ECON 6130

Macroeconomics Notes

Gabe Sekeres

Fall 2024

Contents

1 Mathieu Taschereau-Dumouchel

Class Organization This is the first half of the course, Ryan Chahrour will teach the second half. Ekaterina Zubova and Zheyang Zhu are the teaching assistants. There is no required textbook, but some recommended textbooks for references. Ljundqvist and Sargent, Recursive Macroeconomic Theory; and Stokey, Lucas, and Prescott, Recursive Methods in Economic Dynamics are the two good references. For general discussion, Romer Advanced Macroeconomics and Blanchard and Fisher Lectures on Macroeconomics are solid.

Class Topics:

- Data and Introduction to Macroeconomics
- Endowment Economies with Complete Markets
- Asset Pricing
- Math, Dynamic Programming, and Numerical Methods
- Production and Investment
- Neoclassical Growth Model

Some advice for this class: Focus on understanding rather than the grade you get. Since this is a PhD course, nobody cares about your grades. Focus on *understanding* the material and build your intuition. Also, *question* the material always – is there a problem with it? Does the model fit the data? Is there a better one?

Quick look at the data World GDP per capita over time – essentially no growth until the Industrial Revolution, and exponential growth since then. Mathieu also notes that people had more freedom in general – think the difference between a serf / peasant and a worker. Also note regional dispersion; Western Europe and their offshoots are larger and grew (and are growing) faster than other places.

We will tend to focus on the US economy. Looking at US output since 1900, some notable facts. Recessions don't seem to matter that much for long-term growth. There was a larger downturn after World War II than anything since then, including the Great Recession. Note that US GDP per capita is essentially the same, though the Great Depression looks more significant – there was still population growth, so output doesn't look as large of a difference.

When you take the log of US GDP per capita, it's remarkably linear. It holds a growth rate of approximately two percent from 1820 to present. Why is it two percent? *We have no idea.* It's not necessarily true for other countries, or necessarily true over time.

GDP is a proxy for welfare more generally. We care about welfare, but there's a strong positive correlation between GDP per capita and 'happiness' by different measures. When we move to non-aggregate measures, the distribution of incomes looks like a normal distribution with an extremely long right tail. It's important to think about how policy affects people rather than *average* people.

When you look at US log rGDP, it looks like the Great Recession actually changed the growth rate lower. Rather than a direct interruption, it looks like it fundamentally changed the economy. That's a really interesting fact. Covid, on the other hand, looks extremely temporary – we went back to the same level and trend basically immediately.

Real personal consumption expenditures look to basically mirror each other. Interestingly, in the dot com recession consumption basically did not change. GDP looks more volatile than consumption – not particularly surprising, as people don't like to have big changes in their consumption. People tend to smooth consumption

over time – literally, the curve looks much smoother. Meanwhile, investment is extremely volatile. It varies a ton, has a similar trend but it's a lot wilder.

Unemployment is a weird measure, since we don't count discouraged workers. If you instead look at the employment-population ratio, the Great Recession looks a lot more significant. Worth thinking about workers aging out of the workforce – were there early retirements? More retirements? NY Fed has tried to disentangle this, it seems extremely complicated.

Inflation rate – typically calculated as the percent change in CPI from one year ago. In the 1970s, there were large spikes – oil shocks (OPEC); bad monetary policy, etc. Why was there inflation during / post covid? (i) supply chain disruptions; (ii) fiscal policy; (iii) Fed dropping interest rates; etc.

What do we do about the data? Well, try and understand it. What drives long-run growth, why are some countries richer than others, etc. We want to answer these questions to affect policy, and improve lives. **Example.** The *Phillips Curve* is a negative relationship between inflation and unemployment. This relationship held for many countries during many periods. This seems great! We can increase inflation to decrease unemployment, and help millions of people!

Oops! It does not work. When central banks tried to exploit the relationship, it disappeared! Why? *Observed relationships are the outcome of agent decisions.* Importantly, when the environment changes, agents change their behavior, and these observed relationships can disappear.

1.1 Some Basic Models

So how do we actually learn anything about the macroeconomy? Well, we could run experiments. We could learn something if we had a *lot* of data. Unfortunately, macro data are sparse and experiments infeasible. Instead, *we build models.*

Definition. *Models* are little laboratories to make sense of the economy. We specify how the world works and how agents behave, we anchor these choices with policy-independent foundations, we use data to discipline the model. Using these models, we can test different policies and make predictions.

There is art to building a good model! Not too descriptive; realism is not the objective. The world is too complicated, and we cannot build models axiomatically without *extremely strong* assumptions. We look for regularities in the data, and build models to understand them, focusing on the essential ingredients to explain the data.

Definition. A *model* is a set of entities making decisions under constraints and interacting with each other. The entities are *households*, who maximize their utility under a budget constraint; *firms*, who use technology to transform inputs into outputs; and *governments* who use policies (normative vs positive).

We also need to specify the set of commodities being traded, the information structure, and the market structure.

Model. *(Partial Equilibrium)* Consumption-saving model of a household:

$$\begin{aligned}
\max_{\{c_t\},\{b_{t+1}\}} \sum_{t=0}^{\infty} \beta^t u(c_t) \\
\text{s.t. } & c_t + b_{t+1} \leq y_t + R_t b_t \\
& b_{t+1} \geq -A \\
& c_t \geq 0 \\
& b_0, y_t, R_t \text{ given}
\end{aligned}$$

where $u' > 0$, $u'' < 0$, $\lim_{c_t \to 0} u'(c_t) = \infty \; \forall \, t \in \mathbb{N}$, $0 < \beta < 1$, A is large.

Conclusion This is essentially a class about making models. We will of course focus on macroeconomic models, but the lessons here should apply to different areas of focus – for example, you might use a simulation model in Micro theory; or a model of firms in Industrial Organization.

Last time, we introduced the consumption-saving model of a household. Note that y_t and R_t are given – the income flow and the interest rate. We will endogenize these later, but for now take them as exogenous. Note that the constraint that 'A is large' will not bind – A is large enough to not bind.

How do we solve? First, note that the non-negative consumption constraint will not bind, and the budget constraint will. We will use KKT, with the Lagrangian:

$$\mathcal{L} = \sum_{t=0}^{\infty} \beta^t u(y_t + R_t b_t - b_{t+1})$$

We will optimize with respect to b_{t+1}, because it is the thing chosen in period t. Fix t, and differentiate w/r/t b_{t+1}. Fix $t = 2$, and we get

$$\mathcal{L}' = \beta u'(y_1 + R_1 b_1 - b_2) + \beta^2 u'(y_2 + R_2 b_2 - b_3) R_2 = 0$$

and we get

$$u'(c_1) = \beta u'(c_2) R_2 \quad \text{(and in general: } u'(c_t) = \beta u'(c_{t+1}) R_{t+1})$$

Note that these conditions are necessary, but not sufficient, as we do not have a transversality condition. This will hold in finite time, but not necessarily infinite time. Also note that if $\beta R_{t+1} = 1 \; \forall \, t$, $c_t = c \; \forall \, t \geq 0$.

Example. Solve for c_t, assume that $R_t = R = \frac{1}{\beta} \; \forall \, t \geq 0$, so $c_t = c$. We will rewrite the budget constraint isolating b_t, so we have $b_t = \frac{1}{R}(c + b_{t+1} - y_t)$. We have y_t, R given, and can iterate to find b_0:

$$b_0 = \frac{c - y_0}{R} + \frac{c - y_1}{R^2} + \frac{c - y_2}{R^3} + \cdots = \sum_{t=0}^{\infty} \frac{c - y_t}{R^{t+1}} + \lim_{t \to \infty} \frac{b_{t+1}}{R^{t+1}}$$

$\lim_{t \to \infty} \frac{b_{t+1}}{R^{t+1}} = 0$ is the transversality condition, where if it is nonzero it means that you are not optimizing. We get that:

$$c = \underbrace{(1 - \beta)}_{\text{Propensity to consume}} \underbrace{\left(R \cdot b_0 + \sum_{t=0}^{\infty} \frac{y_t}{R^t} \right)}_{\text{Lifetime income}}$$

This example has essentially been a re-derivation of the Permanent Income Hypothesis (Friedman 1957) – note that we have just shown that people smooth their consumption over time, which is exactly what Friedman predicts. What's incredibly nice is that these results do not depend on the utility function beyond the basic increasing and concave conditions.

Model. *(General Equilibrium)* We have a pure exchange economy, meaning there are no firms, with two agents who live forever and discrete time indexed by $\mathbb{N}$. Each period, agents can trade the unique, nonstorable good. Agent $i \in \{1, 2\}$ has the utility function

$$u(c^i) = \sum_{t=0}^{\infty} \beta^t \log(c_t^i)$$

for some $\beta \in (0, 1)$. Endowments are given by

$$e_t^1 = \begin{cases} 2 & \text{if } t \text{ is even} \\ 0 & \text{if } t \text{ is odd} \end{cases}$$

$$e_t^2 = \begin{cases} 0 & \text{if } t \text{ is even} \\ 2 & \text{if } t \text{ is odd} \end{cases}$$

Definition. A *competitive Arrow-Debreu equilibrium* is a set of prices $\{\hat{p}_t\}_{t=0}^{\infty}$ and allocations $(\{\hat{c}_i^t\}_{t=0}^{\infty})_{i=1,2}$ such that

1. Given $\{\hat{p}_t\}_{t=0}^{\infty}$, for $i = 1, 2$, $(\{\hat{C}_i^t\}_{t=0}^{\infty})_{i=1,2}$ solves

$$\max_{\{c_t^i\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \log(c_t^i)$$

   subject to

$$\sum_{t=0}^{\infty} \hat{p}_t c_t^i \leq \sum_{t=0}^{\infty} \hat{p}_t e_t^i$$
$$c_t^i \geq 0 \text{ for all } t$$

2. Market clearing

$$\hat{c}_t^1 + \hat{c}_t^2 = e_t^1 + e_t^2 \text{ for all } t$$

Definition. Some terms:

- Competitive: no one has any influence over prices
- Arrow-Debreu: this structure of endowment equilibria, where we meet before the process begins and make trades
- Equilibrium: everyone maximizes their utility, and markets clear

How do we solve this model? We solve each agent's individual maximization problem, and then solve for the equilibrium.
Solution. For each agent i, we have:

$$\mathcal{L} = \sum_{t=0}^{\infty} \beta^t \log(c_t^i) - \lambda_i \left(\sum_{t=0}^{\infty} p_t c_t^i - \sum_{t=0}^{\infty} p_t e_t^i \right)$$

The first order conditions give us:

$$\beta^t = p_t c_t^i \lambda_i \ \forall \ t$$

Combining with the budget constraint, we can get $c_t^i(\{p_t\}, \{e_t^i\})$. The first order conditions for t and $t+1$ give us a way to eliminate λ_i. In the simple case:

$$p_{t+1}(c_{t+1}^1 + c_{t+1}^2) = \beta p_t (c_t^1 + c_t^2)$$

and, with market clearing:

$$p_{t+1}(e_{t+1}^1 + e_{t+1}^2) = \beta p_t (e_t^1 + e_t^2)$$

Thus, $p_{t+1} = \beta p_t$, and then $p_t = \beta^t p_0$. We can set $p_0 = 1$, and get that equilibrium prices are $\hat{p}_t = \beta^t$. Also, recall from the FOCs that we have $p_{t+1} c_{t+1}^i = \beta p_t c_t^i$, so we have that $c_{t+1}^i = c_t^i \ \forall \ t$. The values of the endowments are

$$\sum_{t=0}^{\infty} \hat{p}_t e_t^1 = 2 \sum_{t=0}^{\infty} \beta^{2t} = \frac{2}{1-\beta^2}$$

$$\sum_{t=0}^{\infty} \hat{p}_t e_t^2 = 2\beta \sum_{t=0}^{\infty} \beta^{2t} = \frac{2\beta}{1-\beta^2}$$

And thus, equilibrium consumptions are

$$c_t^1 = \frac{2}{1+\beta} \qquad , \qquad c_t^2 = \frac{2\beta}{1+\beta}$$

Some notes on this equilibrium Agent 1 is better off, because they get the endowment in the first period (period 0), so their goods are more valuable. They can leverage this endowment to consume more than agent 2 in every period.

Trade is clearly welfare-improving – in the absence of trade, agent 2 would starve in every even period, and agent 1 would starve in every odd period. Instead, they can smooth their consumption. Indeed, trade is always weakly welfare-improving; you can always consume your endowment if you would not improve from trading.

This allocation *is* socially optimal in the sense that it is Pareto efficient. Some benevolent dictator could not make one better off without making the other worse off. However, note that a utilitarian dictator could improve *total* utility, just not in a Pareto efficient way.

If trade was sequential instead of Arrow-Debreu, the consumption in each period would not change. Note that this is *not* a game – we don't have parameters for that, and people interact only through the market.

1.2 Endowment Economies with Complete Markets

1.2.1 Models of Complete Markets

Model. *(Probabilistic World)* In each period $t \geq 0$, a stochastic event $s_t \in S$ is realized. Denote $s^t = \{s_0, \ldots, s_t\}$ to be a history up to and including time t. The *unconditional* probability of observing s^t is given by the measure $\pi_t(s^t)$. The *conditional* probability of observing s^t given that some s^τ happened is $\pi_t(s^t \mid s^\tau)$. Assume that a given s_0 occurred before trading started, with $P\{s_0\} = 1$.
Model. *(The Economy Environment)* There are I agents indexed by $i = 1, \ldots, I$. Agent i owns a stochastic endowment of goods $y_t^i(s^t)$. Household i values a history-dependent consumption plan $c^i = \{x_t^i(s^t)\}_{t=0}^\infty$ according to

$$U(c^i) = \sum_{t=0}^{\infty} \sum_{s^t} \beta^t u(c_t^i(s^t))\pi_s(s^t)$$

where $u' > 0$, $u'' < 0$, and $\lim_{c \to 0} u'(c) = \infty$.
Definition. A *feasible allocation* satisfies

$$\sum_i c_t^i(s^t) \leq \sum_i y_t^i(s^t) \; \forall \; t, s^t$$

Trading arrangements Suppose that each household evolves in autarky. Their consumption $c_t(s^t) = y_t(s^t)$, which is dependent on s^t. We will study two types of trading arrangements:

1. *Arrow-Debreu securities*: at $t = 0$, households trade claims to consumption at all time $t > 0$, contingent on all possible histories up to time t, s^t. There is no trade at time $t > 0$.
2. *Sequential markets*: trade occurs at each $t \geq 0$. Trades for history s^{t+1}-contingent $t + 1$ goods occur only at node s^t.

Definition. An allocation $\{c^i\}_{i=1}^I$ is *(Pareto) efficient* if there is no feasible allocation $\{\tilde{c}^i\}_{i=1}^I$ such that

$$U(\tilde{c}^i) \geq U(c^i) \; \forall \; i \text{ and } \exists \; i \text{ s.t. } U(\tilde{c}^i) > U(c^i)$$

Proposition 1.1. *An allocation is Pareto efficient if and only if it solves the social planner's problem*

$$\max_{\{c^i\}_i} \sum_{i=1}^{I} \lambda_i U(c^i), \text{ where all } c^i \in \{c^i\}_i \text{ are feasible,}$$

for some non-negative λ_i for all i. The λ's are the Pareto weights.

Solving the planner's problem Lagrangian, where $\theta_t(s^t) \geq 0$ are the Lagrange multipliers:

$$\mathcal{L} = \sum_{t=0}^{\infty} \sum_{s^t} \left(\sum_{i=1}^{I} \lambda_i \beta^t u(c_t^i(s^t))\pi_t(s^t) + \theta_t(s^t) \sum_{i=1}^{I} [y_t^i(s^t) - c_t^i(s^t)] \right)$$

The first order conditions give us:

$$\lambda_i \beta^t u'(c_t^i(s^t))\pi_t(s^t) = \theta_t(s^t)$$

Therefore:

$$c_t^i(s^t) = u'^{-1}(\lambda_i^{-1}\lambda_1 u'(c_t^1(s^t)))$$

and

$$\sum_i u'^{-1}(\lambda_i^{-1}\lambda_1 u'(c_t^1(s^t))) = \sum_i y_t^i(s^t)$$

Note that Agent 1's consumption depends on the aggregate endowment, not their own endowment in particular. It *does* depend on the weight the planner places on them, but not necessarily on their own endowment all that much.

Also note that if aggregate endowments don't change, consumption will not change. Consumption is smoothed across time and across states of the world, given that the aggregate endowment does not change.

Definition. An *Arrow-Debreu equilibrium* is a sequence of allocations $\{c_t^i(s^t)\}_{t=0}^{\infty}$ for all agents i and prices $\{q_t^0(s^t)\}_{t=0}^{\infty}$ such that

1. Given prices, household i's allocation solves its maximization problem

$$\max_{\{c_t^i(s^t)\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \sum_{s^t} \beta^t u(c_t^i(s^t))\pi_t(s^t)$$
$$\text{s.t.} \sum_{t=0}^{\infty} \sum_{s^t} q_t^0(s^t) c_t^i(s^t) \leq \sum_{t=0}^{\infty} \sum_{s^t} q_t^0(s^t) y_t^i(s^t)$$

2. The allocation is feasible (markets clear)

Solving the equilibrium Each agent's FOC is

$$\beta^t u'(c_t^i(s^t))\pi_t(s^t) = \mu_i q_t^0(s^t)$$

where μ_i is the Lagrange multiplier on the budget constraint. Therefore,

$$c_t^i(s^t) = u'^{-1}\left(u'(c_t^1(s^t))\frac{\mu_i}{\mu_1}\right)$$

and

$$\sum_i u'^{-1}\left(u'(c_t^1(s^t))\frac{\mu_i}{\mu_1}\right) = \sum_i y_t^i(s^t)$$

Remark. At the ADE allocation, the shadow prices $\theta_t(s^t)$ are equal to $q_t^0(s^t)$.

Theorem 1.1. (First Welfare Theorem) *Any Arrow-Debreu equilibrium allocation is efficient.*

Proof. (sketch) Set $\lambda_i = \mu_i^{-1}$ and normalize the weights. Need to check the RC, and the shadow prices $\theta_t(s^t) = q_t^0(s^t)$.

From MWG Proposition 5.F.1 (pg. 150), we have that if $y \in Y$ is profit-maximizing for some price vector $p \gg 0$, then y is efficient. It suffices to show that any Arrow-Debreu equilibrium allocation is utility-maximizing. Suppose FSOC that for some consumer, there exists feasible $\{c_n'(s^t)\}_{n=0}^{\infty}$ such that $\sum_{t=0}^{\infty}\sum_{s^t}\beta^t u(c_t'(s^t))\pi_t(s^t) > \sum_{t=0}^{\infty}\sum_{s^t}\beta^t u(c_t(s^t))\pi_t(s^t)$. Since markets clear, and the utility function is strictly increasing, this would imply that $\sum_{t=0}^{\infty}\sum_{s^t} c_t(s^t) < \sum_{t=0}^{\infty}\sum_{s^t} c_t'(s^t)$, and since markets clear we have that $\sum_{t=0}^{\infty}\sum_{s^t} c_t(s^t) = \sum_{t=0}^{\infty}\sum_{s^t} y_t(s^t)$. This implies that $\sum_{t=0}^{\infty}\sum_{s^t} c_t'(s^t) > \sum_{t=0}^{\infty}\sum_{s^t} y_t(s^t)$, meaning that $\{c_t'(s^t)\}_{t=0}^{\infty}$ is not feasible, which is a contradiction. □

Theorem 1.2. (Second Welfare Theorem) *Let $\{c_t^i(s^t,\lambda)\}_{t=0}^{\infty}$ be an efficient allocation for some Pareto weights $\{\lambda^i\}_{i=1}^{\infty}$. Then there exist transfers $\{\tau^i\}_{i=1}^{I}$ such that the allocation is an Arrow-Debreu equilibrium.*

Proof. See MWG Proposition 16.D.1 (pg. 552), which demonstrates that any efficient allocation can be transformed into a price quasi-equilibrium, and then Proposition 16.D.3 (pg. 555-6) which shows that any price quasi-equilibrium can be transformed into a true price equilibrium. □

Definition. *Negishi's Method*, from Negishi (1960), gives us a way to easily find the set of Arrow-Debreu equilibria:

1. Compute all efficient allocations (the social planner's problem)

2. The first welfare theorem says that the set of competitive allocations are contained in the set of efficient allocations

3. Isolate the efficient allocations that are also competitive allocations

Example. Recall the 2-agent economy with endowments $(2, 0)$ in each period. With Pareto weight $\alpha \in [0, 1]$, the social planner's problem is

$$\begin{aligned} \max_{c_1, c_2} \sum_{t=0}^{\infty} \beta^t \left(\alpha \log(c_t^1) + (1-\alpha)\log(c_t^2)\right) \\ \text{s.t. } c_t^i \geq 0 \; \forall \, i, \; \forall \, t \\ c_t^1 + c_t^2 = e_t^1 + e_t^2 \equiv 2, \; \forall \, t \end{aligned}$$

Attach multipliers $\theta_t/2$ to the resource constraints. The FOCs are

$$\frac{\alpha\beta^t}{c_t^1} = \frac{\theta_t}{2} \qquad , \qquad \frac{(1-\alpha)\beta^t}{c_t^2} = \frac{\theta_t}{2}$$

and therefore

$$c_t^1 = \frac{\alpha}{1-\alpha} c_t^2$$

Combining with the resource constraints, we get

$$\begin{aligned} c_t^1(\alpha) &= 2\alpha \\ c_t^2(\alpha) &= 2(1-\alpha) \\ \theta_t &= \beta^t \end{aligned}$$

So there is a continuum of efficient allocations. However, there was a unique solution when we solved this problem earlier. There should be an extra condition that will help us select from the set of efficient allocations. We can add the budget constraints:

$$t^i(\alpha) = \sum_t \theta_t [c_t^i(\alpha) - e_t^i]$$

We will look for $\alpha^\star$ such that $t^1(\alpha^\star) = t^2(\alpha^\star) = 0$. We have

$$\begin{aligned} t^1(\alpha) &= \sum_t \theta_t [c_t^1(\alpha) - e_t^1] = \sum_t \beta^t (2\alpha - e_t^1) = \frac{2\alpha}{1-\beta} - \frac{2}{1-\beta^2} \\ t^2(\alpha) &= \sum_t \theta_t [c_t^2(\alpha) - e_t^2] = \sum_t \beta^t (2(1-\alpha) - e_t^2) = \frac{2(1-\alpha)}{1-\beta} - \frac{2\beta}{1-\beta^2} \end{aligned}$$

Thus, the solution is $\alpha^\star = \frac{1}{1+\beta}$, and the consumptions are

$$c_t^1 = \frac{2}{1+\beta} \qquad \text{and} \qquad c_t^2 = \frac{2\beta}{1+\beta}$$

which is the same solution as when we solved for the Arrow-Debreu Equilibrium.

Example. (Solving the Equilibrium with no Aggregate Uncertainty) Suppose that there is no aggregate uncertainty and $I = 2$. Let the stochastic events $s_t \sim U([0, 1])$ be independent across time. Suppose that the endowments are $y_t^1(s^t) = s_t$ and $y_t^2(s^t) = 1 - s_t$. How do $c_t^i(s^t)$ vary across time? From the FOC, we have that

$$q_t^0(s^t) = \beta^t \pi_t(s^t) \frac{u'(c^i)}{\mu_i}$$

We can use the household budget constraint to write:

$$c^i = (1-\beta)\sum_{t=0}^{\infty}\sum_{s^t}\beta^t\pi_t(s^t)y_t^i(s^t)$$

1.2.2 Asset Pricing

The main idea is that the price of an asset reflects some information about the potential states of the world. Of course, we don't see Arrow-Debreu securities in the real world. However, we do see more complicated securities, which we can recreate using Arrow-Debreu securities.

Definition. Suppose that we have an asset that provides dividends $\{d_t(s^t)\}_{t=0}^{\infty}$, what should the price of this asset be?

$$p_0^0 = \sum_{t=0}^{\infty}\sum_{s^t} q_t^0(s^t)d_t(s^t)$$

What is the price of an asset that pays 1 in each t regardless of s^t? Set $d_t(s^t) = 1$, and we get

$$\sum_{t=0}^{\infty}\sum_{s^t} q_t^0(s^t)$$

What is the price of an asset that pays 1 at period τ regardless of s^τ?

$$\sum_{s^\tau} q_\tau^0(s^\tau)$$

What is the time 0 price of an asset that entitles the owner to dividend stream $\{d_t(s^t)\}_{t\geq\tau}$ if history s^τ is realized?

$$p_\tau^0(s^\tau) = \sum_{t\geq\tau}\sum_{s^t|s^\tau} q_t^0(s^t)d_t(s^t)$$

The units of the price are time 0 goods: $q_0^0(s_0) = 1$. To convert the price into units of time τ, history s^τ consumption goods, we must divide by $q_\tau^0(s^\tau)$:

$$p_\tau^\tau(s^\tau) = \frac{p_\tau^0(s^\tau)}{q_\tau^0(s^\tau)} = \sum_{t\geq\tau}\sum_{s^t|s^\tau}\frac{q_t^0(s^t)}{q_\tau^0(s^\tau)}d_t(s^t)$$

Notice that (using the FOCs), ($q_t^\tau(s^t)$ is the price of one unit of s^t goods in terms of s^τ goods)

$$q_t^\tau(s^t) = \frac{q_t^0(s^t)}{q_\tau^0(s^\tau)} = \frac{\beta^t u'(c_t^u(s^t))\pi_t(s^t)}{\beta^\tau u'(c_\tau^u(s^\tau))\pi_\tau(s^\tau)} = \beta^{t-\tau}\frac{u'(c_t^i(s^t))}{u'(c_\tau^i(s^\tau))}\pi_t(s^t \mid s^\tau)$$

Remember that from Bayes' Law:

$$\pi_t(s^t \mid s^\tau)\times\pi_\tau(s^\tau) = \pi_t(s^t, s^\tau) = \pi_t(s^t)$$

So we can write

$$p_\tau^\tau(s^\tau) = \sum_{t\geq\tau}\sum_{s^t|s^\tau} q_t^\tau(s^t)d_t(s^t)$$

So why did we go to all this trouble? The price of equity in time τ in state s^τ. We have

$$q_{\tau+1}^\tau(s^{\tau+1}) = \beta\frac{u'(c_{\tau+1}^i(s^{\tau+1}))}{u'(c_\tau^i(s^\tau))}\pi_{\tau+1}(s^{\tau+1} \mid s^\tau)$$

Intuitively, what is this quantity and why is it useful? It's the *pricing kernel*. We can write the price at time τ in history s^τ of a claim to a random payoff $\omega(s_{\tau+1})$ as

$$p_\tau^\tau(s^\tau) = \sum_{s_{\tau+1}} q_{\tau+1}^\tau(s^{\tau+1})\omega(s_{\tau+1}) = \mathop{\mathbb{E}}_\tau\left[\beta\frac{u'(c_{\tau+1})}{u'(c_\tau)}\omega(s_{\tau+1})\right]$$

Defining the gross return $R_{\tau+1} \coloneqq \omega(s_{\tau+1})/p_\tau^\tau(s^\tau)$, we can write

$$1 = \mathop{\mathbb{E}}_\tau\left[\beta\frac{u'(c_{\tau+1})}{u'(c_\tau)}R_{\tau+1}\right] \equiv \mathop{\mathbb{E}}_\tau\left[m_{\tau+1}R_{\tau+1}\right]$$

The term $m_{\tau+1}$ is called the *stochastic discount factor*.

Previously, we've done everything in Arrow-Debreu. The math is really nice, but it's not so realistic. We move to a more realistic setting

Model. *(Sequential Trading)* An *Arrow security* (distinct from Arrow-Debreu securities) has that at each date $t \geq 0$, trade occurs in a set of claims to one-period-ahead state-contingent consumption. Are markets complete? Yes, sequentially complete. Define the *natural debt limit*, where q_τ^t are the Arrow-Debreu prices, as

$$A_t^i(s^t) \coloneqq \sum_{\tau=t}^\infty \sum_{s^\tau|s_t} q_\tau^t(s^\tau)y_\tau^i(s^\tau)$$

The intuition is: Household i at time $t-1$ cannot promise to pay more than $A_t^i(s^t)$ at time t in state s^t, otherwise their consumption would be negative. Denote by $\tilde{a}_t^i(s^i)$ the claims to time t consumption, on top of its endowment, that agent i gets at time t in state s^t. Denote by $\tilde{Q}_t(s^{t+1} \mid s^t)$ the price to a claim of one unit of consumption at time $t+1$ in state s^{t+1} when the current history is s^t.

