1 Two elementary results on aggregation of technologies and preferences

In what follows we'll discuss "aggregation". What do we mean with this term? We say that an economy admits aggregation if the behavior of the aggregate equilibrium quantities (e.g., aggregate consumption, investment, wealth,...) and prices (e.g., wage, interest rate, ...) does not depend on the distribution of the individual quantities across agents. In other words, we can aggregate whenever we can define a fictitious "representative agent" that behaves, in equilibrium, as the sum of all individual consumers.

1.1 Aggregating firms with the same technology

Consider an economy with M firms, indexed by $i = 1, 2, ..., M$ which produce a homogeneous good with the same technology $zF(k^i, n^i)$ where z is aggregate productivity. Assume that F is strictly increasing, strictly concave, differentiable in both arguments and constant returns to scale. Can we aggregate these individual firms into a representative firm?

Suppose inputs markets are competitive. Then, each price-taking firm of type i solves

$$\max_{\{k^i,n^i\}} zF\left(k^i, n^i\right) - wn^i - (r+\delta)\, k^i,$$

with first-order conditions

$$\begin{aligned} zF_k\left(k^i, n^i\right) &= r+\delta, \quad (1)\\ zF_n\left(k^i, n^i\right) &= w, \end{aligned}$$

Recall that by CRS, F_k and F_n are homogenous of degree zero, hence:

$$\frac{F_k(k^i, n^i)}{F_n(k^i, n^i)} = \frac{f_k(k^i/n^i)}{f_n(k^i/n^i)}.$$

Dividing through the two first-order conditions, we obtain

$$\frac{f_k(k^i/n^i)}{f_n(k^i/n^i)} = \frac{r+\delta}{w},$$

and using the fact that the left-hand side is a strictly decreasing function of (k^i/n^i), we obtain

$$\frac{k_i}{n_i} = g\left(\frac{r+\delta}{w}\right) = \frac{K}{N}, \text{ for every } i = 1, 2, ..., M$$

where capital letters denote averages: every firm chooses the same capital-labor ratio.

Aggregate production across all firms:

$$
\begin{aligned}
z\sum_{i=1}^{M} F\left(k^i, n^i\right) &= z\sum_{i=1}^{M}\left[F_k\left(k^i, n^i\right)k_i + F_n\left(k^i, n^i\right)n_i\right] \\
&= z\sum_{i=1}^{M}\left[f_K\left(K/N\right)k_i + f_N\left(K/N\right)n_i\right] \\
&= zf_K\left(K/N\right)K + zf_N\left(K/N\right)N \\
&= zF\left(K, N\right)
\end{aligned}
$$

where the last line uses the CRS property of F. This derivation proves the existence of a "representative" firm with technology $zF\left(K, N\right)$. Note that z is the same across firms. With CRS, if one firm is more productive than all the others, it gets all the inputs.

1.2 Aggregating consumers with the same preferences

Consider a version of the neoclassical growth model with N types of consumers indexed by $i = 1, 2, ..., N$ with the same endowments of capital $k_0^i = \kappa$ for all $i's$ and same preferences

$$U\left(c_0^i, c_1^i, ...\right) = \sum_{t=0}^{\infty}\beta^t u\left(c_{it}\right).$$

Assume that markets are competitive, so that every consumer faces the same prices. Then, one would think that since all the N consumers make the same decisions, we can aggregate them into a representative agent, right? Not so quickly... Unless the utility function u is strictly concave, agents may not make the same optimal choices of consumption and leisure.

Let's combine these two results on firms and consumers:

Result 1.0 (trivial aggregation): Suppose that (i) every firm has the same productivity z and the same CRS production function F, where F is strictly increasing and strictly concave; (ii) consumers have the same initial endowments, and same preferences, and their utility function u is strictly increasing and strictly concave. Then, the neoclassical growth model admits a formulation with one representative firm and one representative household.

2 Gorman aggregation

We now study a more interesting case of "demand aggregattion", i.e., aggregation of individual demand curves. Consider a static economy populated by N agents indexed by $i = 1, ..., N$. The commodity space comprises M consumption goods $\left\{c^1, ..., c^M\right\}$ whose price vector is $p = \{p_m\}_{m=1}^M$. Each consumer is endowed with a_i units of wealth, and has utility u_i, strictly increasing and concave over each of the M goods. All goods markets are competitive.

Consider a particular good c^m. Aggregate demand of good c^m is given by the sum of all individual demands, or

$$C^m(p, \{a_i\}) = \sum_{i=1}^{N} c_i^m(p, a_i),$$

which makes it clear that, in general, you need to know the entire distribution of assets across agents to determine aggregate quantities.

When can we, instead, write aggregate demand as $C^m(p, A)$ where $A = \sum_{i=1}^N a_i$? In other words, when does aggregate demand only depend on the aggregate endowment, not on its distribution. In order for this representation to be true, it must be that if we reallocate one dollar of wealth from consumer i to j, the total demand of i and j does not change, or

$$\frac{\partial c_i^m(p, a_i)}{\partial a_i} = \frac{\partial c_j^m(p, a_j)}{\partial a_j}|_{da_j = -da_i} \text{ for all } (i, j) \text{ and for all } m.$$

This condition is true if all agents have the same marginal propensity to consume out of wealth, i.e., if the individual decision rule for consumption (recall, an outcome of optimization) can be written as

$$c_i^m(p, a_i) = \kappa_i^m(p) + \pi^m(p)\, a_i. \tag{2}$$

Condition (2) means that individual Engel curves (individual expenditures as a function of wealth) are linear. Then, aggregate consumption of good c^m is

$$C^m(p, A) = \bar{\kappa}^m(p) + \pi^m(p)\, A$$

where $\bar{\kappa}^m(p) = \sum_{i=1}^I \kappa_i^m(p)$.

This demand aggregation result is due to Gorman (1961), who stated it in terms of indirect utility as follows:

Result 1.0.1 (Gorman aggregation): If (and only if) agents' indirect utility functions can be represented as $v_i(p, a_i) = \alpha_i(p) + \beta(p) a_i$*, then aggregate consumption can be expressed as the choice of a representative agent with indirect utility* $V(p, A) = \bar{\alpha}(p) + \beta(p) A$ *where* $\bar{\alpha}(p) = \sum_{i=1}^{I} \alpha_i(p)$.

That is, for Gorman aggregation what we want is an indirect utility function that can be separated into a term that depends on prices and the consumer's identity but not on her wealth, and a term that depends on a function of prices that is common to all consumers that is multiplied by that consumer's wealth. This indirect utility is said to be of the Gorman form. You can easily prove both directions (iff) of this result using Roy's identity to obtain the demand function from the indirect utility function. Recall that Roy's identity establishes that

$$c^m(p, a_i) = -\frac{\frac{\partial v_i(p,a_i)}{\partial p^m}}{\frac{\partial v_i(p,a_i)}{\partial a_i}}. \tag{3}$$

Gorman aggregation (or demand aggregation) is a very powerful result for a number of reasons. First, beyond giving conditions for aggregations, it also explains how to construct the preferences of the representative agent. Moreover, it requires only (somewhat strong) assumptions on the demand side, such as restrictions on u, but no restriction on financial markets or technology. For example, we don't need complete financial markets. In addition, it is only based on consumer optimization and does not require any equilibrium restrictions.

Quasilinear utility: We now consider an example of Gorman aggregation: agent i with quasilinear utility over two goods (c_1, c_2) solves

$$\begin{aligned} &\max_{\{c_i^1, c_i^2\}} u_i\left(c_i^1\right) + \beta c_i^2 \\ &s.t. \\ p_1 c_i^1 + c_i^2 &= a_i \end{aligned}$$

where p_1 is the relative price –we chose p_2 as the numeraire. The FOCs of this problem are:

$$
\begin{aligned}
u_{i,c}\left(c_i^1\right) &= \lambda_i p_1, \\
\beta &= \lambda_i,
\end{aligned}
$$

where λ_i is the multiplier on the budget constraint. Note the key property of quasi-linear utility. The demand for good 1 is determined by the relative price but not by the endowment level. The demand for good 2 is then determined residually from the budget constraint.

The solution is therefore: $c_i^1 = u_{i,c}^{-1}(\beta p_1)$ and, from the budget constraint, $c_i^2 = a_i - p_1 u_{i,c}^{-1}(\beta p_1)$. The key observation here is that both individual demand functions show constant marginal propensities (respectively zero and one) out of wealth: Gorman aggregation holds. Note that the indirect utility function becomes

$$
\begin{aligned}
v_i(p_1; a_i) &= u_i\left(u_{i,c}^{-1}(\beta p_1)\right) + \beta\left[a_i - p_1 u_{i,c}^{-1}(\beta p_1)\right] \\
&= \alpha_i(p_1) + \beta a_i
\end{aligned}
$$

which is linear in individual wealth a_i and has common coefficient β, so we show again that we satisfy Gorman aggregation.

Another example for which Gorman aggregation holds is that of a homothetic utility function. A utility function $u(x_1, x_2)$ is homothetic if $u(\alpha x_1, \alpha x_2)$ has the same MRS as $u(x_1, x_2)$. In a model where competitive consumers optimize with *homothetic* utility functions subject to a budget constraint, the ratios of any two goods demanded only depends only relative prices, not on income or scale, a helpful proprty for aggregation. In the next section, we explore a model with homothetic preferences.

2.1 Neoclassical growth model with complete markets, heterogeneity in endowments, and homothetic preferences.

We now study a version of the neoclassical growth model with complete markets where consumers are only different in terms of their initial endowments of wealth. There is no individual or aggregate uncertainty. We show that, with homothetic preferences, we obtain demand aggregation. This derivation folows Chatterjee (1994).

2.1.1 The economy

Demographics– The economy is inhabited by N types of infinitely lived agents, indexed by $i = 1, 2, ..., N$. Denote by μ_i the number of agents i and normalize the total number of agents to one, $\sum_{i=1}^{N} \mu_i = 1$, so that averages and aggregates are the same.

Preferences– Preferences are time separable, defined over streams of consumption, and common across agents:

$$U = \sum_{t=0}^{\infty} \beta^t u(c_{it}),$$

where the period utility function u belongs to one of the following three classes: log, power, exponential, i.e.

$$u(c) = \begin{cases} \log(\bar{c} + c) & \text{with } \bar{c} + c > 0, \quad \bar{c} \le 0 \\ \frac{(\bar{c}+c)^{1-\sigma}}{1-\sigma} & \text{with } \bar{c} + c > 0, \quad \bar{c} \le 0 \\ -\bar{c}\exp(-\sigma c) & \text{with } \bar{c} > 0 \end{cases} \tag{4}$$

We impose $\bar{c} \le 0$ for log and power utility to allow for a subsistence level for consumption, and we impose $\bar{c} > 0$ for the exponential case. Note that utility is the same for each type.[1] When period utility belongs to the families in (4), then preferences share a common property. They are *quasi-homothetic*, i.e., they have affine Engel curves in wealth: the wealth-expansion path is linear.[2]

Markets and property rights– There are spot markets for the final good (which can be used for both consumption and investment) and complete financial markets, i.e. there are no constraints on transfers of income across periods. We assign the property rights on capital to the firm and the ownership of the firm to the household.[3] This is a different arrangement of property rights from the one you are used to seeing. Here households own shares of the firm.

We will let the initial level of wealth, at date t, differ across agents. Let a_{it} be the individual wealth of type i at time t. Then, $a_{it} = s_{it} A_t$, where s_{it} is the share of the firm-value owned by consumer i at time t. By summing both sides of this equation

[1] At the end of this derivation think about what happens if $\bar{c}$ is indexed by i.

[2] When $\bar{c} = 0$, preferences are homothetic because the constant in the consumption function becomes zero and Engel curves start at the origin, i.e. they are linear. However, linearity of the wealth-expansion path is not affected by the constant $\bar{c}$.

[3] If we had chosen to model the firm's problem as static (i.e. the firm rents capital services from households), every argument in this lecture would still hold. You should check this, as well as every other claim I make without proving it!

over i and exploiting the fact that $\sum_{i=1}^{N} \mu_i s_{it} = 1$ for every t, we obtain that aggregate household wealth equals the value of the firm (note also that s_{it} can be larger than 1). Besides intertemporal trading, there will be no other securities traded among agents in equilibrium, since there is no risk.

Technology– The aggregate production technology is $Y_t = f(K_t)$ with f strictly increasing, strictly concave and differentiable.

Household's problem– Given complete markets, the maximization problem of household i at time t can therefore be stated as:

$$\max_{\{c_{i\tau}\}} \sum_{\tau=t}^{\infty} \beta^{\tau-t} u(c_{i\tau})$$

$$s.t.$$

$$\sum_{\tau=t}^{\infty} p_\tau c_{i\tau} \leq p_t a_{it} \tag{5}$$

where a_{it} is the wealth of agent i in term of consumption units at time t. Let λ_{it} be the Lagrange multiplier on the individual i time t Arrow-Debreu budget constraint.

Solution– Consider the log-preferences case. From the FOC of the household problem at time t with respect to consumption at time τ, we have:

$$\beta^{\tau-t} u'(c_{i\tau}) = \lambda_{it} p_\tau \quad \Rightarrow \quad \beta^{\tau-t} \left(\frac{1}{\bar{c} + c_{i\tau}} \right) = \lambda_{it} p_\tau \quad \Rightarrow \quad c_{i\tau} = \frac{\beta^{\tau-t}}{\lambda_{it} p_\tau} - \bar{c}. \tag{6}$$

Substituting this FOC into the budget constraint of (5), we can derive an expression for the multiplier λ_{it}:

$$\begin{aligned}
\sum_{\tau=t}^{\infty} p_\tau \left(\frac{\beta^{\tau-t}}{\lambda_{it} p_\tau} - \bar{c} \right) &= p_t a_{it} \\
\frac{1}{\lambda_{it}(1-\beta)} - \bar{c} \sum_{\tau=t}^{\infty} p_\tau &= p_t a_{it} \\
\frac{1}{\lambda_{it}} &= (1-\beta) \left[p_t a_{it} + \bar{c} \sum_{\tau=t}^{\infty} p_\tau \right]
\end{aligned} \tag{7}$$

Let's now substitute the expression on the last line into equation (6) evaluated at $\tau = t$, i.e.

$$c_{it} = \frac{1}{\lambda_{it} p_t} - \bar{c},$$

in order to solve explicitly for c_{it}:

$$\begin{aligned} c_{it} &= \frac{1}{p_t}\left[(1-\beta)\,p_t a_{it} + (1-\beta)\,\bar{c}\sum_{\tau=t}^{\infty} p_\tau\right] - \bar{c} \\ &= \bar{c}\left[(1-\beta)\sum_{\tau=t}^{\infty}\left(\frac{p_\tau}{p_t}\right) - 1\right] + (1-\beta)\,a_{it} \qquad (8) \\ &= \Theta\left(p^t, \bar{c}\right) + (1-\beta)\,a_{it}, \end{aligned}$$

where

$$\Theta\left(p^t, \bar{c}\right) \equiv \bar{c}\left[(1-\beta)\sum_{\tau=t}^{\infty}\left(\frac{p_\tau}{p_t}\right) - 1\right] \qquad (9)$$

is a function of the subsistence level and of the whole price sequence $p^t = \{p_t, p_{t+1}, ...\}$. Thus, we have the optimal individual consumption rule

$$c_{it} = \Theta\left(p^t, \bar{c}\right) + (1-\beta)\,a_{it}, \qquad (10)$$

which is an *affine function* of asset holdings at time t for each type i. We know that (10) implies that we can Gorman-aggregate demand functions.

Even though we have only derived it for the log-case, it is easy to check that this representation of the consumption function holds also for the other two classes of preferences (power and exponential utility).

2.1.2 Equilibrium aggregate dynamics

Denote aggregate variables with capital letters. From (10), we derive easily that aggregate consumption only depends on aggregate variables (prices and aggregate wealth), but it is independent of the distribution of wealth. By summing over i on the LHS and RHS of (10) with weights μ_i we arrive at:

$$C_t = \Theta\left(p^t, \bar{c}\right) + (1-\beta)\,A_t. \qquad (11)$$

Since equilibrium aggregate consumption C_t has the same form as individual optimal consumption choice c_{it}, it is clear that (11) can be obtained as the solution to the following representative agent problem:

$$\begin{aligned} &\max_{\{C_\tau\}} \sum_{\tau=t}^{\infty} \beta^{\tau-t} u\left(C_\tau\right) \\ &s.t. \qquad\qquad (PP) \\ &\sum_{\tau=t}^{\infty} p_\tau C_\tau \leq p_t A_t \end{aligned}$$

which is exactly as in (5) but we have replaced small letters with capital letters. From the FOCs

$$\frac{u'(C_t)}{\beta u'(C_{t+1})} = \frac{p_t}{p_{t+1}}. \tag{12}$$

Let's make some further progress on the solution. To do so, we need to solve for the representative firm's problem.

Firm's problem– The representative firm owns physical capital and makes the investment decision by solving the problem

$$\begin{aligned} A_t &= \max_{\{I_\tau\}} \sum_{\tau=t}^{\infty} \left(\frac{p_\tau}{p_t}\right)[f(K_\tau) - I_\tau] \qquad (13) \\ & s.t. \\ K_{\tau+1} &= (1-\delta) K_\tau + I_\tau, \end{aligned}$$

where p_t is the time t price of the final good. Let's define real profits $\pi_t \equiv f(K_t) - I_t$. Then, it is easy to see that A_t is the value of the firm, i.e. the present value of future profits discounted at rate (p_τ/p_t), the relative price of consumption between time τ and time t. Recall that $p_t/p_{t+1} = (1 + r_{t+1})$ where r is the interest rate.

It is useful to compute the first-order condition (FOC) of the firm problem with respect to K_{t+1} by substituting the law of motion for capital into (13). The problem becomes:

$$\max_{K_{t+1}} \left\{ f(K_t) - K_{t+1} + (1-\delta) K_t + \frac{p_{t+1}}{p_t}[f(K_{t+1}) - K_{t+2} + (1-\delta) K_{t+1}] + \frac{p_{t+2}}{p_{t+1}} ... \right\}$$

with FOC:

$$1 = \frac{p_{t+1}}{p_t}[f'(K_{t+1}) + (1-\delta)] \tag{14}$$

which, incidentally, is exactly the same FOC of a "static" firm who rents capital from households at every date.

From (12) and the FOC for the firm's problem (14), we obtain the familiar Euler equation of the neoclassical growth model

$$u'(C_t) = \beta u'(C_{t+1})[f'(K_{t+1}) + (1-\delta)]. \tag{15}$$

We can state our first important result:

Result 1.1: If preferences are quasi-homothetic, and agents are heterogeneous in initial endowments, the neoclassical growth model with complete markets admit a single-agent *representation. The dynamics of aggregate quantities and prices do not depend on the distribution of endowments: they are the same as in the standard neoclassical growth model with representative agent.*

Two remarks are in order. First, equation (15) governs the dynamics of capital in the representative agent growth model where firms rent capital from households, instead of owning it. Therefore, we have discovered that in complete markets it is irrelevant whether we attribute property rights on capital to firms (and let households own shares of the firms) or to workers (and let firms rent capital from households). Second, equation (15) also governs the dynamics of capital in the social-planner problem. We are still in complete markets, and the Welfare Theorems hold.

Steady-state– The dynamics of the economy will converge to the steady-state values of capital stock satisfying the modified golden rule $f'(K^*) = 1/\beta - (1-\delta)$. Note now that in steady-state $p_{t+1}/p_t = \beta$ for all t, hence from the definition of $\Theta(p^t, \bar{c})$ in (9) we conclude that $\Theta(p^t, \bar{c}) = 0$ and $c_i = (1-\beta) a_i$. In other words, in steady-state, the average propensity to save is β, independently of wealth, for every type of household.

To conclude, in the neoclassical growth model with complete markets and where agents have heterogeneous wealth endowments, the dynamics of the aggregate variables do not depend on the evolution of the wealth distribution. But is the inverse statement true? Does the evolution of the wealth distribution across households (i.e., wealth inequality) depend on the dynamics of aggregate variables (prices and quantities)? We show below that the answer is: yes, it does.

2.1.3 Equilibrium dynamics of the wealth distribution

From the lifetime budget constraint of agent i at time t

$$
\begin{aligned}
p_t c_{it} + \sum_{\tau=t+1}^{\infty} p_\tau c_{i\tau} &= p_t a_{it} && (16)\\
c_{it} + \left(\frac{p_{t+1}}{p_t}\right) a_{i,t+1} &= a_{it} && (17)\\
\frac{a_{i,t+1}}{a_{it}} &= \left(\frac{p_t}{p_{t+1}}\right)\left(1 - \frac{c_{it}}{a_{it}}\right), && (18)
\end{aligned}
$$

which expresses the growth rate of wealth for type i as a function of her consumption-wealth ratio.

By aggregating over types in equation (16), we can obtain an equivalent equation at the aggregate level:

$$\begin{aligned} \sum_i \mu^i c_{it} + \left(\frac{p_{t+1}}{p_t}\right) \sum_i \mu^i a_{i,t+1} &= \sum_i \mu^i a_{it} \\ C_t + \left(\frac{p_{t+1}}{p_t}\right) A_{t+1} &= A_t \\ \frac{A_{t+1}}{A_t} &= \left(\frac{p_t}{p_{t+1}}\right)\left(1 - \frac{C_t}{A_t}\right) \end{aligned}$$

We want to establish conditions under which an individual's share of total wealth will grow over time, i.e. $s_{i,t+1} > s_{it}$. First of all, note that:

$$\frac{s_{i,t+1}}{s_{i,t}} > 1 \quad \Leftrightarrow \quad \frac{a_{i,t+1}}{a_{it}} > \frac{A_{t+1}}{A_t} \quad \Leftrightarrow \frac{c_{it}}{a_{it}} < \frac{C_t}{A_t} \tag{19}$$

Moreover, from equations (10) and (11),

$$\frac{c_{it}}{a_{it}} = \frac{\Theta\left(p^t, \bar{c}\right)}{a_{it}} + (1-\beta), \quad \text{and} \quad \frac{C_t^i}{A_t} = \frac{\Theta\left(p^t, \bar{c}\right)}{A_t} + (1-\beta)$$

and therefore

$$\frac{c_{it}}{a_{it}} < \frac{C_t}{A_t} \quad \Leftrightarrow \quad \frac{\Theta\left(p^t, \bar{c}\right)}{a_{it}} < \frac{\Theta\left(p^t, \bar{c}\right)}{A_t} \quad \Leftrightarrow \quad \Theta\left(p^t, \bar{c}\right)\left(a_{it} - A_t\right) > 0$$

and, thus, summarizing we have the following equivalence (i.e., "if and only if") condition:

$$\frac{s_{i,t+1}}{s_{it}} > 1 \quad \Leftrightarrow \quad \Theta\left(p^t, \bar{c}\right)\left(a_{it} - A_t\right) > 0,$$

which means that whether consumer's i wealth share is increasing or decreasing over time depends on 1) the sign of the constant Θ (equal for everyone) and 2) on her relative position in the distribution. For example, if $\Theta > 0$ then for a consumer whose initial wealth is above average, her share will grow, whereas for a consumer whose initial wealth is below average, her share will fall. And hence the distribution will become more unequal over time. Note that the dynamics of the wealth distribution depend on the entire sequence of prices, hence on the dynamics of aggregate variables in equilibrium.

Two results are immediate. First, in steady-state, $\Theta\left(p^t, \bar{c}\right) = 0$ and $s_{i,t+1} = s_{it}$ since every agent has the same average propensity to save β. Similarly, even out of steay-state, in absence of subsistence level, $\bar{c} = 0$ we still have $\Theta = 0$, and the same prediction is true. Thus, the neoclassical growth model with heterogeneous endowments and homothetic preferences has a sharp prediction for the evolution of inequality.

Result 1.2: In the neoclassical growth model with complete markets, homothetic preferences, heterogeneous endowments, but without subsistence level ($\bar{c} = 0$), the wealth distribution remains unchanged along the transition path, i.e., initial conditions in endowments (and inequality) persist forever.

The intuition is that if $\bar{c} = 0$ then the average propensity to consume, and save, is the same for every agent. Every agent accumulates wealth at the same rate.

