistent with some taxes that satisfy the government's budget constraint. The last step is to solve for the taxes; here we would use equation 26.

19   Introducing uncertainty

Let S be the set of events that could occur in any given period. For now, we will assume that S has finitely many elements, but this is mostly just for notational convenience; the intuition and most results go through if the shock space is countably infinite or continuous. Note that we are also assuming discrete time; considering uncertainty with continuous time makes things significantly more complex, and we will not do so this quarter.

In the simplest possible example, suppose that we flip a coin in each period $t \geq 1$. Letting s_t denote the *realization* of the event at time t, we have

$$s_t \in S = \{H, T\}, \quad \forall t \in \{1, 2, \dots\}.$$

Since s_t is a random variable for each t, we have a *stochastic process*:

Definition 18. *A **stochastic process** is a sequence $\{s_t\}_t$ of random variables, ordered by an integer t (which we will think of as time).*

Suppose we want to characterize the *history* of this stochastic process through period t; we denote this object s^t. For example, after three periods, we may have observed $s^3 = (H, H, T)$. Generally,

$$s^t \equiv (s_1, s_2, \dots, s_t) \in S \times \cdots \times S = S^t, \quad \forall t.$$

This notation can be confusing; a good mnemonic is to use the notational consistency in "$s^t \in S^t$" to help remember that superscripts represent histories.

Any stochastic variable in a model—whether exogenous or endogenous—can only depend on the uncertainty that has already been revealed. That is, if some variable c_t is "determined" at time t, that means that knowing the history s^t must give us enough information to pin down c_t. We often use the notation $c_t(s^t)$ to indicate explicitly that c_t depends on the resolution of uncertainty in the first t periods.

For example, suppose that we make the following bet: I flip a coin each of two days; you pay me \$5 today, and I give you back \$11 tomorrow if my coin flips came up the same both days.[20] Your incomes in the periods are given by the random variables $y_1(s^1)$ and $y_2(s^2)$, where

$$y_1(H) = -5, \qquad\qquad\qquad y_2\big((H, H)\big) = 11,$$
$$y_1(T) = -5; \qquad\qquad\qquad y_2\big((H, T)\big) = 0,$$
$$\qquad\qquad\qquad\qquad\qquad y_2\big((T, H)\big) = 0,$$
$$\qquad\qquad\qquad\qquad\qquad y_2\big((T, T)\big) = 11.$$

---

[19]If you would like a good exercise that allows you to take this example further, consider the following quasilinear felicity function: $u(c, n) = c - a(n)$ for a strictly convex function $a(\cdot)$. Show that the interest rate on government bonds is $1/q_t = \beta^{-1}$, and that no matter the (deterministic) sequence of $\{g_t\}_t$, there is an optimum with tax rates and labor constant over time. Note that these results are specific to this utility function, and require that the government be able to credibly commit in advance to a specific tax plan $\{\tau_t\}_t$.

[20]Suppose I toss a fair coin (and will not run off with your money). Would you take this bet? Your answer will depend on your risk aversion and your discount rate.

19.1 Probability

We denote the (unconditional) probability of observing any particular history $s^t \in S^t$ in periods $1, \ldots, t$ by $\Pr(s^t)$. (The notation $\pi(\cdot)$ is also common.) For $\Pr(\cdot)$ to be a well-defined probability measure, we require that $\Pr(s^t) \geq 0$ for all $s^t \in S^t$, and that

$$\sum_{s^t \in S^t} \Pr(s^t) = 1.$$

In our (fair) coin-tossing example, we have $\Pr(s^t) = 2^{-t}$ for all t and $s^t \in S^t$. Often, we will only allow the first case, where s^t remains possible given s^τ (i.e., s^τ comprises the first τ elements of s^t).

We also consider the probability of observing a particular history s^t *conditional* on having already observed history s^τ for $\tau \leq t$. We denote this probability $\Pr(s^t | s^\tau)$. Returning to coin tossing,

$$\Pr(s^t | s^\tau) = \begin{cases} 2^{-(t-\tau)}, & \text{if } (s_1^t, s_2^t, \ldots, s_\tau^t) = s^\tau; \\ 0, & \text{otherwise} \end{cases}$$

for all t, $\tau \leq t$, $s^t \in S^t$, and $s^\tau \in S^\tau$.

It will often be convenient to consider an additional period, $t = 0$, in which there is no uncertainty. To keep our notation consistent, we use the symbols s_0 and s^0 to represent the (non-stochastic) state at $t = 0$, and extend the definitions of our probability functions as follows:

$$\Pr(s^0) \equiv 1;$$
$$\Pr(s^t | s^0) \equiv \Pr(s^t), \quad \forall t \geq 0.$$

Using this second extension, we can specify the full probability structure with just the conditional probability functions $\Pr(\cdot | \cdot)$.[21]

19.2 Utility functions

Consider an agent whose preferences are time-separable, have exponential discounting, and admit a von Neumann-Morgenstern representation.[22] Her preference can be represented by

$$U(\mathbf{c}) = \sum_t \sum_{s^t \in S^t} \beta^t u\big(c_t(s^t)\big) \Pr(s^t)$$
$$= \mathbb{E}\left[\sum_t \beta^t u(c_t) \right].$$

We rely here on the linearity of the expectation operator; we will *often* do so.

20 Markov chains

Although the general structure laid out above will be useful, we will often be willing to impose more structure on the probability structure. The most common structure we will assume is that $\{s_t\}_t$ forms a *Markov chain*.

---

[21]It is a bit cumbersome to have period zero be "different" from all other periods, and we will not always do so. However, it offers two advantages. The first is that having period zero be non-stochastic allows $s^t \in S^t$, rather than $s^t \in S^{t+1}$. More importantly, having a non-stochastic state allows us to write unconditional probabilities as conditional probabilities, since we can reference/condition on a history (s^0) that is entirely uninformative.

[22]Preferences admit a von Neumann-Morgenstern representation if and only if they satisfy:

1. Continuity: For any x, x', x'' with $x \succsim x' \succsim x''$, there exists $\alpha \in [0, 1]$ such that $\alpha x + (1 - \alpha)x'' \sim x'$;

2. Independence: For any x, x', x'' and $\alpha \in [0, 1]$, we have $x \succsim x' \iff \alpha x + (1 - \alpha)x'' \succsim \alpha x' + (1 - \alpha)x''$; and

3. A "sure thing principle"

(in addition to the usual completeness and transitivity). *Extensive* discussion of these points is conducted in Economics 202.

Definition 19. *A stochastic process $\{s_t\}_t$ satisfies the **Markov property** (or is a **Markov chain**) if for all $t_1 < t_2 < \cdots < t_n$,*

$$\Pr(s_{t_n} = s | s_{t_{n-1}}, s_{t_{n-2}}, \ldots, s_{t_1}) = \Pr(s_{t_n} = s | s_{t_{n-1}}).$$

Although the notation looks odious, the intuition is not bad. A Markov process is one where if one has information about several realizations $(s_{t_{n-1}}, \ldots, s_{t_1})$, only the *latest* realization $(s_{t_{n-1}})$ is useful in helping predict the future.