The objective function of households is unchanged. Using the new notation, their budget constraint is

$$\tilde{c}_t^i(s^t) + \sum_{s^{t+1}} \tilde{a}_{t+1}^i(s^{t+1}, s^t)\tilde{Q}_t(s^{t+1} \mid s^t) \leq y_t^i(s^t) + \tilde{a}_t^i(s^t)$$

To rule out Ponzi schemes, we impose the constraint that

$$-\tilde{a}_{t+1}^i(s^{t+1}) \leq A_{t+1}^i(s^{t+1})$$

This is not the only condition that would work.

Definition. A *sequential trading competitive equilibrium* is a distribution of assets $\tilde{a}_{t+1}^i$ for all i and t, an allocation $\{\tilde{c}^i\}$ for all i, and pricing kernels $\tilde{Q}_t(s^{t+1} \mid s^t)$ such that

- For all i, $\tilde{c}^i$ solves household i's problem
- For all $\{s^t\}_{t=0}^\infty$, we have $\sum_i \tilde{c}_t^i(s^t) = \sum_i y_t^i(s^t)$ and $\sum_i \tilde{a}_{t+1}^i(s^{t+1}, s^t) = 0$.

Solving the Equilibrium The Lagrangian is

$$\mathcal{L}^i = \sum_{t=0}^\infty \sum_{s^t} \left(\beta^t u[\tilde{c}_t^i(s^t)]\pi_t(s^t)\right) + \eta_t^i(s^t)\left(y_t^i(s^t) + \tilde{a}_t^i(s^t) - \tilde{c}_t^i(s^t) - \sum_{s^{t+1}} \tilde{a}_{t+1}^i(s^{t+1}, s^t)\tilde{Q}_t(s^{t+1} \mid s^t)\right)$$
$$+ \left(\sum_{s^{t+1}} \nu_t^i(s^t, s^{t+1})(A_{t+1}^i(s^{t+1}) + \tilde{a}_{t+1}^i(s^{t+1})\right)$$

The First Order Conditions are:

$$\beta^t u'[\tilde{c}_t^i(s^t)]\pi_t(s^t) - \eta_t^i(s^t) = 0$$
$$-\eta_t^i(s^t)\tilde{Q}_t(s^{t+1} \mid s^t) + \nu_t^i(s_t, s^{t+1}) + \eta_{t+1}^i(s^{t+1}, s^t) = 0$$

Since the borrowing constraint will not bind, we can say that $\nu_t^i(s_t, s^{t+1}) = 0$. Further playing with the FOCs, we get

$$\tilde{Q}_t(s^{t+1} \mid s^t) = \beta \frac{u'(\tilde{c}_{t+1}^i(s^{t+1}))}{u'(\tilde{c}_t^i(s^t))}\pi(s^{t+1} \mid s^t)$$

So this pricing kernel is the exact same as in the Arrow-Debreu Equilibrium!

Proposition 1.2. *Let $\{c_t^i(s^t)\}_{t=0}^{\infty}$ be an Arrow-Debreu Equilibrium allocation with associated prices $\{q_t^0(s^t)\}_{t=0}^{\infty}$. Then the pricing kernel given by $q_{t+1}^0(s^{t+1}) = \tilde{Q}_t(s^{t+1} \mid s^t)q_t^0(s^t)$, the consumption $\tilde{c}_t^i(s^t) = c_t^i(s^t)$, and associated assets holdings form a Sequential Trading Equilibrium.*

Proof. See LS chapter 8 □

Remark. The converse is also true. Intuitively, both market structures allow agents to move resources across histories.

Recap of findings:

- The set of equilibria is the same under Arrow-Debreu and Sequential Trading
- Competitive allocations are solutions to a social planner problem (they are Pareto Efficient)
- We can decentralize any Pareto efficient with a set of lump sum transfers
- Pricing kernels allow us to price any securities

1.3 Neoclassical Growth Model

We have previously been assuming that endowments just "fell from the sky" (often called the Robert Lucas Tree Model). This is unsatisfying, both in general and also because it makes it impossible to ask questions about policy.

We are moving to essentially a production economy – the economy is a factory that takes inputs (labor and capital, typically) and produces some consumption goods. There are of course more complicated models, but this is typical. The Neoclassical Growth Model is essentially the most simple growth model. We will use growth as a gateway to study dynamic programming.

Facts about long-run growth Kaldor (1961) popularized the following six stylized facts concerning long run economic growth:

1. Output per capita, Y/N, grows at a constant rate
2. The capital to labor ratio, K/N, grows at a constant rate
3. The interest rate, R, is fairly constant
4. The output to capital ratio, Y/K, is fairly constant
5. The share of value added going to labor and capital is fairly constant
6. There are wide dispersions in Y_i/N_i across countries

These facts generally still hold, with some exceptions. Famously, (5) has been falling recently, meaning less wealth is going to workers and more to capital. This is an issue, as it exacerbates inequality.

Model. *(Discrete Time Neoclassical Growth)* Time is discrete, $t = 0, 1, \ldots$. In each period, three goods are traded: labor services n_t, capital services k_t, and final good output y_t that can be consumed (c_t) or invested (i_t). We have an aggregate production function F, where output $y_t = F(k_t, n_t)$ is either consumed or invested $y_t = c_t + i_t$. Investment increases capital stock which depreciates at rate $\delta > 0$, so $k_{t+1} = (1-\delta)k_t + i_t$. We have a large number of identical, infinitely lived households with utility

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

Each household is endowed with initial capital k_0, and one unit of time each period.

We will concern ourselves with *optimal growth*. We will study the problem of a social planner who maximizes total welfare.

Definition. An allocation $\{c_t, k_{t+1}, n_t\}_{t=0}^{\infty}$ is *feasible* if, for all $t \geq 0$,

$$\begin{aligned} F(k_t, n_t) &= c_t + k_{t+1} - (1-\delta)k_t \\ c_t &\geq 0, k_t \geq 0, 0 \leq n_t \leq 1 \\ k_0 &\text{ given} \end{aligned}$$

Definition. An allocation $\{c_t, k_{t+1}, n_t\}_{t=0}^{\infty}$ is *Pareto efficient* if it is feasible and there is no other feasible allocation $\{\hat{c}_t, \hat{k}_{t+1}, \hat{n}_t\}_{t=0}^{\infty}$ such that

$$\sum_{t=0}^{\infty} \beta^t U(\hat{c}_t) > \sum_{t=0}^{\infty} \beta^t U(c_t)$$

The social planner solves

$$w(k_0) = \max_{\{c_t, k_{t+1}, n_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t U(c_t)$$

subject to

$$\begin{aligned} F(k_t, n_t) &= c_t + k_{t+1} - (1-\delta)k_t \\ c_t &\geq 0, k_t \geq 0, 0 \leq n_t \leq 1 \\ k_0 &\text{ given} \end{aligned}$$

We make the following assumptions.

- Utility:
  - U is continuously differentiable, strictly increasing, strictly concave, and bounded.
  - Inada conditions: $\lim_{c\to 0} U'(c) = \infty$ and $\lim_{c\to\infty} U'(c) = 0$.
  - $\beta \in (0,1)$
- Production function:
  - F is continuously differentiable and homogenous of degree 1, strictly increasing and strictly concave.
  - $F(0,n) = F(k,0) = 0$ for all $k, n > 0$.
  - Inada conditions: $\lim_{k\to 0} F_k(k,1) = \infty$ and $\lim_{k\to\infty} F_k(k,1) = 0$

These assumptions imply that from the structure of U, $n_t = 1$ for all t. We can write

$$f(k) = F(k,1) + (1-\delta)k$$

Since $c_t = f(k_t) - k_{t+1}$, we can write the social planner problem as:

$$w(k_0) = \max \{c_t, k_{t+1}, n_t\}_{t=0}^{\infty} \sum_{t=0}^{\infty} \beta^t U(f(k_t) - k_{t+1})$$

subject to

$$\begin{aligned} 0 \leq k_{t+1} &\leq f(k_t) \\ k_0 &\text{ given} \end{aligned}$$

Why do we care about this problem? Because the Welfare Theorems apply here! By solving the social planner's problem, we can solve for the competitive equilibrium. Specifically, since this is an infinite-dimensional optimization problem, we can use dynamic programming to rewrite it in a much simpler form. We write, recursively, that:

$$\begin{aligned} w(k_0) &= \max_{\{k_{t+1}\}_{t=0}^{\infty} \text{ s.t. } 0 \leq k_{t+1} \leq f(k_t)} \sum_{t=0}^{\infty} \beta^t U(f(k_t) - k_{t+1}) \\ &= \max_{k_1 \text{ s.t. } 0 \leq k_1 \leq f(k_0)} U(f(k_0) - k_1) + \beta \left(\max_{\{k_{t+1}\}_{t=1}^{\infty} \text{ s.t. } 0 \leq k_{t+1} \leq f(k_t)} \sum_{t=1}^{\infty} \beta^{t-1} U(f(k_t) - k_{t+1}) \right) \end{aligned}$$

Intuitively, this looks like we have that

$$w(k_0) = \max_{k_1 \text{ s.t. } 0 \le k_1 \le f(0)} U(f(k_0) - k_1) + \beta(w(k_1))$$

This is a functional problem – the solution to this equation is the function w, not a vector of capital. This is complicated, but much simpler than finding $\{k_{t+1}\}_{t=0}^{\infty}$.

Definition. Denote by $v(\cdot)$ the *value function* for this new formulation of the problem:

$$v(k) = \max_{0 \le k' \le f(k)} \{U(f(k) - k') + \beta v(k')\} \tag{1}$$

Interpretation: $v(k)$ is the discounted lifetime utility of the representative agent, from the current period onward, if the social planner has initial capital stock k and allocated consumption optimally. This is the *recursive formulation* of the planner's problem, where (1) is the *Bellman Equation*, k is a state variable which completely describes the economy today, and k' is called the control variable, which is decided by the planner today. To solve (1), we need a value function v, and a policy function g where $k' = g(k)$.

This approach raises some questions:

1. Under what conditions does a solution to (1) exist, and is it unique? Contraction Mapping Theorem (formally, Banach's Fixed Point Theorem).
2. Under what conditions can we solve (1) and be sure that we've solved the sequential problem (*i.e.* when does $v = w$ and $g(k)$ generate the optimal $\{k_{t+1}\}_{t=0}^{\infty}$)? Bellman's Principle of Optimality.
3. Is there a simple algorithm that allows us to solve (1)? Contraction Mapping Theorem

Example. A simple recursive case. Let $U(c) = \ln c$, $F(k, n) = k^{\alpha} n^{1-\alpha}$, and $\delta = 1$. Then $f(k) = k^{\alpha}$ and

$$v(k) = \max_{0 \le k' \le k^{\alpha}} \{\ln(k^{\alpha} - k') - \beta v(k')\}$$

How do we solve this? Let's go by technique.

1. Guess and verify. We guess that $v(k)$ has the form

   $$v(k) = A + B \ln(k)$$

   for some constants A, B. The maximization problem is now:

   $$\max_{0 \le k' \le k^{\alpha}} \{\ln(k^{\alpha} - k') - \beta(A + B \ln(k'))\}$$

   and the FOC is

   $$k'^{*} = \frac{\beta B k^{\alpha}}{1 + \beta B}$$

   We next plug the optimal back into the Bellman Equation:

   $$\begin{aligned} v(k) &= \max_{0 \le k' \le k^{\alpha}} \{\ln(k^{\alpha} - k') + \beta v(k')\} \\ &= \ln(k^{\alpha} - k'^{*}) + \beta(A + B \ln(k'^{*})) \\ &= \ln\left(\frac{k^{\alpha}}{1 + \beta B}\right) + \beta A + \beta B \ln\left(\frac{\beta B k^{\alpha}}{1 + \beta B}\right) \\ &= -\ln(1 + \beta B) + \beta A + \beta B \ln\left(\frac{\beta B k^{\alpha}}{1 + \beta B}\right) + \alpha \ln(k) + \alpha \beta B \ln(k) \end{aligned}$$

   Was our guess correct? Yes!

   $$B = \alpha(1 + \beta B)$$

$$A = \frac{1}{1-\beta}\left(\frac{\alpha\beta}{1-\alpha\beta}\ln(\alpha\beta) + \ln(1-\alpha\beta)\right)$$

Is the solution unique? Yes, by the Contraction Mapping Theorem! The last step is finding an allocation. Remember that $g(k) = k'$. Thus,

$$g(k) = \frac{\beta B k^\alpha}{1+\beta B} = \alpha\beta k^\alpha$$

How do we interpret this? Saving a constant fraction $\alpha\beta$ of output and consuming what is leftover. From this policy rule, we can construct the whole sequence:

$$\begin{aligned}
k_1 &= g(k_0) = \alpha\beta k_0^\alpha \\
k_2 &= g(k_1) = \alpha\beta k_1^\alpha = (\alpha\beta)^{1+\alpha} k_0^{\alpha^2} \\
k_3 &= g(k_2) = \dots
\end{aligned}$$

2. Value function iteration. We guess an arbitrary function $v_0(k)$, say $v_0(k) = 0$. We solve

$$v_1(k) = \max_{0 \le k' \le k^\alpha} \{\ln(k^\alpha - k') + \beta v_0(k')\}$$

and get the solution $k' = g_1(k) = 0$ for all k. Therefore, $v_1(k) = \ln(k^\alpha) = \alpha\ln(k)$. Then we solve

$$v_2(k) = \max_{0 \le k' \le k^\alpha} \{\ln(k^\alpha - k') + \beta v_1(k')\}$$

We repeat for

$$v_{n+1}(k) = \max_{0 \le k' \le k^\alpha} \{\ln(k^\alpha - k') + \beta v_n(k')\}$$

gt And we get the sequences $\{v_n\}_{n=0}^\infty$ and $\{g_n\}_{n=0}^\infty$. Will these sequences converge to the optimal $v^\star$ and $g^\star$? Yes, by the Contraction Mapping Theorem.

Example. (A numerical example) Since a computer can only deal with finite-dimensional and discrete objects, we can only approximate the value function. This example is from Dirk Krüger's notes. We discretize the space: $k, k' \in K = \{0.04, 0.08, 0.12, 0.16, 0.2\}$, and have value functions $(v_n(0.04), v_n(0.08), v_n(0.12), v_n(0.16), v_n(0.2))$. We also pick values for the parameters. Say $\alpha = 0.3$ and $\beta = 0.6$. We again initially guess $v_0(k) = 0$. We solve the equation

$$v_1(k) = \max_{0 \le k' \le k^{0.3}} \{\ln(k^{0.3} - k') + 0.6 \cdot 0\}$$

and get optimal policy $k'(k) = g_1(k) = 0.04$ for all $k \in K$. Plugging back in:

$$\begin{aligned}
v_1(0.04) &= \ln(0.04^{0.3} - 0.04) = -1.077 \\
v_1(0.08) &= \ln(0.08^{0.3} - 0.04) = -0.847 \\
v_1(0.12) &= \ln(0.12^{0.3} - 0.04) = -0.715 \\
v_1(0.16) &= \ln(0.16^{0.3} - 0.04) = -0.622 \\
v_1(0.20) &= \ln(0.20^{0.3} - 0.04) \quad = -0.55
\end{aligned}$$

We then iterate again and again. We solve

$$v_2(k) = \max_{0 \le k' \le k^{0.3}} \{\ln(k^{0.3} - k') + 0.6 v_1(k')\}$$

Start with $k = 0.04$:

$$v_2(0.04) = \max_{0 \le k' \le 0.04^{0.3}} \{\ln(0.04^{0.3} - k') + 0.6 v_1(k')\}$$

We try different values of k':

$$\begin{aligned}
v_2(0.04) &= \ln(0.04^{0.3} - 0.04) + 0.6(-1.08) = -1.72 \\
v_2(0.04) &= \ln(0.04^{0.3} - 0.08) + 0.6(-0.85) = -1.71 \\
v_2(0.04) &= \ln(0.04^{0.3} - 0.12) + 0.6(-0.72) = -1.77 \\
v_2(0.04) &= \ln(0.04^{0.3} - 0.16) + 0.6(-0.62) = -1.88 \\
v_2(0.04) &= \ln(0.04^{0.3} - 0.20) + 0.6(-0.55) = -2.04
\end{aligned}$$

Therefore, for $k = 0.04$, the optimal choice is $k'(0.04) = g_2(0.04) = 0.08$ and $v_2(0.04) = -1.71$. The table below shows the values of

$$(k^{0.3} - k') + 0.6 v_1(k')$$

for different values of k and k'

(k, k')	0.04	0.08	0.12	0.16	0.20
0.04	-1.72	$-1.71^\star$	-1.77	-1.88	-2.04
0.08	-1.49	$-1.45^\star$	-1.48	-1.55	-1.64
0.12	-1.36	$-1.31^\star$	-1.32	-1.37	-1.44
0.16	-1.27	$-1.21^\star$	-1.21	-1.25	-1.31
0.20	-1.20	-1.13	$-1.13^\star$	-1.16	-1.20

3. Euler equation approach. We are back to the original problem:

$$w(k_0) = \max_{\{k_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t U(f(k_t) - k_{t+1})$$

subject to

$$0 \le k_{t+1} \le f(k_t) \quad ; \quad k_0 \text{ given}$$

We cannot use KKT, because this is an infinite-dimensional object. However, if we assume that there is a final period T, we can solve this. Consider:

$$w^T(k_0) = \max_{\{k_{t+1}\}_{t=0}^{T}} \sum_{t=0}^{T} \beta^t U(f(k_t) - k_{t+1})$$

subject to

$$0 \le k_{t+1} \le f(k_t) \quad ; \quad k_0 \text{ given}$$

We obviously have $k_{T+1} = 0$. The problem is now the optimization of a continuous function in a finite-dimensional space on a compact set, so a solution exists by the Extreme Value Theorem. Since the constraint is convex and U is strictly concave, the first order conditions are necessary and sufficient to characterize an optimum. We write the Lagrangian:

$$\mathcal{L} = U(f(k_0) - k_1) + \cdots + \beta^t U(f(k_t) - k_{t+1}) + \beta^{t+1} U(f(k_{t+1}) - k_{t+2}) + \cdots + \beta^t U(f(k_T) - 0)$$

The FOCs are:

$$\frac{\partial \mathcal{L}}{\partial k_{t+1}} = -\beta^t U'(f(k_t) - k_{t+1}) + \beta^{t+1} U'(f(k_{t+1}) - k_{t+2}) f'(k_{t+1}) = 0$$

We get the Euler Equation:

$$\underbrace{U'(f(k_t) - k_{t+1})}_{\substack{\text{Cost of saving one} \\ \text{more unit of capital}}} = \underbrace{\beta U'(f(k_{t+1}) - k_{t+2})}_{\substack{\text{Discounted benefit from} \\ \text{one more unit of consumption}}} \underbrace{f'(k_{t+1})}_{\substack{\text{Additional productivity} \\ \text{from one more unit of capital}}}$$

This is a system of T second order difference equations with $T+1$ unknowns $\{k_{t+1}\}_{t=0}^{T}$. With $k_{T+1} = 0$, we can solve for this (it can be a huge pain).

Going back to the example with $U(c) = \ln(c)$ and $f(k) = k^\alpha$, we get that the Euler Equation is:

$$\frac{1}{k_t^\alpha - k_{t+1}} = \frac{\beta\alpha k_{t+1}^{\alpha-1}}{k_{t+1}^\alpha - k_{t+2}}$$

so

$$k_{t+1}^\alpha - k_{t+2} = \beta\alpha k_{t+1}^{\alpha-1}(k_t^\alpha - k_{t+1})$$

Here is a trick: Define $z_t \equiv \frac{k_{t+1}}{k_t^\alpha}$. This converts to

$$1 - z_{t+1} = \alpha\beta\left(\frac{1}{z_t} - 1\right)$$

so

$$z_{t+1} = 1 + \alpha\beta - \frac{\alpha\beta}{z_t}$$

We know that $z_T = 0$. Solve backwards from T. Since

$$z_t = \frac{\alpha\beta}{1 + \alpha\beta - z_{t+1}}$$

we get

$$z_t = \alpha\beta\frac{1 - (\alpha\beta)^{T-t}}{1 - (\alpha\beta)^{T-t+1}}$$

and therefore

$$k_{t+1} = \alpha\beta\frac{1 - (\alpha\beta)^{T-t}}{1 - (\alpha\beta)^{T-t+1}}k_t^\alpha$$

and

$$c_t = \frac{1 - \alpha\beta}{1 - (\alpha\beta)^{T-t+1}}k_t^\alpha$$

Finally, notice that

$$\lim_{T\to\infty} k_{t+1} = \alpha\beta k_t^\alpha$$

Plotting z_{t+1} against z_t can give us some insights into the dynamics of the system. Some facts:

- Since $k_{t+1} \geq 0$, $z_t \geq 0$
- $\lim_{z_t\to 0} z_{t+1} = -\infty$
- $\lim_{z_t\to\infty} z_{t+1} = 1 + \alpha\beta > 1$
- $z_{t+1} = 0$ for $z_t = \frac{\alpha\beta}{1+\alpha\beta} < 1$

We define a *steady state* as

$$z_{t+1} = z_t = z$$

There are two steady states in this economy, can find them by factoring:

$$z = 1 + \alpha\beta - \frac{\alpha\beta}{z} \implies (z-1)(z-\alpha\beta) = 0$$

Therefore, $z = 1$ and $z = \alpha\beta$ are steady states.

Now, let's go back to the infinite case. We have

$$w(k_0) = \max_{\{k_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t U(f(k_t) - k_{t+1}) \text{ s.t. } 0 \leq k_{t+1} \leq f(k_t), k_0 \text{ given}$$

The Euler equation is the same:

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

We are missing a terminal condition! So we impose a *Transversality Condition*:

$$\lim_{t\to\infty} \underbrace{\beta^t U'(f(k_t) - k_{t+1}) f'(k_t)}_{\substack{\text{Value in discounted utility terms} \\ \text{of one more unit of capital}}} \underbrace{k_0}_{\substack{\text{capital} \\ \text{stock}}} = 0$$

This essentially means that the shadow value of capital has to converge to zero. Mathematically, it's a condition coming from the use of the Separating Hyperplane Theorem to find optimality conditions in an infinite-dimensional context. See a note on Chris Sims' website for more. This condition can also be $\lim_{t\to\infty} \lambda_t k_{t+1} = 0$, where λ_t is the Lagrange multiplier on $c_t + k_{t+1} = f(k_t)$.

Theorem 1.3. (Sufficiency of the Euler and Transversality Conditions) *Let U, β, and F satisfy our earlier assumptions. Then an allocation $\{k_{t+1}\}_{t=0}^{\infty}$ that satisfies the Euler equations and the transversality condition solves the sequential social planner's problem, for a given k_0*

Proof. (From SLP Theorem 4.15, pg. 98-99) Let k_0 be given; let $\{k_t^\star\}_{t=0}^{\infty} \in \Pi(k_0)$[1] satisfy the Euler equations and the transversality condition, and let $\{k_t\}_{t=0}^{\infty} \in \Pi(k_0)$ be any feasible sequence. It is sufficient to show that the difference D between the objective function in the sequential social planner's problem evaluated at $\{k_t^\star\}$ and $\{k_t\}$ is non-negative.

Since U is continuous, concave, and differentiable, we have that

$$\begin{aligned} D &= \lim_{T\to\infty} \sum_{t=0}^{T} \beta^t [U(k_t^\star, k_{t+1}^\star) - U(k_t, k_{t+1})] \\ &\geq \lim_{T\to\infty} \sum_{t=0}^{T} \beta^t [U_2(k_t^\star, k_{t+1}^\star)(k_t^\star - k_t) + U_1(k_t^\star, k_{t+1}^\star)(k_{t+1}^\star - k_{t+1})] \end{aligned}$$

Since $k_0^\star - k_0 = 0$, rearranging terms gives

$$D \geq \lim_{T\to\infty} \left\{ \sum_{t=0}^{T-1} \beta^t [U_2(k_t^\star, k_{t+1}^\star) + \beta U_1(k_{t+1}^\star, k_{t+2}^\star)](k_{t+1}^\star - k_{t+1}) + \beta^T U_2(k_T^\star, k_{T+1}^\star)(k_{T+1}^\star - k_{T+1}) \right\}$$

Since $\{k_t^\star\}$ satisfies the Euler equation, the terms in the summation are all zero. Therefore, substituting

[1] This notation is defined precisely later. For now, say that it means that the capital stream is feasible.

from the Euler equation into the last term and using the transversality condition gives

$$\begin{aligned} D &\geq -\lim_{T\to\infty} \beta^T U_1(k_T^\star, k_{T+1}^\star)(k_T^\star - k_T) \\ &\geq -\lim_{T\to\infty} \beta^T U_1(k_T^\star, k_{T+1}^\star) k_T^\star = \lim_{T\to\infty} \beta^T U_1(k_T^\star, k_{T+1}^\star) k_0^\star = 0 \end{aligned}$$

Where the last line follows from the fact that U is increasing and capital use is non-negative. It then follows from the transversality condition that $D \geq 0$. □

Note that this does not work for log utility, since it is technically not bounded, but a similar theorem holds for that case. This theorem gives *sufficient* conditions for optimality. The conditions of this theorem are *necessary* for the log-case (Ekelund and Scheinkman (1986)).

Going back to the log example, recall that $U(c) = \ln(c)$ and $f(k) = k^\alpha$. The transversality condition becomes:

$$\lim_{t\to\infty} \beta^t U'(f(k_t) - k_{t+1}) f'(k_t) k_t = \lim_{t\to\infty} \frac{\alpha\beta^t k_t^\alpha}{k_t^\alpha - k_{t+1}} = \lim_{t\to\infty} \frac{\alpha\beta^t}{1 - z_t}$$

The Euler equation is still

$$z_{t+1} = 1 + \alpha\beta - \frac{\alpha\beta}{z_t}$$

How do we solve this? Guess z_0, check if the Transversality Conditions hold. The process was:

(a) If $z_0 < \alpha\beta$: in finite time $z_t < 0$ which violates $k_{t+1} \geq 0$

(b) If $z_0 > \alpha\beta$: we get $\lim_{t\to\infty} z_t = 1$, which violates the Transversality Condition

(c) If $z_0 = \alpha\beta$: then $z_t = \alpha\beta$ for all $t > 0$. This satisfies the Euler Equation and the Transversality Condition:

$$\lim_{t\to\infty} \frac{\alpha\beta^t}{1 - z_t} = \lim_{t\to\infty} \frac{\alpha\beta^t}{1 - \alpha\beta} = 0$$

Theorem 3 tells us that $z_t = \alpha\beta$ is an optimal solution. The log case is essentially the only one that can be computed by hand. Usually, we need computational methods.

Definition. We say that a social planner's optimum or a competitive equilibrium is a *steady state* if $c_t = c^\star$ and $k_{t+1} = k^\star$ for all t. From the Euler Equation:

$$\begin{aligned} U'(f(k_t) - k_{t+1}) &= \beta U'(f(k_{t+1}) - k_{t+2}) f'(k_{t+1}) \\ U'(c_t) &= U'(c_{t+1}) f'(k_{t+1}) \end{aligned}$$

At a steady state:

$$f'(k) = \frac{1}{\beta} \equiv 1 + \rho$$

where ρ is called the time discount rate. Since $f'(k) = F_k(k, 1) + 1 - \delta$, we obtain the *modified golden rule*

$$F_k(k, 1) - \delta = \rho$$

In our example

$$\alpha(k^\star)^{\alpha-1} = \rho + 1 = \frac{1}{\beta} \text{ and } k^\star = (\alpha\beta)^{\frac{1}{1-\alpha}}$$

The planner's optimal sequence will converge to $k^\star$ regardless of k_0. Why is it called the *modified* golden rule? The resource constraint is

$$c_t = f(k_t) - k_{t+1} \Longrightarrow c = f(k) - k$$

Therefore to maximize consumption per capita we need

$$f'(k^g) = 1 \quad \text{and} \quad F_k(k^g, 1) - \delta = 0$$

where k^g is called the *golden rule capital stock*. Why does the social planner find it optimal to pick $k^\star < k^g$ in the long run? Because the agent is impatient.