In presence of a subsistence level, the dynamics are more interesting. We now determine the sign of Θ, through:

Lemma 1.1 (Chatterjee, 1994): *$\Theta(p^t, \bar{c})$ is greater (less) than zero if and only if the economy is converging from below (above) to the steady-state, i.e. if $K_\tau < (>) K^*$.*

Proof: *Suppose the economy grows towards the steady-state, i.e. $K_\tau < K^*$. Then the sequence $\{f'(K_\tau)\}$ is decreasing and, from equation* (15)*, the sequence $\{p_{\tau+1}/p_\tau\}$ is increasing towards β, i.e., $p_{\tau+1}/p_\tau \leq \beta$ for all $\tau \geq t$ where the strict inequality holds at least for some τ. It follows that*

$$p_\tau/p_t = (p_\tau/p_{\tau-1})(p_{\tau-1}/p_{\tau-2}) \dots (p_{t+2}/p_{t+1})(p_{t+1}/p_t) < \beta^{\tau-t}.$$

From the definition of $\Theta(p^t, \bar{c})$ in (8)*, use the above equation to obtain*

$$\Theta\left(p^t, \bar{c}\right) = \bar{c}\left[(1-\beta)\sum_{\tau=t}^{\infty}\left(\frac{p_\tau}{p_t}\right) - 1\right] > \bar{c}\left[(1-\beta)\sum_{\tau=t}^{\infty}\beta^{\tau-t} - 1\right] = 0$$

where the inequality follows from $\bar{c} < 0$. **QED**

The implications for the evolution of the wealth distribution in an economy growing towards the steady-state (the empirically interesting case) are easy to determine, at this point. In the presence of a subsistence level ($\bar{c} < 0$), $\Theta > 0$. From equation (19) this implies that the average propensity to consume (save) declines (increases) with wealth: poor agents must consume proportionately more out of their wealth to satisfy the subsistence level, so wealth inequality increases along the transition. In other words:

Result 1.3: In the neoclassical growth model with complete markets, homothetic preferences, heterogeneous endowments and subsistence level $\bar{c} < 0$, as the economy grows towards the steady-state: (i) the wealth distribution becomes more unequal, as rich agents accumulate more than poor agents along the transition path, and (ii) there is no change *in the ranking of households in the wealth distribution, i.e., initial conditions in wealth (and consumption) ranking persist forever.*

A bit more intuition about this result. First, inequality grows along the transition when the economy is growing. If K_t grows, then r_t falls along the transition. This means that, as the economy converges, households have a temporarily high capital income, but consumption smoothing suggest this temporarily high income should be saved. Who is going to save the most? Those with high initial wealth, and hence high capital income. This is the key force that makes the wealrhy save even more. The opposite logic holds when the economy converges from above.

Second, to see clearly why the initial ranking of shouseholds does not change, recall that from (18) growth in individual wealth is:

$$\frac{a_{i,t+1}}{a_{it}} = \left(\frac{p_t}{p_{t+1}}\right)\left(1 - \frac{c_{it}}{a_{it}}\right) = \left(\frac{p_t}{p_{t+1}}\right)\left[\beta - \frac{\Theta\left(p^t, \bar{c}\right)}{a_{it}}\right]$$

therefore the growth rate of wealth is ordered by a_{it}.

The main conclusion is that in this model there is no economic or social mobility. This is not a good model to understand why some individual are born poor and make it in life, while other are born rich and end up poor as rats. This is just a model of castes. Finally note that absence of economic mobility is sort of implicit in the fact that the CE allocations can be obtained as the result of a planner's problem with fixed individual weights (see below).

Robustness– We now discuss how robust this result is to three of the key assumptions made so far in the analysis: 1) all agents have same discount factor β, 2) all agents are equally productive, 3) markets are complete.

1. When agents have different discount factors, then none of the results hold any longer. Suppose that $\bar{c} = 0$ to simplify the analysis. Then, from (10)

$$c_{it} = \left(1 - \beta_i\right) a_{it},$$

therefore the average propensity to save out of wealth is higher the more patient is the individual and from (19), wealth grows faster for the more patient individuals. In the limit, in steady-state, the most patient type holds all the wealth, and the distribution becomes degenerate.

2. Caselli and Ventura (2000) extend the Gorman aggregation result to a version of the neoclassical growth model where agents also differ in their endowments of efficiency units of labor.

3. In absence of markets and trade (autarky), every consumer has access to her own technology. Each agent i will solve his own maximization problem in isolation

$$\begin{array}{l} \max_{\{ci_\tau\}} \sum_{\tau=t}^{\infty} \beta^{\tau-t} u\left(c_{i\tau}\right) \\ s.t. \\ k_{i,\tau+1} = (1-\delta)\, k_{i\tau} + f\left(k_{i\tau}\right) - c_{i\tau} \\ k_{it} \text{ given} \end{array}$$

with different initial conditions k_{it}. It is easy to see that, independently of initial conditions, each agent will converge to the same capital stock k^*, hence in the long-run the distribution of wealth is perfectly equal. Interestingly, we conclude that less developed financial markets induce less wealth inequality, in the long-run.

2.1.4 Indeterminacy of the wealth distribution in steady-state

One very important implication of the aggregation Result 1.1 is that in steady-state the wealth distribution is indeterminate. From (13), (15) and (10), the set of equations characterizing the steady-state is:

$$\begin{aligned} c_i &= (1-\beta)\, a_i,\ i = 1, 2, ..., N \\ f'\left(K^*\right) &= 1/\beta - (1-\delta)\,, \\ A^* &= \frac{1}{1-\beta}\left[f\left(K^*\right) - \delta K^*\right] \\ \sum_{i=1}^{N} \mu_i a_i &= A^*, \end{aligned}$$

We therefore have $(N+3)$ equations and $(2N+2)$ unknowns $\left(\{c_i, a_i\}_{i=1}^{N}, K^*, A^*\right)$. In other words, the multiplicity of the steady-state wealth distributions is of order $N-1$.[4]

However, suppose we start from a given wealth distribution at date $t = 0$ when the economy has not yet reached its steady-state, then the dynamics of the model are uniquely determined by Results 1.2 and 1.3 and the final steady-state distribution is determined as well. So, let's restate our finding in:

[4] This means that, if $N = 1$ (representative agent), the steady-state is unique. If $N = 2$, there is a continuum of steady-states of dimension 1, and so on.

Result 1.4: In the steady-state of the neoclassical growth model with N agents, heterogeneous initial endowments and homothetic preferences, there is a continuum with dimension $(N-1)$ of steady-state wealth distributions. However, given an initial wealth distribution $\{a_{i0}\}_{i=1}^{N}$ at $t=0$, the equilibrium wealth distribution $\{a_{it}\}_{i=1}^{N}$ in every period t is uniquely determined, and so is the final steady-state distribution.

Example– Consider an economy where $N=2$, where the production technology is $zf(K)$. Then, the set of steady-state equations is

$$\begin{aligned}
c_i &= (1-\beta)\, a_i,\ i=1,2, \\
zf'(K^*) &= 1/\beta-(1-\delta), \\
A^* &= \frac{1}{1-\beta}\left[zf(K^*)-\delta K^*\right], \\
\mu_1 a_1+\mu_2 a_2 &= A^*
\end{aligned}$$

So we have 5 equations, but 6 unknowns $\{a_1, a_2, K^*, A^*, c_1, c_2\}$. We can represent graphically all the possible equilibrium paths between two steady states that differ for their level of technological progress z , say (z_L, z_H). The figure shows that the model has a continuum of steady-state distributions of wealth of dimension one, all consistent with the uniquely determined aggregate capital stock K^*. If we pick an initial distribution in the initial steady-state with productivity z_L, the equilibrium path to the final steady-state with productivity level z_H is uniquely determined.

Finally, in terms of language, this whole section shows that it is important to distinguish "steady-state" from "equilibrium path". In this economy, the equilibrium path is always unique (given initial conditions), but the steady-state is not.

3 The Negishi Approach

Negishi (1960) suggested a method to calculate the competitive equilibrium (CE) prices and allocations of complete markets economies (in particular, economies for which the first welfare theorem holds) with heterogeneous households. This method is particularly useful for those economies where Gorman aggregation does not go through and, hence we do not know how to write the preferences of the representative agent.[5]

[5] In his original paper, Negishi (1960) used this equivalence result to propose a simple way to show existence of competitive equilibria.

From the first welfare theorem, we know that any CE is a Pareto optimum (PO), hence it can be found as the solution to a social planner problem with "some" Pareto weights given to each agent. Suppose we want to compute a particular CE of an economy where agents are initially endowed with heterogeneous shares $\{s_{i0}\}_{i=1}^{N}$ of the aggregate wealth. Can we use the planner problem for this purpose? Negishi showed that the key is to search for the "right" weights given to each type of agent in the social welfare function of the planner. Each set of weights corresponds to a Pareto efficient allocation, the key is to find the set of weights which correspond to our desired CE in the original economy.[6]

3.1 An Example

Consider our neoclassical growth model of section (2.1) with two types of consumers ($N = 2$). The agent's i problem in the decentralized Arrow-Debreu equilibrium can be written as

$$\max_{\{c_{it}\}} \sum_{t=0}^{\infty} \beta^t u\left(c_{it}\right)$$

$$s.t.$$

$$\sum_{t=0}^{\infty} p_t c_{it} \leq p_0 a_{i0}$$

where $a_{i0} = s_{i0} A_0$ is the initial wealth endowment, given at $t = 0$. Let's assign the property rights on capital to the firm, so the firm's problem is exactly the one of the previous section.

From the FOC of the agent of type i, we obtain

$$FOC\left(c_{it}\right) \longrightarrow \beta^t u'\left(c_{it}\right) = \lambda_i p_t,$$

where λ_i is the multiplier on the time zero budget constraint. Thus, putting together the FOC's for the two types:

$$\frac{u'\left(c_{1t}\right)}{u'\left(c_{2t}\right)} = \frac{\lambda_1}{\lambda_2}. \tag{20}$$

[6] Incidentally, the notion of social welfare was introduced by Samuelson (1956).

Now, write down the following Negishi planner problem (NP) for our economy

$$\max_{\{c_{1t}, c_{2t}, K_{t+1}\}} \sum_{t=0}^{\infty} \beta^t \left[\alpha_1 u\left(c_{1t}\right) + \alpha_2 u\left(c_{2t}\right)\right] \tag{NP}$$

$$s.t.$$

$$c_{1t} + c_{2t} + K_{t+1} \leq f\left(K_t\right) + (1-\delta) K_t$$

$$K_0 \text{ given}$$

where (α_1, α_2) are the planner's weights for each type of household in the social welfare function.

The FOC's for this problem are

$$FOC\left(c_{it}\right) \longrightarrow \alpha_i \beta^t u'\left(c_{it}\right) = \theta_t, \tag{21}$$

$$FOC\left(K_{t+1}\right) \longrightarrow u'\left(c_{it}\right) = \beta u'\left(c_{i,t+1}\right)\left[f'\left(K_{t+1}\right) + (1-\delta)\right] \tag{22}$$

where θ_t is the Lagrange multiplier on the planner's resource constraint at time t. Note that putting together the first-order conditions for consumption for the two agents we arrive at

$$\frac{u'\left(c_{1t}\right)}{u'\left(c_{2t}\right)} = \frac{\alpha_2}{\alpha_1}, \tag{23}$$

which tells us that the planner allocates consumption proportionately to the weight it gives to each consumer (with strictly concave utility).[7]

If we want the NP to deliver the same solution as the CE, we need the PO allocations and the CE allocations to be the same. Given strict concavity of preferences, this implies that, putting together (23) and (20):

$$\frac{\alpha_2}{\alpha_1} = \frac{\lambda_1}{\lambda_2}$$

Hence, the relative weights of the planner must correspond to the inverse of the ratio of the Lagrange multipliers on the time-zero Arrow-Debreu budget constraint for the two agents in the CE.

In particular, for log preferences $u\left(c_{it}\right) = \log c_{it}$, we derived in equation (7) that

$$\left(\frac{1}{\lambda^i}\right) = (1-\beta) p_0 a_{i0} \Rightarrow \lambda_i = \left[\frac{1}{(1-\beta) p_0 A_0}\right] \frac{1}{s_{i0}},$$

[7] Note also that the ratio of marginal utility across agents is kept constant in every period (a key features of complete markets allocations, also called *full insurance*).

therefore we obtain that

$$\frac{\alpha_2}{\alpha_1} = \frac{\lambda_1}{\lambda_2} = \frac{s_{20}}{s_{10}}.$$

Imposing the natural and innocuous normalization $\alpha_1 + \alpha_2 = 1$, we can solve explicitly for the two weights: $\alpha_1 = s_{10}$ and $\alpha_2 = s_{20}$, i.e., the weights are exactly equal to the initial wealth shares ("exactly" is because of log utility, in general weights will be proportional to the initial shares). The higher is the initial wealth share s_{i0}, the lower is the multiplier λ_i and the larger is the Pareto weight α_i on the Negishi problem: the planner must deliver more to consumption to the agent who has a large initial share of wealth in the decentralized equilibrium.

Result 1.5: Consider an economy with agents heterogeneous in endowments where the First Welfare Theorem holds. Then, the competitive equilibrium allocations can be computed through an appropriate planner's problem where the relative weights on each agent in the social welfare function are proportional to the relative individual endowments: those agents who initially have more wealth will get a higher weight in the planner's problem.

Now, note that using equation (21) we obtain that

$$\frac{\alpha_i \beta^t u'(c_{it})}{\alpha_i \beta^{t+1} u'(c_{i,t+1})} = \frac{\theta_t}{\theta_{t+1}} \Longrightarrow \frac{u'(c_{it})}{\beta u'(c_{i,t+1})} = \frac{\theta_t}{\theta_{t+1}}.$$

Substituting this last expression into (22), we arrive at a relationship between the sequence of capital stocks and the sequence of Lagrange multipliers on the planner's resource constraint

$$\frac{\theta_t}{\theta_{t+1}} = f'(K_{t+1}) + (1-\delta).$$

Recall that equation (15) dictating the optimal choice of capital for the firm in the CE stated that

$$\frac{p_t}{p_{t+1}} = f'(K_{t+1}) + (1-\delta).$$

Hence, we have

$$\frac{\theta_t}{\theta_{t+1}} = \frac{p_t}{p_{t+1}}, \tag{24}$$

in other words, the Arrow-Debreu equilibrium prices can be uncovered as the sequence of Lagrange multipliers in the Pareto problem: intuitively, the multipliers gives us the shadow value of an extra unit of consumption and, in the CE, prices signal exactly this type of scarcity.[8] Note that equation (24) is true for any well behaved u.

In conclusion, we have uncovered a tight relation between weights of the NP problem and initial endowments in the CE and an equivalence between Lagrange multiplier on the resource constraint of the NP problem and prices in the CE. This strict relationship, that we have uncovered for the log utility case, is true more in general.

3.2 General application of the Negishi method

In general, without specific restrictions on preferences (e.g., CRRA utility), one may not have closed form solutions for the λ's in the CE, so the algorithm is a little more involved. The objective is to compute the CE allocations for an economy with N types of agents and endowment distribution $\{ai_0\}_{i=1}^N$. We can describe the algorithm in four steps:

1. In the Negishi social planner problem (NP), guess a vector of weights $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, ..., \alpha_N\}$. The normalization $\sum_{i=1}^N \alpha_i = 1$ means that $\boldsymbol{\alpha}$ belongs to the N-dimensional simplex

$$\Delta_N = \{\boldsymbol{\alpha} \in \mathbb{R}_+^N : \sum_{i=1}^N \alpha_i = 1\}$$

and the simplex traces out the entire set of Pareto-optimal allocations.

2. From the NP problem compute the sequence of allocations $\left\{\{c_{it}\}_{i=1}^N, K_t\right\}_{t=0}^\infty$ and the implied sequence of multipliers $\{\theta_t\}_{t=0}^\infty$ on the resource constraint in each period t. In practice, at every t, one needs to solve the $N+2$ equations

$$\begin{aligned} \alpha^i \beta^t u'(c_{it}) &= \theta_t,\ i = 1,..,N \\ \frac{\theta_t}{\theta_{t+1}} &= f'(K_{t+1}) + (1-\delta) \\ \sum_{i=1}^N \mu_i c_{it} + K_{t+1} &= f(K_t) + (1-\delta) K_t \end{aligned}$$

in $N+2$ unknowns $\left(\{c_{it}\}_{i=1}^N, K_{t+1}, \theta_{t+1}\right)$. At every t, (K_t, θ_t) are given, therefore the Negishi method simplifies enormously the computation of the equilibrium: the

[8] Using p_0 as the numeraire and imposing the normalizations $p_0 = \theta_0 = 1$, the above relationship implies that $p_t = \theta_t$ so equilibrium prices are exactly equal to the shadow prices of consumption in the planner's problem.

Negishi solution requires solving, for every time t, a small simultaneous system of equations. Recall that, instead, to solve for the CE allocations, at every time t one must set the excess demand function to zero and the excess demand function depends on the entire price sequence—an infinitely dimensional object. To understand, take another look at the consumption allocation (10) where $\Theta(\cdot)$ depends on the entire price sequence from t onward.

Instead of guessing (and iterating over) an infinite sequence of prices, one guesses and iterates over a finite set of weights. With a caveat: even though K_0 is given, θ_0 is not. One has also to guess a value for θ_0. The reason is that, unless you have the right value for θ_0, the system will not be on the saddle-path and capital will diverge. In other words, there is another condition that we need to satisfy in the growth model, the transversality condition:

$$\lim_{t\to\infty}\beta^t u'(c_{it})K_t = 0 \Longrightarrow \lim_{t\to\infty}\frac{1}{\alpha^i}\theta_t\cdot K_t = 0$$

which states that, in the limit, the marginal value of a unit of capital is zero.

3. Exploit the equivalence between prices p_t and multipliers θ_t to verify whether the time-zero Arrow-Debreu budget constraint of each agent holds exactly at the guessed vector of weight $\boldsymbol{\alpha}$. Specifically, for each agent, compute the implicit transfer function $\tau_i(\boldsymbol{\alpha})$ associated with the assumed vector of weights

$$\tau_i(\boldsymbol{\alpha}) = \sum_{t=0}^{\infty}\theta_t c_{it}(\alpha_i) - \theta_0 a_{i0},\ \text{ for every } i = 1, 2, ..., N \tag{25}$$

and if (25) holds for agent i with a "greater (smaller) than" sign, it means that the planner is giving too much (little) weight to agent i. So, in the next iteration reduce (increase) the weight α_i given to agent i. Note one useful property of the transfer functions:

$$\sum_{i=1}^{N}\tau_i(\boldsymbol{\alpha}) = \sum_{i=1}^{N}\left[\sum_{t=0}^{\infty}\theta_t c_{it}(\alpha_i) - \theta_0 a_{i0}\right] = \sum_{t=0}^{\infty}p_t C_t - p_0 A_0 = 0$$

since the discounted present value of resources of the economy cannot be greater than its current wealth. Put differently, recall that from the representative firm problem

$$A_0 = \sum_{t=0}^{\infty}\left(\frac{p_t}{p_0}\right)[f(K_t) - I_t] = \sum_{t=0}^{\infty}\left(\frac{p_t}{p_0}\right)C_t$$

which establishes the same result. To conclude, the individual transfer functions $\tau_i(\boldsymbol{\alpha})$ must sum to zero.

4. Iterate over $\boldsymbol{\alpha}$ until you find the vector of weights $\boldsymbol{\alpha}^*$ that sets *every* individual transfer function $\tau_i(\boldsymbol{\alpha}^*)$ to zero. This vector corresponds to the PO allocations that are affordable by each agent in the CE, given their initial endowment, without the need for any transfer across-agents. Thus, we are computing exactly the CE associated to initial conditions $\{a_{i0}\}_{i=1}^N$.

See also Ljungqvist-Sargent, section 8.5.3, for a discussion of the Negishi algorithm.

4 Aggregation with complete markets

We just learned that the equilibrium allocations of a complete market economy with heterogeneity can be obtained as the solution of a Negishi planner problem. The Negishi approach can be used to prove a more general aggregation result for complete markets economies due to Constantinides (1982) and then refined by Ogaki (2003). Consider the economy described above where we know that the Welfare Theorems hold, and therefore we can solve for the equilibrium allocations through the following Negishi planner problem

$$\begin{aligned} & \max_{\{c_{it},K_{t+1}\}} \sum_{t=0}^{\infty} \beta^t \sum_{i=1}^{N} \alpha_i u(c_{it}) \\ & s.t. \\ K_{t+1} + C_t &= f(K_t) + (1-\delta) K_t \\ C_t &= \sum_{i=1}^{N} \mu_i c_{it} \end{aligned}$$

Now, split the problem in two stages. First, given a sequence of aggregate consumption $\{C_t\}_{t=0}^{\infty}$, consider the static problem of how to allocate C_t across agents at every t

$$\begin{aligned} U(C_t) &= \max_{\{c_{it}\}} \sum_{i=1}^{N} \alpha_i u(c_{it}) \\ & s.t. \\ C_t &= \sum_{i=1}^{N} \mu_i c_{it} \text{ for all } t \end{aligned}$$

where $U(C_t)$ is the indirect utility function of the planner at date t. Second, consider the consumption/investment problem

$$
\begin{aligned}
& \max_{\{C_t, K_{t+1}\}} \sum_{t=0}^{\infty} \beta^t U(C_t) \\
& s.t. \\
K_{t+1} + C_t &= f(K_t) + (1-\delta) K_t \\
& K_0 \text{ given}
\end{aligned}
$$

This second-stage problem shows that the aggregate dynamics of the model can be described by the solution to the problem of a representative agent (RA), but the RA's preferences are different from preferences of the individual consumer, i.e., $U \neq u$. Indeed, one can even allow for different utility functions u^i across agents and show that this result still holds.

The Constantinides aggregation theorem is a very general existence (i.e., existence of a representative agent) result: note that the only restriction on preferences is strict concavity of u, but homotheticity or quasilinearity are not required. However, it requires complete markets, while Gorman aggregation does not. It also suggests an algorithm for constructing the preferences of the RA, but in most cases an analytical solution for U is not attainable, even though we know U exists. In what follows, we show an example with closed-form solution.

4.1 An Example

This example is taken from Maliar and Maliar (2001, 2003). In their model, agents have non-homothetic preferences in consumption and leisure, and are subject to idiosyncratic, but insurable, shocks to labor endowment (i.e., markets are still complete). They show how to recover preferences of a fictitious representative agent whose optimization problem yields, as a solution, the equilibrium aggregate dynamics of the original economy. This representative agent ends up having a *different* (but not so different) utility function from that of individuals populating the economy.

Demographics– The economy is inhabited by a continuum of infinitely lived agents, indexed by $i \in I \equiv [0, 1]$. Denote by μ^i the measure of agents i in the set I and normalize the total number of agents to one, $\int_I d\mu^i = 1$, so that averages and aggregates are the same. Initial heterogeneity is in the dimension of initial wealth endowments.

Uncertainty– Agents are subject to idiosyncratic productivity shocks to skills. Let ε_t^i be the shock of agent i, and suppose shocks are iid, with mean 1, and defined over the set E. This is not necessary, but it simplifies the notation.

Preferences– Preferences are time separable, defined over streams of consumption, given by

$$U = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u\left(c_{it}, 1 - h_t^i\right).$$

where period utility is given by

$$u\left(c_{it}, 1 - h_t\right) = \frac{c_t^{1-\gamma} - 1}{1-\gamma} + \phi \frac{\left(1 - h_t\right)^{1-\sigma} - 1}{1-\sigma} \tag{26}$$

and note that preferences are not quasi-homothetic, unless $\sigma = \gamma$.

Markets and property rights– There are spot markets for the final good (which can be used for both consumption and investment) whose price is normalized to one, and complete financial markets, i.e. agents can trade a full set of state-contingent claims. The agent's portfolio is composed by Arrow securities of the type $a_{t+1}^i\left(\varepsilon\right)$ which pay one unit of consumption at time $t+1$ if the individual's shock is ε and zero otherwise. Let $p_t\left(\varepsilon\right)$ the price of this security and $\int_E p_t\left(\varepsilon\right) a_{t+1}^i\left(\varepsilon\right) d\varepsilon$ the value of such portfolio for agent i.

Technology and firm's problem– The aggregate production technology is $Y_t = Z_t f\left(K_t, N_t\right)$ with f strictly increasing and strictly concave in both arguments and differentiable. The representative firm rents capital from households. N_t is aggregate labor input in efficiency units, i.e. $N_t = \int_I \varepsilon_t^i h_t^i d\mu^i$.

Household problem– For agent i:

$$\begin{aligned}
& \max_{\left\{c_{it}, k_{t+1}^i, a_{t+1}(\varepsilon)\right\}} E_0 \sum_{t=0}^{\infty} \beta^t u\left(c_{it}, 1 - h_t^i\right) && (27) \\
& s.t. \\
c_{it} + k_{t+1}^i + \int_E p_t\left(\varepsilon\right) a_{t+1}^i\left(\varepsilon\right) d\varepsilon = {} & \left(1-\delta\right) k_t^i + w_t \varepsilon_t^i h_t^i + a_t^i\left(\varepsilon_t^i\right) \\
& \left\{k_0^i, a_0^i\right\} \text{ given}
\end{aligned}$$

Equilibrium– This is a complete markets economy. The First Welfare Theorem tells us that the equilibrium is Pareto optimal, so we can use a social planner problem to characterize the equilibrium by applying the Negishi method. The key, as usual, is to find the right weights that guarantee that allocations are affordable for each agent, given their initial endowments.