Note that by the definition of conditional probability (and the trivial fact that $\Pr(s^t | s^{t-1}) = \Pr(s_t | s^{t-1})$),

$$\Pr(s^t) = \Pr(s_t | s^{t-1}) \cdot \Pr(s^{t-1})$$
$$= \Pr(s_t | s^{t-1}) \cdot \Pr(s_{t-1} | s^{t-2}) \cdot \Pr(s^{t-2})$$
$$\vdots$$
$$= \Pr(s_t | s^{t-1}) \cdot \Pr(s_{t-1} | s^{t-2}) \cdots \Pr(s_1 | s^0).$$

Fortunately, the Markov property implies that $\Pr(s_t | s^{t-1}) = \Pr(s_t | s_{t-1})$, so for a Markov chain

$$= \Pr(s_t | s_{t-1}) \cdot \Pr(s_{t-1} | s_{t-2}) \cdots \Pr(s_1 | s_0)$$
$$= \prod_{j=0}^{t-1} \Pr(s_{j+1} | s_j).$$

Similarly, for a Markov chain,

$$\Pr(s^t | s^\tau) = \prod_{j=\tau}^{t-1} \Pr(s_{j+1} | s_j).$$

for all t, $\tau \leq t$, $s^t \in S^t$, and $s^\tau \in S^\tau$ with $s^\tau = (s_1^t, s_2^t, \ldots, s_\tau^t)$. Thus we can specify the entire probability structure of a Markov chain if we know all of its "transition probabilities" $\Pr(s_{t+1} | s_t)$. In general, these transition probabilities may depend on the time t. However, this may not be the case, in which case our Markov chain is time-invariant.

Definition 20. *A **time-invariant Markov chain** is a stochastic process satisfying the Markov property and for which*

$$\Pr(s_{t+1} = j | s_t = i) = \Pr(s_{t+2} = j | s_{t+1} = i)$$

for all t and $(i, j) \in S^2$. This implies that the transition probabilities are constant over time (by induction).

For a time-invariant Markov chain, we can summarize these transition probabilities in a *transition matrix*. Suppose without loss of generality that the shock space $S = \{1, 2, \ldots, m\}$. Then we define the transition matrix P by

$$P_{ij} = \Pr(s_{t+1} = j | s_t = i)$$

for every i and j (and any t: by time-invariance, the choice does not matter). Every row of P must add to one ($\sum_j P_{ij} = 1$ for all i), which means that P can be called a "stochastic matrix."

Note that an iid process (i.e., one for which the realization is distributed independently and identically across periods) is a time-invariant Markov chain, and will have a transition matrix where every row is identical.

20.1 Unconditional distributions

Suppose we have a time-invariant Markov chain and—contravening our earlier notation—also allow s_0 to be stochastic. In particular, let π_0 be a vector whose ith element is $\Pr(s_0 = i)$. Clearly, $\sum_i \pi_{0\,i} = 1$.

Can we say anything about an analogous vector π_1 characterizing the (unconditional) probability distribution of $s_1 \in S$? That is, we seek a vector with elements given by $\pi_{1\,i} = \Pr(s_1 = i)$. By the law of total probability,

$$\pi_{1\,i} = \Pr(s_1 = i) = \sum_j \underbrace{\Pr(s_1 = i | s_0 = j)}_{=P_{ji}} \cdot \underbrace{\Pr(s_0 = j)}_{=\pi_{0\,i}}.$$

This looks like the matrix multiplication algorithm, and indeed is equivalent to stating that $\pi_1 = P'\pi_0$. Similarly, for any t,

$$\pi_{t+1} = P'\pi_t \qquad \text{or equivalently} \qquad \pi'_{t+1} = \pi'_t P, \qquad (27)$$

and induction gives that

$$\pi_t = (P')^t \pi_0 \qquad\qquad \pi'_t = \pi'_0 P^t.$$

20.2   Conditional distributions

Suppose that we are in state i today (i.e., event i occurred); what is the probability that we will be state j tomorrow? This is P_{ij}. But what is the probability we will be in state j in two days? By the law of total probability,

$$\Pr(s_{t+2} = j | s_t = i) = \sum_k \Pr(s_{t+2} = j | s_{t+1} = k, s_t = i) \cdot \Pr(s_{t+1} = k | s_t = i).$$

The first probability on the right-hand side can be simplified by the Markov property:

$$= \sum_k \underbrace{\Pr(s_{t+2} = j | s_{t+1} = k)}_{=P_{kj}} \cdot \underbrace{\Pr(s_{t+1} = k | s_t = i)}_{=P_{ik}}$$

$$= (P^2)_{ij}.$$

It is not hard to see (or show) that more generally, for $\tau \le t$,

$$\Pr(s_t = j | s_\tau = i) = (P^{t-\tau})_{ij}.$$

20.3   Stationary distributions

An unconditional distribution $\Pr(s_t)$ is said to be stationary if it gives the same (unconditional) distribution across states in the following period; by induction, this ensures that the distribution will be the same in *all* subsequent periods for a time-invariant Markov chain. Considering the vector representation of an unconditional distribution $(\pi_{t\,i} = \Pr(s_t = i))$, π_t is a *stationary distribution* (or *invariant distribution*) if $\pi_{t+1} = \pi_t$. Per the law of motion we stated in equation 27, this requires $\pi_t = \pi$ satisfying

$$P'\pi = \pi$$
$$(P' - I)\pi = \mathbf{0}.$$

This should look familiar; we saw something very similar in equation 9. It means that a stationary distribution is any eigenvector associated with a unitary eigenvalue of P', normalized so that the sum of the elements of the eigenvector is one. The fact that P is a stochastic matrix (i.e., one whose rows add to one) ensures that it will have at least one unitary eigenvalue (although it may have more than one).

You should be able intuitively to identify the stationary distribution(s) for an iid process. What about the process with transition matrix $P = \mathbf{I}$? Or

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}? \qquad (28)$$

20.4 Ergodic distributions

Consider starting a Markov chain with initial distribution π_0, and "running the process forward" arbitarily. It seems the distribution across states should be given by $\pi_\infty(\pi_0) \equiv \lim_{t\to\infty} \pi_t = \lim_{t\to\infty} (P')^t \pi_0$. We can call this the "limiting distribution," but should check first whether this limit is even well defined! It turns out that it may or may not be. The clearest illustrations come from considering $P = \mathbf{I}$ (in which case the limiting distribution is the initial distribution), and the process with transition matrix as given in equation 28 (which has no limiting distribution unless $\pi_0 = [\frac{1}{2}, \frac{1}{2}]'$).

Note that any limiting distribution must be stationary.

We will put aside the question of whether a Markov chain has a limiting distribution; suppose for now that it does. In fact, for some Markov chains, the limiting distribution not only exists, but does not depend on the initial distribution.

Definition 21. *A (time-invariant) Markov chain is said to be* ***asymptotically stationary with a unique invariant distribution*** *if all initial distributions yield the same limiting distribution; i.e., $\pi_\infty(\pi_0) = \pi_\infty$ for all π_0. This limiting distribution, π_∞, is called the* ***ergodic distribution*** *of the Markov chain, and it is the only stationary distribution of the Markov chain.*

We state without proof several important results about asymptotically stationary Markov chains with unique invariant distributions.