Model. *(Balanced Growth Path)* What do you think of our *growth* model so far? Not a lot of growth! Let's add in population growth ($N_t = (1+n)^t$) and labor-augmenting technological progress $F(K_t, N_t(1+g)^t)$. What is the utility function now? Either (where c_t is per capita consumption):

(i) Per capita lifetime utility

$$\sum_{t=0}^{\infty} \beta^t U(c_t)$$

(ii) Lifetime utility of the entire dynasty

$$\sum_{t=0}^{\infty} (1+n)^t \beta^t U(c_t)$$

with resource constraint

$$(1+n)^t c_t + K_{t+1} = F(K_t, (1+n)^t(1+g)^t) + (1-\delta)K_t$$

Define

$$\tilde{c}_t = \frac{c_t}{(1+g)^t}$$
$$\tilde{k}_t = \frac{k_t}{(1+g)^t} = \frac{K_t}{(1+n)^t(1+g)^t}$$

We can rewrite the resource constraint as

$$\tilde{c}_t + (1+n)(1+g)\tilde{k}_{t+1} = F(\tilde{k}_t, 1) + (1-\delta)\tilde{k}_t$$

In order to obtain a balanced growth path, we assume CRRA utility $U(c) = \frac{c^{1-\sigma}}{1-\sigma}$

$$\sum_{t=0}^{\infty} \beta^t \frac{c^{1-\sigma}}{1-\sigma} = \sum_{t=0}^{\infty} \tilde{\beta}^t \frac{\tilde{c}^{1-\sigma}}{1-\sigma}$$

where $\tilde{\beta} = \beta(1+g)^{1-\sigma}$. The social planner solves

$$\max_{\{k_{t+1}\}_{t=0}^{\infty}} = \sum_{t=0}^{\infty} \tilde{\beta}^t \frac{(f(\tilde{k}_t) - (1+g)(1+n)\tilde{k}_{t+1})^{1-\sigma}}{1-\sigma}$$

subject to

$$0 \leq (1+g)(1+n)\tilde{k}_{t+1} \leq f(\tilde{k}_t)$$

with k_0 given. A *balanced growth path* is a socially optimal allocation where all variables grow at a constant rate. Here is corresponds to a steady state for $\{\tilde{c}_t, \tilde{k}_{t+1}\}$. From the Euler equations:

$$(1+n)(1+g)(\tilde{c}_t)^{-\sigma} = \tilde{\beta}(\tilde{c}_{t+1})^{-\sigma}\left(F_k(\tilde{k}_{t+1}, 1) + (1-\delta)\right)$$

A steady state on $\{\tilde{c}, \tilde{k}\}$ is

$$(1+n)(1+g) = \tilde{\beta}\left(F_k(\tilde{k}^\star, 1) + (1-\delta)\right)$$

Defining $\tilde{\beta} \equiv \frac{1}{1+\tilde{\rho}}$, we find

$$(1+n)(1+g)(1+\tilde{\rho}) = F_k(\tilde{k}^\star, 1) + (1-\delta)$$

which is (approximately)

$$F_k(\tilde{k}^\star, 1) - \delta \approx n + g + \tilde{\rho}$$

Model. *(Competitive Equilibrium)* So far we have been interested in the social planner's problem. Now we decentralize the Pareto allocation to a competitive equilibrium. We have:

- Arrow-Debreu market structure
- Perfect competition
- Ownership
  - Households own firms (receive their profits)
  - Households own capital (they rent it to firms)
- Goods:
  - Final output y_t: Used for consumption and investment. Its price is p_t (quoted in period 0)
  - Labor n_t: Let w_t be the price of one unit of labor delivered in period t (quoted in period 0) in terms of the period t consumption good. w_t is called the real wage. The nominal wage is $w_t p_t$.
  - Capital services k_t: Let r_t be the rental price of one unit of capital services delivered in period t, quoted in period 0, in terms of the period t consumption good. r_t is the real rental rate, the nominal rental rate is $r_t p_t$.

Firms behave competitively in output and factor markets. The representative firm's problem is, given a sequence of price $\{p_t, w_t, r_t\}_{t=0}^{\infty}$:

$$\pi = \max_{\{y_t, n_t, k_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} p_t(y_t - r_t k_t - w_t n_t)$$

subject to

$$y_t = F(k_t, n_t) \text{ and } y_t, n_t, k_t \geq 0 \text{ for all } t \geq 0$$

Households own capital stock and supply labor and capital services. They decide how much to consume and how much to save (through capital accumulation). Taking prices $\{p_t, w_t, r_t\}_{t=0}^{\infty}$ as given, the representative household solves

$$\max_{\{c_t, i_t, x_{t+1}, k_t, n_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t U(c_t)$$

subject to

$$\begin{aligned}\sum_{t=0}^{\infty} p_t(c_t + i_t) &\leq \sum_{t=0}^{\infty} p_t(r_t k_t + w_t n_t) + \pi \\ x_{t+1} &= (1-\delta)x_t + i_t \\ 0 &\leq n_t \leq 1 \\ 0 &\leq k_t \leq x_t \\ c_t, x_{t+1} &\geq 0 \text{ and } x_0 \text{ given}\end{aligned}$$

Here we are being very careful. We will have $k_t = x_t$.

Definition. A *competitive equilibrium* (Arrow-Debreu) is a set of prices $\{p_t, w_t, r_t\}_{t=0}^{\infty}$ and allocations for the firm $\{y_t^d, n_t^d, k_t\}_{t=0}^{\infty}$ and the household $\{c_t, i_t, x_{t+1}, k_t^s, n_t^s\}_{t=0}^{\infty}$ such that

1. Given prices, the allocation of the representative firm solves the firm's problem
2. Given prices, the allocation of the representative household solves the household's problem
3. Markets clear:

$$\begin{aligned}y_t &= c_t + i_t \quad \text{(Goods market)} \\ n_t^d &= n_t^s \quad \text{(Labor market)} \\ k_t^d &= k_t^s \quad \text{(Capital services market)}\end{aligned}$$

In equilibrium, it must be the case that $k_t = k_t^d = k_t^s$, $n_t = n_t^d = n_t^s$, and all prices are strictly positive, so $p_t, r_t, w_t > 0$.

The firm's problem is static:

$$\max_{k_t, n_t \geq 0} p_t(F(k_t, n_t) - r_t k_t - w_t n_t)$$

Taking the first order conditions, we get that the firm produces such that marginal product is equal to marginal cost:

$$r_t = F_k(k_t, n_t) \quad \text{and} \quad w_t = F_n(k_t, n_t)$$

Using constant returns to scale and Euler's Theorem, we get that firm profits are

$$\pi_t = p_t(F(k_t, n_t) - F_k(k_t, n_t)k_t - F_n(k_t, n_t)n_t) = 0$$

Remark. We need constant returns to scale for this – constant returns imply that $F(\lambda k, \lambda n) = \lambda F(k, n)$. Taking derivatives with respect to λ, we get Euler's Theorem. We will usually assume constant returns to scale. Sometimes we will assume decreasing returns, in which case there will be profits in equilibrium. This could be seen by assuming that there is a scale factor F for every decreasing returns production function. There will be some increase in F in equilibrium, which can be interpreted as profit.

Remark. If firms own their own capital, profits will be positive. The fact that they don't own their own capital is what leads to zero profit in equilibrium. This is still unrealistic, because markups of course exist in real life, but for now we are assuming they don't exist.

Proposition 1.3. *With constant returns to scale, the number of firms is indeterminate.*

Proof. Since we have constant returns, we have that

$$F(\lambda k, \lambda n) = \lambda F(k, n)$$

Differentiating with respect to k, we get that

$$\lambda F_k(\lambda k, \lambda n) = \lambda F_k(k, n) \iff F_k(\lambda k, \lambda n) = F_k(k, n)$$

Now taking $\lambda = \frac{1}{n}$, we get

$$F_k(k/n, 1) = F_k(k, n) = r_t$$

Thus, a large number of small firms that behave optimally are equivalent to a single large firm that behaves optimally. $\square$

Remark. Total output could be produced by one representative firm or n_t firms with one worker each. The only thing that is determinate is total output and the capital / labor ratio.

For the households, we have that $n_t = 1, k_t = x_t$, and $i_t = k_{t+1} - (1-\delta)k_t$. Since the budget constraint holds with equality, we can write

$$\max_{\{c_t, k_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t U(c_t)$$

subject to

$$\sum_{t=0}^{\infty} p_t(c_t + k_{t+1} - (1-\delta)k_t) = \sum_{t=0}^{\infty} p_t(r_t k_t + w_t)$$

where $c_t, k_{t+1} \geq 0$, k_0 given.

Using μ as the Lagrange multiplier of the budget constraint, we get that the FOC are

$$\begin{aligned} \beta^t U'(c_t) &= \mu p_t \\ \beta^{t+1} U'(c_{t+1}) &= \mu p_{t+1} \\ \mu p_t &= \mu(1 - \delta + r_{t+1}) p_{t+1} \end{aligned}$$

Which yield

$$\begin{aligned} \frac{\beta U'(c_{t+1})}{U'(c_t)} &= \frac{p_{t+1}}{p_t} = \frac{1}{1 - \delta + r_{t+1}} \\ \frac{(1 - \delta + r_{t+1})\beta U'(c_{t+1})}{U'(c_t)} &= 1 \end{aligned}$$

Using the previous notation ($f(k) = F(k, 1) + (1-\delta)k$) and the marginal pricing equation $r_t = F_k(k_t, 1) = f'(k_t) - (1-\delta)$, as well as goods market clearing ($c_t = f(k_t) - k_{t+1}$), we obtain

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

Which is exactly the same Euler equation as in the social planner's problem.

The transversality condition for the household is $\lim_{t\to\infty} p_t k_{t+1} = 0$. Using the first order conditions, we get

$$\begin{aligned}
\lim_{t\to\infty} p_t k_{t+1} &= \frac{1}{\mu} \lim_{t\to\infty} \beta^t U'(c_t) k_{t+1} \\
&= \frac{1}{\mu} \lim_{t\to\infty} \beta^{t-1} U'(c_{t-1}) k_t \\
&= \frac{1}{\mu} \lim_{t\to\infty} \beta^{t-1} U'(c_t)(1-\delta+r_t) k_t \\
&= \frac{1}{\mu} \lim_{t\to\infty} \beta^{t} U'(f(k_t)-k_{t+1}) f'(k_t) k_t
\end{aligned}$$

Which is the exact same as the social planner's transversality condition. We have just loosely shown that the welfare theorems hold.

Notice that once we have determined the equilibrium capital stock, we are done:

$$\begin{aligned}
c_t &= f(k_t) - k_{t+1} \\
y_t &= F(k_t, 1) \\
i_t &= y_t - c_t \\
n_t &= 1 \\
r_t &= F_k(k_t, 1) \\
w_t &= F_n(k_t, 1)
\end{aligned}$$

Model. *(Sequential Markets Equilibrium)* In the same setup, the household's problem is

$$\max_{\{c_t, k_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t U(c_t)$$

subject to

$$c_t + k_{t+1} - (1-\delta)k_t = w_t + r_t k_t$$

where $c_t, k_{t+1} \geq 0$ and k_0 is given. The firm's problem is

$$\max_{k_t, n_t \geq 0} F(k_t, n_t) - w_t n_t - r_t k_t$$

Definition. A *sequential markets equilibrium* is streams of prices $\{w_t, r_t\}_{t=0}^{\infty}$, allocations for a representative household $\{c_t, k_{t+1}^s\}_{t=0}^{\infty}$, and allocations for a representative firm $\{n_t^d, k_t^d\}_{t=0}^{\infty}$ such that

1. Given k_0 and $\{w_t, r_t\}_{t=0}^{\infty}$, the household's allocation solves the household's maximization problem
2. For each $t \geq 0$, given (w_t, r_t) firm allocation (n_t^d, k_t^d) solves the firm's maximization problem
3. Markets clear: for all $t \geq 0$,

$$\begin{aligned}
n_t^d &= 1 \\
k_t^d &= k_t^s \\
F(k_t^d, n_t^d) &= c_t + k_{t+1}^s - (1-\delta)k_t^s
\end{aligned}$$

Model. *(Recursive Competitive Equilibrium)* We have:

- State variables (k, K) and control variables (c, k')

- Bellman equation:

$$\begin{aligned} v(k,K) &= \max_{c,k'\geq 0} U(c) + \beta v(k',K') \\ c + k' &= w(K) + (1 + r(K) - \delta)k \\ K' &= H(K) \end{aligned}$$

- $K' = H(K)$ is the aggregate law of motion.
- The solution to this problem is a value function v and policy functions $c = C(k,K)$ and $k' = G(k,K)$
- For firms:

$$w(K) = F_n(K,1) \quad \text{and} \quad r(K) = F_k(K,1)$$

Definition. A *recursive competitive equilibrium* is a value function $v : \mathbb{R}_+^2 \to \mathbb{R}$ and policy functions $C, G : \mathbb{R}_+^2 \to \mathbb{R}_+$ for the representative household, pricing functions $w, r : \mathbb{R}_+ \to \mathbb{R}_+$, and an aggregate law of motion $H : \mathbb{R}_+ \to \mathbb{R}_+$ such that

1. Given w, r, and H, v solves the Bellman equation and C, G are the associated policy functions
2. The pricing functions solve the firm's first order conditions
3. Consistency: $H(K) = G(K,K)$
4. For all $K \in \mathbb{R}_+$,

$$C(K,K) + G(K,K) = F(K,1) + (1-\delta)K$$

1.4 Dynamic Programming

Throughout this section, we will consider problems of the form

$$v(x) = \max_{y \in \Gamma(x)} \{F(x, y) + \beta v(y)\}$$

where x is the set of state variables, y is the set of controls, F is the period return function, and Γ is the constraint set.
Remark. In the neoclassical growth model, $x = k$, $y = k'$, $F(k, k') = U(f(k) - k')$, and $\Gamma(k) = \{k' \in \mathbb{R} : 0 \leq k' \leq f(k)\}$.
Definition. An *operator* T is defined as:

$$(Tv)(x) \equiv \max_{y \in \Gamma(x)} \{F(x, y) + \beta v(y)\}$$

where T takes as an input a function v, and outputs another function Tv. Using this notation, a solution $v^\star$ to our original functional equation is a *fixed point* of the operator T:

$$v^\star = Tv^\star$$

Questions:

1. Under what conditions does T have a fixed point $v^\star$?
2. Under what conditions is $v^\star$ unique?
3. Under what conditions does the sequence $\{v_n\}_{n=0}^\infty$ defined recursively by $v_{n+1} = Tv_n$ with v_0 as a guess converge to $v^\star$?

Answer: Contraction Mapping Theorem

1.4.1 Proving the Contraction Mapping Theorem

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

1. $d(x, y) \geq 0$
2. $d(x, y) = 0 \Leftrightarrow x = y$
3. $d(x, y) = d(y, x)$
4. $d(x, z) \leq d(x, y) + d(y, z)$

Definition. A sequence $\{x_n\}_{n=0}^\infty$ with $x_n \in S \; \forall \; n \in \mathbb{N}$ is said to *converge* to $x \in S$ if for every $\varepsilon > 0$ there exists $N_\varepsilon \in \mathbb{N}$ such that $d(x_n, x) < \varepsilon$ for all $n > N_\varepsilon$. In this case we write $\lim_{n\to\infty} x_n = x$.
Definition. A sequence $\{x_n\}_{n=0}^\infty$ with $x_n \in S \; \forall \; n \in \mathbb{N}$ is called a *Cauchy sequence* if for every $\varepsilon > 0$ there exists $N_\varepsilon \in \mathbb{N}$ such that $d(x_n, x_m) < \varepsilon$ for all $n, m > N_\varepsilon$.
Definition. A metric space (S, d) is said to be *complete* if every Cauchy sequence $\{x_n\}_{n=0}^\infty$ with $x_n \in S$ for all $n \in \mathbb{N}$ converges to some $x \in S$.
Example. Let $X \subseteq \mathbb{R}^I$ and $S = C(X)$ be the set of all continuous and bounded functions $f : X \to \mathbb{R}$. Define the distance $d : C(X) \times C(X) \to \mathbb{R}$ as $d(f, g) = \sup_{x \in X} |f(x) - g(x)|$. This distance is called the sup-norm. Then (S, d) is a complete metric space (the proof is in SLP, Theorem 3.1, pg. 47-49).
Definition. Let (S, d) be a metric space and $T : S \to S$. The function T is a *contraction mapping* if there exists a number $\beta \in (0, 1)$ satisfying

$$d(Tx, Ty) \leq \beta d(x, y) \text{ for all } x, y \in S$$

β is called the *modulus* of the contraction.

Theorem 1.4. (Contraction Mapping Theorem)[2] *Let (S, d) be a complete metric space and suppose that $T : S \to S$ is a contraction mapping with modulus β. Then*

1. *The operator T has exactly one fixed point $v^\star \in S$*
2. *For any $v_0 \in S$ and and $n \in \mathbb{N}$ we have*

$$d(T^n v_0, v^\star) \leq \beta^n d(v_0, v^\star)$$

Lemma 1.1. *Let (S, d) be a metric space and $T : S \to S$ be a contraction mapping. Then T is continuous.*

Proof. (Lemma 1) We need to show: for all $s_0 \in S$ and all $\varepsilon > 0$, there exists $\delta(\varepsilon, s_0)$ such that if $s \in S$ and $d(s, s_0) < \delta(\varepsilon, s_0)$, then $d(Ts, Ts_0) < \varepsilon$. Fix $s_0 \in S$ and $\varepsilon > 0$, and pick $\delta(\varepsilon, s_0) = \varepsilon$. Then

$$d(Ts, Ts_0) \leq \beta d(s, s_0) < \beta\delta(\varepsilon, s_0) < \varepsilon$$

□

Proof. (Contraction Mapping Theorem) We will start with an arbitrary $v_0 \in S$ and consider the sequence $v_n = T^n v_0$. Our candidate for a fixed point is $v^\star = \lim_{n\to\infty} v_n$. We will prove the Contraction Mapping Theorem as follows:

Part 1:

1. First, we will show that $v_n \to v^\star \in S$. Since T is a contraction, we have

$$\begin{aligned} d(v_{n+1}, v_n) &= d(Tv_n, Tv_{n-1}) \leq \beta d(v_n, v_{n-1}) \\ &\leq \beta d(Tv_{n-1}, Tv_{n-2}) \leq \beta^2 d(v_{n-1}, v_{n-2}) \\ &\leq \cdots \leq \beta^n d(v_1, v_0) \end{aligned}$$

   We will now use the Triangle Inequality. For any $m > n$, we have that

$$\begin{aligned} d(v_m, v_n) &\leq d(v_m, v_{m-1}) + d(v_{m-1}, v_n) \\ &\leq d(v_m, v_{m-1}) + d(v_{m-1}, v_{m-2}) + \cdots + d(v_{n+1}, v_n) \\ &\leq \beta^{m-1} d(v_1, v_0) + \beta^{m-2} d(v_1, v_0) + \cdots + \beta^n d(v_1, v_0) \\ &= \beta^n (\beta^{m-n-1} + \beta^{m-n-2} + \cdots + 1) d(v_1, v_0) \\ &\leq \frac{\beta^n}{1-\beta} d(v_1, v_0) \end{aligned}$$

   Therefore, $\{v_n\}_{n=0}^\infty$ is a Cauchy sequence. Since (S, d) is a complete metric space, $\{v_n\}_{n=0}^\infty$ converges in S, so $v_n \to v^\star \in S$.

2. Next, we will establish that $v^\star$ is a fixed point of T. We have that

$$Tv^\star = T(\lim_{n\to\infty} v_n) = \lim_{n\to\infty} Tv_n = \lim_{n\to\infty} v_{n+1} = v^\star$$

3. Now, we will show that the fixed point is unique. Suppose that there exists $\hat{v} \in S$ such that $\hat{v} = T\hat{v}$ and $\hat{v} \neq v^\star$. Then there exists $a > 0$ such that $d(\hat{v}, v^\star) = a$. But then

$$0 < a = d(\hat{v}, v^\star) = d(T\hat{v}, Tv^\star) \leq \beta d(\hat{v}, v^\star) = \beta a$$

[2] Formally, this is the Banach Fixed-point Theorem

which is a contradiction, so $v^\star$ is unique.

Part 2: We proceed by induction. For $n = 0$, the claim holds trivially. Suppose that

$$d(T^k v_0, v^\star) \leq \beta^k d(v_0, v^\star)$$

It suffices to show that

$$d(T^{k+1} v_0, v^\star) \leq \beta^{k+1} d(v_0, v^\star)$$

We have that

$$d(T^{k+1} v_0, v^\star) = d(T(T^k v_0), Tv^\star) \leq \beta d(T^k v_0, v^\star) \leq \beta^{k+1} d(v_0, v^\star)$$

□

Remark. The contraction mapping theorem is extremely powerful. However, it can be hard to show that an operator is a contraction.

Theorem 1.5. (Blackwell's Theorem) *Let $X \subseteq \mathbb{R}^L$ and $B(X)$ be the space of bounded functions $f : X \to \mathbb{R}$ with the distance being the sup-norm. Let $T : B(X) \to B(X)$ be an operator satisfying*

1. *Monotonicity: If $f, g \in B(X)$ are such that $f(x) \leq g(x)\ \forall\ x \in X$, then $(Tf)(x) \leq (Tg)(x)\ \forall\ x \in X$.*
2. *Discounting: Let the function $f + a$, for $f \in B(X)$ and $a \in \mathbb{R}_+$ be defined by $(f + a)(x) = f(x) + a$. There exists $\beta \in (0, 1)$ such that for all $f \in B(X)$, $a \in \mathbb{R}_+$, and $x \in X$,*

$$[T(f + a)](x) \leq [Tf](x) + \beta a$$

Then T is a contraction mapping with modulus β.

Proof. If $f(x) \leq g(x)\ \forall\ x \in X$, we write $f \leq g$. For any $f, g \in B(X)$, $f \leq g + d(f, g)$ where d is the sup-norm. The monotonicity and discounting imply that

$$Tf \leq T(g + d(f, g)) \leq Tg + \beta d(f, g)$$

Reversing f and g gives, by the same logic,

$$Tg \leq Tf + \beta d(f, g)$$

Combining these inequalities, we find that $d(Tf, Tg) \leq \beta d(f, g)$, so T is a contraction. □

Can these theorems help with the Neoclassical Growth Model?

Example. Work in the metric space $B([0, \infty)), d)$ the space of bounded functions with d being the sup-norm. Define an operator

$$(Tv)(k) = \max_{0 \leq k' \leq f(k)} \{U(f(k) - k') + \beta v(k')\}$$

We can verify that $T : B([0, \infty)) \to B([0, \infty))$: Take v to be bounded, and since U is bounded by assumption, Tv is also bounded. First, we will prove monotonicity. Suppose $v \leq w$. Let $g_v(k)$ denote an optimal policy (need not be unique) corresponding to v. Then for all $k \in [0, \infty)$,

$$\begin{aligned}
Tv(k) &= U(f(k) - g_v(k)) + \beta v(g_v(k)) \\
&\leq U(f(k) - g_v(k)) + \beta w(g_v(k)) \\
&\leq \max_{0 \leq k' \leq f(k)} \{U(f(k) - k') + \beta w(k')\} \\
&= Tw(k)
\end{aligned}$$

Next, we will prove discounting. We have that

$$\begin{aligned} T(v+a)(k) &= \max_{0 \le k' \le f(k)} \{U(f(k)-k') + \beta(v(k')+a)\} \\ &= \max_{0 \le k' \le f(k)} \{U(f(k)-k') + \beta v(k')\} + \beta a \\ &= Tv(k) + \beta a \end{aligned}$$

Thus, the neoclassical growth model with bounded utility satisfies Blackwell's Theorem and is therefore a contraction mapping with modulus β. Hence there is a unique fixed point to the functional equation that can be computed from any starting guess v_0 by repeated application of the operator T.

We are interested in the problem of the form

$$h(x) = \max_{y \in \Gamma(x)} f(x,y)$$

Define $G(x) \coloneqq \{y \in \Gamma(x) : f(x,y) = h(x)\}$. Intuitively, what is $G(x)$? What can we say about the properties of h and G?

Definition. Let X, Y be arbitrary sets. A *correspondence* $\Gamma : X \rightrightarrows Y$ maps each element $x \in X$ into a subset $\Gamma(X) \subseteq Y$.

Definition. A correspondence $\Gamma : X \rightrightarrows Y$ is *lower-hemicontinuous* at a point $x \in X$ if $\Gamma(x) \neq \varnothing$ and if for every $y \in \Gamma(x)$ and every sequence $\{x_n\}$ in X converging to $x \in X$ there exists $N \ge 1$ and a sequence $\{y_n\} \in Y$ converging to y such that $y_n \in \Gamma(x_n)$ for all $n > N$.

Definition. A compact-valued correspondence $\Gamma : X \rightrightarrows Y$ is *upper-hemicontinuous* at a point x if $\Gamma(x) \neq \varnothing$ and if for all sequences $\{x_n\} \in X$ converging to $x \in X$ and for all sequences $\{y_n\} \in Y$ such that $y_n \in \Gamma(x_n)$ for all n, there exists a convergent subsequence of $\{y_n\}$ that converges to some $y \in \Gamma(x)$.

Remark. A single-valued correspondence (*i.e.* a function) that is upper-hemicontinuous is continuous.

Definition. A correspondence $\Gamma : X \to Y$ is *continuous* if it is both upper- and lower-hemicontinuous.

Theorem 1.6. (Berge's Theorem of the Maximum) *Let $X \subseteq \mathbb{R}^L$ and $Y \subseteq \mathbb{R}^M$, let $f : X \times Y \to \mathbb{R}$ be a continuous function, and let $\Gamma : X \rightrightarrows Y$ be a compact-valued and continuous correspondence. Then $h : X \to \mathbb{R}$ is continuous and $G : X \rightrightarrows Y$ is nonempty, compact-valued, and upper-hemicontinuous.*

Proof. (From SLP Theorem 3.6, pg. 62). Fix $x \in X$. The set $\Gamma(x)$ is nonempty and compact, and $f(x, \cdot)$ is continuous, meaning that it attains a maximum over $\Gamma(x)$, so the set of maximizers $G(x)$ is nonempty. Moreover, since $G(x) \subseteq \Gamma(x)$, and $\Gamma(x)$ is compact, it follows that $G(x)$ is bounded. Suppose $y_n \to y$ and $y_n \in G(x)\ \forall\ y_n$. Since $\Gamma(x)$ is closed, $y \in \Gamma(x)$. Also, since $h(x) = f(x, y_n)$ for all y_n, since $y_n \in G(x)$, and f is continuous, it follows that $f(x,y) = h(x)$. Thus, $y \in G(x)$, so $G(x)$ is closed. It follows that $G(x)$ is nonempty and compact, for all x.

Next, we will show that $G(x)$ is upper-hemicontinuous. Fix x, and let $\{x_n\}$ be a sequence converging to x. Choose $\{y_n\}$ such that $y_n \in G(x_n)$ for all n. Since Γ is upper-hemicontinuous, there exists a subsequence $\{y_{n_k}\}$ converging to $y \in \Gamma(x)$. Let $z \in \Gamma(x)$. Since Γ is lower-hemicontinuous, there exists a sequence $\{z_{n_k}\} \to z$ with $z_{n_k} \in \Gamma(x_{n_k})$ for all n_k. Since $f(x_{n_k}, y_{n_k}) \ge f(x_{n_k}, z_{n_k})$ for all n_k and f is continuous, it follows that $f(x,y) \ge f(x,z)$. Since this holds for any $z \in \Gamma(x)$, it follows that $y \in G(x)$, and G is upper-hemicontinuous.

Finally, we will show that h is continuous. Fix x, and let $\{x_n\}$ be any sequence that converges to x. Choose $\{y_n\}$ such that $y_n \in G(x_n)$ for all n. Let $\overline{h} = \limsup h(x_n)$ and $\underline{h} = \liminf h(x_n)$. Then there exists a subsequence $\{x_{n_k}\}$ such that $\overline{h} = \lim f(x_{n_k}, y_{n_k})$. But since G is upper-hemicontinuous, there exists a subsequence of $\{y_{n_k}\}$, call it $\{y'_j\}$, converging to $y \in G(x)$. Thus, $\overline{h} = \lim f(x_j, y'_j) = f(x,y) = h(x)$. An analogous argument establishes that $\underline{h} = h(x)$. Thus, $h(x_n)$ converges, and its limit is $h(x)$. $\square$

Example.

$$(Tv)(k) \coloneqq \max_{0 \le k' \le f(k)} \{U(f(k)-k') + \beta v(k')\}$$

Then $x = k$, $y = k'$, $X = Y = \mathbb{R}_+$, $f(x, y) = U(f(x) - y) + \beta v(y)$, and $\Gamma : X \rightrightarrows Y$ is given by $\Gamma(x) = \{y \in \mathbb{R}_+ : 0 \leq y \leq f(x)\}$. Suppose that v is continuous, then the Theorem of the Maximum implies that $Tv(\cdot)$ is a continuous function and that optimal policy $g(\cdot)$ is an upper-hemicontinuous correspondence. If $g(\cdot)$ is a function, then it is continuous.

Example. Compare:

Functional Equation:

$$v(x) \coloneqq \sup_{y \in \Gamma(x)} \{F(x, y) + \beta v(y)\}$$

This has a unique solution $v^\star$ that is approached from any initial guess v^0.