Aggregation?– Given that preferences are not homothetic, we know that Gorman's strong aggregation concept will not hold. But can we, nevertheless, obtain a RA whose choices describe the evolution of the aggregate economy? And how the preferences of the RA look like?

Letting θ_t be the multiplier on the aggregate feasibility constraint, from the FOC with respect to individual i in the Negishi planner problem:

$$\begin{aligned} \alpha^i \left(c_{it}\right)^{-\gamma} &= \theta_t \\ \alpha^i \phi \left(1 - n_t^i\right)^{-\sigma} &= \theta_t w_t \varepsilon_t^i \end{aligned} \tag{28}$$

Rearranging gives

$$\begin{aligned} c_{it} &= \left(\frac{\alpha^i}{\theta_t^i}\right)^{\frac{1}{\gamma}} \\ \left(1 - h_t^i\right) \varepsilon_t^i &= \left(\frac{\phi \alpha^i}{\theta_t w_t}\right)^{\frac{1}{\sigma}} \left(\varepsilon_t^i\right)^{1-1/\sigma} \end{aligned}$$

and note that consumption of individual i is proportional to its weight α^i in the social welfare function. Leisure is directly proportional to its weight (a wealth effect) and inversely proportional to individual productivity: efficiency arguments induce the planner to make high-productivity individuals work harder.

Integrating the two FOCs across agents gives

$$\begin{aligned} C_t &= \int_I c_{it} d\mu^i = \int_I \left(\frac{\alpha^i}{\theta_t}\right)^{\frac{1}{\gamma}} d\mu^i \\ 1 - N_t &= 1 - \int_I \varepsilon_t^i h_t^i d\mu^i = 1 - \int_I \left(\frac{\phi \alpha^i}{\theta_t w_t}\right)^{\frac{1}{\sigma}} \left(\varepsilon_t^i\right)^{1-1/\sigma} d\mu^i \end{aligned} \tag{29}$$

Now, note that

$$c_{it} = \frac{\left(\frac{\alpha^i}{\theta_t}\right)^{\frac{1}{\gamma}}}{\int_I \left(\frac{\alpha^i}{\theta_t}\right)^{\frac{1}{\gamma}} d\mu^i} C_t \;\Rightarrow c_{it} = \frac{\left(\alpha^i\right)^{\frac{1}{\gamma}}}{\int_I \left(\alpha^i\right)^{\frac{1}{\gamma}} d\mu^i} C_t \tag{30}$$

$$\left(1 - h_t^i\right) \varepsilon_t^i = \left(\frac{\phi\alpha^i}{\theta_t w_t}\right)^{\frac{1}{\sigma}} \left(\varepsilon_t^i\right)^{1-1/\sigma} \Rightarrow \left(1 - h_t^i\right) = \left(\frac{\phi\alpha^i}{\theta_t w_t}\right)^{\frac{1}{\sigma}} \left(\varepsilon_t^i\right)^{-1/\sigma}$$

$$\Rightarrow \left(1 - h_t^i\right) = \frac{\left(\alpha^i\right)^{\frac{1}{\sigma}} \left(\varepsilon_t^i\right)^{-1/\sigma}}{\int_I \left(\alpha^i\right)^{\frac{1}{\sigma}} \left(\varepsilon_t^i\right)^{1-1/\sigma} d\mu^i} \left(1 - N_t\right)$$

Now, consider the social welfare function for the Negishi planner who is using weights $\{\alpha^i\}$

$$\int_I \left[\frac{\left(c_{it}\right)^{1-\gamma} - 1}{1-\gamma} + \phi \frac{\left(1 - h_t^i\right)^{1-\sigma} - 1}{1-\sigma}\right] \alpha^i d\mu^i$$

and substitute the two expressions in (30) into the social welfare function:

$$\int_I \alpha^i \frac{\left[\frac{\left(\alpha^i\right)^{\frac{1}{\gamma}}}{\int_I \left(\alpha^i\right)^{\frac{1}{\gamma}} d\mu^i} C_t\right]^{1-\gamma} - 1}{1-\gamma} d\mu^i + \phi \int_I \alpha^i \frac{\left[\frac{\left(\alpha^i\right)^{\frac{1}{\sigma}} \left(\varepsilon_t^i\right)^{-1/\sigma}}{\int_I \left(\alpha^i\right)^{\frac{1}{\sigma}} \left(\varepsilon_t^i\right)^{1-1/\sigma} d\mu^i} \left(1 - N_t\right)\right]^{1-\sigma} - 1}{1-\sigma} d\mu^i$$

$$= \frac{\frac{\int_I \alpha^i \left(\alpha^i\right)^{\frac{1-\gamma}{\gamma}} d\mu^i}{\left[\int_I \left(\alpha^i\right)^{\frac{1}{\gamma}} d\mu^i\right]^{1-\gamma}} C_t^{1-\gamma} - 1}{1-\gamma} + \phi \frac{\frac{\int_I \alpha^i \left(\alpha^i\right)^{\frac{1-\sigma}{\sigma}} \left(\varepsilon_t^i\right)^{1-1/\sigma} d\mu^i}{\left[\int_I \left(\alpha^i\right)^{\frac{1}{\sigma}} \left(\varepsilon_t^i\right)^{1-1/\sigma} d\mu^i\right]^{1-\sigma}} \left(1 - N_t\right)^{1-\sigma} - 1}{1-\sigma}$$

which yields the utility for the RA

$$\frac{C_t^{1-\gamma} - 1}{1-\gamma} + \phi\Phi \frac{\left(1 - N_t\right)^{1-\sigma} - 1}{1-\sigma}$$

where

$$\Phi = \frac{\left[\int_I \left(\alpha^i\right)^{\frac{1}{\sigma}} \left(\varepsilon_t^i\right)^{1-1/\sigma} d\mu^i\right]^{\sigma}}{\left[\int_I \left(\alpha^i\right)^{\frac{1}{\gamma}} d\mu^i\right]^{\gamma}}$$

is independent of t because shocks are iid. Therefore the RA problem which describes the aggregate allocations for this economy becomes:

$$\max_{\{C_t, H_t, K_{t+1}\}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \frac{C_t^{1-\gamma} - 1}{1-\gamma} + \phi\Phi \frac{\left(1 - N_t\right)^{1-\sigma} - 1}{1-\sigma} \tag{31}$$

$s.t.$

$$C_t + K_{t+1} = \left(1 - \delta\right) K_t + Z_t f\left(K_t, N_t\right)$$

K_0 given

Some remarks are in order:

1. We have found a RA, but its preferences are not the ones of the individual agent. Note that preferences of the RA depend on N_t (aggregate efficiency-weighted hours) instead of H_t (aggregate hours), and note that they feature a new weight on leisure Φ. This is the first reason why this is not a Gorman aggregation type of result.

2. The preference shifter Φ, in general, *depends on the distribution of shocks and the initial distribution of endowments*, therefore aggregate quantities do depend on the distribution of exogenous shocks and initial wealth (but not on the time-varying distribution of wealth). This is the second reason why this is, technically, not Gorman's aggregation. Note, however, that because these distributions are exogenous, it is a simple problem. In particular, in the case of no initial wealth heterogeneity $\alpha^i = 1$ for all i, $\Phi = \left[\int_I \left(\varepsilon_t^i\right)^{1-1/\sigma} d\mu^i\right]^\sigma$. Suppose that $\log\ \varepsilon \sim N\left(-v_\varepsilon/2, v_\varepsilon\right)$, then

$$\Phi = \exp\left(\sigma\left(\frac{\sigma-1}{\sigma}\right)\left(\frac{\sigma-1}{\sigma}-1\right)\frac{v_\varepsilon}{2}\right) = \exp\left(\frac{1-\sigma}{\sigma}\frac{v_\varepsilon}{2}\right)$$

   which shows how the variance of the shocks affects the taste for leisure of the fictitious representative agent.

3. Suppose $\gamma = \sigma$. Then utility is quasi-homothetic. If there are no skill shocks, but only differences in endowments, then $\Phi = 1$ and $H_t = N_t$ and the utility of the representative agent is the same as the individual agent. We are back to Gorman's aggregation. If there are idiosyncratic shocks, then $\Phi \neq 1$ and Gorman aggregation fails, which establishes that Gorman aggregation holds only if the unique source of heterogeneity across agents is in initial wealth.

4. Suppose $\gamma \neq \sigma$. Then utility is not quasi-homothetic. Even if there are no skill shocks, but only differences in endowments, then Φ depends on the distribution of endowments and Gorman's aggregation fails.

Finally, note that the assumption that agents can trade a full set of claims contingent on all possible realizations of idiosyncratic labor productivity shocks is not very realistic, as we will argue later in the course.

1 Consumption Insurance

Throughout the chapter we will use the following notation to represent uncertainty. Let $s_t \in S_t$ be the current state of the economy and let $s^t = \{s_0, s_1, ..., s_t\}$ be the history up to time t, with $s^t \in S^t \equiv S_0 \times S_1 \times ... \times S_t$. Let $\pi(s^t)$ the probability of this history occurring. Let $y_t^i(s^t)$ be the individual i realization of endowment/income upon the realization of history s^t, with $\sum_{i \in I} y_t^i(s^t) = Y_t(s^t)$ denoting the aggregate endowment/income.

1.1 Two Benchmarks: Autarky and Complete Markets

If we consider the spectrum of all possible market arrangements, at the two extremes we find autarky and complete markets. Let's analyze these two cases first.

1.1.1 No Risk Sharing in Autarky

The simplest starting point to analyze consumption is an endowment economy where insurance markets to trade across states s_t at a given point in time t are completely absent, and there is no storage technology to transfer resources across periods (e.g., the consumption good is perishable). In this economy, an individual i who receives a random stream of income shocks $\left\{\{y_t^i(s^t)\}_{s^t \in S^t}\right\}_{t=0}^{\infty}$ has no other choice than consuming her income in every state

$$c_t^i\left(s^t\right) = y_t^i\left(s^t\right), \text{ for all } s^t, t \tag{1}$$

and equation (1) is also her budget constraint.

1.1.2 Full Risk Sharing in Complete Markets

AD budget constraint– The diametrically opposite benchmark is an endowment economy with a full set of insurance and financial markets. Here, every agent faces a time-zero Arrow-Debreu budget constraint of the form

$$\sum_{t=0}^{\infty} \sum_{s^t \in S^t} p_t\left(s^t\right)\left[c_t^i\left(s^t\right) - y_t^i\left(s^t\right)\right] = 0, \text{ for all } i \in I. \tag{2}$$

Hence, every possible transfer of income across states and time is possible, as long as the discounted consumption expenditures equal discounted income, for each individual.

SP solution– Using the First Welfare Theorem, we can characterize the complete markets competitive equilibrium allocations as the solution to the Pareto problem

$$\begin{aligned} & \max_{\left\{c_t^i(s^t)\right\}} \sum_{t=0}^{\infty} \sum_{s^t \in S^t} \beta^t \pi\left(s^t\right) \sum_{i \in I} \alpha^i u\left(c_t^i\left(s^t\right)\right) \\ & s.t. \\ \sum_{i \in I} c_t^i\left(s^t\right) &= Y_t\left(s^t\right), \text{ for all } t, s^t \in S^t \end{aligned}$$

with solution

$$\beta^t \pi\left(s^t\right) \alpha^i u'\left(c_t^i\left(s^t\right)\right) = \theta_t\left(s^t\right),$$

which implies that for any pair of agents (i, j) the ratio of marginal utility is constant in every period and every state of the world, i.e.

$$\frac{u'\left(c_t^i\left(s^t\right)\right)}{u'\left(c_t^j\left(s^t\right)\right)} = \frac{\alpha^j}{\alpha^i}, \text{ for all } s^t, (i, j), \tag{3}$$

which is precisely the definition of *full risk-sharing (or full-insurance).*

For example, for the CRRA case, where $u\left(c_t^i\left(s_t\right)\right) = \frac{c_t^i(s_t)^{1-\sigma}}{1-\sigma}$, it is easy to derive that

$$\frac{c_t^i\left(s^t\right)}{c_t^j\left(s^t\right)} = \left(\frac{\alpha^i}{\alpha^j}\right)^{\frac{1}{\sigma}},$$

hence the ratio of consumption allocations is constant. Summing over $i = 1, ..., I$, this implies clearly that

$$c_t^j\left(s^t\right) = \left[\frac{\left(\alpha^j\right)^{\frac{1}{\sigma}}}{\sum_{i=1}^{I}\left(\alpha^i\right)^{\frac{1}{\sigma}}}\right] C_t\left(s^t\right), \tag{4}$$

so, individual consumption tracks perfectly aggregate consumption for every household.[1] Note that this result still holds in a production economy with capital accumulation. In complete markets there is a perfect separation between production of resources and distribution of output across households for consumption. Instead, in autarky, there is perfect correspondance between production and consumption.

[1] In general, full risk sharing does not mean constant consumption over time/across states. In the above model, individual consumption is constant over time and across states only if aggregate consumption is constant. Even if the aggregate endowment is constant, with flexible labor supply and non-separability between consumption and leisure in preferences, consumption may not be constant over time. The only statement that is always true in complete markets is that the ratio of marginal utility for any pair of agents is constant over time.

1.1.3 Empirical Implications

The full risk sharing hypothesis can be tested empirically. Under CRRA preferences, for example, equation (4) implies that the log-change in individual consumption should equal the log-change in aggregate consumption, for every individual, in every period. If we estimate from microdata the relationship

$$\Delta \log c_t^i = \beta_1 \Delta \log C_t + \beta_2 \Delta \log y_t^i + \varepsilon_t^i,$$

where y_t^i is current individual income, then the full risk-sharing hypothesis implies $(\beta_1 = 1, \beta_2 = 0)$. Contrast this prediction with the "autarky hypothesis" which implies $(\beta_1 = 0, \beta_2 = 1)$, i.e. consumption tracks perfectly current income. In general, they are both rejected, albeit the data seem to be much closer to full risk-sharing in many contexts.[2]

Result 2.0: A good model (empirically, at least) for consumption lies between autarky and full risk-sharing, i.e. it must be a model where agents have access to "partial" consumption insurance.

1.1.4 Partial Risk Sharing

It is useful to make a short detour here. How can we model partial consumption insurance? There are two approaches. The first one, which we can call the "endogenous incomplete markets" approach is to model explicitly the frictions that undermine full insurance (e.g., limited enforcement, or private information such as moral hazard and adverse selection). The second is the "exogenous incomplete markets" approach which suggests to model only the contracts/assets that we observe in the data (e.g., stock, bond, housing, etc.). The first approach is deeper than the second, but it yields usually a lot of state-contingent contracts in equilibrium that we do not see in the data, so the second approach is arguably more empirically useful.

Consider an example of the first approach. Suppose that the trades agreed upon at time $t = 0$ cannot be fully enforced. However, consumers who do not honor their contracts are excluded from financial markets forever and will leave in autarky thereon. Then, we can write "participation constraints" for agents, in every node s^t, of the form

$$\sum_{\tau=t+1}^{\infty} \sum_{s^\tau \in S^\tau} \beta^{\tau-t} \pi\left(s^\tau\right) u\left(c_\tau^i\left(s^\tau\right)\right) \geq \sum_{\tau=t+1}^{\infty} \sum_{s^\tau \in S^\tau} \beta^{\tau-t} \pi\left(s^\tau\right) u\left(y_\tau^i\left(s^\tau\right)\right),$$

[2] See Mace (1991), Cochrane (1991).

where we have assumed that the default decision is made after consuming at t.These constraints state that agents always weakly prefers staying in the contract from $t+1$ onward than defaulting. Clearly, these constraints limit the amount of insurance that is available in equilibrium relative to an A-D equilibrium with full enforcement. At the end of the course, we will study economies with limited contract enforcement. For now, we focus on the "exogenous incomplete markets" approach.

1.2 Exogenous Restrictions on Trade: the Bond Economy

Sequential formulation of complete markets– It is useful to start from the sequential formulation version of the Arrow-Debreu constraint (2), i.e., a sequence of budget constraints, for every period t and history s^t, of the form

$$c_t^i\left(s^t\right)+\sum_{s_{t+1}\in S_{t+1}} q_t\left(s_{t+1},s^t\right)a_{t+1}^i\left(s_{t+1},s^t\right)=y_t^i\left(s^t\right)+a_t^i\left(s^t\right), \tag{5}$$

and a no Ponzi scheme condition that rules out excessively high debt at every node.[3] Here, $q\left(s_{t+1},s^t\right)$ is the price at date t and state s^t of an Arrow security that pays one unit of consumption if state s_{t+1} occurs next period. Obviously, this sequential formulation of the complete markets model gives rise to the same conclusion, i.e., full risk sharing.[4]

Bond economy vs complete markets– The key difference between the bond economy and the complete markets model is in the set of securities that the household is allowed to trade. Under the bond economy, agents are restricted to trade only a non state-contingent asset. The budget constraint (5) is replaced by the more restrictive constraint, for every history s^t,

$$c_t^i\left(s^t\right)+q_t\left(s^t\right)a_{t+1}^i\left(s^t\right)=y_t^i\left(s^t\right)+a_t^i\left(s^{t-1}\right),$$

where $q_t\left(s^t\right)$ is the price at date t and state s^t of an asset that pays one unit of consumption next period, *independently* of the realization of the state s_{t+1}, i.e. it is a one-period bond.

[3] The notation $a_t^i\left(s_t,s^{t-1}\right)$ and the somewhat more concise formulation $a_t^i\left(s^t\right)$ are obviously equivalent, since $s^t=\left\{s_t,s^{t-1}\right\}$.

[4] It can be proved that the sequence of constraints in (5) and the no Ponzi-scheme condition

$$\lim_{t\to\infty}\sum_{s_{t+1}\in S_{t+1}} q_t\left(s_{t+1},s^t\right)a_{t+1}^i\left(s_{t+1},s^t\right)\geq 0$$

is equivalent to the Arrow-Debreu constraint (2). See chapter 8 in LS.

In other words, the agent is cut-off from every state-contingent insurance market and has only access to a simple financial instrument to transfer resources over time. The absence of insurance opportunities induces the consumer to hold a certain amount of the bond in order to smooth consumption.

We abstract from borrowing constraints for now, we only impose a No-Ponzi scheme condition stating that in the limit assets cannot be negative, i.e.

$$\lim_{t\to\infty} q_t\left(s^t\right) a^i_{t+1}\left(s^t\right) \geq 0,$$

Optimality implies that the weak inequality above holds with the $=$ sign.

To simplify the notation for the next sections, we assume away fluctuations in the aggregate endowment $Y_t\left(s^t\right)$, either deterministic or stochastic, hence

$$q_t\left(s^t\right) = q \equiv \frac{1}{1+r},$$

where r is the interest rate on a risk-free bond, and reformulate the budget constraint with lighter notation as

$$a_{t+1} = (1+r)\left(y_t + a_t - c_t\right), \tag{6}$$

i.e., we omit the explicit dependence on individual histories. We focus on the problem of a single individual, so we also omit the i subscript.

2 The Permanent Income Hypothesis (PIH)

The strict version of the PIH is a special case of the bond economy with two key assumptions: 1) households have quadratic utility

$$u\left(c\right) = b_1 c_t - \frac{1}{2} b_2 {c_t}^2,$$

and 2) the interest rate on the one-period bond equals the inverse of the discount rate, or $\beta\left(1+r\right) = 1$. Note that $u' > 0$ requires $c_t < b_1/b_2$ and $u'' < 0$ requires $b_2 > 0$.

Consumption as a random walk– From the consumption Euler equation:

$$b_1 - b_2 c_t = \beta\left(1+r\right) E_t\left(b_1 - b_2 c_{t+1}\right) \Rightarrow E_t c_{t+1} = c_t. \tag{7}$$

from which we recover the well known result that consumption is a martingale.[5]

[5]A martingale is a stochastic process (i.e., a sequence of random variables) $\{x_t\}$ which satisfies, at every t, $E_t x_{t+j} = x_t$ for any $j > 0$.

It is useful to note that from the law of iterated expectations and the martingale property:

$$E_t c_{t+2} = E_t \left[E_{t+1} c_{t+2}\right] = E_t c_{t+1} = c_t$$

and, more in general:

$$E_t c_{t+j} = c_t, \text{ for any } j \geq 0. \tag{8}$$

Iterating forward one period on the budget constraint (6), we obtain

$$c_t = y_t + a_t - \frac{1}{1+r} a_{t+1} = y_t + a_t - \frac{1}{1+r} \left[c_{t+1} - y_{t+1} + \frac{a_{t+2}}{1+r} \right]$$

and rearranging

$$c_t + \frac{1}{1+r} c_{t+1} = a_t + y_t + \frac{1}{1+r} y_{t+1} - \left(\frac{1}{1+r} \right)^2 a_{t+2}$$

If we keep iterating J times and use conditional expectations to deal with uncertain future realizations of income (and consumption), we arrive at

$$\sum_{j=0}^{J} \left(\frac{1}{1+r} \right)^j E_t c_{t+j} = a_t + \sum_{j=0}^{J} \left(\frac{1}{1+r} \right)^j E_t y_{t+j} + \left(\frac{1}{1+r} \right)^{J+1} E_t a_{t+J+1}$$

Taking the limit as $J \to \infty$ and using the No Ponzi scheme condition, we arrive at:

$$\begin{aligned} \sum_{j=0}^{\infty} \left(\frac{1}{1+r} \right)^j E_t c_{t+j} &= a_t + \sum_{j=0}^{\infty} \left(\frac{1}{1+r} \right)^j E_t y_{t+j} \\ c_t &= \frac{r}{1+r} \left[a_t + \sum_{j=0}^{\infty} \left(\frac{1}{1+r} \right)^j E_t y_{t+j} \right] = \frac{r}{1+r} \left(a_t + H_t \right) \end{aligned} \tag{9}$$

where the LHS of the second row uses property established in (8), and in the last equality we denoted human wealth, i.e. the expected discounted value of future earnings, with H_t. Recall that financial wealth is a_t. Define permanent income as the annuity value (i.e. $\frac{r}{1+r}$) of total (human and financial) wealth $W_t \equiv (a_t + H_t)$.[6] Therefore, we have the following result:

[6] The annuity value is defined as that portion that, when consumed every period, keeps the asset value constant. Note that (abstracting from y_t):

$$a_{t+1} = (1+r)(a_t - c_t) = (1+r) \left(a_t - \frac{r}{1+r} a_t \right) = a_t.$$

Result 2.1: If preferences are quadratic, and $\beta(1+r) = 1$, then consumption follows a martingale process and equals permanent income, i.e. the annuity value of human and financial wealth.

Certainty equivalence– Notice that, if one solves the non-stochastic version of the PIH problem stated earlier, from the FOC (7) one obtains $c_{t+1} = c_t$ and, by iterating forward on the budget constraint,

$$c_t = \frac{r}{1+r}\left[a_t + \sum_{j=0}^{\infty}\left(\frac{1}{1+r}\right)^j y_{t+j}\right].$$

Compared to equation (9), the above equation suggests that the consumption satisfies *certainty equivalence* in the sense that to obtain the solution of the stochastic problem, one can 1) solve the deterministic problem and 2) substitute conditional expectations of the forcing variables (y_{t+j}) in place of the variables themselves. Put differently, *the variance and higher moments of the income process do not matter for the determination of consumption.* This property descends directly from the linear-quadratic objective function.