Theorem 22. *Let P be a stochastic matrix with $P_{ij} > 0$ for all i and j. Then P is asymptotically stationary, and has a unique invariant distribution.*

Theorem 23. *Let P be a stochastic matrix with $(P^m)_{ij} > 0$ for all i and j for some $m \geq 1$. Then P is asymptotically stationary, and has a unique invariant distribution.*

This means that as long as there is a strictly positive probability of moving from any particular state today to any particular state in one or more steps,[23] then an ergodic distribution exists. Thus if we can find an m for which $(P^m)_{ij} > 0$ for all i and j, we can

1. Note that P must therefore have an ergodic distribution, and that it must be the unique stationary distribution of P, and

2. Solve $(P' - \mathbf{I})\pi_\infty = 0$ for this unique distribution.

21 Risk-sharing properties of competitive markets

Now that we are considering stochastic models, we can ask how agents react in the face of risk. We should not be surprised that when risk-averse agents have access to assets that allow them to insure against uncertainty, they use them. It turns out that when markets are "complete"—that is, agents can trade contingent securities for every state of the world—agents will "perfectly insure," subjecting themselves to no individual (idiosyncratic) risk. There may still be aggregate risk: if the economy is closed and a hurricane strikes everyone, the resource constraint ensures that everyone eats less. However, as long as the aggregate resources available in the economy are non-stochastic (which will often be ensured by a law of large numbers), perfect insurance allows everyone to avoid uncertainty in consumption.

Consider an economy with agents who have identical preferences represented by

$$U\big(\{c_t^i(s^t)\}_{t=0}^\infty\big) = \sum_{t=0}^\infty \mathbb{E}_0\big[\beta^t u\big(c_t^i(s^t)\big)\big].$$

---

[23]This is an example of a "mixing condition." Analogous conditions exist with infinite state spaces, although they are harder to write.

Suppose that agent i receives a stochastic endowment $\{y_t^i(s^t)\}_{t=0}^\infty$ of the (perishable) consumption good. The only assets available to agents are a complete set of state-contingent securities traded at time 0, which (in equilibrium) are priced at $q_t(s^t)$.[24] Agent i's problem is therefore to

$$\max_{\{c_t^i(s^t)\}} \sum_{t=0}^\infty \sum_{s^t \in S^t} \beta^t u\big(c_t^i(s^t)\big) \Pr(s^t) \quad \text{s.t.}$$

$$\sum_{t=0}^\infty \sum_{s^t \in S^t} q_t(s^t)\big[y_t^i(s^t) - c_t^i(s^t)\big] \geq 0.$$

Setting up the Lagrangian

$$\mathcal{L} \equiv \sum_{t=0}^\infty \sum_{s^t \in S^t} \Big[\beta^t u\big(c_t^i(s^t)\big) \Pr(s^t) + \mu^i q_t(s^t)\big[y_t^i(s^t) - c_t^i(s^t)\big]\Big]$$

gives first-order conditions of the form

$$\mu^i q_t(s^t) = \beta^t u'\big(c_t^i(s^t)\big) \Pr(s^t).$$

Dividing the FOCs of two two agents, we have

$$\frac{\mu^i}{\mu^j} = \frac{u'\big(c_t^i(s^t)\big)}{u'\big(c_t^j(s^t)\big)}$$

$$(u')^{-1}\left(\frac{\mu^i}{\mu^j} u'\big(c_t^j(s^t)\big)\right) = c_t^i(s^t).$$

Summing across agents i and noting that the economy's resource constraint, $\sum_i c_t^i(s^t) = \sum_i y_t^i(s^t)$, gives

$$\sum_i (u')^{-1}\left(\frac{\mu^i}{\mu^j} u'\big(c_t^j(s^t)\big)\right) = \sum_i c_t^i(s^t)$$

$$= \underbrace{\sum_i y_t^i(s^t)}_{\equiv \bar{y}(s^t)}. \tag{29}$$

Although ugly, this equality tells us something important: consumption of each agent $c_t^j(s^t)$ depends on the state s^t only through the realization of the aggregate endowment $\bar{y}(s^t)$.

If we are prepared to impose a functional form on $u(\cdot)$, we can go a bit further. Suppose for example that

$$u(c) = \frac{c^{1-\sigma} - 1}{1 - \sigma},$$

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

$$(u')^{-1}(x) = x^{-1/\sigma}.$$

Then equation 29 becomes

$$\sum_i \left(\frac{\mu^i}{\mu^j}\big(c_t^j(s^t)\big)^{-\sigma}\right)^{-1/\sigma} = \bar{y}(s^t)$$

$$c_t^j(s^t) = \bar{y}(s^t) \cdot \frac{\big(\mu^j\big)^{1/\sigma}}{\sum_i \big(\mu^i\big)^{-1/\sigma}}.$$

With complete markets and power utility, each agent's consumption is a constant fraction of the economy's aggregate endowment.

---

[24]Although we solve the Arrow-Debreu problem, identical results obtain with complete sequential markets; there, agents trade each period in contracts that pay off in the subsequent period depending on what state occurs.

22 Perfect and imperfect insurance practice question

This question comes from the Economics 211 midterm examination in 2007. It is based on a model of Doireann Fitzgerald's from 2006.

Question

There is just one period. Suppose there are two countries: A and B. Country A receives an endowment of $y_A(s)$ units of the consumption good as a function of the state of the world s. Country B receives $y_B(s)$. Let $\pi(s) > 0$ denote the probabilities of the state of the world s.

The representative agent in each country has a utility function given by

$$\sum_s \pi(s) \log c(s),$$

where $c(s)$ denotes consumption by the country in state s.

Suppose that the consumption good can be transported at no cost from one country to another.

1. Set up the Pareto problem, letting λ be the planner's weight on Country B. Show that in a Pareto optimal allocation, each country consumes a constant fraction of the total endowment $y_A(s) + y_B(s)$.

2. Suppose that before the state of the world is realized, the countries can trade claims on consumption in a complete asset market. Assume that initially, each country owns the claim on its stochastic endowment. Solve for a competitive equilibrium and show that it is Pareto optimal.

Suppose now that the consumption good is costly to ship across countries. In particular, for a unit of consumption good to arrive in Country B from Country A, $1 + t$ units of consumption have to be shipped from Country A (and similarly from Country B to A), where $t > 0$. We can think of t as capturing a piece of the consumption good that melts away due to transportation costs.[25] Suppose also that there are only two states of nature $s \in \{s_1, s_2\}$ and that the endowments are

$$y_A(s) = \begin{cases} 1, & \text{for } s = s_1 \\ 0, & \text{for } s = s_2 \end{cases}$$

and $y_B(s) = 1 - y_A(s)$. So $y_A(s) + y_B(s)$ is constant. Let $\pi(s_1) = 1/2$.

3. Set up the Pareto problem (let λ be the planner's weight on Country B), and write down the first-order conditions. Show that in a Pareto allocation, the consumption of a country will now depend on its endowment realization, that is, Country A consumption depends on $y_A(s)$ and not only on $y_A(s) + y_B(s)$. How does t affect the sensitivity of consumption to income?

4. What does part 3 imply for tests of complete markets across countries?

Solution

1. The Pareto problem is to

$$\max_{\{c_A(s), c_B(s)\}_{s \in S}} (1 - \lambda) \sum_{s \in S} \pi(s) \log c_A(s) + \lambda \sum_{s \in S} \pi(s) \log c_B(s) \quad \text{s.t.}$$

$$c_A(s) + c_B(s) \leq \underbrace{y_A(s) + y_B(s)}_{\equiv \bar{y}(s)}, \quad \forall s \in S.$$

---

[25] Indeed, this form of transportation costs are often called "iceberg costs," although melting is only one of the analogies that has been used to explain the name.