Sequential Problem:

$$w(x_0) \coloneqq \sup_{\{x_{t+1}\}_{t=0}^\infty} \sum_{t=0}^\infty \beta^t F(x_t, x_{t+1})$$

subject to

$$x_{t+1} \in \Gamma(x_t) \quad \text{and} \quad x_0 \in X \text{ given}$$

Questions: When does $v = w$? When is $\{x_{t+1}\}_{t=0}^\infty$ the same as $y = g(x)$?

Model. *(Principle of Optimality Model)* Define some notation. Let X be the set of possible values that the state x can take, and let correspondence $\Gamma : X \rightrightarrows X$ describes the feasible set of the next period state y given that today's state is x. The graph of Γ is defined as

$$A \coloneqq \{(x, y) \in X \times X : y \in \Gamma(x)\}$$

We have a period return function $f : A \to \mathbb{R}$. The fundamentals of the analysis are the tuple (X, F, β, Γ). For the neoclassical growth model, F and β describe preferences and X, Γ describe technology. Any sequence of states $\{x_t\}_{t=0}^\infty$ is a plan. For a given x_0, the set of feasible plans is

$$\Pi(x_0) \coloneqq \{\{x_t\}_{t=1}^\infty : x_{t+1} \in \Gamma(x_t)\}$$

We need some assumptions for the principle of optimality:

Assumption 1.1. $\Gamma(x)$ is nonempty for all $x \in X$.

Assumption 1.2. For all initial x_0 and all feasible plans $\bar{x} \in \Pi(x_0)$,

$$\lim_{n \to \infty} \sum_{t=0}^n \beta^t F(x_t, x_{t+1})$$

exists (it may be ∞ or $-\infty$).

Theorem 1.7. (Principle of Optimality) *Suppose that (X, F, β, Γ) satisfy the two previous assumptions. Then*

1. *The function w satisfies the functional equation*
2. *If for all $x_0 \in X$ and for all $\bar{x} \in \Pi(x_0)$ a solution v to the functional equation satisfies*

$$\lim_{n \to \infty} \beta^n v(x_n) = 0$$

then $v = w$.

Proof. (From SLP Theorem 4.2 and Theorem 4.3, pg. 71-73). We will prove each part separately.

1. Fix $\beta \in (0, 1)$, and choose some $x_0 \in X$. First, suppose $w(x_0)$ is finite. Then $w(x_0) \geq u(\underline{x})\ \forall\ \underline{x} \in \Pi(x_0)$ and for any $\varepsilon > 0$, $w(x_0) \leq u(\underline{x}) + \varepsilon$ for some $\underline{x} \in \Pi(x_0)$, where $u(c) = \sum_{t=0}^\infty \beta^t F(c_t, c_{t+1})$. To show that $w(x_0)$ is a solution to the functional equation, it suffices to show that two conditions directly hold.

First, fix some $x_1 \in \Gamma(x_0)$ and $\varepsilon > 0$. Then there exists $\underline{x}' \in \Pi(x_1)$ such that $u(\underline{x}') \geq w(x_1) - \varepsilon$. Note also that $\underline{x} = (x_0, \underline{x}') \in \Pi(x_0)$. Thus, it follows that

$$w(x_0) \geq u(\underline{x}) = F(x_0, x_1) + \beta u(\underline{x}') \geq F(x_0, x_1) + \beta w(x_1) - \beta\varepsilon$$

Since ε and x_1 were arbitrarily chosen, we have that $w(x_0) \geq F(x_0, y) + \beta w(y) \; \forall \; y \in \Gamma(x_0)$.

Second, fix some $x_0 \in X$ and $\varepsilon > 0$. We can choose $\underline{x} \in \Pi(x_0)$ such that

$$w(x_0) \leq u(\underline{x}) + \varepsilon = F(x_0, x_1) + \beta u(\underline{x}_1) + \varepsilon$$

It then follows that

$$w(x_0) \leq F(x_0, x_1) + \beta w(x_1) + \varepsilon$$

and since $x_1 \in \Gamma(x_0)$, we have that $w(x_0) \leq F(x_0, y) + \beta w(y) + \varepsilon$ for some $y \in \Gamma(x_0)$.

Thus, if $w(x_0)$ is finite, w satisfies the functional equation. It remains to show that it does when $w(x_0) \in \{-\infty, \infty\}$. Suppose that $w(x_0) = \infty$. It follows that there exists a sequence $\{x^k\} \in \Pi(x_0)$ such that $\lim_{k\to\infty} u(x^k) = \infty$. Since $x_1^k \in \Gamma(x_0) \; \forall \; k$, and

$$u(x^k) = F(x_0, x_1^k) + \beta u(x'^k) \leq F(x_0, x_1^k) + \beta w(x_1^k)$$

for all k, it follows that there exists a sequence $x_1^k \in \Gamma(x_0)$ such that $\lim_{k\to\infty}[F(x_0, x_1^k) + \beta w(x_1^k)] = \infty$.

Finally, assume that $w(x_0) = -\infty$. We have that

$$u(x) = F(x_0, x_1) + \beta u(x') = -\infty \; \forall \; x \in \Pi(x_0)$$

where $x' = (x_1, x_2, \dots)$. Since F is real-valued, it follows that

$$u(x') = -\infty \; \forall \; x_1 \in \Gamma(x_0), \; \forall \; x' \in \Pi(x_1)$$

Thus, $w(x_1) = -\infty \; \forall \; x_1 \in \Gamma(x_0)$.

Thus, if $w(x_0)$ is not finite, w still satisfies the functional equation.

2. Fix some x_0. First, assume that $v(x_0)$ is finite. We have that $v(x_0) \geq F(x_0, y) + \beta v(y) \; \forall \; y \in \Gamma(x_0)$, and for all $\varepsilon > 0$, there exists $y \in \Gamma(x_0)$ such that $v(x_0) \leq F(x_0, y) + \beta v(y) + \varepsilon$. We have that for all $x \in \Pi(x_0)$,

$$\begin{aligned} v(x_0) &\geq F(x_0, x_1) + \beta v(x_1) \\ &\geq F(x_0, x_1) + \beta F(x_1, x_2) + \beta^2 v(x_2) \\ &\vdots \\ &\geq u(x) + \beta^n v(x_n), n = 1, 2, \dots \end{aligned}$$

Taking the limit as $n \to \infty$, we get that for all $x \in \Pi(x_0)$, $v(x_0) \geq \sum_{t=0}^{\infty} \beta^t F(x_t, x_{t+1})$. Next, fix some $\varepsilon > 0$ and choose $\{\delta_t\} \in \mathbb{R}_+$ such that $\sum_{t=1}^{\infty} \beta^{t-1}\delta_t \leq \varepsilon/2$. From above, we can choose $x_1 \in \Gamma(x_0), x_2 \in \Gamma(x_1), \dots$ such that

$$v(x_t) \leq F(x_t, x_{t+1}) + \beta v(x_{t+1}) + \delta_{t+1}, \; t = 0, 1, 2, \dots$$

Then $x = (x_0, x_1, \dots) \in \Pi(x_0)$, and

$$\begin{aligned} v(x_0) &\leq \sum_{t=0}^{n} \beta^t F(x_t, x_{t+1}) + \beta^{n+1} v(x_{n+1}) + (\delta_1 + \beta\delta_2 + \cdots + \beta^n \delta_{n+1}) \\ &\leq u(x) + \beta^{n+1} v(x_{n+1}) + \varepsilon/2, \ n = 1, 2, \dots \end{aligned}$$

Thus, for sufficiently large n, $v(x_0) \leq \sum_{t=0}^{\infty} \beta^t F(x_t, x_{t+1}) + \varepsilon$. Thus, v is a solution to the functional equation, and since such solutions are unique, $v = w$.

□

Intuition: The supremum function from the sequential problem solves the functional equation. Result 2 is key. It states a condition under which a solution to the functional equation is a solution to the sequential problem.

Theorem 1.8. (Principle of Optimality) *Suppose that (X, F, β, Γ) satisfy the two previous assumptions.*

1. *Let $\bar{x} \in \Pi(x_0)$ be a feasible plan that attains the supremum in the sequential problem. Then for all $t > 0$,*

$$w(\bar{x}_t) = F(\bar{x}_t, \bar{x}_{t+1}) + \beta w(\bar{x}_{t+1})$$

2. *Let $\hat{x} \in \Pi(x_0)$ be a feasible plan satisfying, for all $t > 0$,*

$$w(\hat{x}_t) = F(\hat{x}_t, \bar{x}_{t+1}) + \beta w(\hat{x}_{t+1})$$

and

$$\limsup_{t \to \infty} \beta^t w(\hat{x}_t) \leq 0$$

Then $\hat{x}$ attains the supremum in the sequential problem for x_0.

Proof. (From SLP Theorem 4.4 and Theorem 4.5, pg. 75-76). We will prove each part separately.

1. Since $\bar{x}$ attains the supremum, we have that, defining $u(c) = \sum_{t=0}^{\infty} \beta^t F(c_t, c_{t+1})$

$$w(\bar{x}_0) = u(\bar{x}) = F(x_0, \bar{x}_1) + \beta u(\bar{x}') \geq u(x) \ \forall \ x \in \Pi(x_0)$$

In particular, this inequality holds for all plans with $x_1 = \bar{x}_1$. Since $\bar{x}_1, x_2, x_3, \dots \in \Pi(\bar{x}_1) \Rightarrow x_0, \bar{x}_1, x_2, \dots \in \Pi(x_0)$, it follows that $u(\bar{x}) \geq u(x') \ \forall \ x \in \Pi(\bar{x}_1)$. Thus, $u(\bar{x}') = w(\bar{x}_1)$. Substituting, we get that $w(\bar{x}_t) = F(\bar{x}_t, \bar{x}_{t+1}) + \beta w(\bar{x}_{t+1})$ for $t = 0$. Continuing by induction, it holds for all t.

2. Suppose that $\hat{x} \in \Pi(x_0)$ satisfies $w(\hat{x}_t) = F(\hat{x}_t, \bar{x}_{t+1}) + \beta w(\hat{x}_{t+1})$ for all $t > 0$. Then it follows that

$$w(x_0) = u(\hat{x}) + \beta^{n+1} w(\hat{x}_{n+1}) \ \forall \ n = 1, 2, \dots$$

Then, taking the limit as $n \to \infty$, we find that $w(x_0) \leq u(\hat{x}) = \sum_{t=0}^{\infty} \beta^t F(\hat{x}_t, \hat{x}_{t+1})$. Since $\hat{x} \in \Pi(x_0)$, the reverse inequality holds, and $\hat{x}$ attains the supremum in the sequential problem.

□

Model. *(Dynamic Programming with Bounded Returns)* Define the functional equation

$$v(x) = \sup_{y \in \Gamma(x)} \{F(x, y) + \beta v(y)\}$$

with associated operator $T : C(X) \to C(X)$ such that

$$(Tv)(x) = \max_{y \in \Gamma(x)} \{F(x, y) + \beta v(y)\}$$

We will make a number of stronger assumptions on (X, F, β, Γ) to be able to characterize v and g, where

$$g(x) \coloneqq \{y \in \Gamma(x) : v(x) = F(x, y) + \beta v(y)\}$$

is the policy correspondence associated with v.

Assumption 1.3. X is a convex subset of $\mathbb{R}^L$ and the correspondence $\Gamma : X \rightrightarrows X$ is nonempty, compact-valued, and continuous.

Assumption 1.4. The function $f : A \to \mathbb{R}$ is continuous and bounded, and $\beta \in (0, 1)$

Remark. Note that these two assumptions imply our first two assumptions.

Theorem 1.9. (Uniqueness of Solution) *Under Assumptions 3 and 4, the operator T maps $C(X)$ into itself. T has a unique fixed point v and for all $v_0 \in C(X)$,*

$$d(T^n v_0, v) \leq \beta^n d(v_0, v)$$

Further, the policy correspondence g is compact-valued and upper-hemicontinuous.

Proof. (From SLP Theorem 4.6, pg. 79). We have that for each $v \in C(X)$ and $x \in X$, our problem in defining the operator is to maximize the continuous function $F(x, \cdot) + \beta v(\cdot)$ over the compact set $\Gamma(x)$. Thus, the maximum is attained. Since F and v are bounded, clearly Tv is bounded, and since F and v are continuous, and Γ is compact-valued and continuous, it follows from Theorem 1.6 (Berge's Maximization Theorem) that Tv is continuous. Thus, $T : C(X) \to C(X)$.

It then immediately follows that T satisfies the conditions of Theorem 1.5 (Blackwell's Theorem), so it is a contraction, and since $C(X)$ is a Banach space, we have directly from Theorem 1.4 (Contraction Mapping) that T has a unique fixed point v, and the necessary equation holds. The stated properties of g then follow again from Theorem 1.6. □

Assumption 1.5. For fixed y, $F(\cdot, y)$ is strictly increasing in each of its L components.

Assumption 1.6. Γ is monotone in the sense that $x \leq x'$ implies that $\Gamma(x) \subseteq \Gamma(x')$

Theorem 1.10. (Monotonicity of Value Function) *Under Assumptions 3 to 6, the unique fixed point v of T is strictly increasing.*

Proof. (From SLP Theorem 4.7, pg. 80). Let $C'(X) \subset C(X)$ be the set of bounded, continuous, nondecreasing functions on X, and let $C''(X) \subset C'(X)$ be the set of strictly increasing functions. Since $C'(X)$ is a closed subset of the complete metric space $C(X)$, it suffices from Theorem 1.9 and the Contraction Mapping Theorem (Theorem 1.4) to show that $T(C'(X)) \subseteq C''(X)$. The assumptions ensure this. □

Assumption 1.7. F is strictly concave: for all $(x, y), (x', y') \in A$ and $\theta \in (0, 1)$

$$F[\theta(x, y) + (1 - \theta)(x', y')] > \theta F(x, y) + (1 - \theta)F(x', y')$$

Assumption 1.8. Γ is convex in the sense that for $\theta \in [0, 1]$, $x, x' \in X$, $y \in \Gamma(x)$, $y' \in \Gamma(x')$, then

$$\theta y + (1 - \theta)y' \in \Gamma(\theta x + (1 - \theta)x')$$

Theorem 1.11. (Strict Concavity and Unique Policy) *Under Assumptions 3-4 and 7-8, the unique fixed point v is strictly concave and the optimal policy g is a single-valued continuous function.*

Proof. (From SLP Theorem 4.8, pg. 81). Let $C'(X) \subset C(X)$ be the set of bounded, continuous, weakly concave functionf on X, and let $C''(X) \subset C'(X)$ be the set of strictly concave functions. Since $C'(X)$ is a closed subset of the complete metric space $C(X)$, it suffices from Theorem 1.9 and the Contraction Mapping Theorem (Theorem 1.4) to show that $T(C'(X)) \subseteq C''(X)$. To verify this, let $v \in C'(X)$ and let $x_0 \neq x_1$, $\theta \in (0, 1)$, and $x_\theta = \theta x_0 + (1 - \theta)x_1$.

Let $y_i \in \Gamma(x_i)$ attain $(Tv)(x_i)$ for $i = 0, 1$. Then by Assumption 8, $y_\theta = \theta y_0 + (1-\theta)y_1 \in \Gamma(x_\theta)$. It follows that

$$\begin{aligned}(Tv)(x_\theta) \geq F(x_\theta, y_\theta) + \beta v(y_\theta) \\ &> \theta[F(x_0, y_0) + \beta v(y_0)] + (1-\theta)[F(x_1, y_1) + \beta v(y_1)] \\ &= \theta(Tv)(x_0) + (1-\theta)(Tv)(x_1)\end{aligned}$$

where the first line uses the fact that $y_\theta \in \Gamma(x_\theta)$, the second uses the hypothesis that v is concave and the concavity restriction on F (Assumption 7), and the last follows from how y_0 and y_1 were selected. Since x_0 and x_1 were arbitrary, it follows that Tv is strictly concave, and since v was arbitrary, that $T(C'(X)) \subseteq C''(X)$.

Thus, the unique fixed point v is strictly concave. Since F is also concave by assumption, and, for each $x \in X$, $\Gamma(x)$ is convex, it follows that the maximum is attained by a unique y value. Thus, g is single-valued, and its continuity follows from the fact that it is upper hemi-continuous. □

Assumption 1.9. F is continuously differentiable

Theorem 1.12. (Benveniste-Scheinkman (an Envelope Theorem)) *Under assumptions 3-4 and 7-9, if $x_0 \in int(X)$ and $g(x_0) \in int(\Gamma(x_0))$, then the unique fixed point v is continuously differentiable at x_0 with*

$$\frac{\partial v(x_0)}{\partial x_0} = \frac{\partial F(x_0, g(x_0))}{\partial x_0}$$

Proof. (From SLP Theorem 4.10, pg. 84). Any subgradient p of v at x_0 must satisfy

$$p \cdot (x - x_0) \geq v(x) - v(x_0) \geq F(x) - F(x_0)$$

for all x sufficiently close to x_0, where the first inequality uses the definition of a subgradient and the second uses the fact that $F(x) \leq v(x)$, with equality at x_0. Since F is differentiable at x_0, p is unique, and any concave function with a unique subgradient at an interior point x_0 is differentiable at x_0. □

Example. Solving Bellman Equations with Benveniste-Scheinkman:

We have the functional equation

$$v(k) = \max_{0 \leq k' \leq f(k)} U(f(k) - k') + \beta v(k')$$

Taking the first order conditions with respect to k' gives:

$$U'(f(k) - k') = \beta v'(k')$$

Then with Benveniste-Schienkman, we have

$$v'(k) = U'(f(k) - g(k))f'(k)$$

and hence

$$U'(f(k) - g(k)) = \beta f'(g(k))U'(f(g(k)) - g(g(k)))$$

which is the Euler equation we found earlier.

Model. *(Stochastic Growth Model (Markov Process))* Most of what we've done works in a stochastic environment as long as we can summarize the state of the world in a simple way. Here we specify a specific structure to uncertainty that makes our models tractable: *discrete time, discrete state, time homogeneous Markov Processes*. Let

$$\pi(j \mid i) = \mathbb{P}\{s_{t+1} = j \mid s_t = i\}$$

Conditional probabilities of s_{t+1} only depend on the realization of s_t, not s_{t-1} or any other past realization. Time homogeneity means that π is not indexed by time.

Given that $s_{t+1} \in S$ and $s_t \in S$ and S is a finite set, the distribution $\pi(\cdot, \cdot)$ is an $N \times N$ matrix of the form

$$\pi = \begin{pmatrix} \pi_{11} & \cdots & \pi_{1j} & \cdots & \pi_{1N} \\ \vdots & & \vdots & & \vdots \\ \pi_{i1} & \cdots & \pi_{ij} & \cdots & \pi_{iN} \\ \vdots & & \vdots & & \vdots \\ \pi_{N1} & \cdots & \pi_{Nj} & \cdots & \pi_{NN} \end{pmatrix}$$

where an element $\pi_{ij} = \pi(j \mid i) = \mathbb{P}\{s_{t+1} = j \mid s_t = i\}$. Since $\pi_i j \geq 0$ and $\sum_j \pi_{ij} = 1$ for all i, matrix π is called a *stochastic matrix*.

Suppose the probability distribution over states today is given by the N-dimensional column vector $P_t = (p_t^1, \ldots, p_t^N)^T$ where $\sum_i p_t^i = 1$. The probability of being in state j tomorrow is

$$p_{t+1}^j = \sum_i \pi_{ij} p_t^i \quad \equiv \quad P_{t+1} = \pi^T P_t$$

Definition. A *stationary distribution* Π of the Markov chain π is

$$\Pi = \pi^T \Pi$$

A Markov process π has at least one stationary Π: the eigenvector (normalized to 1) of the eigenvalue $\lambda = 1$ of π^T. If only one such eigenvalue exists, then this is a unique stationary distribution. If there are multiple, then there are multiple stationary distributions.

If s_t is a Markov chain, we have

$$\pi(s^{t+1}) = \pi(s_{t+1} \mid s_t) \times \pi(s_t \mid s_{t-1}) \times \cdots \times \pi(s_1 \mid s_0) \times \Pi(s_0)$$

Example. Suppose

$$\pi = \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}$$

for some $p \in (0, 1)$. The unique invariant distribution is $\Pi(s) = 1/2$ for both s.

Suppose

$$\pi = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

then any distribution over the two states is an invariant distribution

Model. *Stochastic Growth Model (Technology)* We incorporate technology as

$$y_t = e^{z_t} F(k_t, n_t)$$

where z_t is a technology stock that has unconditional mean 0 and follows an N-state Markov chain with state space $Z = \{z_1, z_2, \ldots, z_N\}$ and transition matrix $\pi = \left[\pi_{ij}\right]$. Let Π denote a stationary distribution. We also have the standard evolution of capital stock $k_{t+1} = (1 - \delta)k_t + i_t$, a resource constraint $y_t = c_t + i_t$, an endowment k_0 and one unit of time, information, where z_t is publicly observable and $z_0 \sim \Pi$, and preferences:

$$\mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t U(c_t)\right] = \sum_{t=0}^{\infty} \sum_{z^t \in Z^t} \beta^t \pi(z^t) U(c_t(z^t))$$

We can use our new tools to solve this model, with state variables (k, z) and control variable k' we have the

Bellman equation

$$v(k,z) = \max_{0 \le k' \le e^z F(K,1)+(1-\delta)k} \left\{ U(e^z F(K,1) + (1-\delta)k - k') + \beta \sum_{z'} \pi(z' \mid z) v(k', z') \right\}$$

We can also add a labor-leisure choice: $U(c_t, 1-n_t)$. Our new Bellman equation is

$$v(k,z) = \max_{k',n} \left\{ U(e^z F(K,n) + (1-\delta)k - k', 1-n) + \beta \sum_{z'} \pi(z' \mid z) v(k', z') \right\}$$

subject to

$$0 \le k' \le e^z F(K,n) + (1-\delta)k \quad \text{and} \quad 0 \le n \le 1$$

This is the benchmark model of modern business cycle research. See: Cooley and Prescott (1995), "Economic Growth and Business Cycles".

Solving the model: We have three conditions:

The intra*temporal optimality condition*:

$$e^z F_n(k,n) = \frac{U_l(c, 1-n)}{U_c(c, 1-n)}$$

The inter*temporal optimality condition*:

$$U_c(c, 1-n) = \beta \sum_{z'} \pi(z' \mid z) v'(k', z')$$

The *envelope condition*:

$$v'(k,z) = (e^z F_k(k,n) + 1 - \delta) U_c(c, 1-n)$$

Combining, we get

$$U_c(c, 1-n) = \beta \sum_{z'} \pi(z' \mid z)(e^z F_k(k', n') + 1 - \delta) U_c(c', 1-n')$$

Definition. We pursue *calibration* in order to estimate (or choose) parameters of the model so that it could be used for quantitative analysis of the real world and counterfactual analysis. The idea of calibration is:

1. We choose a set of empirical facts that the model should match
2. We choose parameters so that the equilibrium of the model matches the facts

Note: the fact that the model fits the facts cannot be used as a claim of success. We need to evaluate success on other dimensions.

Example. We will calibrate a simple version of the deterministic neoclassical model with population and technology growth. We have the following functional forms:

$$U(c) = \frac{c^{1-\sigma} - 1}{1-\sigma}$$
$$F(K,N) = K^\alpha ((1+g)^t N)^{1-\alpha}$$

Our parameters are technology (α, δ, g), demographics (n), and preferences (β, σ). We will attempt to choose parameters such that the balanced growth path of the model matches long-run average facts for the U.S. economy. Note that we need to decide on period length. Take the period to be one year.

Recall the Kaldor (1961) facts:

1. Output per capita, Y/N, grows at a constant rate
2. The capital to labor ratio, K/N, grows at a constant rate
3. The interest rate, R, is fairly constant
4. The output to capital ratio, Y/K, is fairly constant
5. The share of value added going to labor and capital is fairly constant
6. There are wide dispersions in Y_i/N_i across countries

We will take some parameters directly from the data. Population growth rate in the model is n, in the data $n = 1.1\%$. Growth rate of per-capita GDP in the model is g, in the data $g = 1.8\%$. We can also exploit the balanced growth path relationships:

$$\begin{aligned} w_t &= (1-\alpha)K_t^{\alpha}N_t^{-\alpha}((1+g)^t)^{1-\alpha} \\ \frac{w_t N_t}{Y_t} &= (1-\alpha) \end{aligned}$$

In the U.S., the labor share of income has averaged about $\frac{2}{3}$, so we set $\alpha = \frac{1}{3}$.

To calibrate the depreciation rate δ, we will start with the resource constraint, remembering that $\tilde{x}_t = x_t/(1+g)^t$ and $x_t = X_t/(1+n)^t$. We get that

$$\begin{aligned} \tilde{c} + (1-n)(1+g)\tilde{k} &= F(\tilde{k},1) + (1-\delta)\tilde{k} \\ \tilde{c} + \left[(1-n)(1+g) - (1-\delta)\right]\tilde{k} &= F(\tilde{k},1) \end{aligned}$$

In the model, investment is given by

$$\tilde{i} = \left[(1-n)(1+g) - (1-\delta)\right]\tilde{k}$$

and so

$$\frac{I/Y}{K/Y} = \frac{I}{K} = \frac{\tilde{i}}{\tilde{k}} = (1-n)(1+g) - (1-\delta)$$

In the data, $I/Y \approx 0.2$ and $K/Y \approx 3$. Using our previous parameters, we find that $\delta \approx 4\%$.

We need to pick parameters for the utility function. From the Euler equation with CRRA utility function, we have that

$$(1+n)(1+g)(\tilde{c}_t)^{-\sigma} = (1 + r_{t+1} - \delta)\tilde{\beta}(\tilde{c}_{t+1})^{-\sigma}$$

On the balanced growth path,

$$\begin{aligned} (1+n)(1+g) &= (1+r-\delta)\beta(1+g)^{1-\sigma} \\ \beta(1+g)^{-\sigma} &= \frac{1+n}{1+r-\delta} \end{aligned}$$

We need to find r. The rental rate of capital is

$$r_{t+1} = \alpha K_t^{\alpha-1}\left[(1+g)^t N_t\right]^{1-\alpha} = \alpha\frac{Y_t}{K_t}$$

with $K/Y \approx 3$ and $\alpha \approx 1/3$, we find $r \approx 0.11$. Plugging this back into the first order conditions, we get

$$\beta(1.018)^{-\sigma} = 0.944$$

Note that without growth (*i.e.* when $g = 0$), this relationship pins down β but doesn't tell us anything about σ. With growth, the typical approach is to pick σ from information outside the model. One can estimate σ by taking the log of

$$(1+n)(1+g)(\tilde{c}_t)^{-\sigma} = (1 + r_{t+1} - \delta)\tilde{\beta}(\tilde{c}_{t+1})^{-\sigma}$$

and do the estimation with consumption data. From macroeconomic data, by Hall (1982), we get that $\frac{1}{\sigma} = 0.1$, and from microeconomic consumption data, by Attanasio & Browning (1993), Attanasio & Weber (1995), we get that $\frac{1}{\sigma} \in [0.3, 0.8]$. We will pick $\sigma = 1$.

Summarizing the parameters, we get:

Parameter	Value	Target
g	1.8%	g in data
n	1.1%	n in data
α	0.33	labor share
δ	4%	$\frac{I/Y}{K/Y}$
σ	1	outside evidence
β	0.961	K/Y

Question. How doees the model fare on other moments? See Kydland & Prescott (1982). They found a very solid level of correlation with most parameters, given how parsimonious the model is. However, their estimates of investment were volatile, and their estimates of hours worked were solidly off. This is still an open question – these models do not match fluctuations in employment particularly well.

We will come back to the growth model (in continuous time) later.

1.5 Continuous-Time Growth Theory

I do not see how one can look at figures like these without seeing them as representing possibilities. Is there some action a government could take that would lead the Indian economy to grow like Indonesia's or Egypt's? If so, what exactly? If not, what is it about the nature of India that makes it so? The consequences for human welfare involved in questions like these are simply staggering: Once one starts to think about them, it is hard to think about anything else [Lucas (1988), pg. 5]

Back to the Kaldor (1961) facts (specifically, the first few):

1. Output per capita, Y/N, and capital per worker $k = K/L$, grows at a constant rate
2. The interest rate, R, is fairly constant
3. The output to capital ratio, Y/K, is fairly constant
4. The share of value added going to labor and capital is fairly constant

We have an additional set of stylized facts from Summers & Heston (1991), which are:

1. There is enormous variation of *per capita income* across countries. The poorest have 5% of per capita income in the United States
2. There is enormous variation in the *growth rate* of per capita income across countries
3. Growth rate determines the economic fate of a country over long periods of time. How long does it take for a country to double its per capita GDP if it grows at g% per year? About $70/g$ (Lucas (1988))
4. Countries change their *relative* position in the income distribution

Model. *(The Solow Growth Model)* [Solow 1956] Some preliminary assumptions: a single good and no international trade; factors of production (labor and capital) are fully employed; labor force grows at a rate $n > 0$ so that (with $L(0) = 1$):

$$L(t) = e^{nt}L(0) = e^{nt}$$

Note that

$$\frac{\dot{L}}{L} \equiv \frac{1}{L}\frac{\partial L}{\partial t} = n$$

so that over one unit of time, $\partial t = 1$, we have $L(t+1) = (1+n)L(t)$.