Consumption dynamics– From (9), the change in consumption at time t equals

$$\Delta c_t = c_t - c_{t-1} = c_t - E_{t-1}c_t = \frac{r}{1+r}\left[W_t - E_{t-1}W_t\right],$$

where we have used the random walk property. Now, use the definition of total wealth W_t to define the innovation (i.e. the unexpected change) in permanent income, at time t as

$$\begin{aligned} W_t - E_{t-1}W_t &= a_t - E_{t-1}a_t + \sum_{j=0}^{\infty}\left(\frac{1}{1+r}\right)^j \left[E_t y_{t+j} - E_{t-1}\left(E_t y_{t+j}\right)\right], \qquad (10) \\ &= \sum_{j=0}^{\infty}\left(\frac{1}{1+r}\right)^j \left(E_t - E_{t-1}\right) y_{t+j}, \end{aligned}$$

where we have used the law of iterated expectations $E_{t-1}\left(E_t y_{t+j}\right) = E_{t-1}y_{t+j}$, and the fact that $a_t = E_{t-1}a_t$, since there is no uncertainty at time t about the evolution of wealth next period: just look at the budget constraint (6). Putting together (10) and the expression above for the change in consumption we arrive at

$$\Delta c_t = \frac{r}{1+r}\sum_{j=0}^{\infty}\left(\frac{1}{1+r}\right)^j \left(E_t - E_{t-1}\right) y_{t+j}. \qquad (11)$$

This equation states another useful result:

Result 2.2: under the PIH, the change in consumption between $t-1$ and t *is proportional to the revision in expected earnings due to the new information ("news") accruing in that same time interval.*

2.1 Example with a Specific Income Shock

At this point, to make further progress, we need to make some assumptions on the statistical properties of the labor income process. We choose a specification that is very common in labor economics. We assume that labor income is the sum of two orthogonal components, a permanent component y_t^p which follows a martingale, and a transitory component u_t that is independently distributed over time:

$$\begin{aligned} y_t &= y_t^p + u_t, \\ y_t^p &= y_{t-1}^p + v_t. \end{aligned} \tag{12}$$

Note that v_t is the innovation to the permanent component, independently distributed over time. Assume that $E(v_t) = E(u_t) = 0$ and that the two shocks are orthogonal, $u_t \perp v_\tau$ for all pairs (t, τ).

Using $y_t^p = y_t - u_t$ from the first equation into the second equation, we obtain the representation

$$\begin{aligned} y_t - u_t &= y_{t-1} - u_{t-1} + v_t \\ y_t &= y_{t-1} + u_t - u_{t-1} + v_t \end{aligned} \tag{13}$$

With this structure of shocks in hand, we can express the change in consumption from (11) only as a function of the permanent innovation v_t and the transitory innovation u_t. Focusing on the RHS of (11) and ignoring for now the constant $r/(1+r)$, we have

$$\sum_{j=0}^{\infty} \left(\frac{1}{1+r}\right)^j (E_t - E_{t-1}) y_{t+j} = (E_t - E_{t-1}) \left[y_t + \frac{1}{1+r} y_{t+1} + \left(\frac{1}{1+r}\right)^2 y_{t+2} + \ldots \right] \tag{14}$$

Using (13) into the first term of the RHS of equation (14)

$$\begin{aligned} (E_t - E_{t-1}) y_t &= (E_t - E_{t-1}) [y_{t-1} + u_t - u_{t-1} + v_t] \\ &= u_t + v_t \end{aligned}$$

since all the terms in $t-1$ drop out because $E_t(x_{t-1}) = E_{t-1}(x_{t-1}) = x_{t-1}$. In other words, the unexpected change in income y_t (compared to the one-step-ahead forecast $E_{t-1}y_t$) is the sum of the permanent and the transitory innovations at time t.

Using (13) into the second term of the RHS of (14):

$$\begin{aligned}
(E_t - E_{t-1})\, y_{t+1} &= (E_t - E_{t-1})\left[y_t + u_{t+1} - u_t + v_{t+1}\right] \\
&= u_t + v_t + (E_t - E_{t-1})\left[u_{t+1} - u_t + v_{t+1}\right] \\
&= v_t,
\end{aligned}$$

since all the terms indexed by $t+1$ drop out because $E_t(x_{t+1}) = E_{t-1}(x_{t+1}) = 0$. At this point, it is easy to see that

$$(E_t - E_{t-1})\, y_{t+j} = v_t \quad \text{for any } j > 1$$

In other words, the forecast revision between $t-1$ and t about income beyond time t equals the permanent innovation at time t.

Going back to expression (11), the innovation to permanent income is

$$\begin{aligned}
\Delta c_t &= \frac{r}{1+r}\left[u_t + v_t + \frac{1}{1+r}v_t + \left(\frac{1}{1+r}\right)^2 v_t + ...\right] \\
&= \frac{r}{1+r}\left[u_t + v_t \sum_{j=0}^{\infty}\left(\frac{1}{1+r}\right)^j\right] \\
&= v_t + \frac{r}{1+r}u_t.
\end{aligned}$$

Hence, households adjust their consumption responding to the annuitized change in income. This means that they will respond only weakly to purely transitory shocks (u_t), whereas they will respond one for one to permanent shocks (v_t). Indeed, the former shocks have only a small effect on permanent income, while the latter change permanent income one for one, by definition.

Note that under complete markets, we would have that $\Delta c_t = 0$ since idiosyncratic shocks do not transmit into consumption. Thus, the bond economy is quite close to full insurance with respect to transitory shocks, but very far from it with respect to permanent shocks.

Identification of the shocks through panel data– Suppose that one has panel data on consumption and income for a cohort of households, $i = 1, ..., N$ followed over periods $t = 0, .., T$. Suppose that the variances of (u_t^i, v_t^i) change over time. Let var_t denote the cross-sectional variance (i.e., the variance across individuals of this cohort) at time t. If we are interested in identifying the variances of permanent and transitory shocks, we can simply use income data and the cross-sectional moment restrictions for $t = 1, ..., T$

$$\begin{aligned} var\left(\Delta y_t^i\right) &= var\left(v_t^i\right) + var\left(u_t^i\right) + var\left(u_{t-1}^i\right), \\ cov_t\left(\Delta y_{t-1}^i, \Delta y_t^i\right) &= -var_t\left(u_{t-1}^i\right). \end{aligned}$$

However, an altaernative is to use the restrictions imposed by the theory. Then, note that for $t = 1, ..., T$ we can write:

$$\begin{aligned} var\left(\Delta c_t^i\right) &= \left(\frac{r}{1+r}\right)^2 var\left(u_t^i\right) + var\left(v_t^i\right) \simeq var\left(v_t^i\right), \\ var\left(\Delta y_t^i\right) &= var\left(v_t^i\right) + var\left(u_t^i\right) + var\left(u_{t-1}^i\right), \\ cov\left(\Delta c_t^i, \Delta y_t^i\right) &= \left(\frac{r}{1+r}\right) var\left(u_t^i\right) + var\left(v_t^i\right) \end{aligned}$$

where the approximate equality in the first row holds for r "small". Therefore, it is easy to see that with data on consumption and income one can separately identify the variances of the underlying structural income shocks.

For example, if over a certain period of time we observe the variance of income rising, but the variance of consumption approximately flat, we should conclude that the rise in income uncertainty was mostly transitory. This is an important conclusion for policy, because it suggests that redistributing from the high-income to the low income agents may be largely a waste since agents can self-insure effectively transitory shocks on their own by borrowing and saving.

2.2 When are borrowing constraints binding?

So far, we have ignored the presence of borrowing constraints. We imposed a no-Ponzi scheme condition, but we never checked whether it's actually binding. Is it a good abstraction? The answer,as we show below, depends on the income process.

Wealth dynamics with borrowing constraints– First, note that, from the budget constraint (6)

$$a_{t+1} = (1+r)\left(y_t + a_t - c_t\right),$$

rearranging, we obtain an expression for the change in wealth

$$\Delta a_{t+1} = (1+r)\, y_t + r a_t - (1+r)\, c_t. \tag{15}$$

Substituting into the above equation the optimal consumption choice from (9) reproduced below

$$(1+r)\, c_t = r\left[a_t + \sum_{j=0}^{\infty}\left(\frac{1}{1+r}\right)^j E_t y_{t+j}\right] \tag{16}$$

we obtain

$$\begin{aligned}
\Delta a_{t+1} &= (1+r)\, y_t + r a_t - r\left[a_t + \sum_{j=0}^{\infty}\left(\frac{1}{1+r}\right)^j E_t y_{t+j}\right] = (1+r)\, y_t - r\sum_{j=0}^{\infty}\left(\frac{1}{1+r}\right)^j E_t y_{t+j} \\
&= (1+r)\, y_t - r y_t - r\left[\sum_{j=1}^{\infty}\left(\frac{1}{1+r}\right)^j E_t y_{t+j}\right] \\
&= y_t - \sum_{j=1}^{\infty}\left[\left(\frac{1}{1+r}\right)^{j-1} E_t y_{t+j} - \left(\frac{1}{1+r}\right)^j E_t y_{t+j}\right]
\end{aligned}$$

where the last line uses the simple algebraic relationship

$$\frac{r}{(1+r)^j} = \frac{1}{(1+r)^{j-1}} - \frac{1}{(1+r)^j}.$$

Unfolding the expression in the sum listing first all the positive terms and then all the negative ones:

$$\begin{aligned}
\Delta a_{t+1} &= y_t - \left[E_t y_{t+1} + \left(\frac{1}{1+r}\right) E_t y_{t+2} + \dots - \left(\frac{1}{1+r}\right) E_t y_{t+1} - \left(\frac{1}{1+r}\right)^2 E_t y_{t+2} - \dots\right] \\
&= -\sum_{j=1}^{\infty}\left(\frac{1}{1+r}\right)^{j-1} E_t \Delta y_{t+j}
\end{aligned}$$

Now, suppose that the income process follows a random walk, $y_t = y_{t-1} + \varepsilon_t$, with ε_t iid and $E(\varepsilon_t) = 0$. Then it is easy to see that $\Delta y_{t+j} = \varepsilon_{t+j}$ and $\Delta a_{t+1} = 0$. Therefore, the initial wealth endowment perpetuates itself (i.e., it is constant) so if the individual starts above the borrowing constraint, it will never be binding. The reason for this result is that wealth changes only if the individual is consuming just a part of its income (and saving the remaining part) in order to smooth consumption. With permanent shocks, all the income shock is consumed in every period and the individual also consumes the annuity value of wealth, which therefore remains constant.

However, if the income process is *iid*, we have that $\Delta y_{t+1} = \varepsilon_{t+1} - \varepsilon_t$, $\Delta y_{t+2} = \varepsilon_{t+2} - \varepsilon_{t+1}$, therefore

$$\begin{aligned}
\Delta a_{t+1} &= -\sum_{j=1}^{\infty}\left(\frac{1}{1+r}\right)^{j-1} E_t \Delta y_{t+j} = -\sum_{j=1}^{\infty}\left(\frac{1}{1+r}\right)^{j-1} E_t\left[\varepsilon_{t+j} - \varepsilon_{t+j-1}\right] \\
&= -E_t\left[\varepsilon_{t+1} - \varepsilon_t\right] - \frac{1}{1+r} E_t\left[\varepsilon_{t+2} - \varepsilon_{t+1}\right] - ... \\
&= \varepsilon_t
\end{aligned}$$

since all other terms are zero. This means that wealth follows a random walk and, as a result, any constraint on asset holdings will be binding with probability one sooner or later.

A simpler way to derive the same results is as follows. When income is a unit root ($y_t = y_{t-1} + \varepsilon_t$), from (16) we obtain

$$c_t = \frac{r}{1+r} a_t + y_t$$

since $E_t y_{t+j} = y_t$ which substituted into equation (15) yields $a_{t+1} = a_t$. When income is *iid* ($y_t = \varepsilon_t$) from (16) we have

$$c_t = \frac{r}{1+r}\left(a_t + y_t\right)$$

$E_t y_{t+j} = 0$ which substituted into equation (15) yields $\Delta a_{t+1} = \varepsilon_t$.

To conclude, whether ignoring borrowing constraint is troublesome or not depends on the specific income process. However, in general this result highlights the fact that borrowing constraints cannot be ignored.

1 Precautionary Savings: Prudence and Borrowing Constraints

In this section we study conditions under which savings react to changes in income uncertainty. Recall that in the PIH, when you abstract from borrowing constraints, certainty equivalence implies that "mean preserving spreads" of the income distribution do not impact on saving. But quadratic utility is very special, what happens with more general utility functions?

1.1 Prudence: A two-period model

Consider the simple two-period consumption-saving problem studied by Leland (1968):

$$\max_{\{c_0,c_1,a_1\}} u(c_0) + \beta E[u(c_1)]$$
$$s.t.$$
$$\begin{aligned} c_0 + a_1 &= y_0 \\ c_1 &= Ra_1 + \tilde{y}_1 \end{aligned}$$

where y_0 is given, and income next period $\tilde{y}_1$ is also exogenous but stochastic.[1] If we retain the assumption $\beta R = 1$ to simplify the algebra, the Euler equation gives

$$u'(y_0 - a_1) = E[u'(Ra_1 + \tilde{y}_1)],$$

which is one equation in one unknown, a_1. The LHS is increasing in a_1 since $u'' < 0$, and the RHS is decreasing for the same reason, hence a_1^* is uniquely determined.

FIGURE HERE

Note that current consumption c_0 is determined by the period-zero budget constraint

$$c_0^* = y_0 - a_1^*,$$

hence a rise in savings leads to a fall in current consumption.

[1] Note that the timing of this problem is slightly different from the one adopted in the description of the PIH. There, we assumed that individuals receive income and consume at the beginning of the period and the payments of interests occurs at the end of the period. Here, we assume the payment of interests occurs at the beginning of the period, income is paid at the end of the period and individuals consume at the end of the period. In general, results are robust to this timing, it is a mater of convenience which one to choose.

Mean-preserving spread– What happens to optimal consumption at $t = 0$ if the uncertainty over income next period $\tilde{y}_1$ rises, i.e. as future income becomes more risky? Consider a mean-preserving spread of $\tilde{y}_1$. Define

$$\tilde{y}_1 = \bar{y}_1 + \varepsilon_1,$$

where ε_1 is the stochastic component and $\bar{y}_1$ is the mean. Assume that $E(\varepsilon_1) = 0$ and $var(\varepsilon_t) = \sigma_\varepsilon$. The Euler equation becomes

$$u'(y_0 - a_1) = E\left[u'(Ra_1 + \bar{y}_1 + \varepsilon_1)\right],$$

which shows that *if* u' *is convex*, then by Jensen's inequality, a mean-preserving spread of ε_1 will increase the value of the RHS, which then shifts upward, inducing a rise in a_1^* and a fall in c_0^*. This is an application of the famous result by Rotschild and Stiglitz (1970).

Prudence– The convexity of the marginal utility (or $u''' > 0$) is called "prudence" and is a property of preferences, like risk aversion: risk-aversion refers to the curvature of the utility function, whereas prudence refers to the curvature of the marginal utility function.[2]

Result 2.3: If the marginal utility is convex $(u''' > 0)$, *then the individual is "prudent" and a rise in future income uncertainty leads to a rise in current savings and a decline in current consumption.*

It can be easily seen that any utility function with decreasing absolute risk aversion, i.e. in the DARA class (which includes CRRA), displays positive third derivative. Let $\alpha(c)$ be the coefficient of absolute risk aversion. Then:

$$\alpha(c) = \frac{-u''(c)}{u'(c)} \Rightarrow \alpha'(c) = \frac{-u'''(c)\,u'(c) + \left[u''(c)\right]^2}{\left[u'(c)\right]^2}.$$

Since with DARA $\alpha'(c) < 0$, then we have that

$$-u'''(c)\,u'(c) + \left[u''(c)\right]^2 < 0 \Rightarrow u'''(c) > \frac{\left[u''(c)\right]^2}{u'(c)} > 0.$$

Intuitively, a rise in uncertainty reduces the certainty-equivalent income next period and with DARA effectively increases the degree of risk-aversion of the agent, inducing him to save more.

[2] Precisely, Kimball (1990) defines the index of absolute prudence as the ratio $-u'''(c)/u''(c)$, so in a similar vein to the Arrow-Pratt index of absolute risk-aversion $-u''(c)/u'(c)$.

Prudence is a motive for additional savings in order to take precaution against possible negative realizations of the income shock next period. In this sense, savings induced by prudence are called "precautionary savings" or "self-insurance". In this simple, two-period partial equilibrium model one can define precautionary wealth due to income uncertainty σ_ε as the difference between the optimal asset choice under uncertainty $a_1^*(\sigma_\varepsilon)$ and the optimal asset choice under certainty over next period income, i.e. $a_1^*(0)$.

Saving motives– This is a good time to make a short remark about "saving motives". The saving motive associated to $\beta R > 1$ which pushes the individual to postpone consumption because of patience and/or returns to savings is called *intertemporal motive.* The saving motive of the pure PIH where utility is quadratic (hence uncertainty has no role) and $\beta R = 1$ (hence intertemporal motives are inactive) is called *smoothing motive.* The individual wants to smooth consumption through income shocks. Finally, as explained above, the saving motive associated to future income uncertainty is called *precautionary or self-insurance motive.* We add that in a life-cycle model where the individual faces a retirement period, during the working stage of the life-cycle the individual would have a *life-cycle motive* for saving associated to the desire of smoothing consumption between working life and retirement. In presence of altruism towards their offsprings, we have an additional saving motive aimed at leaving behind some assets as *bequest.*

1.2 Prudence: Multi-period case

Let's generalize the two-period model to a multiperiod model with *iid* income shocks and finite-horizon. In the multi-period case (with time horizon T), the problem of the household can be written, in recursive form, as

$$\begin{aligned} V_t(a_t, y_t) &= \max_{\{c_t, a_{t+1}\}} u(c_t) + \beta E\left[V_{t+1}(a_{t+1}, y_{t+1})\right] \\ & \quad s.t. \\ c_t + a_{t+1} &= Ra_t + y_t \end{aligned}$$

Note that when the income shocks $\{y_t\}$ are *iid*, we can define a unique state variable which is a sufficient statistics for the household choice, "cash in hand" $x_t \equiv Ra_t + y_t$ since (a_t, y_t) always enter additively and current levels of y_t do not provide any information about the future realizations of income shocks. Note that

$$x_{t+1} = Ra_{t+1} + y_{t+1} = R(x_t - c_t) + y_{t+1},$$

which is the law of motion for the new individual state variable. See LS, 17.5.1 for a presentation of cash-in-hand as state variable. This leads to the simpler formulation

$$\begin{aligned} V_t(x_t) &= \max_{\{c_t, x_{t+1}\}} u(c_t) + \beta E\left[V_{t+1}(x_{t+1})\right] \\ & \quad s.t. \\ x_{t+1} &= R(x_t - c_t) + y_{t+1} \end{aligned}$$

From the FOC's with respect to c_t and the constraint, we obtain

$$u'\left(x_t + \frac{y_{t+1} - x_{t+1}}{R}\right) = \beta R E\left[V'_{t+1}(x_{t+1})\right]. \tag{1}$$

Applying the same logic as in the two-period model, it is easy to conclude that precautionary saving arises as long as the derivative of the value function $\left(V'_{t+1}\right)$ is convex, i.e. $V'''_{t+1} > 0$. Sibley (1975) showed that when the time-horizon T is finite, it can be proved that *if* $u''' > 0$, *then* $V'''_t > 0$ *for all* $t = 1, ..., T$. The proof is based on backward induction: in the last period $V'_T = u'$ which is convex by assumption. Then, one can show that V'_{T-1} is also convex, and so on.

1.3 Borrowing Constraints

To isolate the role of borrowing constraints for precautionary saving, we abstract from prudence altogether and focus on the quadratic utility case. To account for the possibility that the borrowing constraint is binding, the Euler equation needs to be modified. Suppose households face a *no-borrowing constraint* $a_{t+1} \geq 0$. Then, (**??**) becomes

$$c_t = \begin{cases} E_t c_{t+1} & \text{if } a_{t+1} > 0 \\ y_t + a_t & \text{if } a_{t+1} = 0 \end{cases}$$

where the first line is just the FOC of the agent when the constraint is not binding, while the second line descends directly from the budget constraint $a_{t+1} = R(y_t + a_t - c_t)$ when the constraint is binding ($a_{t+1} = 0$). The constrained household would like to borrow to finance consumption, but it is not allowed, so it consumes all its resources.

In which scenarios is the borrowing constraint binding? For example, imagine that $a_t = 0$ and that income $y_t = \bar{y} + \varepsilon_t$, where ε_t follows an iid process with mean zero. In this case, we know that optimal unconstrained consumption will be $c_t^* = \bar{y} + \frac{r}{1+r}\varepsilon_t$ while her total resources are $y_t = \bar{y} + \varepsilon_t$. Therefore, if ε_t is negative, the agent, to smooth income, would like to consume $c_t > y_t$ (which she would if she could borrow) but she is constrained at $c_t = y_t$. In general, the constraint is likely to bind whenever y_t follows a mean-reverting process. The above pair of conditions can be written in compound form as

$$c_t = \min\{y_t + a_t, E_t c_{t+1}\} = \min\{y_t + a_t, E_t[\min\{y_{t+1} + a_{t+1}, E_{t+1}c_{t+2}\}]\}.$$

Now, suppose that the uncertainty about income y_{t+1} increases. Very low realizations of income y_{t+1} become more likely, which makes the borrowing constraint more likely to bind in the future and reduces the value of $E_t[\min\{y_{t+1} + a_{t+1}, E_{t+1}c_{t+2}\}]$. This, in turn, reduces the value of $E_t c_{t+1}$. Thus, if the borrowing constraint is not already binding at time t but it may be binding in the future, then agents consume less today.

Intuitively, when agents face borrowing constraints, they fear getting several consecutive bad income realizations which would push them towards the constraint and force them to consume their income *without the ability of smoothing consumption.* To prevent this situation, they save for self-insurance (precautionary motive). Thus, we have an important result: prudence is not strictly necessary for precautionary saving behavior, or:

Result 2.4: Even in absence of prudence (e.g. with quadratic preferences), in presence of borrowing constraints a rise in future income uncertainty leads to a rise in current savings for precautionary reasons and to a decline in current consumption.

Even though we showed this result for quadratic utility, it is a general results that holds for concave utility.

1.3.1 A Natural Debt Limit

This is a good place to discuss how to model debt limits. We started by imposing an exogenous borrowing constraint like $a_{t+1} \geq -\phi$, where ϕ is a parameter (in our previous case $\phi = 0$). However, one may wonder if there is a "natural" borrowing limit that the household faces.

Suppose the income process $\{y_t\}_{t=0}^{\infty}$ is deterministic. Impose non-negativity of consumption throughout the life of the household, i.e. $c_t \geq 0$ for all t and iterate forward on the budget constraint

$$\begin{aligned}
c_t &= a_t + y_t - \frac{a_{t+1}}{1+r} \geq 0 \quad \Rightarrow \quad a_t \geq -y_t + \frac{a_{t+1}}{1+r} \\
a_t &\geq -y_t + \frac{a_{t+1}}{1+r} \geq -y_t + \frac{1}{1+r}\left[-y_{t+1} + \frac{a_{t+2}}{1+r}\right] \geq \ldots \\
a_t &\geq -\sum_{j=0}^{\infty}\left(\frac{1}{1+r}\right)^j y_{t+j}.
\end{aligned}$$

In other words, by imposing this constraint, the household is not allowed to accumulate more debt than what she will ever be able to repay by consuming just zero every period.

If the income process is stochastic, then how can we be sure that whatever the household borrows she will repay *almost surely (i.e. with probability 1)*? Then, we need to substitute y_t at each t with the lowest possible realization of the income shock, call it $y_{\min}$, and we have the *natural debt limit*

$$a_t \geq -\left(\frac{1+r}{r}\right) y_{\min}. \tag{2}$$

This is the loosest possible debt limit. No exogenous borrowing constraint can ever be looser than the natural debt limit. Note however that if $y_{\min} = 0$, then the naturals debt limit is zero!

Inada conditions and natural borrowing limit– Keep in mind an important property: if the utility function satisfies the Inada condition $u(0) = -\infty$, then the consumer will never want to borrow up to the natural debt limit.[3] Suppose she does borrow up to $a_t = -\left(\frac{1+r}{r}\right) y_{\min}$ and suppose the income realization y_t is precisely $y_{\min}$ which has positive probability. From the budget constraint at date t:

$$\begin{aligned}
c_t &= a_t + y_{\min} - \frac{a_{t+1}}{1+r} = -\left(\frac{1+r}{r}\right) y_{\min} + y_{\min} - \frac{a_{t+1}}{1+r} \\
&= -\frac{1}{r} y_{\min} - \frac{a_{t+1}}{1+r} \leq -\frac{1}{r} y_{\min} - \frac{1}{1+r}\underbrace{\left[-\left(\frac{1+r}{r}\right) y_{\min}\right]}_{\text{max that can be borrowed}} \\
&= 0
\end{aligned}$$

which shows that, with positive probability next period the consumer will have to consume zero. However that would lead to an infinitely negative utility, and so the consumer will never to reach that state.

[3] For example, CRRA utility $u(c) = \frac{c^{1-\gamma}-1}{1-\gamma}$ with $\gamma \geq 1$ satisfies the Inada condition $u(0) = -\infty$.

The preferences alone will insure that the *natural* borrowing limit will never bind. In other words, in solving for the optimal consumption you can safely assume interior solutions for the Euler equation. This is not true for ad-hoc debt limits!