Setting up the Lagrangian

$$\mathcal{L} \equiv \sum_{s \in S} \pi(s) \Big[(1 - \lambda) \log c_A(s) + \lambda \log \underbrace{\big(\bar{y}(s) - c_A(s) \big)}_{= c_B(s)} \Big]$$

gives first-order conditions of the form

$$\frac{1 - \lambda}{c_A(s)} = \frac{\lambda}{c_B(s)}.$$

Plugging $c_B(s) = c_A(s) \cdot \lambda/(1-\lambda)$ into the resource constraint gives

$$c_A(s) = (1 - \lambda) \cdot \bar{y}(s),$$
$$c_B(s) = \lambda \cdot \bar{y}(s).$$

2. Let the price of a claim on consumption in state s be $q(s)$. An equilibrium is prices $\{q(s)\}_{s \in S}$ and allocations $\{c_A(s), c_B(s)\}_{s \in S}$ such that the markets clear $(c_A(s) + c_B(s) = y_A(s) + y_B(s)$ for all $s \in S)$, and $\{c_i(s)\}_{s \in S}$ solves the problem of country $i \in \{A, B\}$ given $\{q(s)\}_{s \in S}$.

The problem in country i is to

$$\max_{\{c_i(s)\}_{s \in S}} \sum_{s \in S} \pi(s) \log c_i(s) \quad \text{s.t.}$$
$$\sum_{s \in S} q(s) \big[y_i(s) - c_i(s) \big] \geq 0. \tag{30}$$

Setting up the Lagrangian

$$\mathcal{L} \equiv \sum_{s \in S} \Big[\pi(s) \log c_i(s) + \mu_i q(s) \big[y_i(s) - c_i(s) \big] \Big]$$

gives first-order conditions of the form

$$\frac{\pi(s)}{c_i(s)} = \mu_i q(s). \tag{31}$$

Dividing A and B's first-order conditions gives

$$\frac{c_A(s)}{c_B(s)} = \frac{\mu_B}{\mu_A}.$$

Plugging $c_B(s) = c_A(s) \cdot \mu_A/\mu_B$ into the resource constraint gives

$$c_A(s) = \frac{\mu_B}{\mu_A + \mu_B} \bar{y}(s),$$
$$c_B(s) = \frac{\mu_A}{\mu_A + \mu_B} \bar{y}(s).$$

As these are constant fractions of the total endowment, these allocations are Pareto optimal.

Substituting into A's first-order condition (equation 31) gives

$$\pi(s) = \frac{\mu_A \mu_B}{\mu_A + \mu_B} \bar{y}(s) q(s) \implies q(s) = \pi(s) \frac{\mu_A + \mu_B}{\mu_A \mu_B} \bar{y}(s)^{-1}.$$

52

(This is an example of a more general result show in class: $q(s^t|s^0) \propto \beta^t u'(c(s^t)) \Pr(s^t)$.) We are allowed to normalize one price, which we can do by requiring that $(\mu_A + \mu_B)/(\mu_A \mu_B) = 1$. This gives us equilibrium prices

$$q(s) = \frac{\pi(s)}{\bar{y}(s)}.$$

Substituting into A's (binding) budget constraint (equation 30):

$$\sum_{s \in S} \underbrace{\frac{\pi(s)}{\bar{y}(s)}}_{=q(s)} \Big[\underbrace{y_A(s) - \frac{\mu_B}{\mu_A + \mu_B} \bar{y}(s)}_{=c_A(s)} \Big] = 0$$

$$\sum_{s \in S} \pi(s) \frac{y_A(s)}{\bar{y}(s)} = \frac{\mu_B}{\mu_A + \mu_B} \sum_{s \in S} \pi(s) \frac{\cancel{\bar{y}(s)}}{\bar{y}(s)}.$$

This allows us to pin down consumption:

$$c_A(s) = \Big[\sum_{s \in S} \pi(s) \frac{y_A(s)}{\bar{y}(s)} \Big] \bar{y}(s),$$

$$c_B(s) = \Big[\sum_{s \in S} \pi(s) \frac{y_B(s)}{\bar{y}(s)} \Big] \bar{y}(s).$$

3. The Pareto problem is to

$$\max_{\{c_A(s), c_B(s)\}_{s \in S}} (1 - \lambda) \sum_{s \in S} {}^{1}\!/_{2} \log c_A(s) + \lambda \sum_{s \in S} {}^{1}\!/_{2} \log c_B(s) \quad \text{s.t.}$$

$$c_A(s_1) + (1 + t) c_B(s_1) \leq 1,$$
$$(1 + t) c_A(s_2) + c_B(s_2) \leq 1.$$

Setting up the Lagrangian

$$2\mathcal{L} \equiv (1 - \lambda) \log c_A(s_1) + \lambda \log \Big(\underbrace{\frac{1 - c_A(s_1)}{1 + t}}_{=c_B(s_1)} \Big) + (1 - \lambda) \log c_A(s_2) + \lambda \log \big(\underbrace{1 - (1 + t) c_A(s_2)}_{=c_B(s_2)} \big)$$

gives first-order conditions

$$\frac{1 - \lambda}{c_A(s_1)} = \frac{\lambda}{(1 + t) c_B(s_1)} \implies \qquad (1 + t) c_B(s_1) = \frac{\lambda}{1 - \lambda} c_A(s_1);$$

$$\frac{1 - \lambda}{c_A(s_2)} = \frac{(1 + t)\lambda}{c_B(s)} \implies \qquad (1 + t) c_A(s_2) = \frac{1 - \lambda}{\lambda} c_B(s_2).$$

Substituting into the resource constraints gives

$$c_A(s_1) = 1 - \lambda, \qquad\qquad c_B(s_1) = \frac{\lambda}{1 + t};$$

$$c_A(s_2) = \frac{1 - \lambda}{1 + t}, \qquad\qquad c_B(s_2) = \lambda.$$

Thus $c_A(s_1) > c_A(s_2)$ and $c_B(s_1) < c_B(s_2)$, even though the aggregate endowment is constant across the two states. Hence consumption depends on one's own endowment realization.

4. Suppose that the First Welfare Theorem applies, so that equilibrium is Pareto optimal. Then by part 3, we should not expect to find independence of a country's consumption from its own endowment, even if markets are complete. Hence the fact that empirically, consumption *does* vary with endowment is not by itself sufficient evidence to reject the complete markets hypothesis.