Technology follows $Y(t) = F(K(t), A(t)L(t))$ with F having constant returns to scale, strictly increasing, strictly concave, twice continuously differentiable, $F(0,\cdot) = F(\cdot,0) = 0$, and Inada conditions. We assume that the labor augmenting technological progress grows at rate $g > 0$: $A(t) = e^{gt}$. We define

$$\xi(t) \equiv \frac{Y(t)}{A(t)L(t)} = \frac{F(K(t), A(t)L(t))}{A(t)L(t)} = F\left(\frac{K(t)}{A(t)L(t)}, 1\right) \equiv F(\kappa(t))$$

where

$$\kappa(t) \equiv \frac{K(t)}{A(t)L(t)}$$

We have *capital accumulation*, where

$$\dot{K}(t) = sY(t) + \delta K(t)$$

where $0 < s < 1$ is the *exogeneous* saving rate and $0 < \delta < 1$ is the depreciation rate.

We have the *resource constraint*, where

$$\underbrace{\dot{K}(t) - \delta K(t)}_{\text{Investment}} = Y(t) - C(t)$$

We are basically done with the model. Note that we have no optimizing agents – instead, we have a behavioral assumption on s. Also, technology grows exogenously.

Solving the Solow Model. From the capital accumulation equation, we have that

$$\frac{\dot{K}}{AL} = s\xi - \delta\kappa$$

Also

$$\frac{\dot{K}}{AL} = \frac{\dot{K}}{K}\frac{K}{AL} = \frac{\dot{K}}{K}\kappa$$

But

$$\frac{\dot{\kappa}}{\kappa} = \frac{\dot{K}}{K} - \frac{\dot{L}}{L} - \frac{\dot{A}}{A} = \frac{\dot{K}}{K} - n - g$$

so that

$$\frac{\dot{K}}{AL} = \frac{\dot{K}}{K}\kappa = \left(\frac{\dot{\kappa}}{\kappa} + n + g\right)\kappa \Longrightarrow \dot{\kappa} + \kappa(n+g) = s\xi - \delta\kappa$$

We have found the main equation of the Solow model.
Definition. The *capital accumulation equation in per-effective worker terms* is

$$\dot{\kappa} = sf(\kappa) - (n + g + d)\kappa$$

It is a first order nonlinear ordinary differential equation, and it completely characterizes the economy for any initial condition $\kappa(0) = K(0)$

Once we have the solution $\kappa(t)_{t\in[0,\infty)}$, we can solve the rest of the model:

$$\begin{aligned}
k(t) &= \kappa(t)A(t) = e^{gt}\kappa(t) \\
K(t) &= e^{(n+g)t}\kappa(t) \\
y(t) &= e^{gt}f(\kappa(t)) \\
Y(t) &= e^{(n+g)t}f(\kappa(t)) \\
C(t) &= (1-s)e^{(n+g)t}f(\kappa(t)) \\
c(t) &= (1-s)e^{gt}f(\kappa(t))
\end{aligned}$$

Example. Suppose that the production function is Cobb-Douglas: $f(\kappa) = \kappa^\alpha$. The equation:

$$\dot{\kappa} = s\kappa^\alpha - (n + g + d)\kappa$$

There are two steady-states (where $\dot{\kappa} = 0$): $k^\star = 0$ and

$$k^\star = \left(\frac{s}{n + g + \delta}\right)^{\frac{1}{1-\alpha}}$$

To solve the equation, we define $v(t) = \kappa(t)^{1-\alpha}$. Then

$$\dot{v} = \frac{(1-\alpha)\dot{\kappa}}{\kappa^\alpha}$$

Dividing both sides by $\kappa^\alpha/(1-\alpha)$:

$$(1-\alpha)\frac{\dot{\kappa}}{\kappa^\alpha} = (1-\alpha)s - (1-\alpha)n + g + \delta)\kappa^{1-\alpha} \Longrightarrow \dot{v} = (1-\alpha)s - (1-\alpha)(n+g+\delta)v$$

The general solution to the homogeneous equation is

$$v_g(t) = Ce^{-(1-\alpha)(n+g+\delta)t}$$

for some constant C. A particular equation for the nonhomogeneous equation is

$$v_p(t) = \frac{s}{n+g+\delta} \equiv v^\star = (\kappa^\star)^{1-\alpha}$$

Therefore, all solutions to this equation are of the form

$$v(t) = v_g(t) + v_p(t) = v^\star + Ce^{-(1-\alpha)(n+g+\delta)t}$$

Using the initial condition $v(0) = \kappa(0)^{1-\alpha}$:

$$v(t) = v^\star + (v(0) - v^\star)e^{-(1-\alpha)(n+g+\delta)t}$$

Going back to the original notation, we have

$$\kappa = \left[\frac{s}{n+g+\delta} + \underbrace{\left(\kappa(0)^{1-\alpha} - \frac{s}{n+g+\delta}\right)e^{-(1-\alpha)(n+g+\delta)t}}_{\text{Goes } \to 0 \text{ as } t\to\infty}\right]^{\frac{1}{1-\alpha}}$$

Notice that we will converge to the steady state regardless of the starting point of the economy.
Question. What do we do if we don't have Cobb-Douglas? Graphical analysis.
Question. Does the Solow Model fit the data well? Let's see.
Example. If the wage is per unit of labor L, we see that

$$r(t) = F_K\left(\frac{K(t)}{L(t)A(t)}, 1\right) = F_K(\kappa(t), 1)$$
$$w(t) = A(t)F_L\left(\frac{K(t)}{L(t)A(t)}, 1\right) = A(t)F_L(\kappa(t), 1)$$

At the balanced growth path (*i.e.* the steady state in κ), r is constant and w grows at the same rate as technology. The capital share $r(t)K(t)/Y(t)$ is constant since Y and K grow at the same rate $n+g$. The balanced growth path of the Solow model represents the four Kaldor facts we talked about earlier. Solow won the Nobel prize for this work in 1989.
Question. What about the Summers-Heston facts? Can we explain the large difference in per-capita income across countries?
Answer. Suppose that all countries have the same technology, same population growth, and same savings rate. Then they all converge to the same steady state and per-capita income converges to $y(t) = A(t)f(\kappa^\star)$. If we observe y to be different across countries, some of them have not converged yet. This implies that poorer countries should be growing faster than richer countries:

$$\gamma_y(t) = \frac{\dot{y}(t)}{y(t)} = g + \frac{f'(\kappa(t))\dot{\kappa}(t)}{f(\kappa(t))} = g + \frac{f'(\kappa(t))}{f(\kappa(t))}(sf(\kappa(t)) - (n+g+\delta)\kappa(t))$$

Since $\frac{f'(\kappa)}{f(\kappa)}$ and $sf(\kappa) - (n+g+\delta)\kappa$ are decreasing in κ, countries further away from the balanced growth

path should converge faster. What do the data say? Not really!
Definition. The idea that countries should converge to the same path is called *absolute convergence*.
Question. Should we discard the Solow model? What if the parameters (s, n, δ) are different?
Definition. The idea that countries have different parameters that they converge under is called *conditional convergence*. Now, the fast-growing countries are the ones farthest from *their own* balanced growth path. The per-capita incomes should grow at the same rate once they reach their balanced growth path.

To test conditional convergence:

1. Compute the steady-state output per worker that a country should possess in an initial period (say, 1960) given (s, n, δ) observed in the country's data
2. Then measure actual GDP per worker and build the difference. The difference indicates how far the country is from its balanced growth path
3. Plot the difference against the growth rate of GDP per worker from the initial period to the current period
4. Countries that are further away from their balanced growth path should grow faster

The data seem to confirm the Solow model (see Jones & Vollrath (1998) fig. 3.8). See Mankiw, Romer, & Weil (1992) for a test of the Solow model.
Remark. Constant savings rate is a strong behavioral assumption that is not derived from maximizing agents. Relaxing it leads to the Cass-Koopmans-Ramsey model.
Remark. Technological progress is modeled exogenously. Relaxing that leads to the endogenous growth model.
Model. *(Cass-Koopmans-Ramsey Growth Model)* We carry over most of the assumptions from Solow: population grows at a rate $n > 0$, we have a production function $F(K, AL)$, technology A grows at a rate $g > 0$, and aggregate capital stock evolved according to $\dot{K} = F(K, AL) - \delta K - C$. Notice that we have consumption here instead of an exogenous savings rate. We can write the equation in per-effective labor terms:

$$\dot{\kappa} = f(\kappa) - \zeta - (n + \delta + g)\kappa$$

where $\zeta = C/(AL)$ and $\kappa = K/(AL)$. We have a representative agent with utility

$$u(c) = \int_0^\infty e^{-\rho t} \underbrace{\frac{(c(t))^{1-\sigma}}{1-\sigma}}_{U(c)} dt$$

where $\rho > 0$ is a discount factor. By using our other notation:

$$\begin{aligned} e^{-\rho t} U(c(t)) &= e^{-\rho t} \frac{(c(t))^{1-\sigma}}{1-\sigma} \\ &= e^{-\rho t} \frac{(\zeta(t) e^{gt})^{1-\sigma}}{1-\sigma} = e^{-(\rho - g(1-\sigma))t} \frac{(\zeta(t))^{1-\sigma}}{1-\sigma} \end{aligned}$$

We assume that $\rho > g(1-\sigma)$ and define $\hat{\rho} = \rho - g(1-\sigma)$:

$$u(c) = \int_0^\infty e^{-\hat{\rho} t} U(\zeta(t)) dt = \int_0^\infty e^{-\hat{\rho} t} \frac{(\zeta(t))^{1-\sigma}}{1-\sigma} dt$$

The Social Planner Problem is

$$\begin{aligned}
&\max_{(\kappa,\zeta)\geq 0} \int_0^\infty e^{-\hat{\rho}t} U(\zeta(t))dt \\
&\text{s.t.} \quad \dot{\kappa}(t) = f(\kappa(t)) - \zeta(t) - (n+\delta+g)\kappa(t) \\
&\qquad \kappa(0) = \kappa_0
\end{aligned}$$

This problem can be solved using Pontryagin's Maximum Principle (see Pontryagin (1962) or Intriligator's (2002) "Mathematical Optimization and Economic Theory"), where $\kappa(t)$ is the state variable, $\zeta(t)$ is the control, and $\lambda(t)$ denotes the co-state variable associated with $\kappa(t)$.

Solving the model. We write the Hamiltonian

$$\mathcal{H}(t,\kappa,\zeta,\lambda) = e^{-\hat{\rho}t} U(\zeta(t)) + \lambda(t)[f(\kappa(t)) - \zeta(t) - (n+\delta+g)\kappa(t)]$$

Sufficient conditions for a solution are:

$$\frac{\partial \mathcal{H}(t,\kappa,\zeta,\lambda)}{\partial \zeta(t)} = 0 \qquad \dot{\lambda}(t) = -\frac{\partial \mathcal{H}(t,\kappa,\zeta,\lambda)}{\partial \kappa(t)} \qquad \lim_{t\to\infty} \lambda(t)\kappa(t) = 0$$

Plugging in we find the four equations:

$$\begin{aligned}
e^{-\hat{\rho}t} U'(\zeta(t)) &= \lambda(t) \\
\dot{\lambda}(t) &= -(f'(\kappa(t)) - (n+\delta+g))\lambda(t) \\
\lim_{t\to\infty} \lambda(t)\kappa(t) &= 0 \\
\dot{\kappa}(t) &= f(\kappa(t) - \zeta(t) - (n+\delta+g)\kappa(t)
\end{aligned}$$

Differentiating with respect to t and combining it, we get

$$\frac{\dot{\lambda}}{\lambda} = \frac{\dot{\zeta} U''(\zeta)}{U'(\zeta)} - \hat{\rho}$$

Combining with the second equation:

$$\dot{\zeta}\frac{\zeta U''(\zeta)}{U'(\zeta)} = -\zeta(f'(\kappa) - (n+\delta+g+\hat{p}))$$

Because of CRRA, the first fraction on the left hand side is equal to $-\sigma$:

$$\dot{\zeta}(t) = \frac{1}{\sigma}\zeta(t)(f'(\kappa(t)) - (n+\delta+g+\hat{\rho}))$$

Therefore, any allocation (κ,ζ) that satisfies the following system is Pareto optimal:

$$\begin{aligned}
\dot{\zeta}(t) &= \frac{1}{\sigma}\zeta(t)(f'(\kappa(t)) - (n+\delta+g+\hat{\rho}) \\
\dot{\kappa}(t) &= f(\kappa(t)) - \zeta(t) - (n+\delta+g)\kappa(t)
\end{aligned}$$

with $\kappa(0) = \kappa_0$ and the Transversality Condition. We have a steady state ($\dot{\zeta} = \dot{\kappa} = 0$) at

$$f'(\kappa^\star) = n+\delta+g+\hat{\rho} \qquad \zeta^\star = f(\kappa^\star) - (n+\delta+g)\kappa^\star$$

Question. How do we study the economy out of steady-state?

Phase diagram. We have

$$\dot{\zeta}(t) = \frac{1}{\sigma}\zeta(t)\left(f'(\kappa(t)) - (n+\delta+g+\hat{\rho})\right)$$
$$\dot{\kappa}(t) = f(\kappa(t)) - \zeta(t) - (n+\delta+g)\kappa(t)$$

We first find the two isocline:

$$\dot{\zeta} = 0 \Rightarrow f'(\kappa(t)) - (n+\delta+g+\hat{\rho}) = 0 \quad \text{and} \quad \dot{\kappa} = 0 \Rightarrow f(\kappa(t)) - \zeta(t) - (n+\delta+g)\kappa(t) = 0$$

In the (κ, ζ) plan, the first one is a vertical line at $\kappa =$ constant. The second one is strictly concave in κ.

We have an initial condition $\kappa(0) = \kappa_0$. We need in general two conditions to pin down a path for the system. The question is how the planner should pick $\zeta(0)$. We argue two things:

1. For a given $\kappa(0) > 0$, any choice $\zeta(0)$ of the planner that leads to a path not converging to the steady state $(\kappa^\star, \zeta^\star)$ cannot be optimal
2. There is a unique stable path leading to the steady state. This is called saddle-path stability of the steady state and the unique stable path is called a *saddle path*

For point 1: Paths going toward point E lead to negative consumption in finite time, and paths converging to north-west hit $\zeta = 0$ which means $\kappa = 0$ forever. Because of Inada, this cannot be optimal.

For point 2: We would need to study the dynamics around the steady state by using local approximation (we skip this here, but it is true).

Combining the results: Only the paths converging to the unique steady state can be optimal solutions. Locally, around the steady state, the converging path is unique (the saddle path). Given an initial $\kappa(0)$, the planner picks the $\zeta(0)$ that puts the economy on the saddle path.

What is the behavior of the economy once it has reached the steady state? Per capita variables grow at rate g, and aggregate variables grow at rate n:

$$c(t) = e^{gt}\zeta^\star$$
$$k(t) = e^{gt}\kappa^\star$$
$$y(t) = e^{gt}f(\kappa^\star)$$
$$C(t) = e^{(g+n)t}\zeta^\star$$
$$K(t) = e^{(g+n)t}\kappa^\star$$
$$Y(t) = e^{(g+n)t}f(\kappa^\star)$$

The long-run behavior of this model is identical to the Solow model; where the economy converges to a balanced growth path.

Remark. We can understand the Cass-Koopmans-Ramsey model as a micro foundation for the Solow model.

The second issue with the Solow model was the *exogenous* growth in technology. Here we can move away from this with the basic AK endogenous growth model:

Model. *(Endogenous Growth Model)* There is no technological progress. Production function is $Y(t) = AK(t)$ with A fixed, and population grows at a rate $n > 0$. The preferences of the representative agent are given by

$$U(c) = \int_0^\infty e^{-\rho t}\frac{c(t)^{1-\sigma}}{1-\sigma}dt$$

with a budget constraint of $c(t) + \dot{a}(t) + na(t) = w(t) + (r(t)+\delta)a(t)$, where $a = A/L$ is per capita asset

holdings with $a(0) = k_0$ given. We also need a no-ponzi-scheme condition. The firm's problem is

$$\max_{K(t),L(t)\geq 0} AK(t) - r(t)K(t) - w(t)L(t)$$

Definition. A *sequential markets equilibrium* is a set of allocations for the household $(c(t), a(t))_{t\in[0,\infty)}$, allocations for the firm $(K(t), L(t))_{t\in[0,\infty)}$, and prices $(r(t), w(t))_{t\in[0,\infty)}$ such that

1. Given prices, the allocation $(c(t), a(t))_{t\in[0,\infty)}$ solves the household's problem
2. Given prices, the allocation $(K(t), L(t))_{t\in[0,\infty)}$ solves the firm's problem
3. $L(t) = e^{nt}$, $L(t)a(t) = K(t)$, and $L(t)c(t) + \dot{K}(t) + \delta K(t) = AK(t)$

Solving the model. We can solve the household's problem like Cass-Koopman. We have solutions:

$$\dot{c} = \frac{1}{\sigma}(r - (n + \delta + \rho)c) \quad \text{and} \quad \gamma_c = \frac{\dot{c}}{c} = \frac{1}{\sigma}(A - (n + \delta + \rho))$$

Consumption per capita is *always* growing at the same rate (not just at the steady state!). Integrating, we find

$$c(t) = c(0)e^{\frac{1}{\sigma}(A-(n+\delta+\rho))t}$$

We make the following assumptions on the parameters: $A - (n + \delta + \rho) > 0$ so that we have positive growth; and $\frac{1-\sigma}{\sigma}\left[A - (n + \delta) - \frac{\rho}{1-\sigma}\right] \equiv \phi < 0$ so that the integral of discounted consumption is finite.

What is the behavior of capital per capita? From the resource constraint, we have

$$\dot{K}(t) + \delta K(t) + C(t) = AK(t) \Longrightarrow c(t) + \dot{k}(t) = Ak(t) - (n + \delta)k(t)$$

so that

$$\gamma_k(t) = \frac{\dot{k}(t)}{k(t)} = A - (n + \delta) - \frac{c(t)}{k(t)}$$

In a *balanced growth path*, γ_k is constant so that c and k must grow at the same rate:

$$\gamma_k = \gamma_c = A - (n + \delta + \rho)$$

Question. What if we are not at the balanced growth path?

We have that

$$\dot{k}(t) = -c(0)e^{\frac{1}{\sigma}(A-(n+\delta+\rho))t} + Ak(t) - (n + \delta)k(t)$$

Exercise. Solve this equation! (Use the general and particular solution, and use the transversality condition to pin down the constraint)
Solution. The solution is:

$$k(t) = -\frac{c(0)}{\phi}e^{\frac{1}{\sigma}(A-(n+\delta+\rho))t} = -\frac{c(t)}{\phi}$$

Remark. The capital stock is *always* proportional to consumption. Since consumption *always* grows at a constant rate, so does capital. The initial condition $k(0) = k_0$ pins down $c(0) = -\phi k(0)$ and $y(0) = Ak(0)$. All variables grow at the same rate $\gamma_c = \gamma_k = \gamma_y$.
Question. Why do we care about this model?
Answer. In the Solow and Cass-Koopman model, parameters like s, n, and δ affect per capita income *level* but not *growth rate*. In the AK model, the growth rate is determined by all of these parameters! There is no convergence in the AK model, and (a key result), no decreasing returns to capital.

2 Ryan Chahrour

Course Overview The *optional* text (which is extremely useful) is Applied Computational Economics and Finance, by Mario Miranda and Paul Fackler. Ryan also has a set of materials on numerical methods on his website, which is a great resource. This course will largely cover the same model the entire time, but we will explore the techniques used to solve it very deeply.

We will cover the (tentative) topics:

1. A labor search model (with numerical introductions)
2. Perturbation techniques
3. The perfect-foresight shooting model
4. Interpolation and extrapolation
5. The value function iteration
6. Projection methods
7. Business Cycles
8. Modern Topics in Macroeconomics

In the last week, we will cover some topics on climate change in macroeconomics and neural networks and machine learning.

The problem sets will be coding-based. In general, you should use Matlab, as it is the language Ryan is most familiar with. Ryan does not want to see code, you should structure the answers like a technical appendix rather than a codebase.

2.1 The Labor Search Model

This model can be thought about in both of the ways we tend to think about models – as a social planner optimizing, and as firms and agents optimizing.
Model. *(Planner's Labor Search Model)* The planner is optimizing over the following. We have a CRRA utility function

$$U(\{C_t\}) \equiv \sum_{t=0}^{\infty} \beta^t \frac{C_t^{1-\sigma}}{1-\sigma}$$

where σ is the coefficient of relative risk aversion. We also have a Cobb-Douglas production function

$$Y_t = A_t K_t^{\alpha} N_t^{1-\alpha} \tag{2}$$

where N_t is the number of people employed (note not hours!), and K_t is capital. We have a law of motion for capital

$$K_{t+1} = (1-\delta_k)K_t + I_t \tag{3}$$

Importantly, we have a new *law of motion for labor*:

$$N_t = (1-\delta_n)N_{t-1} + M(V_t, S_t) \tag{4}$$

where $M(V_t, S_t)$ is a matching function, where V_t are firm vacancy postings and S_t are searching workers. We assume that M has constant returns to scale, and we define δ_n as the (exogenous) separation rate. Finally,

we have an aggregate resource constraint (instead of firms making production decisions):

$$Y_t = C_t + I_t + \phi_n V_t \tag{5}$$

where ϕ_n is a parameter to constrain how valuable a unit of investment is relative to a single worker. Specifically, ϕ_n could be thought of as the cost of posting an open position. The social planner solves the following problem, subject to the above constraints (2) to (5):

$$\max_{\{C_t, I_t, V_t, Y_t, N_t, K_{t+1}\}} \mathbb{E}\left[U(\{C_t\})\right] \equiv \max_{\{C_t, I_t, V_t, Y_t, N_t, K_{t+1}\}} \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t \frac{C_t^{1-\sigma}}{1-\sigma}\right]$$

We could eliminate some variables and constraints from this problem, but we'll wait to do that until later.

We will consider two methods for deriving the optimality conditions. These two methods will suggest different approaches to a solution.

2.2 Lagrangian Approach (Social Planner)

The Lagrangian of this model is

$$\begin{aligned}\mathcal{L} = \mathbb{E}\Bigg[\sum_{t=0}^{\infty} \beta^t \Bigg\{ & \frac{C_t^{1-\sigma}}{1-\sigma} + \lambda_{1,t}(A_t K_t^{\alpha} N_t^{1-\alpha} - Y_t) \\ & + \lambda_{2,t}((1-\delta_k)K_t + I_t - K_{t+1}) \\ & + \lambda_{3,t}((1-\delta_n)N_{t-1} + M(V_t, S_t) - N_t) \\ & + \lambda_{4,t}(Y_t - C_t - I_t - \phi_n V_t) \Bigg\}\Bigg]\end{aligned}$$

with choice variables $\{C_t, I_t, V_t, N_t, K_{t+1}\}$. The necessary first order conditions are:

$$\begin{aligned}
(C) \quad & 0 = C_t^{-\sigma} - \lambda_{4,t} \\
(I) \quad & 0 = \lambda_{2,t} - \lambda_{4,t} \\
(V) \quad & 0 = \lambda_{3,t} M_v(V_t, S_t) - \phi_n \lambda_{4,t} \\
(Y) \quad & 0 = -\lambda_{1,t} + \lambda_{4,t} \\
(N) \quad & 0 = \lambda_{1,t} A_t K_t^{\alpha}(1-\alpha) N_t^{-\alpha} - \lambda_{3,t} + \beta\, \mathbb{E}[\lambda_{3,t+1}(1-\delta_n)] \\
(K) \quad & 0 = -\lambda_{2,t} + \mathbb{E}[\lambda_{1,t+1} A_{t+1} \alpha K_{t+1}^{\alpha-1} N_{t+1}^{1-\alpha} + \lambda_{2,t+1}(1-\delta_k)]
\end{aligned}$$

Note that we've been careful to write these such that all Lagrange multipliers are positive. The first order conditions imply that $\lambda_{1,t} = \lambda_{2,t} = \lambda_{4,t} = C_t^{-\sigma}$ and $\lambda_{3,t} = C_t^{-\sigma}\phi_n / M_v(V_t, S_t)$. The simplified first order conditions are

$$1 = \beta\, \mathbb{E}\left[\left(\frac{C_{t+1}}{C_t}\right)^{-\sigma} \left(A_{t+1}\alpha \left(\frac{K_{t+1}}{N_{t+1}}\right)^{\alpha-1} + (1-\delta)\right)\right] \tag{6}$$

$$\frac{\phi_n}{M_v(V_t, S_t)} = A_t \alpha \left(\frac{K_t}{N_t}\right)^{\alpha} + \beta\, \mathbb{E}\left[\left(\frac{C_{t+1}}{C_t}\right)^{-\sigma} \frac{\phi_n}{M_v(V_{t+1}, S_{t+1})}(1-\delta_n)\right] \tag{7}$$

as well as the constraints (2) to (5).

Equation (6) is often called the *capital Euler equation*, or sometimes the *intertemporal Euler equation*. This

equation equates the cost of one unit of investment with the benefits that it brings – notably, the benefits in the next period.

Equation (7) is called the *labor Euler equation*, or sometimes the *vacancy posting condition*. It equates the cost of hiring an additional worker (the cost of the additional posting ϕ_n divided by the marginal increase in the number of jobs due to that posting, $M_v(V_t, S_t)$ with the returns that that worker brings. Note that since the worker starts producing in the current period, their marginal product is in terms of today's technology. The second term is the value they bring in the future, multiplied by the probability that they will not separate.

Remark. These expectations are slightly odd objects. Specifically, we consider the expectation of the object in the next period given everything that has happened in the current period. For a variable x, we can say that

$$\mathbb{E}[x_{t+1}] \equiv \mathbb{E}[x_{t+1} \mid \text{everything denoted } t \text{ or earlier}]$$

In our case, we will have that $\mathbb{E}[x] = \mathbb{E}[x \mid \{A_t, K_t, N_{t-1}\}]$

In both (6) and (7), returns that happen in the future are discounted to be in terms of today's consumption by using $M_{t+1} \equiv \beta \left(\frac{C_{t+1}}{C_t}\right)^{-\sigma}$. We often call M_{t+1} the *stochastic discount factor*. It is extremely important in macroeconomics models, especially macrofinance. Note that M_{t+1} is in no way related to the matching function $M(\cdot)$. This notation is confusing, but it's (annoyingly) standard in the literature.

The Matching Function We've left the matching function $M(V_t, S_t)$ as an arbitrary function. Most of the literature following Mortensen and Pissarides (1994) assumes that the matching function satisfies the standard form of a neoclassical production function. From now on, we will assume that

$$M(V_t, S_t) = \chi V_t^{\varepsilon} S_t^{1-\varepsilon}$$

This functional form is nice, but it has a massive annoyance: the probability that a worker ends up in a given vacancy post is not necessarily in $[0, 1]$ (or, relatedly, that the number of people hired is less than the number of people looking for jobs). There are other examples of micro-founded matching functions in den Haan, Ramey, & Watson (2000), though they are harder to work with.