2 The Income Fluctuation Problem

We are now ready to formally examine the general problem

$$\begin{aligned} & \max_{\{c_t, a_{t+1}\}_{t=0}^{\infty}} E_0 \sum_{t=0}^{\infty} \beta^t u(c_t) \\ & s.t. \\ c_t + a_{t+1} &= Ra_t + y_t \\ a_{t+1} &\geq 0 \end{aligned}$$

where we combine general preferences (i.e., we just impose $u' > 0, u'' < 0$) with a no-borrowing constraint. The value for the interest rate is always *exogenously given* and constant. Note the slight difference in timing of the budget constraint with respect to the problem stated when we studied the PIH. We explained that it only depends on the timing of the consumption choice.

The first order necessary condition for optimality is

$$u'(c_t) = \beta R E_t [u'(c_{t+1})] + \lambda_t, \tag{3}$$

where $\lambda_t > 0$ is the multiplier on the no-borrowing constraint. Condition (3) implies the Euler equation

$$u'(c_t) \geq \beta R E_t [u'(c_{t+1})]. \tag{4}$$

We want to understand whether the optimal consumption sequence $\{c_t\}$ is bounded above or whether it will be diverging as $t \to \infty$. This characterization is important because, if the consumption sequence is bounded, since $c'(a) > 0$ then the endogenous state space for assets $[0, \bar{a}]$ is compact, i.e. there exists an upper bound $\bar{a}$ which is finite. This is a crucial requirement for proving the existence of an equilibrium in economies populated by many agents who chooses their optimal consumption by solving income fluctuations problems. If the consumption sequence diverges, then $\bar{a} \to \infty$. This means that, in such an economy, there will be an infinite supply of assets which prevents existence of an equilibrium.

The convergence properties of the consumption sequence will depend on the value of the term βR. We always examine three separate cases: βR above, equal to or below one.

We start from a problem where income fluctuations are deterministic, i.e., perfectly foreseen. Next we move to stochastic income fluctuations.

2.1 Deterministic Income Fluctuations

Case $\beta R > 1$: Without uncertainty, since u is strictly concave, the Euler equation (4) implies

$$u'(c_t) \geq \beta R u'(c_{t+1}) > u'(c_{t+1}) \Rightarrow c_{t+1} > c_t,$$

thus consumption grows indefinitely. Since borrowing is limited at zero, assets must grow to finance consumption, so also assets diverge. The individual is "too patient" or/and the rate of return on savings is too high. Both forces push her to accumulate too much wealth.

Case $\beta R = 1$: From the Euler Equation, $u'(c_t) \geq u'(c_{t+1})$. Ideally, households want perfectly smooth consumption ($c_{t+1} = c_t$) when the liquidity constraint is not binding. If it is binding, then $c_{t+1} > c_t$. Overall, consumption is a *nondecreasing* sequence. One can prove that the existence of the borrowing constraint affects the problem only until a given time τ and vanishes thereafter. Until $t = \tau$, consumption will grow and then it remains constant thereafter. How is τ determined? Define

$$h_t = \frac{r}{1+r} H_t = \frac{r}{1+r} \sum_{j=0}^{\infty} \left(\frac{1}{1+r} \right)^j y_{t+j}$$

as the annuity value of the of human wealth (i.e., the discounted present value of future income). Then τ satisfies

$$\tau = \arg\sup_t h_t.$$

So, consumption and assets converge to a finite value. See LS (16.3.1) for a formal proof of this result.

To gain some intuition, consider a sequence for y_t that increases until time t^* and then is decreasing from there on. It is easy to see that $\tau < t^*$. Households would like to borrow initially against future higher income, but they cannot so the constraint will bind for a while and they will keep increasing consumption until τ and thereafter they will follow a constant consumption path where they will save at the beginning to finance future consumption when income is low.

Clearly, there may be some paths for $\{y_t\}$ such that h_t is always increasing, in which case the individual will always be constrained (or, $\tau \to \infty$). However, assuming that the sequence $\{y_t\}$ is bounded above (reasonable, since y_t is individual income) is enough to guarantee that h_t has a maximum for finite τ.

Case $\beta R < 1$**:** Consider the simple case where the endowment sequence is constant at y. Let's write the above problem in DP form with cash-in-hand as a state variable, i.e., $x \equiv Ra + y$. Then:

$$\begin{aligned} V(x) &= \max_c \{u(c) + \beta V(x')\} \\ & \quad s.t. \\ x' &= R(x-c) + y \\ x' &\geq y \end{aligned}$$

If at time t the liquidity constraint is binding, then next period $a_{t+1} = 0$, and $c_{t+1} = y$. Since y is constant, this situation perpetuates and c is constant forever: the individual is always constrained. In this case the consumption sequence is bounded.

The interesting case is when the liquidity constraint is not binding and wealth is positive. From the envelope condition of the above problem,

$$u_c(c) = V_x(x),$$

which differentiated wrt to x gives

$$u_{cc}(c)\frac{dc}{dx} = V_{xx}(x) \Longrightarrow \frac{dc}{dx} = \frac{V_{xx}(x)}{u_{cc}(c)} > 0, \tag{5}$$

thus, consumption is increasing in cash-in-hand, under concavity of V.[4]

Next, we want to show that, as long as the borrowing constraint is not binding (i.e. if $x' > y$), then $x' < x$. When $x' > y$, the Euler Equation holds with equality, and:

$$\begin{aligned} u_c(c) &= \beta R V_x(x') \\ V_x(x) &= \beta R V_x(x') < V_x(x') \Longrightarrow x' < x, \end{aligned}$$

[4]Here we are assuming the concavity of V but we know that under certain conditions the Bellman operator preserves strict concavity of u.

where the second line follows from the envelope condition, from $\beta R < 1$ and from the concavity of V. Therefore, if we start from a positive level of assets a_0, assets will be optimally decumulated over time and consumption will decline because of (5). Because $dc/dx > 0$, since x declines over time, the consumption sequence is bounded above.

Finally, we can also show that when x reaches y from above, then $x' = y$ and $c = y$, i.e. in the limit as assets get depleted and reach zero (and cash in hand equals y), the consumption sequence converges to the constant endowment stream.[5] We prove it *by contradiction*. Suppose that at the point where $x = y$, $c < x$, hence $x' > y$ Then, the FOC holds with equality and

$$\begin{aligned} u_c(c) &= \beta R V_x(x') \\ V_x(y) &= \beta R V_x(R(x-c)+y) < V_x(R(x-c)+y) < V_x(y), \end{aligned}$$

where the second line uses the envelope condition, the fact that $\beta R < 1$ and the strict concavity of the value function. The second line contains the contradiction. Thus, when $\beta R < 1$ consumption converges to a finite value.

It is not difficult to generalize this result, and show that in this case the consumption sequence will converge even if the y_t sequence is time-varying.

Taking stock- How can we take stock of these results? In the deterministic case, the desire to save is increasing in patience (β) and the interest rate (R). When βR is too large, households' desire to accumulate is "too strong" and both assets and consumption diverge to infinity. To prevent divergence, we need $\beta R \leq 1$.

2.2 Stochastic Income Fluctuations

We now turn to the stochastic case. Here there is an additional motive for saving, the *precautionary motive*, due to 1) prudence (for some utility functions) and 2) the interaction between risk-aversion (i.e., aversion to consumption fluctuations) and the borrowing constraint. Given the extra saving motive that pushes consumption upward over time, we should expect that the condition under which $\{c_t\}$ converges will be more stringent than in the deterministic case. It turns out that we will need $\beta R < 1$.

A useful supermartingale– Multiply both sides of

$$u'(c_t) \geq \beta R E_t[u'(c_{t+1})]$$

[5] You should try to prove that the strictly decreasing cash in hand sequence will reach y.

by $(\beta R)^t$ and define $M_t \equiv (\beta R)^t u'(c_t) > 0$. Then equation (4) can be written as

$$M_t \geq E_t M_{t+1}$$

which asserts that M_t follows a *supermartingale.* Since M_t is non-negative, by the supermartingale convergence theorem (Doob, 1995), this stochastic process converges almost surely to a non-negative finite limit $\bar{M}$, i.e.,

$$\lim_{t \to \infty} M_t = \bar{M} < \infty. \tag{6}$$

Case $\beta R > 1$**:** According to the convergence theorem above, $(\beta R)^t u'(c_t)$ has a finite limit. Since $(\beta R)^t \to \infty$, then marginal utility $u'(c_t)$ can only converge to $u'(\bar{c}) = 0$ or, given the Inada condition, $c_t \to \infty$. Since debt is limited (at zero, in our benchmark, but could be any finite limit), this large consumption cannot be financed by debt but must be financed by saving. And hence divergence of consumption means $a_t \to \infty$, which in turn implies that there is no upper bound in the asset space. This is the same result we found for the certainty case.

Case $\beta R = 1$**:** We can give a simple proof of this result for general income process if we assume that $u''' > 0$. From the Euler Equation

$$u'(c_t) \geq E_t [u'(c_{t+1})].$$

From convexity of the marginal utility, by Jensen's inequality

$$u'(c_t) \geq E_t [u'(c_{t+1})] > u'(E_t (c_{t+1})),$$

and by concavity of u, we have that $E_t (c_{t+1}) > c_t$, so consumption will always tend to ratchet upward over time which prevents the consumption sequence from converging almost surely to a finite number.

Another simple argument is available for the case where y is iid. The proof is by contradiction. Suppose there exists an upper bound for cash-in-hand $\bar{x}$. Let $\bar{y}$ be the highest possible realization of y, and $\bar{x} = \bar{y} + Ra'(\bar{x})$ be the maximum amount of cash in hand. Using the envelope condition into the Euler equation

$$V_x(\bar{x}) \geq E_t [V_x (y_{t+1} + Ra_{t+1}(\bar{x}))] > V_x(\bar{y} + Ra_{t+1}(\bar{x})) = V_x(\bar{x})$$

where the second inequality descends from the strict concavity of the value function, and the last equality, which states the contradiction, comes from the definition of cash in hand. See LS (16.5) for a more complete formal proof of this result.

Finally, note that the convergence result with $\beta R = 1$ of the deterministic case does not hold in the stochastic case.

Case $\beta R < 1$**:** Consider the case of *iid* income shocks. Let x be cash in hand. From the Euler Equation:

$$u_c\left(c\left(x\right)\right) = \beta R E\left[u_c\left(c\left(x'\right)\right)\right] = \beta R \frac{E\left[u_c\left(c\left(x'\right)\right)\right]}{u_c\left(c\left(\bar{x}'\right)\right)} u_c\left(c\left(\bar{x}'\right)\right), \tag{7}$$

where $x' = Ra'\left(x\right) + y'$ is cash in hand next period given that today's cash in hand is x and given the current income realization is y. Let $\bar{x}' = Ra'\left(x\right) + \bar{y}$ be the cash in hand associated to the maximum realization of income next period, given that today's cash in hand is x. Suppose that the limit

$$\lim_{x \to \infty} \frac{E\left[u_c\left(c\left(x'\right)\right)\right]}{u_c\left(c\left(\bar{x}'\right)\right)} = 1. \tag{8}$$

Then, for x large enough, since $\beta R < 1$, the Euler equation (7) yields

$$u_c\left(c\left(x\right)\right) = \beta R u_c\left(c\left(\bar{x}'\right)\right) < u_c\left(c\left(\bar{x}'\right)\right).$$

Concavity of u and monotonicity of c wrt x (proved in the deterministic case earlier) implies that

$$c\left(\bar{x}'\right) < c\left(x\right) \Rightarrow \bar{x}'\left(x\right) < x$$

thanks to the fact that $c_x\left(x\right) > 0$. And we would be done, because we have demonstrated that cash in hand does not increase forever: when x is large enough $x' < x$ for sure.

Therefore, we only need to establish conditions under which the limit in (8) holds. Consider $u_c\left(c\left(x'\right)\right)$ and compute a first-order Taylor approximation around $x' = \bar{x}'$:

$$u_c\left(c\left(x'\right)\right) \simeq u_c\left(c\left(\bar{x}'\right)\right) + u_{cc}\left(c\left(\bar{x}'\right)\right) c_x\left(\bar{x}'\right)\left(x' - \bar{x}'\right).$$

Taking expectations of both sides

$$\begin{aligned} E\left[u_c\left(c\left(x'\right)\right)\right] &\simeq u_c\left(c\left(\bar{x}'\right)\right) - u_{cc}\left(c\left(\bar{x}'\right)\right) E\left[\bar{x}' - x'\right] c_x\left(\bar{x}'\right) \qquad (9) \\ &= u_c\left(c\left(\bar{x}'\right)\right) - u_{cc}\left(c\left(\bar{x}'\right)\right) E\left[\bar{y} - y'\right] c_x\left(\bar{x}'\right) \end{aligned}$$

where in the first line we use the fact that $\bar{x}'$ is deterministic since it is implied by the specific income realization $\bar{y}$. And in the second line we use the fact that $x' \equiv Ra' + y'$ which implies $\bar{x}' - x' \equiv \bar{y} - y'$.

Dividing equation (9) by $u_c(c(\bar{x}'))$ we obtain:

$$\frac{E[u_c(c(x'))]}{u_c(c(\bar{x}'))} \simeq 1 + \alpha(c(\bar{x}'))[\bar{y} - E(y')]\, c_x(\bar{x}'),$$

where $\alpha(c)$ is the coefficient of absolute risk aversion at consumption level $c(\bar{x}')$. Since both $[\bar{y} - E(y')]$ and $c_x(\bar{x}')$ are positive and finite, a sufficient condition for the limit in (8) to hold is

$$\lim_{x \to \infty} \alpha(c(x)) = 0. \tag{10}$$

If, for example, *absolute risk aversion is monotonically decreasing with asset holdings* then condition (10) holds. The faster α decreases, the smaller the upper bound on the asset space. The intuition is clear: DARA means that the agent is less worried about income uncertainty as she gets rich because she becomes less risk averse, so she will consume more and accumulate less. This force offsets, and eventually overcomes, precautionary accumulation as wealth increases.

Note that CRRA utility has DARA, so it satisfies condition (10). Remember that DARA is a sufficient condition and that this result holds for *iid* shocks. Huggett (1993) generalizes this result to a 2-state Markov chain for the income process (but only for the CRRA utility case).

We conclude by summarizing our findings in:

Result 2.5: In presence of borrowing constraints and uncertain income, the condition $\beta R < 1$ is necessary for the optimal consumption sequence and for the asset space to be bounded. Moreover, when $\beta R < 1$, if income shocks are iid and absolute risk-aversion is decreasing (DARA utility), then the asset space is bounded. More in general, even with Markov shocks, as long as absolute risk aversion decreases fast enough with c, the state space will remain bounded.

3 Notes

Leland (1968) and Sandmo (1970) showed in a two-period setting that a positive third derivative is needed to obtain precautionary savings. Sibley (1975) extended the result to a multi-period setting. Kimball (1990) further generalized these results. Schechtman and Escudero (1977) characterized the income fluctuation problem with *iid* shocks. Huggett (1993) generalizes their proofs to the case of a 2-state Markov chain for the income process and CRRA utility. Chamberlain and Wilson (2000) extend it to any arbitrary endowment sequence that is sufficiently stochastic.

References

[1] Chamberlain, Gary and Charles Wilson (2000); Optimal Intertemporal Consumption under Uncertainty, *Review of Economic Dynamics*

[2] Huggett, Mark (1993); The risk-free rate in heterogeneous-agent incomplete-insurance economies, *Journal of Economic Dynamics and Control*

[3] Kimball, Miles (1990); Precautionary Saving in the Small and in the Large, *Econometrica*

[4] Leland, H. (1968); Savings and Uncertainty: The Precautionary Demand for Saving, *Quarterly Journal of Economics*

[5] Sandmo Agnar (1970); The Effect of Uncertainty on Saving Decisions, *Review of Economic Studies*.

[6] Schechtman J. and V. Escudero (1977); Some Results on An Income Fluctuation Problem, *Journal of Economic Theory*

[7] Sibley David (1975); Permanent and Transitory Income Effects in a Model of Optimal Consumption with Wage Income Uncertainty, *Journal of Economic Theory*

1 Discretization of a continuous stochastic process

In many of the problems we studied in the previous chapter, we postulated that agents face a continuous stochastic income process. A typical example would be assuming that income is given by

$$Y_t = \exp(y_t),$$

where y_t follows a first-order autoregressive process of the class

$$y_t = \rho y_{t-1} + \varepsilon_t, \tag{1}$$

where $|\rho| < 1$ and ε_t is a white noise process with variance σ_ε^2. Assume that ε_t is drawn from a distribution G. Let $\Pr\{\varepsilon_t \leq \bar{\varepsilon}\} = G(\bar{\varepsilon}) = F(\bar{\varepsilon}/\sigma_\varepsilon)$, where F is the "standardized" version of $G(\varepsilon_t)$ with unit variance. This $AR(1)$ process for y_t is covariance-stationary with mean zero and variance $\sigma_y^2 = \sigma_\varepsilon^2/(1-\rho^2)$.

Solving the household consumption-saving problem with a continuous shock can be done, but it is very costly, computationally. So it's useful to learn how to approximate discretely a continuous process through a finite-state Markov chain that will mimic closely the underlying process. To discretize the continuous process in (1) we need two ingredients: (i) the points on the finite state space and (ii) the transition probabilities.

1.1 Tauchen's method

Let $\tilde{y}$ be the discrete-valued process that approximates y and let $\{y_1, y_2, ..., y_N\}$ be the finite set of possible realizations of $\tilde{y}$.

Choice of points. Tauchen (1986) suggests to select a maximum value y_N as a multiple m (e.g., $m = 3$) of the unconditional standard deviation, i.e.

$$y_N = m\left(\frac{\sigma_\varepsilon^2}{1-\rho^2}\right)^{\frac{1}{2}},$$

and let $y_1 = -y_N$ (assuming G is symmetric), and $\{y_2, y_3, ..., y_{N-1}\}$ be located in a equi-spaced manner over the interval $[y_1, y_N]$. Denote with d the distance between successive points in the state space.

Transition probabilites. Let

$$\begin{aligned}
\pi_{jk} &= \Pr\{\tilde{y}_t = y_k | \tilde{y}_{t-1} = y_j\} = \Pr\{y_k - d/2 < \rho y_j + \varepsilon_t \leq y_k + d/2\} \\
&= \Pr\{y_k - d/2 - \rho y_j < \varepsilon_t \leq y_k + d/2 - \rho y_j\}
\end{aligned}$$

be the generic transition probability.

Then, if $1 < k < N-1$, for each j choose

$$\pi_{jk} = F\left(\frac{y_k + d/2 - \rho y_j}{\sigma_\varepsilon}\right) - F\left(\frac{y_k - d/2 - \rho y_j}{\sigma_\varepsilon}\right),$$

while for the boundaries of the interval $k = 1$ and $k = N$ choose:

$$\begin{aligned} \pi_{j1} &= F\left(\frac{y_1 + d/2 - \rho y_j}{\sigma_\varepsilon}\right), \\ \pi_{jN} &= 1 - F\left(\frac{y_N - d/2 - \rho y_j}{\sigma_\varepsilon}\right). \end{aligned}$$

Clearly, as $d \to 0$ (and therefore $N \to \infty$), the approximation becomes better and better until it converges to the true continuous process y_t. It is useful to notice that other integration rules (e.g., Gaussian quadrature or the collocation methods) could lead to a more efficient placement of the points on the the interval $[y_1, y_N]$.

Accuracy of approximation: To assess the adequacy of the approximation, note that the approximate process $\tilde{y}$ admits a representation

$$\tilde{y}_t = \tilde{\rho}\tilde{y}_{t-1} + \tilde{\varepsilon}_t,$$

where $\tilde{\rho} = cov(\tilde{y}_{t-1}, \tilde{y}_t) / var(\tilde{y}_t)$ and $\tilde{\sigma}_\varepsilon = \left(1 - \tilde{\rho}^2\right) var(\tilde{y}_t)$. The unconditional second moments of the $\tilde{y}_t$ distribution can just be computed from the stationary distribution of $\tilde{y}_t$. In turn, the stationary distribution can be computed as follows. Let Θ_N be the transition matrix of dimension $(N \times N)$ with generic element π_{jk}. Then, the stationary distribution π^* solves

$$\pi^* = \Theta_N \pi^* \to (I - \Theta_N)\pi^* = 0$$

which is a linear system of N equations. Using this method, one can directly compare ρ to $\tilde{\rho}$ and σ_ε to $\tilde{\sigma}_\varepsilon$ and gauge the quality of the approximation. Alternatively, one can simulate a long sample for $\{\tilde{y}_t\}$ (for example a sample of 10,000 draws with a burn-in period of 1,000 draws to avoid initial conditions to affect the outcome), compute some key statistics and compare them with those in the original process.

Multivariate processes: Tauchen describes how to approximate also a multivariate process. This strategy that can be used to approximate higher-order autoregressive processes as well. For example, consider the $AR(2)$ process

$$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + \varepsilon_t.$$

Define a column vector $Z_t = [y_t \quad y_{t-1}]'$. Then one can write the AR(2) process above in multivariate form as

$$Z_t = \begin{bmatrix} \rho_1 & \rho_2 \\ 0 & 1 \end{bmatrix} Z_{t-1} + [\varepsilon_t \quad 0]'.$$

1.2 Rouwenhorst method

This is the best method to discretize a continuous stochastic process, in particular those with very high persistence, near unit root, typical in macro. It is accurate and fast. Consider again the $AR(1)$ process above, but this time we do not need to make any distributional assumption. We want to approximate it through a discrete-space Markov chain $\tilde{y}$ over a symmetric and evenly-spaced state space $\{y_1, y_2, ..., y_N\}$ with $-y_1 = y_N = \psi$. Construct the Markov chain Θ_N recursively as a function of three parameters only (p, q, ψ) as follows:

- **Step 1** For $N = 2$, define Θ_2 as

$$\Theta_2 = \begin{bmatrix} p & 1-p \\ 1-q & q \end{bmatrix}$$

Step 2 For $N \geq 3$, define recursively Θ_N as the $N \times N$ matrix

$$\Theta_N = p \begin{bmatrix} \Theta_{N-1} & \mathbf{0} \\ \mathbf{0}' & 0 \end{bmatrix} + (1-p) \begin{bmatrix} \mathbf{0} & \Theta_{N-1} \\ 0 & \mathbf{0}' \end{bmatrix} + (1-q) \begin{bmatrix} \mathbf{0}' & 0 \\ \Theta_{N-1} & \mathbf{0} \end{bmatrix} + q \begin{bmatrix} 0 & \mathbf{0}' \\ \mathbf{0} & \Theta_{N-1} \end{bmatrix}$$

and note that, for example, for $N = 3$

$$\begin{aligned} \Theta_3 &= p \begin{bmatrix} p & 1-p & 0 \\ 1-q & q & 0 \\ 0 & 0 & 0 \end{bmatrix} + (1-p) \begin{bmatrix} 0 & p & 1-p \\ 0 & 1-q & q \\ 0 & 0 & 0 \end{bmatrix} \\ &+ (1-q) \begin{bmatrix} 0 & 0 & 0 \\ p & 1-p & 0 \\ 1-q & q & 0 \end{bmatrix} + q \begin{bmatrix} 0 & 0 & 0 \\ 0 & p & 1-p \\ 0 & 1-q & q \end{bmatrix} \end{aligned}$$

which shows that the third line sums to 2, which justifies step 3

Step 3 Divide all but the top and bottom rows by 2 so that their elements sum to one.

One can show that this Markov chain converges to the invariant binomial distribution

$$\lambda_i^{(N)} = \begin{pmatrix} N-1 \\ i-1 \end{pmatrix} s^{i-1} (1-s)^{N-1}$$

where

$$s = \frac{1-p}{2-(p+q)}.$$

Recall that *a binomial distribution gives the discrete probability distribution of obtaining exactly $i-1$ successes out of $N-1$ Bernoulli trials (where the result of each Bernoulli trial is true with probability s and false with probability $1-s$).*

We can compute analytically all the moments of this distribution, for example

$$\begin{aligned}
E(\tilde{y}_t) &= \frac{(q-p)\,\psi}{2-(p+q)} \\
var(\tilde{y}_t) &= \psi^2\left[1-4s(1-s)+\frac{4s(1-s)}{N-1}\right] \\
Corr(\tilde{y}_t,\tilde{y}_{t+1}) &= p+q-1 \\
E[\tilde{y}_{t+1}|y_k] &= (q-p)\,\psi+(p+q-1)\,y_k \\
& \quad\ etc...
\end{aligned}$$

For any continuous stochastic process $\{y_t\}$ we can compute analytically a number of unconditional and conditional moments and we can use a method of moment estimator to choose the three parameters of the Markov chain (p,q,ψ) that minimize the distance between true and approximated selected moments.