23 Asset pricing with complete markets

For our purposes, the term "asset" refers to a contractually-guaranteed right to delivery of consumption goods, with the amount of delivery conditional on the history of the world s^t. The notation can be a little bit tricky, but there are three main ways that we denote the prices of assets:

1. $q^\tau(s^t)$: This is the price of an asset that delivers one unit of consumption at history s^t, where the price is paid in history-s^τ consumption goods (for $\tau \leq t$, and assuming that $s_1^\tau = s_1^t, \ldots, s_\tau^\tau = s_\tau^t$). This notation captures the price of Arrow-Debreu securities, $q^0(s^t)$; these are actually sufficient to pin down all $q^\tau(s^t)$ according to

$$q^\tau(s^t) = \frac{q^0(s^t)}{q^0(s^\tau)}.$$

   The Arrow-Debreu prices are pinned down (up to a normalization) by any agent's consumption, since the first-order conditions of the consumer problem

$$\max_{\{\{c_t(s^t)\}_{s^t \in S^t}\}_t} \mathbb{E}\left[\sum_t \beta^t u\big(c_t(s^t)\big)\right] \quad \text{s.t.}$$

$$\sum_t \sum_{s^t \in S^t} q^0(s^t)c_t(s^t) \leq B$$

   are of the form

$$q^0(s^t) = \lambda^{-1}\beta^t u'\big(c_t(s^t)\big)\pi(s^t).$$

   Thus

$$q^\tau(s^t) = \frac{q^0(s^t)}{q^0(s^\tau)} = \frac{\lambda^{-1}\beta^t u'\big(c_t(s^t)\big)\pi(s^t)}{\lambda^{-1}\beta^\tau u'\big(c_\tau(s^\tau)\big)\pi(s^\tau)}$$

$$= \beta^{t-\tau}\frac{u'\big(c_t(s^t)\big)}{u'\big(c_\tau(s^\tau)\big)}\pi(s^t|s^\tau). \tag{32}$$

2. $p^0(s^\tau)$: This is the price of an asset that delivers $d(s^t)$ units of consumption at every history s^t for $t \geq \tau$ *if* history s^τ is achieved, where the price is paid in time-zero consumption goods. This is a "redundant" asset; i.e., it could be created with a suitable combination of Arrow-Debreu securities, which determines the asset's price:

$$p^0(s^\tau) = \sum_{t \geq \tau}\sum_{s^t|s^\tau} q^0(s^t)d(s^t).$$

   In the simplest case, where $s^\tau = s^0 \equiv s_0$, the asset delivers $d(s^t)$ at every s^t; the price (paid in time-zero consumption goods) is

$$p^0(s_0) = \sum_t \sum_{s^t} q^0(s^t)d(s^t).$$

3. $p^\tau(s^\tau)$: This is the price of an asset that delivers $d(s^t)$ units of consumption at every history s^t for $t \geq \tau$ *if* history s^τ is achieved, where the price is paid in history-s^τ consumption goods. (This is sometimes called the price of the "tail asset.") To convert from a price measured in time-zero consumption goods, we must divide by the time-zero price of history-s^τ consumption goods:

$$p^\tau(s^\tau) = \frac{p^0(s^\tau)}{q^0(s^\tau)} = \sum_{t \geq \tau}\sum_{s^t|s^\tau} \underbrace{\frac{q^0(s^t)}{q^0(s^\tau)}}_{=q^\tau(s^t)} d(s^t). \tag{33}$$

54

Suppose that $d(\cdot)$ is such that $d(s^t) = 0$ for all $t \neq \tau + 1$; that is, the asset can only pay off in the period after it is "purchased." By equations 32 and 33,

$$
\begin{aligned}
p^\tau(s^\tau) &= \sum_{s_{\tau+1}} q^\tau(s^{\tau+1}) d(s^{\tau+1}) \\
&= \sum_{s_{\tau+1}} \beta \frac{u'\big(c_{\tau+1}(s^{\tau+1})\big)}{u'\big(c_\tau(s^\tau)\big)} \pi(s^{\tau+1}|s^\tau) d(s^{\tau+1}) \\
&= \mathbb{E}_\tau \bigg[\underbrace{\beta \frac{u'\big(c_{\tau+1}(s^{\tau+1})\big)}{u'\big(c_\tau(s^\tau)\big)}}_{\equiv m_{\tau+1}(s^{\tau+1})} d(s^{\tau+1}) \bigg],
\end{aligned}
$$

where $m_{\tau+1}$ is called the *stochastic discount factor* (note that it sort of tells us how much the consumer discounts consumption payouts in $s^{\tau+1}$ when valuing them in units of history-s^τ consumption). This also gives an expression that functions as a stochastic intereuler: defining the (stochastic) return on the asset as $R_{\tau+1}(s^{\tau+1}) \equiv {d(s^{\tau+1})}/{p^\tau(s^\tau)}$,

$$
1 = \mathbb{E}_\tau \big[m_{\tau+1} R_{\tau+1} \big].
$$

We can also apply the law of iterated expectations to ensure that an analogous result holds with unconditional expectations:

$$
1 = \mathbb{E} \big[m_{\tau+1} R_{\tau+1} \big].
$$

24 Introducing incomplete markets

Thus far, we have discussed models where agents can trade contingent claims for consumption at every possible history of the world. There are important insights to be gleaned from such models, but they also make strong predictions about perfect insurance that clearly don't hold up in the real world. We therefore start to consider economies where only a limited set of assets trade.

There are a number of reasons we might expect markets to be incomplete. One possibility is that the state of future states is so large and complex that either bounded rationality or transaction costs interfere with the issuance of a full set of contingent contracts.

Another potential problem is that people may not actually keep their promises. (We touched on a related problem in our discussion of time-inconsistency in optimal taxation.) An agent may commit to pay, and then fail to. She might commit to reveal some private information, but lack a means to prove that her revelation is honest. Or she might commit to take some private action, but lack a means to prove that she is actually taking it.

In the simplest models we discuss, we will exogenously impose a particular form of incompletness on markets. Typically, we will only allow consumers to trade risk-free securities. Although we won't explicitly discuss why we make these impositions, you should typically have stories about transaction costs, bounded rationality, or commitment issues (including private information and hidden actions) in mind.

Incompleteness of markets can introduce significant complications to our analysis. For example,

- There may no longer be an equivalence between sequential and date-zero trading. We need to solve the economy as it actually exists.

- Representative agents may no longer exist.

- When developing a recursive formulation for agents' problems, the state space is typically much more complicated. Perfect insurance under complete markets means that we often need only include, for example *aggregate* asset holdings and today's shock realization. With incomplete markets, asset holdings vary across agents, so we need to track the full *distribution*.

25 Econometrics of incomplete markets practice question

This question comes from the Economics 211 midterm examination in 2007.

Question

Suppose an agent consumes two goods every period: bananas and newspapers. Let the per-period utility function given consumption of bananas c_b and consumption of newspapers c_n be given by $u(c_b, c_n)$. Let us assume that this utility satisfies all standard properties (strictly concave, strictly increasing, and differentiable).

The agent lives for two periods. Let p_1 and $p_2(s)$ denote the price of newspapers in units of bananas in periods 1 and 2, respectively, where s represents a stochastic state of the world that realizes in period 2. The agent receives an endowment in period 1 equal to y_1 units of bananas, and receives $y_2(s)$ units of bananas in period 2. The probabilities of states of the world are denoted by $\pi(s) > 0$. The agent maximizes expected discounted utility, where the discount factor is given by β.

Suppose the agent can save in a riskless bond that returns R units of bananas in period 2 irrespective of the state of the world. Suppose that the agent cannot borrow. Assume for the questions that follow that the solution to the agent problem is interior—i.e., the borrowing constraint does not bind.

Suppose that utility is separable between bananas and newspapers and takes the following form:

$$u(c_b, c_n) = \frac{c_b^{1-\gamma_b}}{1-\gamma_b} + \frac{c_n^{1-\gamma_n}}{1-\gamma_n}.$$

1. Show that the following equation holds:

$$\beta \mathbb{E}\left[R \left(\frac{c_{b2}}{c_{b1}} \right)^{-\rho} \right] = 1 \tag{34}$$

   for some ρ, where c_{b1} and c_{b2} are consumption of bananas in periods 1 and 2, respectively.