With this functional form and our assumption that $S_t = 1$, we get that

$$M_v(V_t, S_t) = \varepsilon \chi V_t^{\varepsilon - 1}$$

Exogenous Processes We finally need to make some assumptions on how the variables tend to move. We could assume that any of the parameters are time-varying, and generally people have studied all of them in DSGE models. For now, we'll assume that technology is the only thing moving exogenously. We will assume that

$$\log(A_{t+1}) = \rho \log(A_t) + \sigma_a \epsilon_{t+1}$$

Steady State We are almost always going to want to identify a steady state before solving the model numerically. To find such a steady state, we can write our necessary equilibrium conditions again, dropping

the subscripts. Our assumptions about the movement of technology imply that $A = 1$. We have that

$$
\begin{aligned}
Y &= K^{\alpha} N^{1-\alpha} \\
I &= \delta_k k \\
N &= \frac{\chi}{\delta_n} V^{\varepsilon} \\
Y &= C + I + \phi_n V \\
(K^{ss}) \qquad 1 &= \beta \left(\alpha \left(\frac{K}{N} \right)^{\alpha-1} + (1-\delta_k) \right) \\
(V^{ss}) \qquad \frac{\phi_n}{\varepsilon \chi V^{\varepsilon-1}} &= (1-\alpha) \left(\frac{K}{N} \right)^{\alpha} + \beta \frac{\phi_n}{\varepsilon \chi V^{\varepsilon-1}} (1-\delta_n)
\end{aligned}
$$

Using K^{ss}, we can get that the steady-state capital to labor ratio is

$$
\frac{K}{N} = \left(\frac{\beta^{-1} - 1 + \delta_k}{\alpha} \right)^{\frac{1}{\alpha-1}}
$$

Using V^{ss}, we can additionally get that

$$
V = \left(\frac{\varepsilon \chi}{\phi_n} \frac{(1-\alpha)}{1-\beta(1-\delta_n)} \left(\frac{K}{N} \right)^{\alpha} \right)^{\frac{1}{1-\varepsilon}}
$$

Using these, along with the conditions above, lets us identify the steady state levels of capital and labor.
Remark. It's always good to try and find the steady state. First, we can often catch mistakes in the dynamic equations. Second, we can see peculiarities in the model that would be hard to see otherwise. For example, in this model it's possible to find values of the parameters that would lead to negative consumption. That's strange, and it's worth thinking about whether it's a fundamental issue with the model or only occurs at strange parameter values so not worth worrying about.

2.3 Log-Linearization Method

Algorithm. *(Log Linearization of the Model)* Define: $X_t = [K_t, N_{t-1}, A_t]$ and $Y_t = [GDP_t, I_t, C_t, V_t]$. We have the model equation (in *General Model Form*)

$$
\mathbb{E}\left[F(X_t, Y_t, X_{t+1}, Y_{t+1})\right] = 0
$$

Remark. This essentially uses the Euler equations and the Law of Iterated Expectations to bring all of our equations together into one.

We have a number of strategies for solving this model (where solving means finding policy functions $Y_t = g(X_t)$ and $X_{t+1} = h(X_t, \varepsilon_{t+1})$). In general, these are called *perturbation techniques*, but we will think about linearization.

Approximation with (first-order) Taylor Expansions We will generally linearize around the steady state – since we have a non-stochastic steady state, it makes to linearize around that state. The naive version of the first-order expansion about the steady state is

$$
\mathbb{E}[\underbrace{F(X^{ss}, Y^{ss}, X^{ss}, Y^{ss})}_{=0} + F_X(\underbrace{X_t - X^{ss}}_{\hat{x}_t}) + F_Y(\underbrace{Y_t - Y^{ss}}_{\hat{y}_t}) + F_{X'}(\underbrace{X_{t+1} - X^{ss}}_{\hat{x}_{t+1}}) + F_{Y'}(\underbrace{Y_{t+1} - Y^{ss}}_{\hat{y}_{t+1}})] = 0
$$

So this simplifies to

$$\mathbb{E}[F_X \hat{x}_t + F_Y \hat{y}_t + F_{X'} \hat{x}_{t+1} + F_{Y'} \hat{y}_{t+1}] = 0$$

We want to find solutions (called *undetermined coefficients*, *i.e.* guess and check)

$$\hat{y}_t = g\hat{x}_t \qquad ; \qquad \hat{x}_{t+1} = h\hat{x}_t + \eta\varepsilon_{t+1}$$

Remark. We are not thinking about g and h as functions anymore – since the model is linearized, we consider only linear solutions to it (*i.e.* matrices).
Remark. We often log-linearize rather than just linearize. This is just for percentage analysis, it doesn't change the analysis. Consider defining $\ell_t \equiv \log(L_t)$, so $\hat{\ell}_t = \ell_t - \ell^{ss}$. We then have that

$$X_t = [K_t, N_{t-1}, A_t] \implies X_t = [k_t, n_{t-1}, a_t]$$

As an example, consider the capital evolution equation $K_{t+1} = (1-\delta_k)K_t + I_t$. We want to log-linearize it. First, we have that

$$\exp(k_{t+1}) = (1-\delta_k)\exp(k_t) + \exp(i_t)$$

At the steady state, we have that the first order Taylor approximation is

$$\begin{aligned}\exp(k^{ss})(k_{t+1} - k^{ss}) &= (1-\delta_k)\exp(k^{ss})(k_t - k^{ss}) + \exp(i^{ss})(i_t - i^{ss}) \\ \hat{k}_{t+1} &= (1-\delta_k)\hat{k}_t + \frac{I^{ss}}{K^{ss}}\hat{i}_t\end{aligned}$$

and since $\delta_k = \frac{I^{ss}}{K^{ss}}$, we have that

$$\hat{k}_{t+1} = (1-\delta_k)\hat{k}_t + \delta_k \hat{i}_t$$

Going back to the general approximation, the general strategy is to plug a 'guess' into the linearized model equations, and simplify. First, note that

$$\hat{y}_{t+1} = g\hat{x}_{t+1} = gh(\hat{x}_t + \eta\varepsilon_{t+1})$$

which implies that

$$F_X \hat{x}_t + F_Y g\hat{x}_t + F_{X'} h\hat{x}_t + F_{Y'} gh\hat{x}_t = 0$$

Collecting like terms, we get that

$$(F_X + F_Y g + F_{X'} h + F_{Y'} gh)\hat{x}_t = 0$$

This is a matrix quadratic, and we are finding g, h such that this holds for all $\hat{x}_t$. We want solutions that do not imply divergent behavior in the economy. Ryan will provide code (called `gx_hx.m`) that will select solutions that don't diverge.
Definition. This solution gives us the *impulse responses*: we can get the difference in expected variable values conditional on a shock happening versus conditional on a shock not happening. Think of it mathematically as

$$\mathbb{E}[Y_{t+j} \mid \varepsilon_t^i = 1] - \mathbb{E}[Y_{t+j} \mid \varepsilon_t^i = 0]$$

At $j = 0$, we call this the *impact effect*.
Example. Consider the following example, starting at $j = 0$:

$$g\left(X_{t-1} + \eta\begin{bmatrix}0\\1\\0\end{bmatrix}\right) - g\left(X_{t-1} + \eta\begin{bmatrix}0\\0\\0\end{bmatrix}\right) \implies g\eta\begin{bmatrix}0\\1\\0\end{bmatrix}$$

Linearity implies that X_{t-1} does not matter, $\varepsilon_t^{m\neq i}$ does not matter, and the response scales linearly with

the shock size.

At $j = 1$, we have that this becomes:

$$X_{t+1} = h\left(hX_{t-1} + \eta \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right) + \eta\varepsilon_{t+1}$$

so we have that

$$\mathbb{E}[Y_{t+1} \mid \varepsilon_t^i = 1] - \mathbb{E}[Y_{t+1} \mid \varepsilon_t^i = 0] = gh\eta \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + gh^2\eta X_{t-1} + g\eta\varepsilon_{t+1} - gh\eta \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = gh^1\eta \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

In general, for arbitrary j, we have that

$$\mathbb{E}[Y_{t+j} \mid \varepsilon_t^i = 1] - \mathbb{E}[Y_{t+j} \mid \varepsilon_t^i = 0] = gh^j\eta \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

Remark. Some coding notes, recalling that $X_t = [K_t, N_{t-1}, A_t]$ and $Y_t = [C_t, \ldots, N_t]$. First, note that this model contains all of N_{t-1}, N_t, and N_{t+1}. We will find it useful to define $N = N_t$ and $NL = N_{t-1}$, and impose the additional constraint that $NL' = N$.

Next, note that it might be useful to split X_t into the endogenous states and the exogenous states. We have that, defining X_t^1 as the endogenous states and X_t^2 as the exogenous states,

$$\begin{bmatrix} X_{t+1}^1 \\ X_{t+1}^2 \end{bmatrix} = \begin{bmatrix} h_{1,1} & h_{2,2} \\ 0 & \rho \end{bmatrix} \begin{bmatrix} X_t^1 \\ X_t^2 \end{bmatrix} + \begin{bmatrix} 0 \\ \sigma \end{bmatrix} \varepsilon_t$$

Finally, note that $e^{-12} \equiv 0$ for doubles in Matlab. e^{-8} or e^{-10} or such are not equivalent to zero.

Second Moments. We will often want to care about the variance of consumption / etc. We will assume that $\text{cov}(\varepsilon_t) = \Sigma_{\varepsilon_t} = I$ (*i.e.* that the covariance matrix is the identity matrix, since the shocks are independent and identically distributed). We could consider the equation

$$\hat{X}_{t+1} = h\hat{X}_t + \eta\varepsilon_{t+1} \Longrightarrow \text{cov}(\hat{X}_t)$$

Note that covariance is unconditional in this model, so we have $\text{cov}(\hat{X}_{t+1}) = \text{cov}(\hat{X}_t) = \Sigma_X$. Taking the covariance of each side, we get that

$$\Sigma_X = \mathbb{E}[(h\hat{X}_t + \eta\varepsilon_{t+1})(h\hat{X}_t + \eta\varepsilon_{t+1})'] = \mathbb{E}[h\hat{X}_t\hat{X}_t{}'h'] + \mathbb{E}[\eta\varepsilon_{t+1}\varepsilon_{t+1}'\eta']$$

thus,

$$\Sigma_X = h\,\mathbb{E}[\hat{X}_t\hat{X}_t']h' + \eta I\eta' = h\Sigma_X h' + \eta\eta'$$

We have a recursive expression for the covariance matrix of X. If we can identify a fixed point, we can identify the covariance.

Remark. We could simply set Σ_X to 0 and iterate, and that method is actually fairly good. We could also solve using the Matlab program `mom.m`, which speeds it up a fair bit. However, we can also solve this using linear algebra.

We can linearize this as follows:

$$\text{vec}(\Sigma_X) = \underbrace{(h \otimes h)}_{\substack{\text{Kronecker} \\ \text{Product}}} \text{vec}(\Sigma_X) + \text{vec}(\eta\eta')$$

Thus,

$$\text{vec}(\Sigma_X) = (I - h \otimes h)^{-1}\text{vec}(\eta\eta')$$

What if we instead wanted Σ_Y? We have that

$$\Sigma_Y = \text{cov}(\hat{Y}_t) = \text{cov}(g\hat{X}_t) = \mathbb{E}[g\hat{X}_t\hat{X}_t'g'] = g\Sigma_X g'$$

Suppose that we're interested in the *auto-covariance* of a variable, defined as $\text{cov}(\hat{Y}_t, \hat{Y}_{t-1})$. We have that

$$\text{cov}(Y_t, Y_{t-1}) = \text{cov}(g(hX_{t-1} + \varepsilon_t), gX_{t-1}) = \mathbb{E}[g(hX_{t-1} + \varepsilon_t)(gX_{t-1})'] = gh\Sigma_X g'$$

and following that,

$$\text{cov}(Y_t, Y_{t-j}) = gh^j\Sigma_X g'$$

2.3.1 An Old Macro Definition

The following is a simplification of Klein (2000). Define $\hat{X}_{t+1,t} \equiv \mathbb{E}[X_{t+1}]$, for any variable X. We have that

$$\underbrace{\begin{bmatrix} F_{X'} & F_{Y'} \end{bmatrix}}_{A} \begin{bmatrix} \hat{X}_{t+1,t} \\ \hat{Y}_{t+1,t} \end{bmatrix} + \underbrace{\begin{bmatrix} F_X & F_Y \end{bmatrix}}_{B} \begin{bmatrix} \hat{X}_t \\ \hat{Y}_t \end{bmatrix} = 0$$

Proposition 2.1. (Ordered QZ Decomposition) *(Generalization of Eigenvalues / Eigenvectors). For any A, B there exist some Q, Z, S, T such that*

$$QAZ = S \qquad \textit{and} \qquad QBZ = T$$

We will enforce the following forms in Matlab:

$$S = \begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{bmatrix} \qquad T = \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix} \qquad Z = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix}$$

where the top row of each matrix is associated with eigenvalues less than 1, and the bottom row is associates with eigenvalues greater than 1. There should be N_X eigenvalues less than 1, and N_Y eigenvalues greater than 1, where N_X is the number of state variables and N_Y is the number of jump variables. The relevant eigenvalues are $\text{diag}(S)/\text{diag}(T)$.

Define:

$$\begin{bmatrix} X_t^\star \\ Y_t^\star \end{bmatrix} = Z^{-1} \begin{bmatrix} \hat{X}_t \\ \hat{Y}_t \end{bmatrix} \Longrightarrow \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \begin{bmatrix} X_t^\star \\ Y_t^\star \end{bmatrix} = \begin{bmatrix} \hat{X}_t \\ \hat{Y}_t \end{bmatrix}$$

Proposition 2.2. $X_t^\star = Z_{11}^{-1}\hat{X}_t - Z_{11}^{-1}Z_{12}Y_t^\star$

Proof. We have that

$$\begin{bmatrix} Z_{11} & Z_{12} \end{bmatrix} \begin{bmatrix} X_t^\star \\ Y_t^\star \end{bmatrix} = \hat{X}_t \Longrightarrow Z_{11}X_t^\star + Z_{12}Y_t^\star = \hat{X}_t$$

□

Substituting this back, we get that

$$AZ\begin{bmatrix}X^\star_{t+1,t}\\ Y^\star_{t+1,t}\end{bmatrix}+BZ\begin{bmatrix}X^\star_t\\ Y^\star_t\end{bmatrix}=0$$

Then left-multiplying by Q, we get that

$$QAZ\begin{bmatrix}X^\star_{t+1,t}\\ Y^\star_{t+1,t}\end{bmatrix}+QBZ\begin{bmatrix}X^\star_t\\ Y^\star_t\end{bmatrix}=0\Longrightarrow\begin{bmatrix}S_{11} & S_{12}\\ 0 & S_{22}\end{bmatrix}\begin{bmatrix}X^\star_{t+1,t}\\ Y^\star_{t+1,t}\end{bmatrix}+\begin{bmatrix}T_{11} & T_{12}\\ 0 & T_{22}\end{bmatrix}\begin{bmatrix}X^\star_t\\ Y^\star_t\end{bmatrix}=0$$

where the bottom block becomes:

$$Y^\star_t=-T_{22}^{-1}S_{22}Y^\star_{t+1,t}$$

Note that this is recursive, and if the eigenvalues of the matrix $-T_{22}S_{22}$ are less than 1 (which is true since they are the inverses of the eigenvalues greater than 1 above), then

$$Y^\star_t=-0\cdot Y^\star_{t+n,t}=0$$

and thus, we have that

$$X^\star_t=Z_{11}^{-1}\hat{X}_t$$

Now we have that $S_{11}X^\star_{t+1,t}=-T_{11}X^\star_t$, which implies that

$$X^\star_{t+1,t}=-S_{11}^{-1}T_{11}X^\star_t\qquad\text{where }\ \text{eigs}(-S_{11}^{-1}T_{11})<1$$

and

$$Z_{11}^{-1}\hat{X}_{t+1,t}=-S_{11}^{-1}T_{11}Z_{11}^{-1}\hat{X}_t\Longrightarrow\hat{X}_{t+1,t}=\underbrace{-Z_{11}S_{11}^{-1}T_{11}Z_{11}^{-1}}_{h}\hat{X}_t$$

$$\hat{Y}_t=Z_{21}X^\star_t=\underbrace{Z_{21}Z_{11}^{-1}}_{g}\hat{X}_t$$

This is coded in `gx_hx_alt.m`.

2.4 Shooting Method

Remark. We are going to solve the exact equations of our non-linear model in this case (so no approximations). However, we do make a major assumption:
Assumption 2.1. Agents have perfect foresight
Definition. An *MIT Shock* is a sudden change that agents in the economy did not perceive as possible *ex ante*. This being called an MIT shock is a reference to the saltwater-freshwater divide. It's a stylized experiment, and not micro-founded, but it allows us to approach non-linearities in a principled way.
Remark. This shooting method is very common in modern non-heterogenous agent models.
Model. *Real Business Cycles* (*n.b.* this is the model the example code all uses) We will approach a simplified version of this model:

$$Y_t=A_tK_t^\alpha\qquad;\qquad U(C_t)=\log(C_t)$$

We are in a perfect foresight world (meaning no uncertainty over states, and no expectations), so we get the Euler equation

$$\frac{C_{t+1}}{C_t}=\beta\left(A_{t+1}\alpha K_{t+1}^{\alpha-1}+(1-\delta_k)\right)$$

and the resource constraint

$$A_tK_t^\alpha=C_t+K_{t+1}-(1-\delta_k)K_t\Longrightarrow K_{t+1}=A_tK_t^\alpha+(1-\delta_k)K_t-C_t$$

For now, we fix $A_t = 1 \; \forall \; t$.
Definition. A *phase diagram* is a two-dimensional figure that describes the movement of a model. Figure 1 is a simple one in the RBC model:

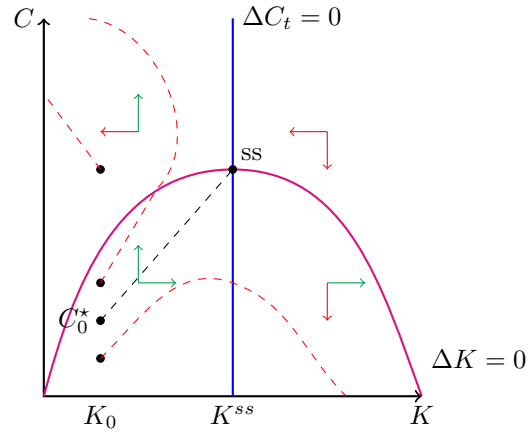


Figure 1: Phase Diagram for RBC Model

The steps to generating this are:

1. Find the line where $\Delta C_t = 0$, fix K^{ss} at that level
2. Suppose that $K_{t+1} = K_t$ ($\Delta K_t = 0$), so $C_t = K_t^\alpha - \delta_k K_t$
3. Fix K_0, and see where the economy would go at any initial C_0. There exists a unique $C_0^\star$ such that $(K_0, C_0^\star)$ is *saddle path stable*. It is the unique stable equilibrium.

Remark. Intuitively, we are thinking about testing every possible C_0, and figuring out which paths end up at the steady state (and satisfy the model equations). We can also work this out algorithmically.
Algorithm. *(Shooting Method)* Fix some K_0. Guess C_0, which gives you K_1 and C_1. Repeat this T times for "fairly large" T, ending with K_T, C_T. We then check if the economy is close to the steady state. Imagine that we are making a function $K^T(K_0, C_0)$, and checking whether

$$K^T(K_0, C_0) - K^{ss} \equiv 0$$

Note well the concept of "fairly large" – we need enough periods to get to the steady state, but not so many that we end with `Inf` or `NaN` or something like that. We will use `fsolve` in Matlab to work through this. This function implements Newton-Raphson to find the optimal guess of C_0. It will have issues around T, and the usual numerical issues for nonlinear equations, but is fairly consistent.
Remark. "This is Angry Birds!" – Ryan Chahrour, Ph.D.

We can also implement the opposite method!
Algorithm. *(Reverse Shooting Method)* We will guess some (K_T, C_T) close to the steady state, and reverse the logical process – get (K_{T-1}, C_{T-1}), then (K_{T-2}, C_{T-2}), all the way to (K_0, C_0). We now have a function $C_0(C_T)$, and solve the equation

$$K_0^\alpha - C_0(C_T) + K_1 - (1 - \delta_k)K_0 \Longleftarrow \texttt{fzero}$$

This algorithm is nice because you tend to not have the `Inf` / `NaN` issues that the normal shooting method runs into.

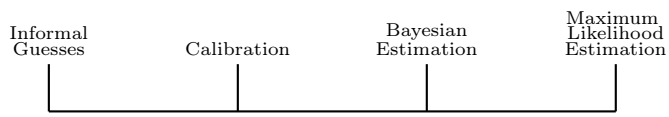
However, Ryan has his own preferred approach to shooting.

Algorithm. *(Chahrour Shooting)* Fix K_0, and fix $C_T = C^{ss}$ (an initial and a terminal condition). We will guess an *entire sequence* of K's and C's. We want to guess

$$\begin{pmatrix} C_0 & C_1 & C_2 & \cdots & \boxed{C_T} \\ \boxed{K_0} & K_1 & K_2 & \cdots & K_t \end{pmatrix}$$

Where K_0 and C_T are fixed. We will check the Euler equation and the resource constraint at every time t, and will solve $F(\{C\}, \{K\})$ using (something similar to) `fsolve`. Intuitively, we are pinning down the starting and ending points, and we correct the path at every point – rather than going through the whole path and seeing how correct it is, we make small changes every step. Since the starting point is precise, we know that it will not go to infinity or zero. It is *very* numerically stable, though it is slow.

Matching Models with Data: a short, somewhat unrelated comment Imagine that you have a spectrum of how related the model is with the data. This might look something like



Informal guesses are informal guesses. Calibration has some parameters based on outside evidence, and some chance to match a few moments of the data. Bayesian Estimation is more formal, uses a likelihood function, but naturally incorporates priors. Maximum likelihood estimation has no priors, and is all based on the data.

2.5 Lagrangian Approach (Decentralized)

Model. *(Decentralized Labor Search)* Before this, we've been working on the social planner's problem. Often, we can reference the Welfare Theorems to say that this is equivalent to the decentralized model. We cannot do this here, so we will work through the households, firms, and market clearing conditions to find a solution.

Households Our households are relatively simple, because they don't make a labor choice. They solve the Lagrangian

$$\max_{\{C_t, \Gamma_t\}} \mathbb{E} \sum_{t=0}^{\infty} \beta^t \left\{ \frac{C_t^{1-\sigma}}{1-\sigma} + \lambda_t \left[W_t N_t + \Gamma_{t-1}(D_t + P_t) - C_t - \Gamma_t P_t \right] \right\}$$

where C_t is consumption, σ is the coefficient of relative risk aversion, λ_t is the Lagrange multiplier on the budget constraint, W_t is the wage rate, N_t is the amount of labor supplied by the household, Γ_t is the number of shares of the company owned, D_t is the dividend per share, P_t is the price of a share, and β is the discount factor. Everything is indexed to time t.

The household's first order conditions are

$$C_t^{-\sigma} = \lambda_t \tag{C}$$
$$\lambda_t P_t = \beta \, \mathbb{E} \left[\lambda_{t+1} (D_{t+1} + P_{t+1}) \right] \tag{Γ}$$

The second FOC becomes

$$P_t = \beta\,\mathbb{E}\left[\underbrace{\left(\frac{C_{t+1}}{C_t}\right)^{-\sigma}}_{\substack{\text{Stochastic}\\ \text{Discount Factor}}}(D_{t+1}+P_{t+1})\right]$$

Note that $\Gamma_t = 1$ for a representative agent model.

Firms The firm's problem assumes that they take wages and the probability that a vacancy gets filled as given. They solve the Lagrangian

$$\begin{aligned}\max_{\{Y_t,I_t,V_t,K_{t+1},N_{t+1}\}} \mathbb{E}\sum_{t=0}^{\infty}\beta^t\Bigg\{&\frac{\lambda_t}{\lambda_0}\left[Y_t - W_tN_t - I_t - \phi_nV_t\right]\\ &+\frac{1}{\lambda_0}\Gamma_{1,t}\left[(1-\delta_n)N_{t-1}+V_tQ_t-N_t\right]\\ &+\frac{1}{\lambda_0}\Gamma_{2,t}[(1-\delta_k)K_t+I_t-K_{t+1}]\\ &+\frac{1}{\lambda_0}\Gamma_{3,t}\left[AK_t^{\alpha}N_t^{1-\alpha}-Y_t\right]\Bigg\}\end{aligned}$$

The firm's first order conditions are

$$\lambda_t = \Gamma_{3,t} \tag{Y}$$

$$\lambda_t = \Gamma_{2,t} \tag{I}$$

$$\Gamma_{1,t} = \lambda_t\frac{\phi_n}{Q_t} \tag{V}$$

$$\frac{\Gamma_{2,t}}{\lambda_0} = \beta\,\mathbb{E}\left[\frac{\Gamma_{3,t+1}}{\lambda_0}A_{t+1}\alpha\left(\frac{K_{t+1}}{N_{t+1}}\right)^{\alpha-1}+\frac{1}{\lambda_0}\Gamma_{2,t+1}(1-\delta_k)\right] \tag{K}$$

$$0 = -\frac{\lambda_t}{\lambda_0}W_t - \frac{1}{\lambda_0}\Gamma_{1,t} + \frac{1}{\lambda_0}\Gamma_{3,t}A_t\left(\frac{K_t}{N_t}\right)^{\alpha}(\alpha-1)+\frac{\beta}{\lambda_0}\mathbb{E}[\Gamma_{1,t+1}(1-\delta_n)] \tag{N}$$

We can reorganize these first order conditions as

$$\underbrace{\frac{\phi_n}{Q_t}}_{\text{Cost}} = \underbrace{A_t\left(\frac{K_t}{N_t}\right)^{\alpha}(1-\alpha)-W_t}_{\text{Net Benefit, } MPL_t - W_t} + \beta\,\mathbb{E}\left[\underbrace{\frac{\lambda_{t+1}}{\lambda_t}(1-\delta_n)\frac{\phi_n}{Q_{t+1}}}_{\text{Expected Savings Tomorrow}}\right] \tag{Decentralized}$$

Recall that the Social Planner's Problem gave:

$$\frac{\phi_n}{M_v(\cdot)} = A_t\left(\frac{K_t}{N_t}\right)^{\alpha}(1-\alpha)+\beta\,\mathbb{E}\left[\frac{\lambda_{t+1}}{\lambda_t}(1-\delta_n)\frac{\phi_n}{M_v(\cdot)}\right] \tag{Social Planner}$$

where $M_v(\cdot) = \chi\varepsilon V_t^{\varepsilon-1}S_t^{1-\varepsilon}$. Consider the differences in these two equations. We have that the probability of a vacancy being matched is defined as

$$Q_t = \frac{\text{Matches}}{\text{Vacancies}} = \frac{\chi V_t^{\varepsilon}S_t^{1-\varepsilon}}{V_t} = \chi\left(\frac{S_t}{V_t}\right)^{1-\varepsilon}$$

while the marginal increase in the number of jobs created by a marginal posting is

$$M_v(V_t, S_t) = \chi\varepsilon \left(\frac{S_t}{V_t}\right)^{1-\varepsilon}$$

If we plug these into our above Euler equations, we get that

$$\frac{\phi_n}{\chi}\left(\frac{V_t}{S_t}\right)^{1-\varepsilon} = MPL_t - W_t + \beta\,\mathbb{E}\left[\frac{\lambda_{t+1}}{\lambda_t}(1-\delta_n)\frac{\phi_n}{\chi}\left(\frac{V_{t+1}}{S_{t+1}}\right)^{1-\varepsilon}\right] \qquad \text{(Decentralized)}$$

$$\frac{\phi_n}{\chi\varepsilon}\left(\frac{V_t}{S_t}\right)^{1-\varepsilon} = MPL_t + \beta\,\mathbb{E}\left[\frac{\lambda_{t+1}}{\lambda_t}(1-\delta_n)\frac{\phi_n}{\chi\varepsilon}\left(\frac{V_{t+1}}{S_{t+1}}\right)^{1-\varepsilon}\right] \qquad \text{(Social Planner)}$$

The first difference is that the social planner can internalize the fact that the probability of a successful matching decreases when an additional vacancy is posted (ε), but the firm in the decentralized economy cannot do that.

The second difference is that in the decentralized economy, the firm gets lower benefits than the marginal social benefits ($-W_t$), as they only get to keep the marginal product after paying the wage, while the social planner just experiences the entire marginal product as a benefit.

These two models coincide only if

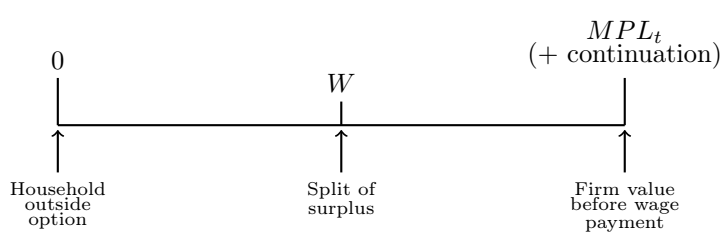
$$\varepsilon MPL_t = MPL_t - W_t$$

Market Clearing In a market clearing setting, we will necessarily have that for the shares, $\Gamma_t = 1$. Treating Γ_t as a choice variable is useful only for pricing the firm's shares. We will also have that the matching function from before holds when the market clears – reprinting it, we have that

$$Q_t = \frac{\text{Matches}}{\text{Vacancies}} = \frac{\chi V_t^{\varepsilon} S_t^{1-\varepsilon}}{V_t} = \chi\left(\frac{S_t}{V_t}\right)^{1-\varepsilon}$$

We also technically have that output clears, meaning that $Y_t = C_t + I_t + \phi_n V_t$. The first and third of these conditions are largely trivial, the real relevant one is Q_t.