For example, consider the $AR(1)$ process $\{y_t\}$ above and note that

$$\begin{aligned}
E(y_t) &= 0 \rightarrow \frac{(q-p)\,\psi}{2-(p+q)} = 0 \rightarrow q = p \\
Corr(y_t,y_{t+1}) &= \rho \rightarrow p+q-1 = \rho \rightarrow p = q = \frac{1+\rho}{2} \\
var(y_t) &= \sigma_y^2 \rightarrow \psi^2\left[1-4s(1-s)+\frac{4s(1-s)}{N-1}\right] = \sigma_y^2 \rightarrow \psi = \sigma_y\sqrt{N-1}.
\end{aligned}$$

Therefore, in this example, one can replicate exactly unconditional mean, variance, and first-order auto correlation of the original continuous process. Finally, since we know that the binomial converges to a Normal distribution for $N \rightarrow \infty$, this method is particularly apt at approximating log-normal (in levels) income processes, but it works with other distributions as well.

2 Global Solution Methods for the Income Fluctuation Problem

Local approximation methods, like linear quadratic (LQ) approximations, construct functions that match well the properties of the original function around a particular point. In the stochastic growth model, for example, there is a natural point around which the approximation could be taken: the deterministic steady-state level of capital k^* and the mean of the shock (usually normalized to 1). Fluctuations due to aggregate productivity shocks move the system in a small neighborhood of k^*.

When solving the consumption-saving problem with uncertainty, additional issues arise. First, choosing a particular point on the asset grid where to approximate individual behavior is much less obvious. Second, quantitatively, individual income uncertainty is much larger than aggregate uncertainty, which means that the system moves over a large interval in the state space. Third, in LQ approximations the marginal utility is linear and the role of uncertainty as a motive for saving is artificially reduced (in fact, it completely disappears in absence of borrowing constraints).

For all these reasons, we need to resort to *global* solution methods for the consumption-saving problem. There are several ways to attack the problem. One can iterate directly on the value function (value function iteration), or on the Euler equation (policy function iteration). Moreover, one can choose among three alternative ways to approximate the policy function or the value function: 1) *discretization* means that the state is assumed to take only values on a pre-specified grid; 2) *interpolation* is a procedure that finds a function that goes exactly through the grid points and defines the function outside the grid by interpolating locally between two adjacent grid points; 3) L^p*-approximation* is any procedure that finds a function close to the original one in the sense of a L^p norm, so it's a more global interpolation method.[1]

Consider the problem in its recursive formulation, with states (a, y). In general, with two state variables, we must approximate/interpolate a function in two dimension, which

[1] A distance between two functions evaluated over the N points $\{x_k\}_{k=1}^N$ in the L^p norm is

$$||f(x) - \hat{f}(x)||_p = \left(\sum_{k=1}^{N} |f(x_k) - \hat{f}(x_k)|^p \right)^{1/p}.$$

is a hard problem.[2] This is where the discretization of the income process comes in handy. We maintain that asset holdings a are a continuous variable and discretize the income process. In other words, we need to approximate/interpolate N functions $a'(a, y_j)$, for $j = 1, ..., N$.

Suppose we have already discretized the income process following Tauchen's method. Let $y \in Y$, where y is income in level, and Y is the finite set of values the Markov chain y can take. And let $\pi(y'|y)$ be the discretized transition function. Then, we can write the income fluctuation problem in recursive form as:

$$\begin{aligned} V(a, y) &= \max_{\{c,a'\}} u(c) + \beta \sum_{y' \in Y} \pi(y'|y) V(a', y') \\ & \quad s.t. \\ c + a' &\leq Ra + y \\ a' &\geq -\phi \end{aligned}$$

The Euler equation, once we substitute the budget constraint, reads

$$u_c(Ra + y - a') - \beta R \sum_{y' \in Y} \pi(y'|y) u_c(Ra' + y' - a'') \geq 0.$$

Equality holds if $a' > -\phi$. This is a second-order stochastic difference equation. The objective is to find a decision rule for asset holding $a'(a, y)$, i.e., an invariant function of the states (a, y) that satisfies the Euler Equation. The strategy will be to guess a decision rule for a'', and solve for a'. At the solution, the two policy functions must be identical.

We present here in detail the three variants of the policy-function iteration method: discretization, interpolation, and approximation.

2.1 Discretization

1. Construct a grid on the asset space $\{a_1, a_2, ..., a_M\}$, with $a_1 = -\phi$. The best way to construct the grid varies problem by problem. In general, one must put more points where the policy function is expected to display more curvature (i.e., where it is far from linear). In our case, this happens for values of a near the debt constraint $-\phi$.

[2] If the income process is *iid*, remember that you can reduce the dimensionality of the state space to one variable, cash in hand $x = Ra + y$. In general, whenever you can reduce the dimensionality of the state-space you should do so. The numerical complexity of the problem increases geometrically with the dimension of the state space, e.g., the points on the grid to be evaluated grow from N to N^2 to N^3, and so on.

2. Guess an initial decision rule for a'' on the grid points (i.e., a $M \times N$ matrix):

$$\{\hat{a}_0(a_1, y_1), \hat{a}_0(a_1, y_2), ..., \hat{a}_0(a_1, y_N); \hat{a}_0(a_2, y_1), ..., \hat{a}_0(a_2, y_N); ..., \hat{a}_0(a_M, y_N)\}$$

by making sure that each decision rule maps into a point on the asset grid. The subscript denotes the iteration number, and "0" denotes the first iteration. A reasonable guess for $a_0''(a_i, y_j)$ would be

$$\hat{a}_0(a_i, y_j) = a_i,$$

for example, we know that this exactly true when the utility function is quadratic and shocks follow a random walk.

3. Determine if the liquidity constraint is binding. For each point (a_i, y_j) on the grid, check whether

$$u_c(Ra_i + y_j - a_1) - \beta R \sum_{y' \in Y} \pi(y'|y_j) u_c(Ra_1 + y' - \hat{a}_0(a_1, y')) > 0.$$

If this inequality holds, then it means that the borrowing constraint binds, set $a_0'(a_i, y_j) = a_1$ and repeat this check for the next grid point. If the equation instead holds with the $<$ inequality, it means that there is an interior solution (it is optimal to save, or to borrow less than ϕ, for the household) and we proceed to the next step.

4. Determine $a_0'(a_i, y_j)$. Find the pair of adjacent grid points (a_k, a_{k+1}) such that

$$\begin{aligned}
\delta(a_k) &\equiv u_c(Ra_i + y_j - a_k) - \beta R \sum_{y' \in Y} \pi(y'|y_j) u_c(Ra_k + y' - \hat{a}_0(a_k, y')) < 0 \\
\delta(a_{k+1}) &\equiv u_c(Ra_i + y_j - a_{k+1}) - \beta R \sum_{y' \in Y} \pi(y'|y_j) u_c(Ra_{k+1} + y' - \hat{a}_0(a_{k+1}, y')) > 0,
\end{aligned}$$

This means that $a_0'(a_i, y_j) \in (a_k, a_{k+1})$.[3] Since we only work with points on the grid, set

$$a_0'(a_i, y_j) = \arg \min_{i \in \{k, k+1\}} |\delta(a_i)|$$

5. Check convergence by comparing the guess $\hat{a}_0(a_i, y_j)$ to the solution $a_0'(a_i, y_j)$ and stop if, for each pair (a_i, y_j) on the grid, $a_0'(a_i, y_j) = \hat{a}_0(a_i, y_j)$.

[3] In other words, a_k is smaller than $a^*(a_i, y_j)$ because $u'(c)$ is too small and a_{k+1} is larger than $a^*(a_i, y_j)$ because $u'(c)$ is too large, relative to the optimal choice.

6. If convergence is achieved, stop. Otherwise, go back to point 3 with the new guess $\hat{a}_1 (a_i, y_j) = a'_0 (a_i, y_j)$.

This is a very simple and fast method because it avoids the calculation of the policy function outside the grid points, but at the same time it can be imprecise, unless M is a very large number, but in this case the method becomes computationally costly.

2.2 Policy Function Iteration with Linear Interpolation

1. Construct a grid on the asset space $\{a_1, a_2, ..., a_M\}$ with $a_1 = -\phi$.

2. Guess an initial vector of decision rules for a'' on the grid points, call it $\hat{a}_0 (a_i, y_j)$.

3. For each point (a_i, y_j) on the grid, check whether the borrowing constraint binds, as before. I.e., for each point (a_i, y_j) on the grid, check whether

$$u_c (Ra_i + y_j - a_1) - \beta R \sum_{y' \in Y} \pi (y'|y_j) u_c (Ra_1 + y' - \hat{a}_0 (a_1, y')) > 0.$$

   If this inequality holds, then it means that the borrowing constraint binds, set $a'_0 (a_i, y_j) = a_1$ and repeat this check for the next grid point. If the equation instead holds with the $<$ inequality, it means that we have an interior solution (it is optimal to save, or to borrow less than ϕ, for the household) and we proceed to the next step.

4. For each point (a_i, y_j) on the grid, use a *nonlinear equation solver* to look for the solution a^* of the nonlinear equation

$$u_c (Ra_i + y_j - a^*) - \beta R \sum_{y' \in Y} \pi (y'|y_j) u_c (Ra^* + y' - \hat{a}_0 (a^*, y')) = 0, \qquad (2)$$

   and notice that the equation solver will try to evaluate many times the function $\hat{a}_0$ outside the grid points for assets $\{a_1, a_2, ..., a_M\}$. Hence, the need for the linear interpolation.

   (a) Assume that, for every value y', the function $\hat{a}_0 (a, y')$ is *piecewise linear* between the asset grid points. So, every time the equation solver above requires evaluating the function on a^* which lies between grid points, do as follows.

First, find the pair of adjacent grid points $\{a_k, a_{k+1}\}$ such that $a_k < a^* < a_{k+1}$, and then compute

$$\hat{a}_0(a^*, y') = \hat{a}_0(a_k, y') + (a^* - a_k)\left(\frac{\hat{a}_0(a_{k+1}, y') - \hat{a}_0(a_k, y')}{a_{k+1} - a_k}\right)$$

(b) If the solution of the nonlinear equation in (2) is a^*, then set $a'_0(a_i, y_j) = a^*$ and iterate on the next new grid point.

5. Check convergence by comparing $a'_0(a_i, y_j) - \hat{a}_0(a_i, y_j)$ through some pre-specified norm. For example, declare convergence at iteration n when

$$\max_{i,j}\{|a'_n(a_i, y_j) - \hat{a}_n(a_i, y_j)|\} < \varepsilon$$

for some small number ε which determines the degree of tolerance in the solution algorithm.

6. If convergence is achieved, stop. Otherwise, go back to point 3 with the new guess $\hat{a}_1(a_i, y_j) = a'_0(a_i, y_j)$.

Piecewise linear interpolation is fast, and obviously preserves positivity, monotonicity and concavity. However, the optimal policy obtained is not differentiable everywhere.

2.3 Policy function Iteration with Chebishev Approximation

There are several families of polynomials with good properties for function approximation, e.g., Legendre, Chebyshev, Laguerre, Hermite. In what follows, we use Chebyshev since in most of the common applications, they have the best properties (see below and Judd, chapter 6).

Chebishev Polynomials Suppose we want to approximate a function $f(x)$ over the interval $[-1, 1]$, through a polynomial function

$$\hat{f}(x) = \sum_{p=0}^{N} \kappa_p T_p(x) \tag{3}$$

where N is the order of the polynomial approximation, $T_p(x)$ is called the basis functions, i.e. it is a polynomial of order p, and κ_p are the coefficients that weight the various polynomials.

The Chebishev polynomials are a family of basis functions that is very useful for this type of approximations. They are given by the simple formula

$$T_p(x) = \cos(p \arccos(x)),$$

which can be simply constructed sequentially (please verify) as

$$\begin{aligned} T_0(x) &= 1, \\ T_1(x) &= x, \\ T_{p+1}(x) &= 2xT_p(x) - T_{p-1}(x), \quad \text{for } p > 1. \end{aligned}$$

The polynomial $T_p(x)$ with $p > 0$ has p zeros located at the points $x_k = -\cos\left(\frac{2k-1}{2p}\pi\right)$, with $k = 1, 2, .., p$ and it has$(p+1)$ extrema. All the maxima equal 1 and all the minima equal -1, so the range of the Chebishev polynomials is $[-1, 1]$. It is precisely this property that makes the Chebishev polynomials so useful in the approximation of functions. Note that the Chebishev polynomials satisfy the orthogonality condition among them

$$\sum_{k=1}^{M} T_q(x_k) T_p(x_k) = 0, \qquad \text{for } p \neq q,$$

which makes them basis functions.

One can prove that, if the κ_p coefficients are defined as

$$\kappa_p = \frac{\sum_{k=1}^{M} f(x_k) T_p(x_k)}{\sum_{k=1}^{M} T_p(x_k)^2}, \qquad p = 0, ..., N$$

then the approximation formula (3) is exact for those x_k equal to every zero of $T_N(x)$. Note that the expression for κ_p is that of an OLS estimator that solves

$$\min_{\{\kappa_p\}_{p=0}^{N}} \left\{ \sum_{k=1}^{M} \left[f(x_k) - \sum_{p=0}^{N} \kappa_p T_p(x_k) \right]^2 \right\}$$

i.e. the vector of κ_p minimizes the sum of squared residuals between the true function and the approximated function.

As we increase the order of the approximating Chebishev polynomial toward $N = \infty$, we get closer to the true function. However, even for relatively small N, this procedure yields very good approximations. In particular, notice that since the T_p functions are all bounded by one in absolute value, if the coefficients κ_p decline fast with p (and they do), then the largest omitted term in the error has order $(N+1)$. Moreover, $T_{N+1}(x)$ is an oscillatory function with $(N+1)$ extrema distributed smoothly over the interval. This smooth spreading out of the error is a very important property of optimal approximations![4]

The Algorithm

1. Compute the M Chebishev interpolation nodes on the normalized interval $[-1,1]$ which are given by the simple formula

$$x_k = -\cos\left(\frac{2k-1}{2M}\pi\right), \; k = 1, ..., M,$$

   i.e. they are the points where the Chebishev polynomials are exactly equal to zero.

   (a) Fix the bounds of the asset space $\{-\phi, a_{\max}\}$. Transform the Chebishev nodes over $[-1,1]$ into a grid over the $[-\phi, a_{\max}]$ interval, by setting

$$a_k = -\phi + (x_k + 1)\left(\frac{a_{\max}+\phi}{2}\right), \; k = 1, ..., M$$

   In particular note that for $k = M$ and M large, $x_k \simeq 1$ and $a_M \simeq a_{\max}$; for $k = 1$ and M large, $x_k \simeq -1$ and $a_k \simeq -\phi$.[5]

2. Guess an initial vector of decision rules for a'' on the grid point, call it $\hat{a}_0(a_i, y_j)$

3. For each point (a_i, y_j) on the grid, check whether the borrowing constraint binds. If not, continue.

4. For each point (a_i, y_j) on the grid, use a nonlinear equation solver to look for the solution a^* of the nonlinear equation

$$u_c(Ra_i + y_j - a^*) - \beta R \sum_{y' \in Y} \pi(y'|y_j)\, u_c(Ra^* + y' - \hat{a}_0(a^*, y')) = 0,$$

   and notice that the equation solver will try to evaluate many times $\hat{a}_0$ off the grid. Here, we use the Chebishev approximation.

[4] The best polynomial approximation is given by the *minimax* polynomial (see Judd, chapter 6) which has the property that the sign of the error term should alternate between points (the so-called Equioscillation Theorem).

[5] To understand this transformation, note that the reverse transformation is

$$x_k = 2\frac{a_k + \phi}{a_{\max} + \phi} - 1.$$

(a) Choose an order $N < M$ for the Chebishev polynomials. Compute the Chebishev coefficients

$$\kappa_p^0 = \frac{\sum_{k=1}^M \hat{a}_0\left(a_k, y_j\right) T_p\left(x_k\right)}{\sum_{k=1}^M T_p\left(x_k\right)^2},\ p = 0, ..., N \tag{4}$$

(b) Use the coefficients κ_p^0 for the evaluation outside the grid as follows:

$$\hat{a}_0\left(a^*, y'\right) = \sum_{p=0}^N \kappa_p^0 T_p\left(2\frac{a^* + \phi}{a_{\max} + \phi} - 1\right).$$

(c) If the solution of the nonlinear equation in (2) is a^*, then set $a_0'\left(a_i, y_j\right) = a^*$ and iterate on a new grid point from step 4.

5. To check convergence, in this case, it is better to compare iteration after iteration the Chebishev coefficients in order to get a sense of how globally distant are the policy functions obtained as a solution in successive iterations. For example, at iteration n, one could use criterion

$$\min_p \left|\kappa_p^n - \kappa_p^{n-1}\right| < \varepsilon.$$

As explained, the Chebishev approximation can be extremely good. It is smooth and differential everywhere, however keep in mind that it may not preserve concavity.

2.4 The Endogenous Grid Method

This is a newly developed method that, when it can be applied (not always), is much faster than the traditional method just described because it does not require the use of a nonlinear equation solver. The essential idea of the method is to construct a grid on a', next period's asset holdings, rather than on a, as is done in the standard algorithm. The method also requires the policy function for consumption to be at least weakly monotonic in current asset holdings.

Consider the above income fluctuation problem and write the Euler equation as follows:

$$u_c\left(c\left(a, y\right)\right) \geq \beta R \sum_{y' \in Y} \pi\left(y'|y\right) u_c\left(c\left(a', y'\right)\right).$$

Equality holds if $a' > -\phi$. We now solve for the consumption policy instead that for the next-period assets policy, i.e. we look for a decision rule for consumption $c\left(a, y\right)$, which is an invariant function of the states that satisfies the Euler Equation and that does not violate the borrowing constraint. We are going to start from a guess $\hat{c}_0(a, y)$ and will iterate on the Euler Equation until the decision rule for consumption that we solve for is essentially identical to the one in the previous iteration. The algorithm involves the following steps:

1. Construct a grid on (a, y) where $a \in G_A = \{a_1, ..., \bar{a}\}$ with $a_1 = -\phi$ and $y \in Y = \{y_1, ..., y_N\}$.

2. Guess a policy function $\hat{c}_0(a_i, y_j)$. A good initial guess is to set $\hat{c}_0(a_i, y_j) = ra_i + y_j$ which is the solution under quadratic utility if income follows a random walk.

3. Here's the key difference. Instead of iterating over pairs of $\{a_i, y_j\}$, we iterate over pairs $\{a'_i, y_j\}$. Fix y_j and iterate over all values of a'_i on the grid. For any pair $\{a'_i, y_j\}$ on the mesh $G_A \times Y$ construct the RHS of the Euler equation [call it $B(a'_i, y_j)$]

$$B(a'_i, y_j) \equiv \beta R \sum_{y' \in Y} \pi(y'|y_j) u_c(\hat{c}_0(a'_i, y'))$$

   where the RHS of this equation uses the guess $\hat{c}_0$.

4. Use the Euler equation to solve for the value $\tilde{c}(a'_i, y_j)$ that satisfies

$$u_c(\tilde{c}(a'_i, y_j)) = B(a'_i, y_j)$$

   and note that it can be done analytically, i.e. for $u_c(c) = c^{-\gamma}$ we have $\tilde{c}(a'_i, y_j) = [B(a'_i, y_j)]^{-\frac{1}{\gamma}}$. This is the step where the algorithm becomes much more efficient than the traditional one. First, we do not require a nonlinear equation solver, which takes a lot of computing time and could introduce numerical instabilities. Second, we only compute the expectation in step 3 once. In the traditional algorithm the expectation is computed within the nonlinear equation solver multiple times.

5. From the budget constraint, solve for $a^*(a'_i, y_j)$ such that

$$\tilde{c}(a'_i, y_j) + a'_i = Ra^*_i + y_j$$

   which implicitly gives the function $c(a^*_i, y_j) = \tilde{c}(a'_i, y_j)$. $a^*(a'_i, y_j)$ is the value of assets today that would lead the consumer to have a'_i assets tomorrow if his income shock was y_j today. Note that this function in general is not defined on the grid points of G_A. This is the endogenous grid and it changes on each iteration.

6. Let a_1^* be the value of asset holdings that induces the borrowing constraint to bind next period, i.e., the value for a^* that solves that equation at the point a_1', the lower bound of the grid.

7. Now we need to update our guess. To get the new guess $\hat{c}_1(a_i, y_j)$ on grid points $a_i > a_1^*$ we can use simple interpolation methods using values for $\left\{c\left(a_n^*, y_j\right), c\left(a_{n+1}^*, y_j\right)\right\}$ on the two most adjacent values $\left\{a_n^*, a_{n+1}^*\right\}$ that include the given grid point a_i. It is possible that some points a_i are beyond a_N^*, the upper bound of the endogenous grid. Then, we just use linear interpolation methods to obtain a new guess of the policy $\hat{c}_1$ on those grid points. To define the new guess of the consumption policy function on grid values $a_i < a_1^*$, then we use the budget constraint

$$\hat{c}_1(a_i, y_j) = Ra_i + y_j - a_1'$$

since we cannot use the Euler equation as the borrowing constraint is binding for sure next period. The reason is that we found it was binding at a_1^*, therefore a fortiori it will be binding for $a_i < a_1^*$.

8. Check convergence by comparing $\hat{c}_{n+1}(a_i, y_j)$ to $\hat{c}_n(a_i, y_j)$. For example declare convergence at iteration $n+1$ when

$$\max_{i,j}\left\{\left|\hat{c}_{n+1}(a_i, y_j) - \hat{c}_n(a_i, y_j)\right|\right\} < \varepsilon$$

for some small ε which determines the degree of tolerance in the solution algorithm.

3 Simulation

Once we solved for the policy function $a'(a, y)$ and obtained the consumption function $c(a, y)$ residually from the budget constraint (or viceversa if using the endogenous grid method), we can simulate paths for the variables of interest and compute statistics like the saving rate, the variance of consumption, the correlation between consumption change and income change, etc... This also allows us to do comparative statics exercises, i.e., asking questions such as what happens to average savings when uncertainty rises? Or, how do savings respond to changes in the interest rate? And so on.

One key step in the simulation is drawing paths of the income process. Suppose, for simplicity that we have a 2-state Markov chain with transition probabilities $\{\pi_{LL}, \pi_{LH}, \pi_{HL}, \pi_{HH}\}$ over two realizations (y_L, y_H). Then, we implement the following algorithm:

1. Draw a long sequence of T realizations from a uniform distribution on $[0,1]$, i.e. $\{\alpha_1, \alpha_2, ..., \alpha_T\}$. Note that $\pi_{sL} + \pi_{sH} = 1$, with $s \in \{L, H\}$.

2. Start from $y_0 = y_L$. If $\alpha_1 < \pi_{LL}$, then set $y_1 = y_L$. Otherwise, set $y_1 = y_H$. Suppose $\alpha_1 > \pi_{LL}$ so that $y_1 = y_H$. Next, if $\alpha_2 < \pi_{HL}$, then set $y_2 = y_L$. Otherwise, set $y_2 = y_H$. And continue until you have mapped the α-sequence into a y-sequence.

To simulate series for consumption and asset holdings, start from an initial condition (a_0, y_0) and, from the policy functions, compute

$$\begin{aligned} a_1 &= a'(a_0, y_0), a_2 = a'(a_1, y_1), ... \\ c_1 &= c(a_0, y_0), c_2 = c(a_1, y_1), ... \end{aligned}$$

One important trick to keep in mind when doing comparative statics (e.g., to study the effect of a rise in r on the saving rate) is that the statistics of interest (e.g., average savings) must be computed based on simulations run *with the same seed of the random number generator for the* α sequence, i.e., one has to use the same paths for the shocks in order to minimize the simulation variance.

It is useful to discard a certain number of initial observations (the so called burn-in period) to avoid dependence of the result from the initial condition (a_0, y_0) and allow draws to occur from the stationary distribution.

4 Checking Accuracy of the Numerical Solution

Suppose that we are using the piecewise linear interpolation. How can we determine when the solution is accurate enough that no more points in the grid are needed? Or, if we are using the Chebishev approximation method, how do we determine that increasing the order of the polynomial would not lead to any significant improvement in accuracy?