Suppose that an econometrician observes the interest rate on the risk-free bond, but only has information about the number of bananas consumed by the agent in periods 1 and 2. She does not observe the agent's endowment, his consumption of newspapers, nor the price of newspapers (p_1 and p_2).

2. Assuming that the econometrician observes several of these agents at different points in time (with possibly different risk-free interest rates), can she estimate γ_b? What about γ_n? Ignoring non-linearities, write down the regression that could be used.

Suppose for the rest of the question that the utility is non-separable and takes the following form:

$$u(c_b, c_n) = \frac{(c_b^\alpha c_n^{1-\alpha})^{1-\gamma}}{1-\gamma} \tag{35}$$

for some $\alpha \in (0, 1)$.

3. Show that the econometrician in part 2 can estimate γ as long as $p_2(s) = p_1$ (the price of newspapers is constant) by estimating the same Euler equation as in part 2. Can she estimate α?

Solution

1. The consumer's problem is to

$$\max_{c_{b1},c_{n1},\{c_{b2}(s),c_{n2}(s)\}_{s\in S},a} u(c_{b1},c_{n1}) + \beta \sum_{s\in S}\pi(s)u\big(u(c_{b2}(s),c_{n2}(s))\big) \quad \text{s.t.}$$

$$c_{b1} + p_1 c_{n1} + a \leq y_1;$$
$$c_{b2}(s) + p_2(s)c_{n2}(s) \leq y_2(s) + Ra, \forall s;$$
$$a \geq 0.$$

Setting up the Lagrangian (assuming that $a \geq 0$ does not bind, and substituting in for c_{b1} and $c_{b2}(s)$ using the budget conditions)

$$\mathcal{L} \equiv u(y_1 - p_1 c_{n1} - a, c_{n1}) + \beta \sum_{s\in S}\pi(s)u\big(y_2(s) + Ra - p_2(s)c_{n2}(s), c_{n2}(s)\big)$$

gives first-order conditions of the form

$$p_1 u_1(c_{b1},c_{n1}) = u_2(c_{b1},c_{n1})$$
$$p_2 u_1\big(c_{b2}(s),c_{n2}(s)\big) = u_2\big(c_{b2}(s),c_{n2}(s)\big)$$
$$u_1(c_{b1},c_{n1}) = \beta R\mathbb{E}\big[u_1(c_{b2},c_{n2})\big].$$

Noting that the functional form of u gives $u_1(c_b,c_n) = c_b^{-\gamma_b}$, the banana intereuler becomes

$$c_{b1}^{-\gamma_b} = \beta R\mathbb{E}[c_{b2}^{-\gamma_b}]$$
$$1 = \beta\mathbb{E}\left[R\left(\frac{c_{b2}}{c_{b1}}\right)^{-\gamma_b}\right].$$

2. Define ε by

$$\beta R\left(\frac{c_{b2}}{c_{b1}}\right)^{-\gamma_b} \equiv e^\varepsilon;$$

therefore, $\mathbb{E}[e^\varepsilon] = 1$. Taking logs,

$$\log\beta + \log R - \gamma_b\log\left(\frac{c_{b2}}{c_{b1}}\right) = \varepsilon,$$

$$\log R = -\log\beta + \gamma_b\log\left(\frac{c_{b2}}{c_{b1}}\right) + \varepsilon.$$

Thus the econometrician can estimate $\hat\gamma_b$ using OLS, since $e^{\mathbb{E}[\varepsilon]} \approx \mathbb{E}[e^\varepsilon] = 1 \implies \mathbb{E}[\varepsilon] \approx 0$.

The intereuler for newspaper consumption is

$$1 = \beta\mathbb{E}\left[R\frac{p_1}{p_2}\left(\frac{c_{n2}}{c_{n1}}\right)^{-\gamma_n}\right].$$

As p_1, $p_2(s)$, and endowments are unobserved, the econometrician cannot recover c_{n2}/c_{n1}. Hence, she cannot estimate γ_n. (Note that I do not know if we can actually *prove* that γ_n is not estimable.)

3. Using the new functional form of u, the banana intereuler becomes

$$\alpha c_{b1}^{\alpha(1-\gamma)-1} c_{n1}^{(1-\alpha)(1-\gamma)} = \beta R \mathbb{E}\left[\alpha c_{b2}^{\alpha(1-\gamma)-1} c_{n2}^{(1-\alpha)(1-\gamma)}\right]$$

$$1 = \beta R \mathbb{E}\left[\left(\frac{c_{b2}}{c_{b1}}\right)^{\alpha(1-\gamma)-1} \left(\frac{c_{n2}}{c_{n1}}\right)^{(1-\alpha)(1-\gamma)}\right].$$

Since period-utility is Cobb-Douglas, expenditure on newspapers is a constant fraction of expenditure on bananas. This implies $c_{b2}/c_{b1} = (p_2 c_{n2})/(p_1 c_{n1})$, hence

$$= \beta R \mathbb{E}\left[\left(\frac{c_{b2}}{c_{b1}}\right)^{-\gamma} \left(\frac{p_1}{p_2}\right)^{(1-\alpha)(1-\gamma)}\right].$$

Hence, if $p_1 = p_2$, γ can be estimated exactly as in part 2. However, the econometrician cannot estimate α from this intereuler. (As above, I do not know if we can actually *prove* that α is not estimable.)

26 Hall's martingale hypothesis with durables practice question

This question comes from an old Economics 211 problem set. It is not solved in these notes, but is a good practice problem for you to work through!

Question

Hall (1978) showed that consumption should follow an AR(1) process, and that no other variable known at time t should influence the expected consumption at time $t + 1$. Mankiw (1982) generalized Hall's results for the case of durable consumption as follows. Suppose consumers have the following preferences over durable goods and non-durables:

$$\sum_{t=0}^{\infty} \sum_{s^t \in S^t} \pi(s^t) \beta^t u\big(K(s^t)\big),$$

where

$$K(s^t) = (1 - \delta)K(s^{t-1}) + c(s^t),$$

K represents the stock of the durable good, and c the acquisition of new durables. Consumers have access to a riskless bond. Labor income is the only source of uncertainty, and the gross interest rate is constant and equal to R. The evolution of assets is given by

$$A(s^t) = RA(s^{t-1}) + y(s^t) - c(s^t).$$

1. Show that the first-order condition for optimality implies (ignoring the possibility of a binding borrowing constraint) that

$$u'\big(K(s^t)\big) = \beta R \sum_{s^{t+1}} \pi(s^{t+1}|s^t) u'\big(K(s^{t+1})\big).$$

2. If u is quadratic, show that

$$c(s^{t+1}) = \gamma_1 + \big(\alpha - (1 - \delta)\big)K(s^t) + \varepsilon(s^{t+1}),$$

where $\alpha = (\beta R)^{-1}$, $\mathbb{E}[\varepsilon(s^{t+1})|s^t] = 0$, and γ_1 is a constant.

3. If u is quadratic, also show that

$$\alpha c(s^t) = \gamma_2 + \big(\alpha - (1-\delta)\big)K(s^t) + (1-\delta)\varepsilon(s^t),$$

and hence

$$c(s^{t+1}) = \gamma_3 \alpha c(s^t) + \varepsilon(s^{t+1}) - (1-\delta)\varepsilon(s^t)$$

where γ_2 and γ_3 are constants.