Nash Bargaining of Wages Above, we identified that the decentralized economy coincides with the social planner's problem only if $\varepsilon MPL_t = MPL_t - W_t$. Let's think more about when this might happen. In general, in search models there is a surplus for a match, for both the worker and the firm. Let's assume that the worker gets zero earnings for not having a job. Any wage between 0 and MPL_t (+ continuation) would be consistent with equilibrium. We have that the continuum of possible wages is



We will think of a particular wage paradigm, known as *Nash bargaining*, in which workers and firms bargain over the current wage, taking as given the wage that will be realized in future periods.

First, we need to calculate the surplus that the firms and the workers receive for a given equilibrium wage.

Denote $\bar{W}_t(w_t)$ to be the workers' surplus and $J_t(w_t)$ to be the firms' surplus. Given relative bargaining power η and $(1-\eta)$ for both sides, the Nash bargaining solution maximizes:

$$\max_{w_t} [\bar{W}_t(w_t)]^{\eta} [J_t(w_t)]^{1-\eta}$$

which captures the idea that firms and workers share the surplus of a match according to their bargaining powers.

Let's find closed forms for the worker and firm surplus. For the firm, their outside option is zero, because they will pay no wage and earn no profit. If they do make an agreement, their value is

$$J_t(w_t) = \underbrace{MPL_t - w_t}_{\text{Today's profits}} + \underbrace{\beta(1-\delta_n)\,\mathbb{E}\left[\left(\frac{C_{t+1}}{C_t}\right)^{-\sigma} J_{t+1}\right]}_{\text{Future value if not separated}}$$

Note that the worker's benefits of not being matched should be zero – we think that they get no wage and no dividends. We are assuming that $S_t = 1$, so the additional unemployed worker does not affect the stock tomorrow. Their benefits if they do get matched are

$$\bar{W}_t(w_t) = \underbrace{w_t}_{\text{Today's wage}} + \underbrace{\beta(1-\delta_n)\,\mathbb{E}\left[\left(\frac{C_{t+1}}{C_t}\right)^{-\sigma} \bar{W}_{t+1}\right]}_{\text{Future value if not separated}} + \underbrace{\beta\delta_n\,\mathbb{E}\left[\left(\frac{C_{t+1}}{C_t}\right)^{-\sigma} \bar{U}_{t+1}\right]}_{\text{Unemployment benefits; }=0}$$

Taking first order conditions of the problem with respect to w_t, we get that

$$\eta \bar{W}_t(w_t)^{\eta-1} J_t(w_t)^{1-\eta} + \bar{W}_t(w_t)^{\eta}(1-\eta) J_t(w_t)^{-\eta} = 0$$

which rearranges to

$$\bar{W}_t(w_t) = \frac{\eta}{1-\eta} J_t(w_t)$$

Substituting the analytic forms for $\bar{W}$ and J, and using this equation again, we get that

$$\frac{\eta}{1-\eta}\left\{MPL_t - w_t + \beta(1-\delta_n)\,\mathbb{E}\left[\left(\frac{C_{t+1}}{C_t}\right)^{-\sigma} J_{t+1}(w_t)\right]\right\} = w_t + \beta(1-\delta_n)\,\mathbb{E}\left[\left(\frac{C_{t+1}}{C_t}\right)^{-\sigma} \underbrace{\frac{\eta}{1-\eta} J_{t+1}(w_t)}_{\bar{W}_{t+1}(w_t)}\right]$$

Simplifying, we get that

$$\eta(MPL_t - w_t) = (1-\eta) w_t \Longrightarrow \boxed{w_t^{\star} = W_t = \eta \cdot MPL_t + (1-\eta)\cdot 0}$$

which is the solution to the Nash bargaining problem. The decentralized problem and the social planner's problem coincide only if $\eta = 1 - \varepsilon$, which we call the *Hosios Condition* (after Hosios (1990)).

Intuition. Remember that this is not a First Welfare Theorem (Theorem 1.1) economy. The fact that there can be efficiency here is actually somewhat strange – we would not expect that. However, we can think about this as a sort of balancing act between the private benefit of hiring a worker and the social benefit of hiring a worker. We can think about the following conditions:

$$\begin{aligned}
\text{Private Benefit} &= MPL_t - W_t & < \quad & \text{Social Benefit} = MPL_t \\
\text{Private Cost} &= \frac{\phi_n}{\chi}\left(\frac{V_t}{S_t}\right)^{1-\varepsilon} & < \quad & \text{Social Cost} \quad = \frac{\phi_n}{\chi\varepsilon}\left(\frac{V_t}{S_t}\right)^{1-\varepsilon}
\end{aligned}$$

The first inequality would indicate that firms tend to under-hire, and the second would indicate that firms tend to over-hire. The Hosios Condition is the exact point where they exactly cancel out, and the firm hires at exactly at the social optimum. When $\eta > 1-\varepsilon$, firms under-hire (because worker bargaining power is too high), and when $\eta < 1-\varepsilon$, firms over-hire (because firm bargaining power is too high).

2.6 Value Function Approach (Social Planner)

Roadmap. We have seen the same problem through the social planner's perspective and the decentralized perspective, using the same strategy – solving the Lagrangian. We will now move to the same problem(s), solved through value function iteration. We've also seen two main techniques – log-linearization, which is linear but allows for uncertainty, and the shooting method, which is nonlinear but assumes perfect foresight. Both of these techniques can solve both the planner's problem and the decentralized problem, but say nothing about the value function.

Value function iteration is both nonlinear and allows for uncertainty, but can only solve the planner's problem – it works through the value function, which cannot be used in a decentralized way.

After value function iteration, we will explore projection methods, which are nonlinear, allow for uncertainty, and work for both the planner's problem and the decentralized problem; at the cost of time of implementation. We will explore that through the Lagrangian approach.

Model. *Value Function (Social Planner)* The social planner has the Bellman Equation:

$$V(K_t, N_{t-1}, A_t) = \max_{K_{t+1}, N_t} U_t(C_t) + \beta\,\mathbb{E}\left[V(K_{t+1}, N_t, A_{t+1})\right]$$

subject to

$$\begin{aligned}
C_t &= Y_t - I_t - \phi_n V_t \\
&= A_t K_t^{\alpha} N_t^{1-\alpha} - (K_{t+1} - (1-\delta_k)K_t) - \phi_n \left(\frac{N_t - (1-\delta_n)N_{t-1}}{\chi}\right)^{\frac{1}{\varepsilon}}
\end{aligned}$$

where the second equality uses the fact that $Y_t = A_t K_t^{\alpha} N_t^{1-\alpha}$, $I_t = K_{t+1} - (1-\delta_k)K_t$, and rearranging $N_t = (1-\delta_n)N_{t-1} + \chi V_t^{\varepsilon} S_t^{1-\varepsilon}$, we get that

$$V_t = \left(\frac{N_t - (1-\delta_n)N_{t-1}}{\chi}\right)^{\frac{1}{\varepsilon}}$$

Substituting into the planner's Bellman equation, it becomes

$$\begin{aligned}
V(K_t, N_{t-1}, A_t) = \max_{K_{t+1}, N_t} U_t &\left(A_t K_t^{\alpha} N_t^{1-\alpha} - (K_{t+1} - (1-\delta_k)K_t) - \phi_n \left(\frac{N_t - (1-\delta_n)N_{t-1}}{\chi}\right)^{\frac{1}{\varepsilon}}\right) \\
&+ \beta\,\mathbb{E}\left[V(K_{t+1}, N_t, A_{t+1})\right]
\end{aligned}$$

The first order conditions are:

$$
\begin{aligned}
(K_{t+1}) \qquad & 0 = -U'(C_t) + \beta\,\mathbb{E}[V_K(K_{t+1}, N_t, A_{t+1})] \\
(N_t) \qquad & 0 = U'(C_t)\left[-\frac{\phi_n}{\chi\varepsilon}\left(\frac{N_t - (1-\delta_n)N_{t-1}}{\chi}\right)^{\frac{1-\varepsilon}{\varepsilon}} + A_t\left(\frac{K_t}{N_t}\right)^{\alpha}(1-\alpha)\right] + \beta\,\mathbb{E}[V_N(K_{t+1}, N_t, A_{t+1})]
\end{aligned}
$$

Envelope Theorem Aside. We need to take the derivatives of $V_K(\cdot)$ and $V_N(\cdot)$. The Envelope Theorem tells us that we can ignore the chain rule in taking these, and proceed with just the obvious terms. We will do a quick proof for $V_K(\cdot)$.

Suppose that we know the policy functions already, and call them $K_{t+1} = \mathbf{K}(K_t, N_{t-1}, A_t)$ and $N_t = \mathbf{N}(K_t, N_{t-1}, A_t)$. Our value function is now:

$$
\begin{aligned}
V(K_t, N_{t-1}, A_t) =& U\left(A_t K_t^{\alpha}\mathbf{N}(\cdot)^{1-\alpha} - (\mathbf{K}(\cdot) - (1-\delta_k)K_t) - \phi_n\left(\frac{\mathbf{N}(\cdot) - (1-\delta_n)N_{t-1}}{\chi}\right)^{\frac{1}{\varepsilon}}\right) \\
&+ \beta\,\mathbb{E}[V(\mathbf{K}(\cdot), \mathbf{N}(\cdot), A_{t+1})]
\end{aligned}
$$

Note that we don't have the max operator anymore! We can take the derivative with respect to K_t, using all of the appropriate chain rules. We get that

$$
\begin{aligned}
V_K(K_t, N_{t-1}, A_t) = U'(C_t)\cdot & \left[A_t\left(\frac{K_t}{N_t}\right)^{\alpha-1}\alpha + (1-\delta_k)\right] \\
& - U'(C_t)\cdot\frac{\partial\mathbf{K}}{\partial K_t} + \beta\,\mathbb{E}\left[V_K(K_{t+1}, N_t, A_{t+1})\right]\frac{\partial\mathbf{K}}{\partial K_t} \\
& + U'(C_t)\left[\frac{\phi_n}{\varepsilon\chi}\left(\frac{N_t - (1-\delta_n)N_{t-1}}{\chi}\right)^{\frac{1-\varepsilon}{\varepsilon}} + A_t\left(\frac{K_t}{N_t}\right)^{\alpha}(\alpha-1)\right]\frac{\partial\mathbf{N}}{\partial K_t} \\
& + \beta\,\mathbb{E}\left[V_N(K_{t+1}, N_t, A_{t+1})\right]\frac{\partial\mathbf{N}}{\partial K_t}
\end{aligned}
$$

However, from our first order conditions above, if we are at an optimum, we have that the second line should evaluate to zero! Also, the third and fourth lines should evaluate to zero! We finally have that

$$
V_K(K_t, N_{t-1}, A_t) = U'(C_t)\cdot\left[A_t\left(\frac{K_t}{N_t}\right)^{\alpha-1}\alpha + (1-\delta_k)\right]
$$

Back to FOCs. Using the Envelope Theorem, we have that

$$
\begin{aligned}
V_K(K_t, N_{t-1}, A_t) &= U'(C_t)\cdot\left[A_t\left(\frac{K_t}{N_t}\right)^{\alpha-1}\alpha + (1-\delta_k)\right] \\
V_N(K_t, N_{t-1}, A_t) &= U'(C_t)\cdot\left[\frac{\phi_n}{\varepsilon\chi}\left(\frac{N_t - (1-\delta_n)N_{t-1}}{\chi}\right)^{\frac{1-\varepsilon}{\varepsilon}}(1-\delta_n)\right]
\end{aligned}
$$

Combining the first order conditions and the Envelope Theorem results, we get that

$$U_t'(\cdot) = \beta\,\mathbb{E}\left[U_{t+1}'(\cdot)\left(A_{t+1}\left(\frac{K_{t+1}}{N_{t+1}}\right)^{\alpha-1}\alpha + (1-\delta_k)\right)\right]$$

$$U_t'(\cdot)\left[\frac{\phi_n}{\varepsilon\chi}\left(\frac{N_t-(1-\delta_n)N_{t-1}}{\chi}\right)^{\frac{1-\varepsilon}{\varepsilon}}\right] = U_t'(\cdot)\left[A_t\left(\frac{K_t}{N_t}\right)^{\alpha}(1-\alpha)\right]$$
$$+\beta\,\mathbb{E}\left[U_{t+1}'(\cdot)\left(\frac{\phi_n}{\varepsilon\chi}\left(\frac{N_{t+1}-(1-\delta_n)N_t}{\chi}\right)^{\frac{1-\varepsilon}{\varepsilon}}(1-\delta_n)\right)\right]$$

Rearranging and imposing the functional form for the matching function, these clearly become the capital and labor Euler equations we found above in the planner's problem.

2.7 Mathieu Interlude (Labor Search Formalities)

Motivation. We spent a lot of time last quarter on the neoclassical growth model. It has some distinct drawbacks: it assumes a unique labor market with a market clearing wage; it does not capture the process of finding a job particularly well (think search, vacancy posting, matching, bargaining, etc.); and there is no unemployment! We turn to a new class of models that take these details seriously, for which Peter Diamond, Christopher Pissarides, and Dale Mortensen won the 2010 Nobel Prize.

For this section, we reference Ljundqvist and Sargent chapters 6 and 26, as well as Roger, Shimer, & Wright (2005) and Pissarides (2000).

Math preliminaries Let p be a random variable with CDF $F(P) = \mathbb{P}\{p \leq P\}$. Assume that $F(0) = 0$, $F(\infty) = 1$, and F continuous from the right. Assume that p is bounded above with probability 1 – so $\exists\, B$ s.t. $F(B) = 1$. Recall that the mean of p is given by

$$\mathbb{E}[p] = \int_0^B p dF(p)$$

Letting $u = 1 - F(p)$ and $v = p$, we can integrate by parts and get that

$$\mathbb{E}[p] = \int_0^B p dF(p) = \int_0^B (1-F(p))dp = B - \int_0^B F(p)dp$$

Now consider two independent random variables $p_1, p_2 \sim F$ and consider the event $\{(p_1 \leq p) \cap (p_2 \leq p)\}$. This event happens with probability $(F(p))^2$, and this is equivalent to the event $\{\max\{p_1, p_2\} \leq p\}$. Using our previous result, we have that

$$\mathbb{E}[\max\{p_1, p_2\}] = B - \int_0^B F(p)^2 dp$$

Generalizing with n independent draws, we get that

$$M_n \equiv \mathbb{E}[\max\{p_1, p_2, \ldots, p_n\}] = B - \int_0^B F(p)^n dp$$

Stigler Model

Model. *Stigler Search Model* (from Stigler (1961)) This is a partial equilibrium model of an agent looking for a job. We have a risk-neutral agent who samples i.i.d. wages from some distribution $F(w)$ with the earlier assumptions. Their *ex ante* decision of how many wages to gather is n, and getting a wage offer has cost c. How many offers should they ask for?
Definition. The *expected gain* from an additional draw is

$$\begin{aligned} G_n = M_n - M_{n-1} &= B - \int_0^B F(p)^n dp - \left(B - \int_0^B F(p)^{n-1} dp \right) \\ &= \int_0^B F(p)^{n-1} dp - \int_0^B F(p) F(p)^{n-1} dp \\ &= \int_0^B F(p)^{n-1} (1 - F(p)) dp \end{aligned}$$

Note that G_n decreases with n, and $\lim_{n\to\infty} G_n = 0$. The optimal rule is to pick n such that $G_n \geq c > G_{n+1}$.
Question. What's weird with this model? Static search. We tend to think of searching for a job as a sequential thing, not a single time drawing a large number of wages. McCall (1970) fixes this.

Consider a class of distributions $F(p, r) = \mathbb{P}\{P_r \leq p)$ indexed by $r \in \mathbb{R}$. $F(p, r)$ is differentiable with regard to r for all $p \in [0, B]$, and there exists $B \in \mathbb{R}$ such that $F(B, r) = 1$, and $F(0, r) = 0 \ \forall \ r \in \mathbb{R}$. Since $\mathbb{E}[p] = B - \int_0^B F(p, r) dp$, two distributions with the same $\int_0^B F(p, r) dp$ have the same mean.
Definition. We say that a distribution r_2 is a *mean-preserving spread* of a distribution r_1 if

1. Identical means condition:
$$\int_0^B (F(\theta, r_1) - F(\theta, r_2)) d\theta = 0$$

2. Single-crossing property: There exists $\hat{\theta} \in (0, B)$ such that
$$F(\theta, r_2) - F(\theta, r_1) \leq 0 \ (\geq 0) \text{ when } \theta \leq \ (\geq) \ \hat{\theta}$$

This is illustrated in Figure 2.

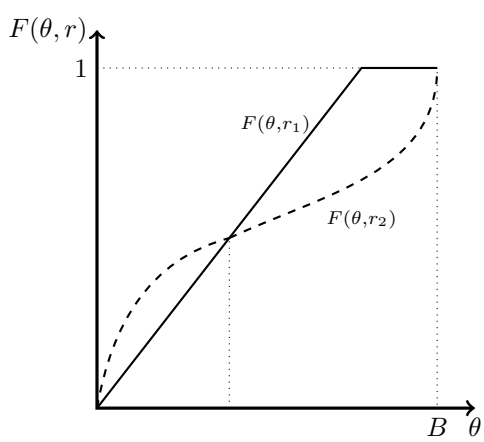


Figure 2: Mean-Preserving Spreads

Properties 1 and 2 imply that

$$\int_0^y (F(\theta, r_2) - F(\theta, r_1))d\theta \geq 0 \; \forall \; y \in [0, B]$$

For infinitesimal changes in r, an increase in r is said to represent a *mean-preserving increase in risk* if

$$\int_0^B F_r(\theta, r)d\theta = 0 \qquad \text{and} \qquad \int_0^y F_r(\theta, r)d\theta \geq 0 \; \forall \; \theta \in [0, B]$$

where $F_r(\theta, r) = \frac{\partial F(\theta,r)}{\partial r}$.

Model. *McCall Labor Search* (from McCall (1970)) An agent searches for a job, taking market conditions as given. Each period the agent draws one offer $w \sim F(W)$, where $F(0) = 0$ and $F(B) = 1$ for some $B < \infty$. The agent can accept or reject the offer. If she rejects, she gets c today and draws another offer tomorrow. If she accepts, she receives w per period forever.

The agent maximizes

$$\mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t y_t\right]$$

where $y_t = c$ if she is unemployed and $y_t = w$ if she is employed at wage w.

Solving the model. Denote by $v(w)$ the expected value of an offer w for an agent who is deciding whether or not to accept the offer or reject it. If she behaves optimally, we have

$$v(w) = \max\left\{\frac{w}{1-\beta}, c + \beta \int v(w')dF(w')\right\}$$

The solution of this Bellman equation is of the form

$$v(w) = \begin{cases} \frac{\bar{w}}{1-\beta} = c + \beta \int v(w')dF(w') & \text{if } w \leq \bar{w} \\ \frac{w}{1-\beta} & \text{if } w > \bar{w} \end{cases}$$

where $\bar{w}$ is called the *reservation wage*. The agent's value function is illustrated in Figure 3.

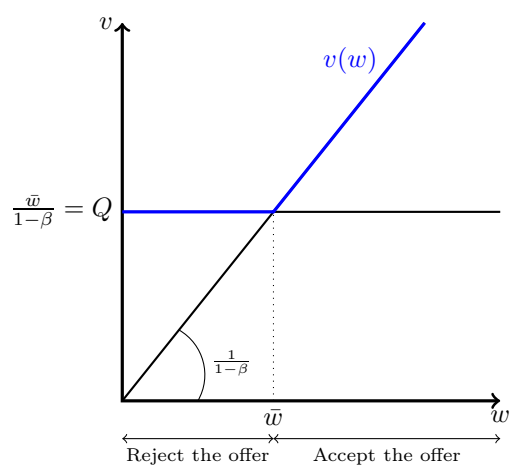


Figure 3: Value per Wage Offer

How do we find $\bar{w}$? At $w = \bar{w}$, the agent is indifferent, so

$$\frac{\bar{w}}{1-\beta} = c + \beta \int_0^{\bar{w}} \frac{\bar{w}}{1-\beta} dF(w') + \beta \int_{\bar{w}}^{B} \frac{w'}{1-\beta} dF(w')$$

or

$$\frac{\bar{w}}{1-\beta} \int_0^{\bar{w}} dF(w') + \frac{\bar{w}}{1-\beta} \int_{\bar{w}}^{B} dF(w') = c + \beta \int_0^{\bar{w}} \frac{\bar{w}}{1-\beta} dF(w') + \beta \int_{\bar{w}}^{B} \frac{w'}{1-\beta} dF(w')$$

or

$$\bar{w} \int_0^{\bar{w}} dF(w') - c = \frac{1}{1-\beta} \int_{\bar{w}}^{B} (\beta w' - \bar{w}) dF(w')$$

Adding $\bar{w} \int_{\bar{w}}^{B} dF(w')$ to both sides, we get

$$\begin{aligned} \bar{w} - c &= \frac{\beta}{1-\beta} \int_{\bar{w}}^{B} (w' - \bar{w}) dF(w') \\ &= \beta \, \mathbb{E}\left[\frac{w' - \bar{w}}{1-\beta} \middle| w' \geq \bar{w} \right] \mathbb{P}\{w' \geq \bar{w}\} \end{aligned}$$

where the left hand side is the cost of searching one more time with offer $\bar{w}$ in hand, and the right hand side is the surplus from searching one more time and getting a better offer. These two things must be equal if the agent is optimizing: $MR = MC$.

We can define the function

$$h(w) := \frac{\beta}{1-\beta} \int_{w}^{B} (w' - w) dF(w')$$

Note that:

$$\begin{aligned} h(0) &= \mathbb{E}\left[\frac{\beta w}{1-\beta} \right] \\ h(B) &= 0 \\ h(w) &\text{ is differentiable} \\ h'(w) &= -\frac{\beta}{1-\beta}(1 - F(w)) < 0 \\ h''(w) &= \frac{\beta}{1-\beta} F'(w) > 0 \end{aligned}$$

$h(w)$ is illustrated in Figure 4.

Question. What happens when the environment changes? Let's manipulate the equations a bit more.

$$\begin{aligned} \bar{w} - c &= \frac{\beta}{1-\beta} \int_{\bar{w}}^{B} (w' - \bar{w}) dF(w') + \frac{\beta}{1-\beta} \int_0^{\bar{w}} (w' - \bar{w}) dF(w') - \frac{\beta}{1-\beta} \int_0^{\bar{w}} (w' - \bar{w}) dF(w') \\ &= \frac{\beta}{1-\beta} \mathbb{E}[w] - \frac{\beta}{1-\beta} \bar{w} - \frac{\beta}{1-\beta} \int_0^{\bar{w}} (w' - \bar{w}) dF(w') \end{aligned}$$

or

$$\bar{w} - (1-\beta)c = \beta \, \mathbb{E}[w] - \beta \int_0^{\bar{w}} (w' - \bar{w}) dF(w')$$

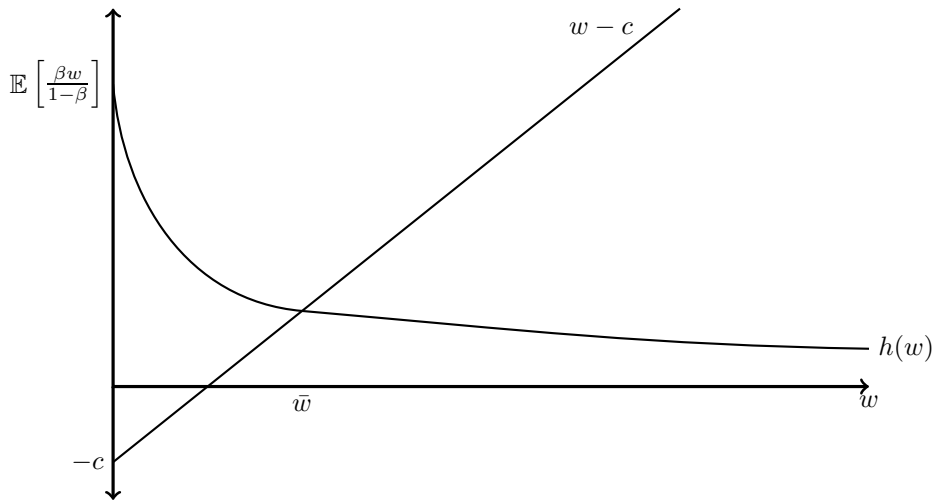


Figure 4: Finding the Reservation Wage, part 1

Integrating by parts, we get

$$\bar{w} - c = \beta(\mathbb{E}[w] - c) + \beta \int_0^{\bar{w}} F(w')dw' \equiv \beta(\mathbb{E}[w] - c) + \beta g(\bar{w})$$

This relationship is illustrated in Figure 5.

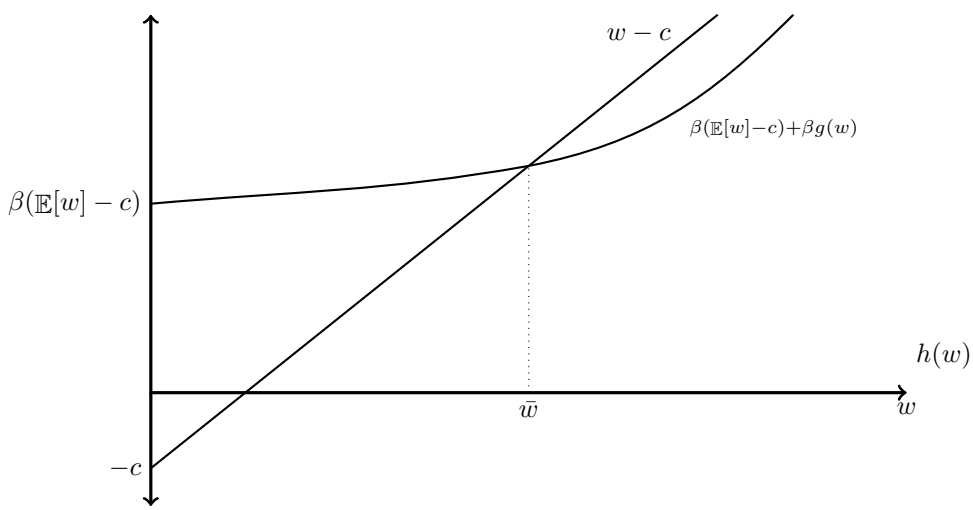


Figure 5: Finding the Reservation Wage, part 2

Question. What happens when c increases? Both curves move to the right, the reservation wage increases.
Question. What happens when there is a mean-preserving increase in risk? $g(\bar{w})$ increases, meaning that the reservation wage increases. Intuition: there are more bad jobs, but we don't care about those. Instead, the agent sees the upside of the better jobs that are available.

Problems with the model. Firm behavior isn't considered, but it gets weird. Workers follow a reservation wage strategy, so firms do not gain anything for posting $w > \bar{w}$. Firms also do not hire anyone when they

post $w < \bar{w}$. Therefore, $F(w)$ will have a unit mass at $\bar{w}$ (Rothschild (1974)). Moreover, from Diamond (1971),

$$\begin{aligned}
\bar{w} - c &= \beta(\mathbb{E}[w] - c) + \beta \int_0^{\bar{w}} F(w')dw' \\
\bar{w} - c &= \beta(\bar{w} - c) \\
\bar{w} &= c
\end{aligned}$$

Intuition? Firms know that they can post at $\bar{w}$ and get that everyone will accept. Then workers will re-optimize, and $\bar{w}$ will decrease, and the process will repeat. We will unravel, until $\bar{w} = c$, the theoretical minimum.

2.8 Value Function Iteration

Numerically, we are going to be finding a value function (and associated policy functions) that satisfy

$$V(K_t, N_{t-1}, A_t) = \max_{K_{t+1}, N_t} U_t(C_t) + \beta \, \mathbb{E}\left[V(K_{t+1}, N_t, A_{t+1})\right]$$

We need to make some choices:

1. How should we summarize the state space? Traditionally, we use a discrete grid. We will use an even grid for K_t and N_{t-1}, and will use either the Tauchen process (from Mathieu's section) or the Rouwenhorst Method (in the sample code, and see Kopecky and Suen (2009) for more).
2. How should we compute expectations? Using the Rouwenhorst approximation weights, essentially.
3. How do we solve for the maximum? We simply optimize over the discrete grid.

2.9 Projection Methods

We will begin talking about projection methods, a functional approximation of the problem. Projection methods, as talked about above, are slow but very precise, and deal with nonlinearities nicely.