4.0.1 den Haan-Marcet Test

Den Haan and Marcet (1994) devise a simple test based on Hansen J test of overidentifying restrictions. We present here the test applied to the consumption-saving problem. The consumption Euler equation

$$u_c(c_t) = \beta R E_t [u_c(c_{t+1})]$$

implies that the residual

$$\varepsilon_{t+1} = u_c(c_t) - \beta R u_c(c_{t+1})$$

should not be correlated with any variable dated t and earlier, since the expectation at time t is conditional on everything observable up to then. Therefore, we should have, for every t

$$E_t[\varepsilon_{t+1} \otimes h(z_t)] = 0, \tag{5}$$

where the symbol $\otimes$ denotes element-by-element product. The term $h(z_t)$ is a $(r \times 1)$ vector of functions of z_t, which can include all the variables in the information set of the agent at time t like $\{y_j, c_j, a_j\}_{j=0}^{t}$. Clearly, this is true only for the *exact* solution. A badly approximated solution will not satisfy this property. This is the key idea of the test proposed by den Haan and Marcet.

One can obtain an estimate of the LHS of (5) through a simulation of length S of the model and the construction of

$$B_S = \frac{\sum_{t=1}^{S} \hat{\varepsilon}_{t+1} \otimes h(\hat{z}_t)}{S},$$

where $\hat{\varepsilon}_{t+1}$ and $\hat{z}_t$ are the simulated counterparts. It can be shown that, under mild conditions, $\sqrt{S}B_S \xrightarrow{d} N(0, V)$, and one can construct the appropriate quadratic form for the test statistics

$$SB_S'\hat{V}_S^{-1}B_S \xrightarrow{d} \chi_r^2$$

where $\hat{V}_S^{-1}$ is the inverse of some consistent estimate of V.

Some remarks are in order. First, the test does not require any knowledge of the true solution, which is a big advantage. Second, given a certain level of approximation in the solution, we can always find a number S large enough so that the approximation fails the accuracy test. This is not a problem when comparing solution methods, since one can fix the same S for both, or one can look for the smallest S such that the method fails the test and compare these thresholds. However, when we want to judge the accuracy of our unique solution, how large should S be? It's not clear: Den Haan and Marcet pick S to be 20 times larger than the typical sample period available.

4.0.2 Euler Equation Error Analysis

The numerically approximated Euler equation is

$$u_c(c_t) - \beta R E_t[u_c(c_{t+1})] \simeq 0.$$

The above equation will be very close to zero when evaluated on gridpoints, but it may be quite far from zero outside the grid.

One can define the relative approximation error ε_t as that value such that the equation holds exactly at t

$$u_c\left(c_t\left(1-\varepsilon_t\right)\right)=\beta R E_t\left[u_c\left(c_{t+1}\right)\right]$$

Let g denote the inverse function of the marginal utility, then we have

$$\varepsilon_t=1-\frac{g\left(\beta R E_t\left[u_c\left(c_{t+1}\right)\right]\right)}{c_t}.$$

For example an error of 0.01 means that the agent is making a mistake equivalent to \$1 for every \$100 consumed when choosing consumption and saving in period t. Santos (2000) proves that, the implied cost in terms of household welfare is the square of the Euler equation error, i.e., the error in the value function is of the order ε^2.

By running a long (but finite) simulation, one can compute a string of Euler equation errors and report either the maximum or the average of the errors in absolute value. This procedure may not explore the entire state space, but it will explore the region of the state space where the agent is more likely to find herself.

Alternatively, one can explore the entire state space by choosing a grid for assets different from the one used in the computation and calculate for each point a_i on this new grid, and for every y_j the Euler equation error $\varepsilon\left(a_i, y_j\right)$. Let $a^*\left(a, y\right)$ be the policy function obtained in the numerical solution. Then, compute:

$$u_c\left(\left(Ra_i+y_j-a^*\left(a_i, y_j\right)\right)\left(1-\varepsilon\left(a_i, y_j\right)\right)\right)-\beta R E_y\left[u_c\left(Ra^*\left(a_i, y_j\right)+y'-a^*\left(a^*\left(a_i, y_j\right), y'\right)\right)\right]=0.$$

Usually, absolute errors are reported in base 10 logarithm, so a value of -2 means an error of \$1 for every \$100, a value of -3 means an error of \$1 for every \$1,000, and so on. Acceptable average errors should be around values of 4 and above. See Aruoba, Fernandez-Villaverde and Rubio-Ramirez (2006) for a comparison of various global solution methods of the growth model based on Euler Equation error analysis.

1 Economies with Idiosyncratic Risk and Incomplete Markets: Stationary Equilibrium

We are interested in building a class of models whose equilibria feature a nontrivial endogenous distribution of income and wealth across agents in order to analyze questions such as:

1. What is the fraction of aggregate savings due to the precautionary motive?

2. How much of the observed wealth inequality can one explain through uninsurable earnings variation across agents?

3. What are the redistributional implications of various fiscal policies? How are inequality and welfare affected by such policies? It is key to have an equilibrium model to answer policy questions, because changes in policy affect equilibrium prices.

4. Can we generate a reasonable equity premium (i.e., excess return of stocks over a risk-free bonds), once we introduce a risky asset? This model has an additional source of uninsurable risk, which is idiosyncratic, and if this risk comoves with aggregate risk households require an even larger risk premium to hold stocks.

5. How large are the welfare losses from individual-level labor market risk (e.g., unemployment)?

The model is constructed around three building blocks: 1) the "income-fluctuation problem", 2) the aggregate neoclassical production function, and 3) the equilibrium of the asset market. We focus on the *stationary equilibrium*, for now, i.e., an economy without aggregate shocks.

Income fluctuation problem– This is the problem we studied in the previous chapter. Individuals are subject to exogenous income shocks. These shocks are not fully insurable because of the lack of a complete set of Arrow-Debreu contingent claims. There is only a risk-free asset (i.e., and asset with non-state contingent rate of return) in which the individual can save/borrow, and that the individual faces a borrowing (liquidity) constraint. A continuum of such agents subject to different shocks will give rise to a wealth distribution. Integrating wealth holdings across all agents will give rise to an *aggregate supply of capital.*

Aggregate production function: profit maximization of the competitive representative firm operating a CRS technology will give rise to an *aggregate demand for capital.*

Equilibrium in the asset market: When we let demand and supply interact in an asset market, an *equilibrium interest rate* will arise endogenously. Notice that if a full set of Arrow-Debreu contingent claims were available, the economy would collapse to a representative agent model with a stationary amount of savings such that $(1+r)\beta = 1$. With uninsurable risk, the supply of savings is larger (r is lower) because of precautionary saving, and consequently $(1+r)\beta < 1$. We like this because we know that this is a necessary condition for the income-fluctuation problem to have a bounded consumption sequence as solution.

1.1 The Economy

Demographics: the economy is populated with a continuum of measure one of infinitely lived, ex-ante identical agents.

Preferences: the individual has time-separable preferences over streams of consumption

$$U(c_0, c_1, c_2, ...) = E_0 \sum_{t=0}^{\infty} \beta^t u(c_t),$$

where the period utility function $u(c_t)$ satisfies $u' > 0, u'' < 0$ and the discount factor $\beta \in (0,1)$. The expectation is over future sequences of shocks, conditional to the realization at time 0. The individual supplies labor inelastically.

Endowment: each individual has a stochastic endowment of efficiency units of labor $\varepsilon_t \in E \equiv \left\{\varepsilon^1, \varepsilon^2, ..., \varepsilon^{N-1}, \varepsilon^N\right\}$. The shocks follow a Markov process with transition probabilities $\pi(\varepsilon', \varepsilon) = \Pr(\varepsilon_{t+1} = \varepsilon' \mid \varepsilon_t = \varepsilon)$. Shocks are *iid* across individuals. We assume a law of large numbers to hold, so that $\pi(\varepsilon', \varepsilon)$ is also the fraction of agents in the population subject to this particular transition.[1] We assume that the Markov transition is well-behaved, so there is a unique invariant distribution $\Pi^*(\varepsilon)$. As a result, the aggregate

[1] There are some tricky issues with laws of large numbers in this setting. Please, refer to Judd (1985) and Uhlig (1996) for a discussion.

endowment of efficiency units

$$H_t = \sum_{i=1}^{N} \varepsilon_i \Pi^* (\varepsilon_i), \text{ for all } t$$

is constant over time, i.e., there is no aggregate uncertainty. Note in particular, that H_t is exogenously determined.

Budget constraint: For individual i at time t, the budget constraint reads

$$c_t + a_{t+1} = (1 + r_t) a_t + w_t \varepsilon_t,$$

where c_t is current consumption, a_{t+1} is next period wealth, $(1 + r_t)$ is the gross interest rate and w_t is the wage rate at period t. Wealth is held in the form of a one-period risk-free bond whose price is one and whose return, next period, will be $(1 + r_{t+1})$, independently of the individual state (i.e., r_{t+1} does not depend on the realization of ε_{t+1}). In this sense, the asset a is non state-contingent.

Liquidity constraint: At every t, agents face the borrowing limit

$$a_{t+1} \geq -b$$

where b is exogenously specified. Alternatively, we could assume agents face the "natural" borrowing constraint, which is the present value of the lowest possible realization of her future earnings.

Technology: The representative competitive firm produces with CRS production function $Y_t = F(K_t, H_t)$ with decreasing marginal returns in both inputs and standard Inada conditions. Physical capital depreciates geometrically at rate $\delta \in (0, 1)$.

Market structure: final good market (consumption and investment goods), labor market, and capital market are all competitive.

Aggregate resource constraint: The aggregate feasibility condition in this economy reads:

$$F(K_t, H_t) = C_t + I_t = C_t + K_{t+1} - (1 - \delta) K_t,$$

where capital letters denote aggregate variables.

1.2 Stationary Equilibrium

We are now ready to define the stationary equilibrium of this economy through the concept of *Recursive Competitive Equilibrium* (RCE). Most of the requirement of this RCE definition will be standard (agents optimize, markets clear). Moreover, in the stationary equilibrium of this economy we require the distribution of agents across states to be invariant.[2] This probability measure will permanently reproduce itself. It is in this sense that the economy is in a rest-point, i.e., a steady state.

1.2.1 Some Mathematical Preliminaries

The individual is characterized by the pair (a, ε) –the individual states. Let λ be the distribution of agents over states. We would like this object to be a *probability measure,* so we need to define an appropriate mathematical structure. Let $\bar{a}$ be the maximum asset holding in the economy, and for now assume that such upper bound exists. Define the compact set $A \equiv [-b, \bar{a}]$ of possible asset holdings, and the countable set E as above.

Let the the state space S be the Cartesian product $A \times E$, and let the σ-algebra Σ_s be defined as $B_A \otimes P(E)$ where B_A is the Borel sigma-algebra on A and and $P(E)$ is the power set of E. The space (S, Σ_s) is a measurable space. Let $\mathcal{S} = (\mathcal{A} \times \mathcal{E})$ be the typical subset of Σ_s. For any element of the sigma algebra $\mathcal{S} \in \Sigma_s$, $\lambda(\mathcal{S})$ is the measure of agents in the set $\mathcal{S}$.

How can we characterize the way individuals transit across states over time? I.e. how do we obtain next period distribution, given this period distribution? We need a transition function. Define $Q((a, \varepsilon), \mathcal{A} \times \mathcal{E})$ as the (conditional) probability that an individual with current state (a, ε) transits to the set $\mathcal{A} \times \mathcal{E}$ next period, formally $Q : S \times \Sigma_s \to [0, 1]$, and

$$Q((a, \varepsilon), \mathcal{A} \times \mathcal{E}) = I_{\{a'(a,\varepsilon) \in \mathcal{A}\}} \sum_{\varepsilon' \in \mathcal{E}} \pi(\varepsilon', \varepsilon) \tag{1}$$

where $I_{\{.\}}$ is the indicator function, and $a'(a, \varepsilon)$ is the optimal saving policy. Then Q is our transition function and the associated T^* operator yields

$$\lambda_{n+1}(\mathcal{A} \times \mathcal{E}) = T^*(\lambda_n) = \int_{A \times E} Q((a, \varepsilon), \mathcal{A} \times \mathcal{E}) \, d\lambda_n. \tag{2}$$

[2] However, individuals move up and down in the earnings and wealth distribution, so "social mobility" can be meaningfully defined. Recall that with complete markets, there is no social mobility: initial rankings persist forever.

where I have used the notation $d\lambda_n$ as short for $\lambda_n(da, d\varepsilon)$.

Let us now re-state the problem of the individual in recursive form, i.e. through dynamic programming

$$\begin{aligned} v(a,\varepsilon;\lambda) &= \max_{c,a'}\left\{u(c)+\beta\sum_{\varepsilon'\in E}\pi(\varepsilon',\varepsilon)v(a',\varepsilon';\lambda)\right\} \\ & \quad s.t. \\ c+a' &= R(\lambda)\,a+w(\lambda)\,\varepsilon \\ a' &\geq -b \end{aligned} \tag{3}$$

where, for clarity, we have made explicit the dependence of prices from the distribution of agents (although, strictly speaking this dependence is *redundant* in a stationary environment and it can be omitted since the probability measure λ is constant). We are now ready to proceed to the definition of equilibrium.

1.2.2 Definition of Stationary RCE

A **stationary recursive competitive equilibrium** is a value function $v : S \to \mathbb{R}$; policy functions for the household $a' : S \to \mathbb{R}$, and $c : S \to \mathbb{R}_+$; firm's choices H and K; prices r and w; and, a stationary measure λ^* such that:

- given prices r and w, the policy functions a' and c solve the household's problem (3) and v is the associated value function,
- given r and w, the firm chooses optimally its capital K and its labor H, i.e., $r+\delta = F_K(K,H)$ and $w = F_H(K,H)$,
- the labor market clears: $H = \int_{A\times E}\varepsilon d\lambda^* = \sum_{i=1}^{N}\varepsilon_i\Pi^*(\varepsilon_i)$,
- the asset market clears: $K = \int_{A\times E} a'(a,\varepsilon)d\lambda^*$,
- the goods market clears:[3] $\int_{A\times E} c(a,\varepsilon)d\lambda^* + \delta K = F(K,H)$,
- for all $(\mathcal{A}\times\mathcal{E}) \in \Sigma_s$, the invariant probability measure λ^* satisfies

  $$\lambda^*(\mathcal{A}\times\mathcal{E}) = \int_{A\times E} Q((a,\varepsilon), \mathcal{A}\times\mathcal{E})\, d\lambda^*,$$

  where Q is the transition function defined in (1).

[3] This condition is redundant by Walras law.

1.3 Existence and Uniqueness of the Stationary Equilibrium

Characterizing the conditions under which an equilibrium exists and is unique boils down, like in every general equilibrium model, to show that the excess demand function (i.e., a function of the price) in each market is continuous, strictly monotone and intersects "zero". Equilibrium in the labor market is trivial: aggregate labor supply is exogenous and labor demand is strictly decreasing in wages. By Walras law, if we prove that the equilibrium in the asset market exists and is unique, we are done.

Demand for capital– Consider first the aggregate demand of capital. From the optimal choice of the firm, we obtain

$$K(r) = F_k^{-1}(r + \delta).$$

Note that for $r = -\delta$, then $K \to +\infty$, while for $r \to +\infty$, $K \to 0$. Moreover, the supply of capital is a continuous, strictly decreasing function of the interest rate r. For example, if $F(K, H) = K^\alpha H^{1-\alpha}$, then

$$K(r) = \left(\frac{\alpha H}{\delta + r}\right)^{\frac{1}{1-\alpha}}.$$

Supply of capital– If we could show that the aggregate supply function

$$A(r) = \int_{A \times E} a'(a, \varepsilon; r) d\lambda_r^*$$

is continuous in r and crosses the aggregate demand function, then we would prove existence. Suppose first $(1 + r)\beta = 1$, i.e. $r = \frac{1}{\beta} - 1$, then we know by the super-martingale converge theorem that the aggregate supply of assets goes to infinity, i.e. $A(\frac{1}{\beta} - 1) \to \infty$. For $r = -1$ the individual would like to borrow until the limit, as every unit of capital saved will vanish, so $A(-1) \to -b$.[4]

Thus, if $A(r)$ is continuous, it will cross the aggregate demand and we would have an existence result.

Standard results in dynamic programming ensures us that, if u is continuous, $u' > 0$ and $u'' < 0$, the solution to the household problem is unique and that the policy function $a'(a, \varepsilon; r)$ is continuous in r (by the Theorem of the Maximum). To establish continuity

[4]For values of the interest rate $r < 0$, the agent may still want to hold some wealth for precautionary reasons. It depends on the exact parametrization.

of λ_r^* in r we would like to apply Theorem 12.13 in SLP. However, the Theorem requires existence and uniqueness of an invariant distribution for given r.

To establish the existence and uniqueness of the invariant distribution, from Theorem 12.12 in SLP, we know that we need to verify four properties: compactness of the state space and Q with Feller property imply existence. Q satisfying monotonicity and the monotone-mixing condition (MMC) yields uniqueness. We verify these properties one at the time.

- *Compactness:* When $\beta(1+r) < 1$ and preferences display decreasing absolute risk aversion (or asymptotitcally bounded relative risk aversion) we showed that an upper bound $\bar{a}$ on the asset space exists–recall our discussion in the previous chapter– so the state space is a compact subset of $\mathbb{R}^2$.

- *Feller property of Q:* The Feller property requires that the associated operator T maps continuous and bounded functions into themselves. For Q it is easily verified, because $a'(a,\varepsilon)$ is continuous and bounded since the domain of the asset space is compact. In particular, we can apply Theorem 9.14 in SLP which states that if E is countable and $P(E)$ is the sigma-algebra on E, A is compact and a' is continuous, then Q has the Feller property.

- *Monotonicity of Q:* Monotonicity of Q requires that for every increasing function f, the function Tf is also increasing. Suppose that the Markov process has two possible states, $E \equiv (\varepsilon_L, \varepsilon_H)$. Assume that $\pi(\varepsilon_H, \varepsilon_H) \geq \pi(\varepsilon_H, \varepsilon_L)$ and $\pi(\varepsilon_L, \varepsilon_L) \geq \pi(\varepsilon_L, \varepsilon_H)$, i.e. the Markov chain is monotone.[5] Recall that $a'(a,\varepsilon)$ is an increasing function. Then it is easy to see that Q is monotone. Let $f(a', \varepsilon')$ be an increasing function. Applying the definition in SLP, we want to show that the conditional expectation
$$h(a,\varepsilon) = Tf = \sum_{\varepsilon' \in E} \int_A f(a', \varepsilon') Q((a,\varepsilon), da' \times \varepsilon')$$

[5] A Markov chain ε is monotone iff, for any increasing function $f(\varepsilon')$ the conditional expectation
$$\mathbb{E}[f(\varepsilon')|\varepsilon] = \sum_{\varepsilon'} f(\varepsilon') \pi(\varepsilon', \varepsilon) d\varepsilon'$$
is increasing in ε. Note that this restriction on the Markov chain, with multiple states, corresponds to positive autocorrelation in the income process.

is monotonically increasing. This is easy to see. Intuitively, a higher pair (a, ε) increases the probability of being in state $(a', \varepsilon') > (a, \varepsilon)$ next period. Thus, more weight is put on the region of the domain where f is high (since f is increasing).

- *MMC:* Suppose the household starts from $(\bar{a}, \varepsilon_{\max})$ and receives a long stream of the worst realization of the shock $\varepsilon_{\min}$. If the ε process is stationary (i.e., mean reverting) then, she will keep decumulating wealth until she reaches some neighborhood of the lower bound. The reason for decumulation is that the household knows that this income realization is well below average, his permanent income is higher and consumption is dictated by permanent income. Suppose now that the household starts with $(-b, \varepsilon_{\min})$ and receives a long stream of the best shock $\varepsilon_{\max}$. Then, she will accumulate wealth until she reaches some neighborhood of the upper bound. The reason for accumulation is similar: the household realizes that this good realization is "transitory" and her expected income is below the current income, so she saves a fraction of these lucky draws.

At this point, we can apply Theorem 12.13 in SLP. This proves existence of the equilibrium.

If, in addition, we could show that $A(r)$ is strictly increasing, we would prove uniqueness. Unfortunately, there are no results on the monotonicity of the aggregate supply of capital with respect to r, so uniqueness is never guaranteed. Intuitively, a higher r has both income and substitution effects on savings: the relative dominance between the two could switch at a certain level of assets, so $a'(a, \varepsilon; r)$ may not be monotone. Even if we make sure that preferences are such that one of the two effects always dominates, it is very hard to assess what a change in r does to the distribution of assets.

One can use the computer to plot aggregate supply as a function of the interest rate on a fine grid for a reasonably large range of values of r to check its slope.

FIGURE

This figure plots the equilibrium in the asset market graphically.

1.4 An Algorithm for the Computation of the Equilibrium

How do we compute, in practice, this equilibrium? The algorithm that can be used is a fixed point algorithm over the interest rate.

1. Fix an initial guess for the interest rate $r^0 \in \left(-\delta, \frac{1}{\beta} - 1\right)$, where these bounds follow from our previous discussion. The interest rate r^0 is our first candidate for the equilibrium (the superscript denotes the iteration number).

2. Given the interest rate r^0, obtain the wage rate $w(r^0)$ using the CRS property of the production function (recall that H is given exogenously with inelastic labor supply).

3. Given prices $(r^0, w(r^0))$, you can now solve the dynamic programming problem of the agent (3) to obtain $a'(a, \varepsilon; r^0)$ and $c(a, \varepsilon; r^0)$. We described several solution methods.

4. Given the policy function $a'(a, \varepsilon; r^0)$ and the Markov transition over productivity shocks $\pi(\varepsilon', \varepsilon)$, we can construct the transition function $Q(r^0)$ and, by successive iterations over (2), we obtain the fixed point distribution $\lambda(r^0)$, conditional on the candidate interest rate r^0.

   (a) The easiest method to implement this step, in practice, is by simulation of a large number of households N (say 10,000) and track them over time, like survey data do. Initialize each individual in the sample with a pair (a_0, ε_0) and, using the decision rule $a'(a, \varepsilon)$ and a random number generator that replicates the Markov chain $\pi(\varepsilon', \varepsilon)$, update their pair of individual states at every period t.

   (b) For every t, compute a set of cross-sectional moments J_t^N which summarize the distribution of assets (e.g., mean, variance, various percentiles). Stop when J_t^N and J_{t-1}^N are close enough. At that point, the cross-sectional distribution has converged. We know that for any given r, a unique invariant distribution will be reached for sure.

5. Compute the aggregate demand of capital $K(r^0)$ from the optimal choice of the firm who takes as given r^0, i.e.

$$K(r^0) = F_k^{-1}(r^0 + \delta)$$

6. Compute the integral

$$A\left(r^{0}\right)=\int_{A\times E}a'(a,\varepsilon;r^{0})d\lambda(a,\varepsilon;r^{0})$$

which gives the aggregate supply of assets. Clearly, this can be easily done by exploiting the model-generated data from the invariant distribution obtained in step 4. It's just an average over the artificial cross-section of households.

7. Compare $K\left(r^{0}\right)$ with $A\left(r^{0}\right)$ to verify the asset market clearing condition. If $A\left(r^{0}\right)>(<)\,K\left(r^{0}\right)$, then the next guess of the interest rate should be lower (higher), i.e. $r^{1}<(>)\,r^{0}$. To obtain the new candidate r^{1} a good choice is, for example,

$$r^{1}=\frac{1}{2}\left\{r^{0}+\left[F_{K}(A\left(r^{0}\right),H)-\delta\right]\right\}$$

This method is called bi-section method. Note that r^{0} and $F_{K}(A\left(r^{0}\right),H)-\delta$ are, by construction, on opposite sides of the steady-state interest rate r^{*}.

8. Update your guess to r^{1} and go back to step 1). Keep iterating until one reaches convergence of the interest rate, i.e., until

$$\left|r^{n+1}-r^{n}\right|<\varepsilon,$$

for ε small. Typically, you need less than 10 iterations for convergence.

9. All the equilibrium statistics of interest, like aggregate savings, inequality measures, etc. can be computed from the simulated data in step 4.

1.5 Calibration of the model

To solve the model numerically, one needs first to choose values for the parameters. Here's some guidance on how to pick values. Suppose you set the model's period to one year.

Technology– With Cobb-Douglas production function, pick the capital share α to be equal to 1/3. Set the depreciation rate δ to 6%.