Acquisition of durables therefore follows an $\mathrm{ARMA}(1,1)$ process.

4. Argue that the same result will obtain if non-durable consumption goods are introduced into the model, with the agent's utility flow given by a separable function

$$u(K, z) = u_K(K) + u_z(z),$$

where K is the stock of durables and z is the consumption of non-durables.

27 General equilibrium in incomplete markets

27.1 A constant absolute risk aversion example

Consider an economy populated by a large number of *ex ante* identical agents. That is, each faces the same stochastic distribution of shocks, but is affected by her shock's idiosyncratic realization. Each agent has constant absolute risk aversion; her preferences are represented by

$$U\big(\{c(s^t)\}_t\big) = \mathbb{E}\left[\sum_t \beta^t u\big(c(s^t)\big)\right] = -\frac{1}{\gamma}\mathbb{E}\left[\sum_t \beta^t e^{-\gamma c(s^t)}\right],$$

where s^t is the history of the consumer's shocks (*not* the history of the world).

Each consumer receives an i.i.d., normally-distributed stochastic endowment each period:

$$y(s^t) \sim N(\bar{y}, \sigma^2).$$

Note that by a law of large numbers, there is no aggregate uncertainty. Since there is no "worst" possible shock, we cannot apply a natural borrowing constraint ($\sigma = -\infty$); we do need to prevent over-borrowing, but will do so through an (unstated) no-Ponzi condition. Furthermore, we must allow consumption to be negative.

The only markets available are for risk-free bonds. The consumer problem is therefore to

$$\max_{\{a(s^t)\}_t} \mathbb{E}\left[\sum_t \beta^t u\big(R_{t-1}a(s^{t-1}) + y(s^t) - a(s^t)\big)\right].$$

Given the stationarity of the problem, we can also write this as a functional equation:

$$V(x) = \max_a \big[u(x - a) + \beta \mathbb{E} V(Ra + y')\big],$$

where x is the "cash on hand" after the realization of the stochastic shock.

Earlier in the term, we discussed several techniques for solving problems like this one. We will proceed by "guessing and verifying;" fortunately, we will start with some very good guesses:

$$V(x) = -\frac{1}{\gamma}e^{-\hat{A}x - \hat{B}},$$

$$c(x) = Ax + B, \text{ and}$$

$$a(x) = (1 - A)x - B.$$

If these guesses are correct, the value function becomes

$$\underbrace{-\frac{1}{\gamma}e^{-\hat{A}x-\hat{B}}}_{=V(x)} = \underbrace{-\frac{1}{\gamma}e^{-\gamma Ax-\gamma B}}_{=u(c)} + \beta\mathbb{E}\Big[\underbrace{-1/\gamma \cdot e^{-\hat{A}(1-A)Rx+\hat{A}BR-\hat{A}y'-\hat{B}}}_{=V(Ra+y')}\Big]$$

$$= -\frac{1}{\gamma}e^{-\gamma Ax-\gamma B} - \frac{\beta}{\gamma}e^{-\hat{A}(1-A)Rx+\hat{A}BR-\hat{B}}\mathbb{E}\big[e^{\hat{A}y'}\big].$$

Since $\hat{A}y' \sim N(-\hat{A}\bar{y}, \hat{A}^2\sigma^2)$, we have $\mathbb{E}[e^{\hat{A}y'}] = e^{-\hat{A}\bar{y}+\hat{A}^2\sigma^2/2}$.

$$= -\frac{1}{\gamma}e^{-\gamma Ax-\gamma B} - \frac{\beta}{\gamma}e^{-\hat{A}(1-A)Rx+\hat{A}BR-\hat{B}-\hat{A}\bar{y}+\hat{A}^2\sigma^2/2}$$

$$e^{-\hat{A}x-\hat{B}} = e^{-\gamma Ax-\gamma B} + \beta e^{-\hat{A}(1-A)Rx+\hat{A}BR-\hat{B}-\hat{A}\bar{y}+\hat{A}^2\sigma^2/2}$$

$$e^{-\hat{B}} = e^{-(\gamma A-\hat{A})x-\gamma B} + \beta e^{(\hat{A}-\hat{A}R+\hat{A}AR)x-\hat{B}-\hat{A}\bar{y}+\hat{A}^2\sigma^2/2}.$$

Since the left-hand side does not depend on x, the right-hand side also must not;[26] this implies $\gamma A - \hat{A} = 0$ and $\hat{A} - \hat{A}R + \hat{A}AR = 0$, or

$$A = \frac{R-1}{R},$$

$$\hat{A} = \gamma\frac{R-1}{R}.$$

Thus if we can pin down R, we know A and $\hat{A}$.[27]

The envelope condition is that

$$V'(x) = u'\big(c(x)\big)$$

$$\frac{\hat{A}}{\gamma}e^{-\hat{A}x-\hat{B}} = e^{-\gamma Ax-\gamma B}$$

$$\log A - \cancel{\hat{A}x} - \hat{B} = \cancel{\hat{A}x} - \gamma B$$

$$\log A + \gamma B = \hat{B}. \tag{36}$$

Thus if we can pin down R and B, we know A, $\hat{A}$, and $\hat{B}$.

Finally, we have the first-order condition:

$$u'\big(c(x)\big) = \beta R\mathbb{E}V'\big(Ra(x)+y'\big)$$

$$e^{-\gamma Ax-\gamma B} = \beta R\mathbb{E}\big[\hat{A}/\gamma \cdot e^{-\hat{A}(Ra(x)+y')-\hat{B}}\big]$$

$$= \beta RAe^{-\hat{A}Ra(x)-\hat{B}}\mathbb{E}\big[e^{-\hat{A}y'}\big]$$

$$= \beta RAe^{-\hat{A}Ra(x)-\hat{B}-\hat{A}\bar{y}+\hat{A}^2\sigma^2/2}.$$

Plugging in $a(x) = (1-A)x - B$ gives

$$= \beta RAe^{-\hat{A}R(1-A)x+\hat{A}RB-\hat{B}-\hat{A}\bar{y}+\hat{A}^2\sigma^2/2}.$$

---

[26]Here we act as if the fact that the *sum* of the right-hand-side terms does not depend on x means that neither term depends on x.

[27]If we want to confirm that our functional form guesses are correct, we should return—after pinning down B and $\hat{B}$—to confirm that $e^{-\hat{B}} = e^{-\gamma B} + \beta e^{-\hat{B}-\hat{A}\bar{y}+\hat{A}^2\sigma^2/2}$. We will not do this.