The theoretical object we are concerned with is $h(x)$, the continuous policy function. We will approximate it as

$$\hat{h}(x) = \sum_{n=1}^{N} \underbrace{a_n}_{\text{Weights}} \underbrace{b_n(x)}_{\text{Basis functions}}$$

These basis functions will end up being simple functions in the mathematical sense as well, which is lovely! Mainly, we care about them being numerically simple. We will ideally solve the problem

$$\min_{\{a_n\}} \int_{\mathcal{X}} (h(x) - \hat{h}(x))^2 dx$$

Practically, we will minimize

$$\min_{\{a_n\}} \sum_{i \in \mathcal{X}} w_i (h(x_i) - \hat{h}(x_i))^2$$

Special Case. When $|\mathcal{X}| = N$, we will often be able to find an exact solution, where the minimization problem goes to zero. This is called *interpolation*.

We will make two key choices: the basis functions and the loss function (or interpolation points). We are assuming least squares above, but this would also include the weights w_i.

Choosing Basis Functions. We have three common choices for basis function:

1. Taylor Polynomial. In the scalar case, we have:

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_N x^N$$

   Positives: this is differentiable. Negative: slow-spanning.

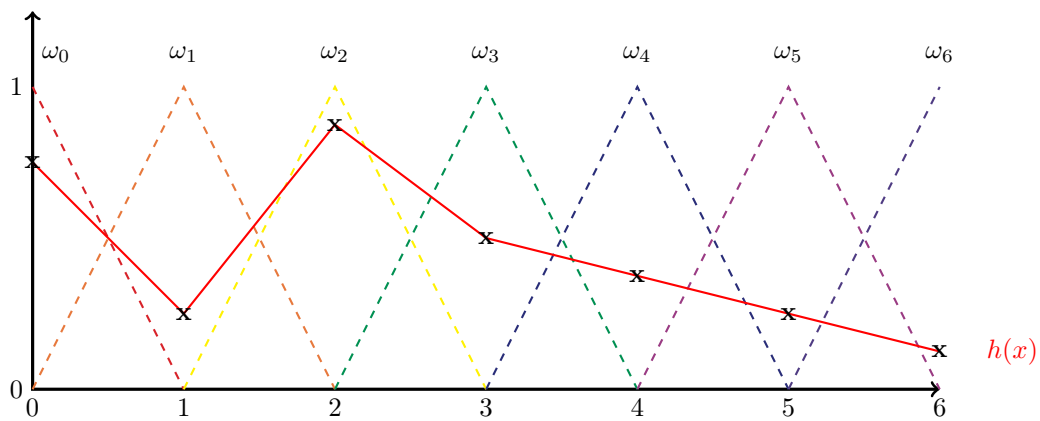
2. Chebyshev Polynomials. These are polynomials which are composed of trigonometric transformations. There's not a nice closed form for arbitrary h, but they have some nice properties generally.

   Positives: also differentiable, and quick-spanning. Negative: weird behavior off-grid.

3. Finite element basis functions. These only have local support, and are represented with linear 'hat' functions. They are functions which are linear, and a maximum of two of them are non-zero at any one time. We can imagine that we know $h(x)$ at $x = \{0, 1, 2, \ldots, 6\}$, and we define h as a weighted average of the relevant basis functions ω_i with weights b_i:

$$\min_{\{b_i\}} \sum_{x_j} \left(h(x_j) - \sum_{i=0}^{6} a_i \omega_i(x_j) \right)^2$$

   where $\sum_{i=0}^{6} a_i \omega_i(x_j) = \hat{h}(x_j)$. It's easiest to think of these graphically.



   As you can see, with finite points we can easily solve the optimization problem by setting $a_i = h(x_i)$. Finding the weights $\{a_i\}$ for any basis function $w_i(x)$ on grid points $x_0, x_1, \ldots, x_n$, where n is not necessarily equal to the number of basis functions. We want that

$$\begin{bmatrix} h(x_0) \\ h(x_1) \\ \vdots \\ h(x_n) \end{bmatrix} = \begin{bmatrix} w_0(x_0) & w_1(x_0) & \cdots \\ w_0(x_1) & w_1(x_1) & \cdots \\ \vdots & \cdots & \ddots \\ w_0(x_n) & w_1(x_n) & \cdots \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \end{bmatrix} + \begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \varepsilon_2 \\ \vdots \end{bmatrix}$$

   where ε_i represents the error. If this was an interpolation case, the matrix would be square and ε_i would go to zero and we would have a unique fit. In the general case, this is a least squares problem.

Remark. (1) and (2) are global basis functions, that are non-zero for all x. (3) is local, and will only be non-zero for some x.
Remark. Ryan has a nice example in his sample code of doing this with a strange (sort of logistic?) function.

Some Thoughts on Projection. Recall our generic macro model function from before:

$$\mathbb{E}\left[f(X_t, Y_t, X_{t+1}, Y_{t+1})\right] = 0$$

Our goal is to find policy functions $Y_t = g(X_t)$, $X_{t+1} = h(X_t, \varepsilon_{t+1})$. These functions are smooth, infinite-dimensional (actually, uncountably dimensional) objects. However, they do satisfy

$$\mathbb{E}\left[f(X_t, g(X_t), h(X_t, \varepsilon_{t+1}), g(h(X_t, \varepsilon_{t+1})))\right] = 0$$

meaning that we can convert the model into terms of just the state variables and the shocks. The idea of the projection strategy is that we will replace $g(\cdot)$ and $h(\cdot)$ with approximations:

$$\hat{g}(\cdot) = \sum a_n^g b_n(X_t) \qquad \text{and} \qquad \hat{h}(\cdot) = \sum a_n^h b_n(X_t)$$

Thus,

$$\mathbb{E}\left[f(X_t, \hat{g}(X_t), \hat{h}(X_t, \varepsilon_{t+1}), (\hat{g} \circ \hat{h})(X_t, \varepsilon_{t+1}))\right] = 0$$

Our goal is to find $\{a_n^g, a_n^h\}$ such that this is satisfied *as well as possible.*

We need to think about a few things:

1. Unlike above, we now do not know the true $h(\cdot)$, $g(\cdot)$. Instead, they are implicitly defined by the macro model function above. We want to use that equation to get feedback on how good our approximations are.
2. We need a metric to actually get any feedback. Ryan proposes least squares on some grid of states – note that we need to declare a grid here! This choice of metric is where the term *projection* is coming from.
   (a) We can't use OLS, since the problem is nonlinear. Instead we will use the function lsqnonlin in Matlab.
   (b) If n(states) $= n$(parameters), then this is often called *collocation*. Note that the states is the number of grid points, and the number of parameters is the subscript on a.
3. Computing $\mathbb{E}[\,\cdot\,]$ is *hard*. Ryan proposes we use Gaussian-Hermite Quadrature.
4. We know the part of $h(\cdot)$ that corresponds to exogenous processes (*e.g.* A_t). We will just use the "true" exogenous process for A_t and only find $\hat{h}(\cdot)$ for endogenous states.

Definition. *Gaussian-Hermite Quadrature* (GHQ, link) is a method to compute expected values. Thinking back, we would ideally want to integrate f with respect to the density of the exogenous shocks. However, that would be incredibly hard to do – it's a complicated object. GHQ is a way of approximating that integral, especially when the density of the shocks is a Gaussian density. It will take a grid over ε, and evaluate the integral over those points. Generically, the approximation for

$$\int f(\varepsilon)\phi(\varepsilon)d\varepsilon$$

is essentially a Riemann sum, or a trapezoidal integration approximation. The Gaussian-Hermite strategy will (cleverly) choose the grid such that the integral is exact for polynomials of order $2N - 1$ (or less) if N is the number of grid points. Ryan has code that does this.

2.10 Projection Methods Application (RBC Model)

Recall that the Real Business Cycle model is described by $\log(C_t) - \chi h_t$. This can be reduced to the equations:

$$C_t^{-1} = \beta\,\mathbb{E}\left[C_{t+1}^{-1}\left(A_{t+1}\alpha\left(\frac{K_{t+1}}{h_{t+1}}\right)^{\alpha-1} + 1 - \delta_k\right)\right] \qquad \text{(Capital FOC)}$$

$$C_t = A_t(1-\alpha)\left(\frac{K_t}{h_t}\right)^{\alpha}\chi^{-1} \qquad \text{(Labor FOC)}$$

Our goal is to find the policy functions $h(A_t, K_t)$ and $K(A_t, K_t)$. Consider the following (pseudocode) algorithm:

Algorithm. The process is as follows:

1. **Initialization:** We first choose a grid on the state space $\{A^i\}$ and $\{K^i\}$. We will use the basis functions, and guess initial weights a_n^h, a_n^K to compute $\hat{h}(\cdot) = \sum a_n^h b_n(A, K)$ and $\hat{K}(\cdot) = \sum a_n^K b_n(A, K)$. We will then use Gaussian-Hermite Quadrature to get $\{\varepsilon^m\}$ and weights $\{p^m\}$.
2. **Residual Function:** We will introduce a function computing the residuals based on $\{a_n^h\}$, $\{a_n^K\}$, and other fixed parameters.
   (a) This function will loop over i, which indexes all possible combinations of A and K, and for each point in the initial state grid we will compute $h_i = \hat{h}(A^i, K^i)$ and $(K^i)' = \hat{K}(A^i, K^i)$.
   (b) Then, we will compute C^i using the resource constraints, and we can move on to potential futures, with our computed present values. Still in the initial loop, we will loop over m, which indexes the integration nodes from GHQ. We will have $(A^{i,m})' = \exp(\rho\log(A^i) + \varepsilon^m)$, $(K^{i,m})' = (K^i)'$, and $(h^{i,m})' = \hat{h}((A^{i,m})', (K^{i,m})')$. Finally, we will compute $(C^{i,m})'$ using the resource constraints, and end the m loop.
   (c) Now, back in the i loop, we will compute expectations from the Capital and Labor FOC. We will compute the following:
   $$\text{RHS}_1^i = \beta\sum_m p^m\left[(C^{i,m})'\left((A^{i,m})'\alpha\left(\frac{(K^{i,m})'}{(h^{i,m})'}\right)^{\alpha-1} + 1 - \delta_k\right)\right]$$
   and we have that the computed residuals are r_1^i and r_2^i, which are defined as:
   $$r_1^i \coloneqq (C^i)^{-1} - \text{RHS}_1^i$$
   $$r_2^i \coloneqq C^i - A^i(1-\alpha)\left(\frac{K^i}{h^i}\right)^{\alpha}\chi^{-1}$$
   (d) Finally, we will end the loop, and return the vectors $\{r_1^i\}$, $\{r_2^i\}$.
3. **Solve:** Use a function solver (in Matlab, fsolve) to find the $\{a_n^h, a_n^K\}$ that solve the residual function.

2.11 Business Cycles and Macroeconomic News

Consider the following quote, from Robert Lucas: "One is led by the facts to conclude that, with respect to the qualitative behavior of comovements among series, *business cycles are all alike*."

Going back in time, for the 1980s and 1990s, 'surprise technology' was the main business cycle shock. Barro & King (1984) consider the planner's problem of the real business cycle model, and solve

$$\max \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t \left(\log(C_t) - \chi L_t + \lambda_t(r_t K_t + w_t L_t - C_t - I_t)\right)\right]$$

The first order conditions give us

$$\frac{1}{C_t} = \lambda_t \qquad (C_t)$$

$$\chi = \lambda_t w_t = \lambda_t A_t (1-\alpha) K_t^{\alpha} L_t^{-\alpha} \qquad (L_t)$$

Combining, we get

$$C_t = \frac{1}{\chi} A_t (1-\alpha) K_t^{\alpha} L_t^{-\alpha} \qquad \text{(FOC Constraint)}$$

This equation is a very powerful argument for considering technology shocks as business cycle shocks. Note that K_t is predetermined. If we wanted C_t and L_t to go in the same direction is by changing A_t – everything else is exogeneous. If A_t increases, then L_t could increase and C_t would still increase.

We want C and L to move together – from Lucas above (and also just qualitative intuition). The only way to do that in rational expectations models is by technology changing.

In the 1980s and early 1990s, macroeconomics was mostly about calibrating the process by which A_t changed, and seeing if that could explain elements such as the variance, covariance, etc. of C_t, L_t, and I_t. Breaking the first order condition is hard, because we can't do it in a microfounded way – this is our straitjacket, and it very much limits how we think about business cycles.

In 1994, John Cochrane wrote Shocks, in which he talks about being unhappy with macroeconomics. His argument is that (i) technology is precisely measurable, and there is no empirical evidence that A_t moves at the same time as C_t, L_t, I_t; and (ii) all of the other candidate shocks (monetary policy, oil prices, government spending, etc.) that we could measure have their own issues. His conclusion is that maybe something like macroeconomic 'news' could do the trick – anticipation of future changes. There are two main issues with this: it's incredibly hard to measure the news, and doing so would still necessitate changing the models.

In the late 1990s, there is the development of *New Keynesian Macroeconomics*. The really big innovation is that new Keynesianism breaks the FOC constraint in a (at least somewhat) micro-founded way: sticky prices. In the New Keynesian literature, at least one of the following holds:

$$w \neq \text{marginal product of labor} \qquad \text{or} \qquad w \neq \text{marginal disutility of labor}$$

Sticky wages lead to the first, and sticky prices lead to the second. Next semester, Chris Neimark will go through this. This literature tends to focus on monetary policy shocks which, ironically, means that surprise A_t shocks generally do not give comovement the way we want. The culmination of this is Christiano, Eichenbaum, and Evans (2005), which matched a lot of shocks and frictions. This paper is extremely important and successful, and also a strawman for people arguing against this literature.

Another paper is Beaudry and Portier (2006), which uses structural autoregression to identify evidence of news shocks. They are using stock prices, and essentially looking for news that moves stock prices today, as evidence for macroeconomic movement tomorrow. This provides some optimism for Cochrane's measurement challenge.

Model. *(Jaimovich and Rebelo Model)* (Jaimovich and Rebelo (2009)) This model showed how to get comovement from productivity news for the first time outside of the new Keynesian models – they show it in a *real* model. They use Greenwood-Herkowitz-Huffman (like) preferences, which break the separability

between labor and consumption as follows:

$$U(C, L) = \frac{(C - \psi L)^{1-\sigma}}{1-\sigma}$$

They also include in their model variable capacity utilization, where they add a variable μ_t representing the household's choice of depreciation – essentially, you can work your capital harder, causing faster depreciation. The production function becomes $A_t(\mu_t K_t)^\alpha L_t^{1-\alpha}$.

Finally, they have investment adjustment costs, where

$$K_{t+1} = (1-\delta_k)K_t + I_t\left(1 - \phi\left(\frac{I_t}{K_t} - \delta_k\right)\right)$$

Essentially, in this model we learn about productivity ahead of time, anticipate the technology growth, and the economy responds.

Taking first order conditions, we get that

$$\lambda_t = (C_t - \psi L_t)^{-\sigma} \qquad (C_t)$$

$$\lambda_t w_t = \underbrace{(C_t - \psi L_t)^{-\sigma}}_{\substack{\text{Marginal}\\ \text{disutility}\\ \text{of } L_t}} \psi \qquad (L_t)$$

Combining, using the marginal product of labor, we get

$$(C_t - \psi L_t)^{-\sigma}\psi = (C_t - \psi L_t)^{-\sigma} A_t(1-\alpha)(\mu_t K_t)^\alpha L_t^{-\alpha} \implies \psi = A_t(1-\alpha)(\mu_t K_t)^\alpha L_t^{-\alpha}$$

Since we are eliminating consumption, we are no longer in our straitjacket! Additionally, we can actually change μ_t at time t, and we are no longer bound by our predetermined capital.

Intuition. We no longer have a direct inverse relationship between C_t and L_t. We hold A_t fixed today, but imagine that at some point in the future it will increase. Suppose that A_{t+h} will increase, and further suppose that this is public information. We will have C_t increase because of consumption smoothing. Can we argue why L_t would also increase? Our investment adjustment costs do two things: they make I_t rise, so that the increase in investment for time h is smooth rather than a large discrete jump; and they make the value of existing capital K_t fall – since we will be raising investment in the future, people will be raising investment today, but not storing more capital. Instead, they will raise the 'burn rate' of capital μ_t; making it worth less today. Higher μ_t implies higher marginal product of labor, implies an increase in L_t.

Other ways productivity news shocks can matter. (in particular, drive a business cycle).

1. New Keynesian frictions can do this (talked about above)
2. Labor search model can also do this! Consider the model we've been thinking about this whole class:

$$\frac{\phi_n}{M_V(V_t)} = \text{MPL}_t + \beta\,\mathbb{E}\left[\left(\frac{C_{t+1}}{C_t}\right)^{-\sigma}\frac{\phi_n}{M_V(V_{t+1})}(1-\delta_n)\right]$$

This has some really nice features! Namely, there's no straitjacket here. We could do recursive substitution, and get that

$$\frac{\phi_n}{M_v(\cdot)} = \sum_{h=0}^{\infty}\beta^t\left(\frac{C_{t+h}}{C_t}\right)^{-\sigma}\text{MPL}_{t+h}(1-\delta_n)^h$$

Instead of an impossibility theorem, we have a race: the discount factor will decrease as the marginal product of labor (in the future) increases. Whether A_{t+h} drives N_t up depends on parameters. In principle, this can absolutely work! Unfortunately, in most calibrations, it's a very small effect.

3. *Decentralized* labor search model (with some source of sticky wage) can solve many of the above problems – it can attain some fairly large effects from an expected productivity shock. To do this, replace MPL_{t+h} with $\text{MPL}_{t+h} - w_{t+h}$ (*i.e.* internalize the externality). Ryan showed this recently in Chahrour, Chough, and Potter (2023).

2.12 Modern Topics: Climate Change

Remark. Macroeconomics is really good at thinking about general equilibrium effects, and aggregate effects. With regard to climate change, this should be extremely useful, but there hasn't been much research on climate change and macroeconomics.

First, just note that atmospheric carbon dioxide has increased rapidly recently (last $\approx 50 - 100$ years). This is very correlated with emissions, and is a signal of human effects on climate change. We will focus on carbon here.

A very relevant quote from Ryan: "When you have little data, you need more model." Broadly, there are three keys to evaluating the *macroeconomic* effects of climate change. We need (i) a model of how carbon dioxide accumulates in the world, (ii) a damage function, and (iii) discounting (specifically, precise measures of discounting). All of these things are controversial, especially (i) and (ii). Our goal is to provide some evidence on each element.

First, we think about the carbon cycle. Like our unemployment model, we will think about flows (and specifically sources and sinks). The path we will think about most is water – human activity sends water into the atmosphere, which is then absorbed by plants, soil, and specifically water. There is some more controversy here, people are not so convinced that water absorbs as much carbon as it once did – Ryan believes it may be essentially saturated. We should, for our lifetimes, think about carbon released into the atmosphere as being there (practically) forever.

The other key input into the classical model is a damage function, which maps carbon admissions into economic outcomes. We will answer the broad question of how do changes in atmospheric carbon affect global output (*i.e.* well-being). This admits two sub-questions: how does carbon affect temperature, and how does temperature affect economic activity? There is an historical relationship for the first sub-question:

$$T_t = \lambda \log_2(S_t/\bar{S})$$

where $\lambda \approx 3.0°C$. The implications are that carbon has an immediate effect on temperature, where doubling of atmospheric carbon will increase global temperature by $3.0°C$. There's lots of uncertainty about this relationship – Barnett, Brock, and Hansen (2021) talk about this uncertainty.

The second sub-question is very difficult. We're fairly sure it's convex, so the first degree increase is less harmful than the fifth, for example. We will look at a basically linear damage function (really, an exponential over a very small range, which is practically linear).

Finally, we will think about the discount rate. Today, we'll set it to 0, so today and tomorrow are worth the same. This is unusual, but thinking about the exact discount rate is kind of philosophical and not so well-posed.

Model. *Climate Change* (from Golosov et al. (2014)) We have some assumptions: The usual utility function: $\mathbb{E}_t \sum_{t=0}^{\infty} \beta^t U(C_t)$, intermediate goods E as well as one final good, and a standard resource constraint: $C_t + K_{t+1} - (1-\delta)K_t = Y_t$. We have sectors, with the final sector denoted by $i = 0$, so

$$Y_t = F_{0,t}(K_{0,t}, N_{0,t}, \mathbf{E}_{0,t}, S_t)$$

where $E_{0,t}$ is a vector of energy use, and S_t is the temperature. For all $i > 0$ intermediate (or energy) sectors, we split those sectors into groups $i = 1, \ldots, I_g - 1$ 'brown' energy sectors and $y = I_g, \ldots, I$ 'green' energy sectors. For all of the brown sectors, we will have a finite stock of that energy source, so $R_{i,t+1} = R_{i,t} - E_{i,t} \geq 0$. Classically, this is a cake eating problem. For both sectors, we will have a production function where

$$E_{i,t} = F_{i,t}(K_{i,t}, N_{i,t}, \underbrace{\mathbf{E_{i,t}}}_{\{E_{ij,t}\}}, R_{i,t})$$

In the green sectors, we have no dependence on $R_{i,t}$. The brown sectors use only $R_{i,t}$. Note that one baked-in assumption is that there is no technological progress to creating green energy. As we use brown energy, it will get *relatively* cheaper, but not *absolutely* cheaper. We have the following market clearing conditions:

$$\sum_{i=0}^{I} K_{i,t} = K_t$$
$$\sum_{i=0}^{I} N_{i,t} = N_t$$
$$\sum_{i=0}^{I} E_{ij,t} = E_{j,t} \text{ (an element of } \mathbf{E_{i,t}})$$

We have some special assumptions for the applied version of this model. First, we assume that

$$F_0(\cdot) = [1 - D_t(S_t)]\tilde{F}_0(K_{0,t}, N_{0,t}, \mathbf{E_{0,t}})$$

where $1 - D_t(S_t) = \exp(-\gamma(S_t - \bar{S}))$, where $\bar{S}$ is the carbon concentrations pre-industry – think of it as the carbon concentrations when we use no brown energy. We have a resource constraint for S_t:

$$S_t - \bar{S} = \underbrace{\sum_{i=0}^{t}(1 - d_t) \sum_{i=0}^{I_g - 1} E_{i,t}}_{\substack{\text{Discounted sum of total} \\ \text{brown energy usage} \\ \text{until time } t}}$$

Finally, we have that

$$\Lambda_t = Y_t \, \mathbb{E}\left[\sum_{j=0}^{\infty} \beta^j \gamma_{t+j} \underbrace{(1 - d_j)}_{\substack{\text{Decay rate of} \\ \text{emissions}}}\right]$$

where this equation captures the three key factors we talked about above. Λ_t is essentially the marginal cost of additional pollution (in the paper, they call it the *marginal externality*). This is relevant to the social planner, but not to the household. Imagine that we are not internalizing this externality today, but if the economy were to decide to emit one more unit of carbon dioxide, what would that cost the social planner in the future.

Conclusions of Golosov et al. The authors calibrated the model and identified / estimated parameters. They conclude that (i) the Laissez-faire economy uses *way* too much carbon, (ii) coal is a bigger problem than oil – there's so much of it in stock!, and (iii) carbon tax is one solution ($25 per ton for a "standard" discount, $220 for a "low" discount, and $2,000 for a "worst case" scenario). The second point is fascinating, and it might be overstated by their model, but it should qualitatively hold.

Note that with optimal taxes, the optimal taxed versus non-taxed is barely different. However, coal massively rises when not taxed, while when taxed it is essentially zero. The damage path looks very similar to the coal plot – if we take the optimal tax path it will cost $\approx 1.5\%$ of GDP versus $\approx 10\%$ of GDP in 100 years. This seems low, and it might not be capturing the most relevant channels, but it does provide a sort of lower bound.

There's a lot more to be studied here. A big challenge is that energy producers actually stand to gain from climate change – think Canada, Russia, northern European countries. Also, incorporating technological change may boost the benefits of a carbon tax, but it may also reduce the costs of climate change. Interacting those two seems interesting. Migration may also play a role. All in all, there's a lot that economists can do here.

2.13 Modern Topics: Machine Learning and Neural Networks

Definition. Ryan's definition of what we can describe as *artificial intelligence*:

1. A particular approach to functional approximation. This approach is described below, but specifically is extremely good at handling lots of states, and lots of parameters. This class of approximators is called *neural networks*.

2. These procedures use some sort of *stochastic optimization technique*, meaning that they handle lots of parameters well (and often are relatively fast). This is the training step.

Model. We have a true function $h(x)$, which we will attempt to approximate with $\hat{h}(x;\theta)$. We define $\{x_\ell\}$ as the set of points where we know $h(x_\ell)+\varepsilon_\ell$, where ε_ℓ represents observational noise. For now, assume that $\varepsilon_\ell = 0$. We look to solve

$$\min_\theta \underbrace{\sum_{\ell=1}^{L} \Big(h(x_\ell) - \hat{h}(x_\ell;\theta\Big)^2}_{\varepsilon(\theta)}$$

In the basis function case, we could solve this solution using ordinary least squares. In general, we won't have the nice convexity features we need for that to hold. Most of the techniques we'd use to solve this involve *gradient descent*[3], on $\nabla\varepsilon(\theta)$. Here, we will essentially take a subsample $\tilde{\ell}$ from the ℓ's, compute $\nabla\varepsilon_{\tilde{\ell}}(\theta)$ for the subsample, and update using that as an approximation for $\nabla\varepsilon(\theta)$. Note that we have (necessarily) randomization here.

Remark. Some advantages of this approach: It's less likely to get stuck on local optima, as the randomization will mean we try directions that do not 'look' optimal, but may be closer to global optima. When we have lots of data, this approach can be significantly faster than computing the gradient at each point. However, this approach can be 'wobbly' – hard to reproduce, sensitive to choice of $\tilde{\ell}$.

Algorithm. We have $h(X)$, where $X = [X_1, \ldots, X_n, \ldots, X_{N_0}]$. We have N_0 state variables. Recall that in the basis function approach, we had that

$$\hat{h}(x;a) = \sum_{i=1}^{I} a_i b_i(x)$$

A challenge of that approach is coming up with good basis functions in a principled way. The neural network approach is similar, but different – we will perform an affine transformation of X, plug it into an activation function, and then repeat the process some number of times. Mathematically, for a single layer, we will have

$$\hat{h}(X;\theta,\beta) = \theta_1^2 + \sum_{n_1=1}^{N_1} g(\theta_{n_1}^1 + X\beta_{n_1}^1)$$

[3]See Boyd and Vandenberghe for extensive treatment of these techniques.

where $N_1 2$ is the number of nodes in the (hidden) layer. We can think of this like follows:

$$\begin{aligned} \text{Node 1:} \quad & g(\theta_1^1 + X\beta_1^1) \\ \text{Node } n_1\text{:} \quad & g(\theta_{n_1}^1 + X\beta_{n_1}^1) \\ \text{Node } N_1\text{:} \quad & g(\theta_{N_1}^1 + X\beta_{N_1}^1) \end{aligned}$$

In OLS, for example, Node 1 would be the only one, and $g(\cdot)$ would be he sum of squares function. Here is an illustration of what a larger neural net would look like, in Figure 6

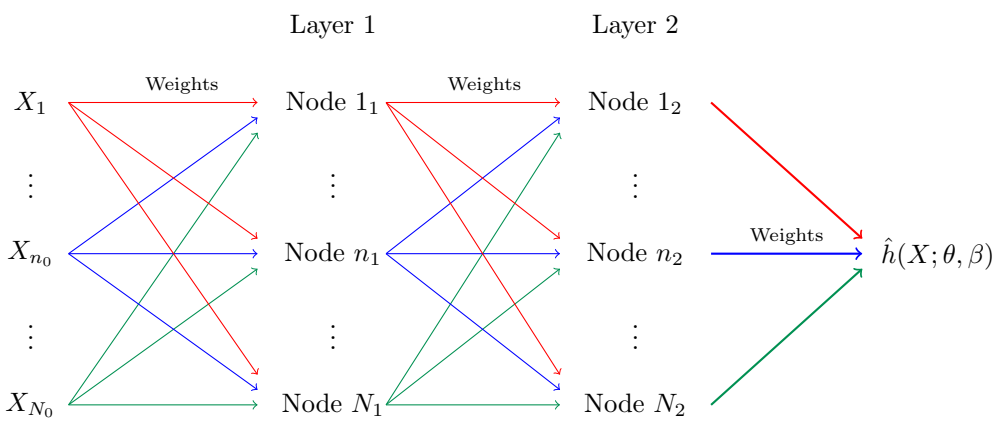


Figure 6: A Two-Layer Neural Network

Remark. Ryan has a really lovely example of this in the Real Business Cycle model, you can work through the code yourself.
Remark. Andrew Ng's CS 229 Lecture Notes have an excellent treatment of this in particular and machine learning in general. See Chapter 7.2.