Taking logarithms and cancelling terms using equation 36,

$$-\gamma Ax - \gamma B = \log(\beta R) + \log A - \hat{A}R(1-A)x + \hat{A}RB - \check{B} - \hat{A}\bar{y} + \hat{A}^2\sigma^2/2$$

$$-\gamma \frac{R-1}{R}x = \log(\beta R) - \gamma\frac{R-1}{R}R\frac{1}{R}x + \gamma\frac{R-1}{R}RB - \gamma\frac{R-1}{R}\bar{y} + \gamma^2\frac{(R-1)^2}{R^2}\cdot\frac{\sigma^2}{2}$$

$$B = -\frac{\log(\beta R)}{\gamma(R-1)} + \frac{\bar{y}}{R} - \gamma\frac{R-1}{R^2}\cdot\frac{\sigma^2}{2}.$$

Thus we can write A, $\hat{A}$, B, and $\hat{B}$ as a function of R. Plugging what we know into the consumption equation gives that

$$c(x) = Ax + B = \underbrace{\frac{R-1}{R}x + \frac{\bar{y}}{R}}_{\text{Permanent income}} - \underbrace{\frac{\gamma(R-1)\sigma^2}{2R^2}}_{\text{Precautionary savings}} - \underbrace{\frac{\log(\beta R)}{\gamma(R-1)}}_{\text{Slope (?)}}.$$

Using this expression and a market clearing condition, we could (in theory) pin down the equilibrium interest rate. One way to do this is to consider the evolution of the agents' cash-in-hand:

$$\begin{aligned}
x' &= Ra(x) + y' \\
&= R\big[(1-A)x - B\big] + y' \\
&= x - RB + y' \\
&= x + R\frac{\log(\beta R)}{\gamma(R-1)} + \frac{\gamma(R-1)\sigma^2}{2R} + (y' - \bar{y}).
\end{aligned}$$

This is a random walk with innovations $(y' - \bar{y})$ and a drift:

$$\mathbb{E}x' = x + \underbrace{R\frac{\log(\beta R)}{\gamma(R-1)} + \frac{\gamma(R-1)\sigma^2}{2R}}_{\text{Drift}}.$$

Note that if $\beta R = 1$, the drift is positive: cash-in-hand (and hence assets) diverge in expectation to positive infinity. In fact, assets must diverge in expectation (to positive or negative infinity) unless the drift term equals zero, or

$$\beta = \frac{1}{R^*}\exp\left(-\gamma^2\left(\frac{R^*-1}{R^*}\right)^2\cdot\frac{\sigma^2}{2}\right).$$

The right-hand side is decreasing in R, so no drift implies a unique value of R^* with $\beta R^* < 1$. It is straightforward to consider comparative statics of this equilibrium interest rate in terms of β, γ, and σ^2.

One final note: this economy does *not* have a steady-state. The distribution of wealth follows a random walk, which means that its unconditional variance increases without bound over time. We can think of this as the gap between the richest and poorest agents—who, recall, are *ex ante* identical—growing endlessly.

27.2 Aiyagari (1994) model

Aiyagari's model is similar in several respects to the CARA example given above. The following differences are important:

- The distribution of shock realizations has bounded support. That means we can define a $y_{\min}$ and $y_{\max}$ with $\Pr(y \notin [y_{\min}, y_{\max}]) = 0$.[28] Because there is a "worst" shock, we can define a natural borrowing

---

[28] We will further insist that $y_{\min}$ and $y_{\max}$ be the highest and lowest values, respectively, for which $\Pr(y \notin [y_{\min}, y_{\max}]) = 0$. This can be written as requiring that for all $\varepsilon > 0$, $\Pr(y \in [y_{\min}, y_{\min} + \varepsilon]) > 0$ and $\Pr(y \in [y_{\max} - \varepsilon, y_{\max}]) > 0$.

constraint: an agent can never borrow so much that he can never repay her debt, even if she receives the worst possible shock from now on; i.e., $a(s^t) \geq -\phi$, where

$$\phi \equiv \frac{y_{\min}}{R-1}.$$

- The agents' coefficients of relative risk aversion must be bounded above. In our earlier example consumers had CARA, which implied *increasing* relative risk aversion. The result was that as they got richer, they saved an increasing fraction of their wealth. We were able to keep average asset holdings bounded, but only for one particular value of $R^* < 1/\beta$; by bounding the coefficient of relative risk aversion, we can keep (expected) assets from growing arbitrarily for all $R < 1/\beta$. (This also relies on the existence of an upper bound on possible shocks.)

The model can be more convenient to analyze when values are renormalized as follows in terms of the borrowing constraint:

	Original	Renormalized
Assets	a	$\hat{a} \equiv a + \phi$
Cash-in-hand	x	$z \equiv x + \phi$
Cash-in-hand evolution	$x' = Ra + y'$	$z' = R\hat{a} + \tilde{y}'$
Endowment	y	$\tilde{y} \equiv y - (R-1)\phi$
Consumption	$c = x - a$	$c = z - \hat{a}$

Although the endowment renormalization looks a bit strange, it takes a form implied by our definitions of $\hat{a}$ and z, along with a desire for z to evolve similarly to how x does.

The representative consumer's problem is characterized by the following functional equation:

$$V(z) = \max_{\hat{a} \geq 0} \left[u(z - \hat{a}) + \beta \mathbb{E} V(\underbrace{R\hat{a} + \tilde{y}'}_{=z'}) \right],$$

which implies the first-order condition

$$u'\big(c(z)\big) \geq \beta R \mathbb{E} V'(R\hat{a}(z) + \tilde{y}').$$

If the borrowing constraint does not bind, this takes the form of our standard intereuler. When the borrowing constraint does bind, the consumer would like to consume more (i.e., lower the left-hand side through c) and save less (i.e., raise the right-hand side through $\hat{a}$). We define (with apologies for the unfortunate notation choice) $\hat{z}$ as the level of cash-in-hand at which the borrowing constraint barely binds.

If $z \leq \hat{z}$, the agent borrows up to the max, so $\hat{a}(z) = 0$ and $c(z) = z$. If $z \geq \hat{z}$, the FOC holds with equality:

$$u'\big(z - \hat{a}\big) = \beta R \mathbb{E} V'(R\hat{a} + \tilde{y}').$$

As we consider going from z up to $\bar{z} > z$, we would need $\hat{a}$ to increase the same amount to keep the left-hand side constant. But this would lower the right-hand side somewhat; thus $\hat{a}$ must increase less than one-for-one with z.

28   Where we have been: a brief reminder

Broadly speaking, the material we have covered in the second half of the term (excluding our initial discussion of optimal taxation) is focused on the the introduction of uncertainty to our models. There are dimensions along which our discussions have varied:

- **Complete vs. incomplete markets**. Under complete markets, agents can buy and sell a complete set of contingent securities. We typically think of these securities as trading in a time-zero (Arrow-Debreu) market, but this is mainly a mathematical convenience—we showed that there is an "equivalence" between the allocations and prices that obtain at equilibrium in this market and in a sequence of period-by-period markets. When agents have access to complete markets, they have the *ability* to hedge all idiosyncratic shocks.[29] We showed that they take advantage of this ability, so that individual consumption can only depend on the history of the world through the history of aggregate shocks.

- **Partial equilibrium vs. general equilibrium**. We have considered two types of economies. The first is small "open" economies, where trade takes place with foreigners; in these markets, asset prices are exogenously set by the global market, and the openness of the market means economy-wide resource constraints don't have as much bite as in closed economies. In closed economies, the price of assets (and notably, therefore, interest rates) are set to clear the asset markets given the economy-wide resource constraint. That is, prices arise endogenously, through general equilibrium.

- **Idiosyncratic risk only vs. aggregate risk**. As discussed when complete markets are available, there is a sense in which only aggregate shocks matter. However, when markets are imcomplete, both idiosyncratic and aggregate shocks are important. We started by considering incomplete markets with only idiosyncratic risk, and then introduced additional aggregate risk.

---

[29]Agents may also be able to hedge aggregate risk if the economy is open or there is a storage technology.