Preferences– Typically, we work with CRRA utility. Let γ be the coefficient of relative risk aversion. Typical values, in this type of applications, range between 1 and 5. As for the discount rate β, it should be chosen so that the aggregate wealth-income ratio replicates the one for the U.S. economy which is around 3.[6] However, this means that the parameter is not calibrated externally, but internally which is computationally painful. So, you could do the following. Imagine that you're in complete markets, then you know that

$$\begin{aligned}
\alpha K^{\alpha-1}H^{1-\alpha}-\delta &= \left(\frac{1}{\beta}-1\right)\Rightarrow\alpha\left(\frac{Y}{K}\right)-\delta=\frac{1}{\beta}-1\Rightarrow \\
\beta &= \frac{1}{1+\alpha\left(\frac{Y}{K}\right)-\delta}=\frac{1}{1+0.33\,(0.33)-0.06}=0.951.
\end{aligned}$$

In other words, this value for β would give you a K/Y ratio of 3 in complete markets. With incomplete markets the same β gives you a slightly larger capital-output ratio because of the extra precautionary capital accumulation, so one should set β slightly smaller.

Labor income process– We want to calibrate the labor endowment shocks to replicate the typical dynamics of individual earnings in the U.S. economy. The right source of data for this purpose are panel-data with information on labor income, such as the *Panel Study of Income Dynamics.* A decent approximation to U.S. individual earnings dynamics is an AR(1) process like

$$\ln y_t=\rho\ln y_{t-1}+\varepsilon_t,\qquad \text{with } \varepsilon_t\sim N\left(0,\sigma_\varepsilon\right)$$

where the autocorrelation coefficient is $\rho=0.95$ and the standard deviation of the shocks is $\sigma_\varepsilon=0.20$ at a annual frequency. More sophisticated estimates include a transitory component to capture measurement error, as well as less persistent shocks, and a fixed individual component to capture the effect of education, ability, etc.

Borrowing constraint– If the natural borrowing constraint is not a good choice for the problem at hand, one could calibrate the borrowing constraint in order to match, say, the fraction of agents with negative net worth which is around 13% in the U.S. economy. The difficulty is that this strategy requires, again, an internal calibration.

1.6 Notes

This class of model with a continuum if agents facing individual income shocks and trading a risk-free asset (money, in the original model) was introduced by Bewley (1986).

[6] To be precise, the size of the K/Y ratio depends on whether we include residential capital. If we do, then the K/Y ratio is closer to 4.

Laitner (1992) studies one version of this economy with altruism. Huggett (1993) analyzes the equilibrium in an endowment economy where agents trade an asset in zero net supply. Aiyagari (1994) generalizes the model to a production economy with an aggregate production function (the model described here). Huggett (1996) presents an OLG version of this model. The chapter by Rios-Rull in the book edited by Cooley (1995) contains a careful description of how to compute equilibria in this class of economies. So does Section 7 in Herr and Maussner.

1 Some Applications of Bewley Models

1.1 Precautionary Savings

The previous graph allows to examine the amount of savings that would occur in the economy with full insurance. Under full insurance, the Euler equation implies $(1+r)\beta = 1$, hence the supply of capital is infinitely elastic at $r = 1/\beta - 1$. The point where this horizontal line crosses $K(r)$ represents the stock of capital of the neoclassical deterministic growth model, K^{FI}. The magnitude $\left(K^{*} - K^{FI}\right)$ is the amount of aggregate capital accumulated for self-insurance. To express this comparison in terms of saving rates, note that with Cobb-Douglas production function with capital share α,

$$r + \delta = \alpha \left(\frac{Y}{K}\right) = \alpha\delta \left(\frac{Y}{\delta K}\right) = \frac{\alpha\delta}{s} \Rightarrow s = \frac{\alpha\delta}{r+\delta},$$

hence there is a one-to-one mapping between the equilibrium interest rate and the saving rate: differences between r^{*} and $1/\beta - 1$ translate directly into the precautionary saving rate.

1.1.1 How much saving for self-insurance in the US?

Aiyagari (1994) calibrates the model to replicate certain key facts of the U.S. economy.[1] He finds that with log utility and *iid* shocks, the precautionary saving rate is approximately zero, i.e. $r^{*} \simeq 1/\beta - 1$ (probably a lower bound). The reason is that agents with low risk aversion who face income shocks with low persistence are not too concerned about the missing insurance markets and do accumulate little capital for self-insurance. With $CRRA$ utility, risk-aversion parameter equal to 5, and autocorrelation of the income shocks of 0.9, the size of precautionary savings 14% of aggregate output (probably an upper bound). These households are much more concerned about self-insurance since with high persistence income can remain low for a long time, plus for them consumption fluctuations are very costly.

These are both extreme parameterizations. A reasonable estimate, in a model economy calibrated to the U.S. with risk aversion around 2, the precautionary saving rate would be

[1] One attractive feature of this class of neoclassical growth models with idiosyncratic earnings risk is that they can be calibrated exactly as in the same way as the representative agent model, except for the stochastic process that governs earnings risk. This process is parametrized by using empirical microeconomic studies on earnings dynamics based on panel data surveys, like the Panel Study of Income Dynamics (PSID).

5% of aggregate income, i.e. roughly 1/4 of total aggregate savings. Note the importance of equilibrium considerations here: in the income fluctuation problem, one can always set r close enough to $1/\beta - 1$ in order to generate any desired amount of savings (even $+\infty$). The equilibrium model imposes more discipline because, given demand for capital, there is only one value of the interest rate consistent with the consumption/saving decisions of the households.[2]

1.1.2 Comparative Statics on Precautionary Saving

The Figure depicting the equilibrium is helpful to perform some comparative statics on the equilibrium. We are interested in the effects on the borrowing constraint b, risk aversion, the persistence of the shock and the variance of the shock.

Borrowing limit– Suppose we increase b, i.e. we slacken the liquidity constraint and increase the maximum amount that can be borrowed by the individual. Graphically, the asset supply curve $A(r)$ shifts upward, with $K(r)$ constant, which leads to a rise in the interest rate and a reduction of precautionary savings. The interest rate increases because more individuals have negative wealth, so the supply of capital falls at any given r. Note here an important point: agents can both save and borrow for self-insurance. The availability of a generous borrowing limit reduces the need for precautionary saving.

Risk aversion- The $K(r)$ curve is unaffected by changes in preferences. If we raise risk aversion, the $A(r)$ curve shifts downward: individuals are more concerned about consumption smoothing, so they cumulate higher buffer-stock savings: for any given r, $A(r)$ is larger. This leads to a lower equilibrium interest rate.

Changes in the income process– Suppose we increase the variance of the uninsurable income shock. $K(r)$ is unchanged, but the supply of capital $A(r)$ would go up (curve shifts down), as individuals cumulate more savings to cope with the higher uninsurable uncertainty of their income. Increasing the persistence of the shock has a similar effect.

[2] In particular, the logic that the closer is $1+r$ to $1/\beta$, the higher are savings is reversed in equilibrium because $(1+r)\beta = 1$ is the complete-markets benchmark which has the lowest saving rate (no saving for self-insurance).

1.2 Understanding Wealth Inequality

The Bewley models contain a theory of consumption and wealth inequality. In the model, agents are ex-ante equal and become different as time goes by due to variation in the realization of their income shocks. It is a theory of inequality based largely on luck. In response to shocks, they choose optimally how much to consume and how much to save. Hence, different paths of shocks induce different levels of consumption and wealth across agents. Note that, in the absence of endogenous labor supply, the model has nothing to say on earnings (wages times hours worked) inequality.

For a full description of facts on earnings and wealth inequality, one should refer to a recent paper by Budria, Diaz-Gimenez, Quadrini and Rios-Rull (2002). The following table, reproduced from the paper above, provides the key statistics for the US (from the *Survey of Consumer Finances*, 1998). The key fact to observe is that wealth is much more unequally distributed than earnings. Both distributions are skewed (mean>median), but the wealth distribution much more so that the earnings distribution: the top 1% of the wealth distribution owns 30% of the US wealth.

U.S. SCF (1998). Values of earnings and wealth are in thousands of 1998 $

	Mean	$\frac{Mean}{Median}$	Gini	CV	Q1	Q3	Q5	Top 1%	Top 5%	Bottom 5%	$Share$ $of\ top$ 1%
Earnings	21.1	1.57	0.61	2.65	-0.7	20.7	101.9	491	136	0.0	7.5%
Wealth	47.4	4.03	0.80	6.53	-2.3	51	770	5,988	1,150	-4.7	31%

After the model is calibrated and simulated, we can use it as a measurement tool for following question: how far can we go in explaining wealth inequality when idiosyncratic earnings shocks are the only source of heterogeneity among households? The typical answer is that the standard model generates too much asset holdings at the bottom and too little at the top: the Gini generated by the model economy is around 0.4 –much smaller that the data value 0.8.

This standard model needs to be modified to 1) introduce an extra incentive for the rich to accumulate capital, 2) reduce the incentives for the poor to save for self-insurance purposes.

Inequality at the bottom– Modelling carefully the welfare state goes a long way in generating the right amount of asset holdings at the bottom. Once we introduce public insurance schemes (e.g., social security, housing benefits, child benefits, unemployment insurance), the incentives for private self-insurance are much reduced because some of these benefits are means-tested. See, for example, Hubbard, Skinner and Zeldes (1995).

Inequality at the top– To improve the quantitative explanation of inequality at the top, several alternatives have been pursued. 1) Quadrini (2000) explores the role of entrepreneurship. Implicitly, entrepreneurs have a higher return on their investment, hence a stronger incentive to accumulate. Empirically, a large fraction of wealth at the top is held by entrepreneurs. 2) Krusell and Smith (1997) study heterogeneity in discount factors. A Markov process regulates transitions between two levels of patience $\{\beta_L, \beta_H\}$. In the β_H state, households are more patient and save more. Small differences in β lead to a jump in the wealth Gini. 3) De Nardi (2003) studies the role of bequest. If (rich) households have a stronger bequest motive than poor households, this represents an additional reason to save. 4) Castaneda, Diaz-Jimenez and Rios-Rull (2003) add to an otherwise standard income process a very high realization of earnings (roughly 200 times larger than the mean) which occurs with a very low probability. They argue that income data are top-coded, so one does not observe these realizations in the data, even though they exist.

1.3 A Bewley Model with Entrepreneurship

We have argued that, empirically, a large fraction of wealth at the top is held by entrepreneurs. Here we want to write down a model where we have both entrepreneurs (individuals owning their own business) and workers (individuals working for someone else) and argue that it can replicate the upper tail in the wealth distribution. We follow Quadrini (2000) and Kitao (2008).

Demographics– Agents are infinitely lived. Every period agents choose an occupation. Entrepreneurs run their own business, and workers supply labor in the market. Entrepreneurs can manage one project which combines her managerial ability, capital and labor.

Endowments– Each agent is endowed with labor productivity $\varepsilon \in E$ and entrepreneurial ability $\theta \in \Theta$. The joint process is regulated by a Markov chain $\Gamma_{\varepsilon,\theta}$.

Preferences– They are standard, time-separable preferences defined over streams of consumption, with discount factor β

$$E_0 \sum_{t=0}^{\infty} \beta^t u(c_t)$$

Technology– There are two production sectors. The corporate sector is CRS with production $Y = F(K, H)$. Capital is rented by the corporate firms at the risk-free rate r and depreciates at rate δ. The non-corporate sector is composed by many entrepreneurs who run their own project according to the production function $y = f(k, n, \theta) = \theta k^{\nu_1} n^{\nu_2}$, with $\nu_1 + \nu_2 < 1$. This formulation mirrors the classic span of control model developed by Lucas (1978), i.e. entrepreneurial ability determines the size of the firm. Also, note that $(1 - \nu_1 - \nu_2)$ is the share of output retained by the entrepreneurs as rents or profits from his managerial skills.

Financial markets– Workers cannot borrow to finance consumption. Households with savings are indifferent between lending directly to the corporate sector and lending to the banking sector. Both sectors pay an interest rate r. The banking sector allocates funds to entrepreneurs. The banking sector is competitive and has an operation cost ϕ units of the final good for any unit of capital intermediated. Therefore, entrepreneurs can borrow to finance their project at rate $r_d = r + \phi$. Entrepreneurs can borrow only up to a fraction d of their assets. Hence, they can invest in their project up to $(1 + d)a$.

Problems of the worker and entrepreneur– The problem of the worker is written as

$$\begin{aligned}
V^w(a, \varepsilon, \theta) &= \max_{c, a', i} u(c) + \beta \{iE[V^w(a', \varepsilon', \theta')] + (1 - i) E[V^e(a', \varepsilon', \theta')]\} \\
& \quad s.t. \\
c + a' &= w\varepsilon + Ra \\
a' &\geq 0
\end{aligned}$$

and the problem of the entrepreneur is written as:

$$\begin{aligned}
V^e(a, \varepsilon, \theta) &= \max_{c, a', i} u(c) + \beta i \{E[V^w(a', \varepsilon', \theta')] + (1 - i) E[V^e(a', \varepsilon', \theta')]\} \\
& \quad s.t. \\
c + a' &= \pi(a, \varepsilon, \theta) \\
a' &\geq 0
\end{aligned}$$

where $\pi(a,\varepsilon,\theta)$ are profits after payments of factors of production and loans and are determined as:

$$\begin{aligned}
\pi(a,\varepsilon,\theta) &= \max_{k,n} f(k,n,\theta) + (1-\delta)k - (1+\tilde{r})(k-a) - w\max\{n-\varepsilon,0\} \\
& \quad s.t. \\
k &\leq (1+d)a \\
\tilde{r} &= \begin{cases} r & k \leq a \\ r+\phi & \text{if } k > a \end{cases}
\end{aligned}$$

Note that, if $a \geq k$, then the entrepreneur has some excess saving that she invests at the risk free rate while, if $k > a$, then she borrows the difference at rate $r + \phi$. Another important implication of the way we wrote the problem of the entrepreneur is that an entrepreneur cannot sell her efficiency units in the labor market, which amplifies the risk of entrepreneurship, in case θ turns out to be very low. A limit of this model is that entrepreneurs never make an operating loss within a period, as they can always choose $k = n = 0$ and earn the risk free rate on saving a.[3]

Why are entrepreneurs richer than workers? Individuals with high entrepreneurial ability have access to a saving/investment technology with higher return than workers (who save at rate r) thanks to this decreasing return technology f, and therefore they accumulate wealth faster.

1.4 Role of Redistributive Taxation

In this model-economy where some of the earnings risk is uninsurable because of market incompleteness, there could be scope for public insurance, i.e. government intervention through taxation and redistribution from the rich-lucky to the poor-unlucky.

In a model with exogenous labor supply, suppose that the government (as in a Ramsey-style optimal taxation problem) chooses a labor income tax τ and a lump-sum subsidy t in order to maximize the ex-ante welfare of the households. Clearly, the optimal tax rate in the Aiyagari economy would be $\tau = 1$. The government would tax away all income and redistribute equally across all agents. This policy would achieve the first-best because

[3] If there was a fixed entry cost of becoming entrepreneurs and a lower operating cost that must be paid every period, entrepreneurs may choose to stay in business even though, for one or more periods, $\pi < 0$. This will happen if θ is mean reverting.

taxation entails no distortions and no loss of efficiency and, at the same time, generates full insurance.

A more interesting economy is one with an endogenous margin of labor/leisure choice where there is a *trade-off between insurance and efficiency.* To evaluate this trade-off, we develop a variation of the benchmark Aiyagari economy with endogenous labor supply which follows Floden and Linde (2001).

Model– First of all, period utility is given by $u(c, l)$ i.e., we introduce leisure $l \in (0, 1)$ in order to have a margin where distortions matter. This means that we will have an optimal policy for labor supply $h(a, \varepsilon)$. Notice also that the agents might be using their elastic labor supply to self-insure. Take an agent who is liquidity constrained and has a low realization of the productivity shock: to keep his consumption high, he could intensify his labor supply.

The new budget constraint reads

$$c + a' = (1 + r)\, a + (1 - \tau)\, w\varepsilon h + t,$$

where τ is a flat earnings tax, and t is the lump-sum transfer of the government. Note that the tax scheme is progressive because the least productive agents pay less taxes but get the same transfer. In other words, the average tax rate faced by a household $(\tau - t/w\varepsilon h)$ is increasing in labor income $w\varepsilon h$.

With leisure, the new equilibrium condition in the labor market becomes

$$N = \int_{A\times E} \varepsilon h(a, \varepsilon) d\lambda^*$$

where N is aggregate labor demand and the RHS is aggregate labor supply.

The government budget constraint (balanced in equilibrium) reads

$$T = \tau w N,$$

where T denotes aggregate transfers (and in equilibrium it equals t).

The definition of the recursive competitive equilibrium for this economy is very similar to the benchmark case (the chief differences are the existence of a decision rule for leisure and of the balanced budget condition of the government).

Computation– Suppose the tax rate τ is a parameter of the problem. We guess $\{r^0, N^0\}$. Given r^0, if the production function is CRS, we get w^0. From the government budget constraint and from our guess of H^0, we recover t and we have all the inputs needed for solving the household problem. The rest of the computation algorithm is the same as before, with the caveat that we need to check two equilibrium conditions, asset market clearing

$$K^0\left(r^0, N^0\right) = \int_{A\times E} ad\lambda_{(r^0,N^0)}$$

and labor market clearing:

$$N^0 = \int_{A\times E} \varepsilon h(a, \varepsilon; r^0, N^0) d\lambda_{(r^0,N^0)}$$

Question– Floden-Linde ask the following question: what is the level of government redistribution that maximizes welfare? What welfare gains does such redistribution imply for individuals compared to the pure laissez-faire, no-redistribution benchmark?

In an economy with heterogeneous agents there is not a unique welfare function, it all depends on what weights are assigned to each type. Floden-Linde assume an equal-weight social welfare function, i.e. they solve

$$\max_{\tau} W(\tau) = \int_{A\times E} V^*(a, \varepsilon; \tau) d\lambda^*_{\tau},$$

where $V^*(a, \varepsilon; \tau)$ is the value function associated to the competitive equilibrium indexed by τ. It is a version of the Ramsey taxation problem with incomplete markets.

Intuitively, for low levels of redistribution, welfare is low because individuals have a large amount of undesired consumption fluctuations; for very high levels of taxes, consumption insurance is very high but at the same time heavy distortions on labor supply are imposed. So, there will be an interior level of τ, call it τ^*, that maximizes welfare.

Results– When the model is calibrated to the U.S. economy, they find that $\tau^*_{US} = .27$. The welfare gain from this level of redistribution increase annual consumption by 5.6% per year, compared to the no redistribution-case where $\tau = 0$.

Floden and Linde examine also the case of Sweden, a country that traditionally has heavy government intervention and generous welfare programs. They calibrate the same model to Sweden. The major difference is the wage process: shocks are much less variable and less persistent than the U.S., so wage fluctuation in Sweden are more insurable through precautionary savings. The key reason, perhaps, is that unions and other wage compressing institutions reduce wage volatility already before taxes. It shouldn't come as a surprise then that they find an "optimal" tax rate $\tau^*_{Sweden} = .03$, i.e., very low. Essentially, labor endowment fluctuations in Sweden are small and individuals can largely self-insure against them. The amount of actual government transfers in Sweden is much larger than 3%, so in this sense there is "too much" public insurance in Sweden: as the optimum amount is exceeded the tax-induced distortions from actual redistribution can be quite costly.

Finally, keep in mind that the authors only considered a flat tax on labor income. Often governments use capital income tax for redistribution, which is much more distortionary, but also more progressive since the rich have a higher capital-income to labor-income ratio in this model.

As far as transfers are concerned, it would be more efficient to condition the transfer on ε (i.e. agent with low ε would receive more), but Floden-Linde assume that ε is private information, hence unobservable to the government, so the government cannot condition on ε. A more sophisticated approach to optimal taxation (called the Mirleesian approach) would ask the question of what is the optimal tax scheme that maximizes welfare, given this private information constraint. In other words, the planner would not tie its hands to tax with a flat tax and redistribute lump sum.

1.5 Optimal Quantity of Debt

Aiyagari and Mc Grattan (1998) study the quantity of government debt that maximizes welfare (same social welfare function as Floden and Linde) in the U.S. economy. The economy is like the one in Floden and Linde, except for the government sector.

Every period the government has two type of outlays: transfers T, and interest payments on the stock of existing public (one-period) debt B. These outlays need to be financed by distortionary taxes on labor income at rate τ. The government budget constraint reads

$$T + (1 + r) B = B' + \tau w H,$$

where in the stationary equilibrium $B' = B$. Here capital letters denote aggregate quantities.

Government debt is an additional risk-free asset and, by no arbitrage, it must carry the same rate of return as capital in equilibrium. Debt has a number of negative and positive effects on the equilibrium. On the negative side, first of all, debt is costly because financing interest payments on debt requires distortionary taxes. Second, public debt crowds-out productive capital because some of the savings are shifted away from productive capital into unproductive debt. Note that the equilibrium condition in the asset market is now

$$K(r) + B = A(r) \Rightarrow K(r) = \tilde{A}(r) \equiv A(r) - B.$$

Note that, in the graphical representation of the stationary equilibrium, the aggregate demand is still $K(r)$. The aggregate supply shifts to the left by an amount B (it's as if the effective borrowing constraint shifts), so the interest rate rises unambiguously.

FIGURE

The rise in the equilibrium interest rate means that government debt has an advantage as well. An increase in debt is effectively like introducing a looser borrowing constraint: the government enhances liquidity by providing additional means for consumption smoothing, besides claims to physical capital. Thus, increases in debt raise the return on assets, and make assets cheaper to hold. Recall that the closer the equilibrium interest rate to β, the more efficient the economy. Put differently, with incomplete markets agents are forced to hold assets for self-insurance which is costly because it reduces their consumption. The higher the equilibrium interest rate, the lower this cost. Note however that this argument holds if the borrowing limit is exogenous. It would be greatly weakened if we had assumed that agents can borrow up to the natural borrowing limit, since a higher r has no effect on an exogenous borrowing constraint, but it tightens the natural borrowing limit.

After calibrating their model, Aiyagari and McGrattan conclude that the optimal quantity of debt is very close to the actual one for the U.S. economy, i.e., around 2/3 of GDP.

References

[1] Aiyagari, Rao and Ellen Mc Grattan (1998); The Optimum Quantity of Debt, *Journal of Monetary Economics.*

[2] Budria Santiago, Javier Diaz-Gimenez, Vincenzo Quadrini, Victor Rios-Rull (2002); Updated Facts on the US Distributions of Earnings, Income and Wealth, *Minneapolis Fed Quarterly Review*

[3] Castaneda, Ana, Javier Diaz-Jimenez and Jose-Victor Rios-Rull (2003); Accounting for earnings and wealth inequality, *Journal of Political Economy*

[4] De Nardi, Mariacristina (2004); Wealth Inequality and Intergenerational Links, *Review of Economic Studies*

[5] Floden Martin and Jesper Linde (2001); Idiosyncratic Risk in the U.S. and Sweden: Is there a Role for Government Insurance?, *Review of Economic Dynamics*

[6] Hubbard, Skinner and Zeldes (1995); Precautionary saving and social insurance, *Journal of Political Economy*

[7] Kitao, Sagiri (2004); Entrepreneurship, taxation and capital investment, *Review of Economic Dynamics*

[8] Krusell, Per and Tony Smith (1997); Income and wealth heterogeity, portfolio choice, and equilibrium asset returns, *Macroeconomic Dynamics*

[9] Quadrini Vincenzo (2000); Entrepreneurship, Saving and Social Mobility, *Review of Economic Dynamics*

1 Constrained Efficiency in the Neoclassical Growth Model with Idiosyncratic Shocks

We have learned that the equilibrium allocations in the neoclassical growth model with idiosyncratic risk display *over-accumulation* of capital relative to the *first best* where $R\beta = 1$. The reason is that, given the lack of perfect insurance markets, agents save more for self-insurance, the capital stock goes up, the wage rises, and the interest rate falls below the discount rate, i.e. $R\beta < 1$.

In other words, incomplete-markets equilibrium allocations are Pareto inefficient, or, not first-best. An unconstrained planner achieves Pareto efficiency because it has more freedom in allocating resources than is provided by the system of incomplete markets: it can make state-contingent transfers across agents. Put it differently, the unconstrained planner effectively reintroduce the missing markets.

While first-best is an important and useful benchmark, one can argue that the interesting question is —not so much whether a new economic structure with state contingent transfer can do better— but whether the market performs efficiently relative to the set of allocations achievable with the this same economic structure. We want to assume that the planner cannot overcome directly the frictions imposed by the missing markets, but faces the same constraints on trade and transfer of resources across agents as in the original environment. This is the concept of *constrained efficiency*.

How do we investigate constrained optimality of the equilibrium allocations in the Aiyagari model? We must solve the problem of a planner that instructs consumption and saving decisions to each agent —by choosing a consumption policy function— while facing the same technology and asset structure (i.e., only a risk-free bond), and hence the same set of constraints, that agents face in the decentralized equilibrium.

We will find that the competitive equilibrium is constrained inefficient. The reason is the presence of a so called "pecuniary externality", i.e. each agents' decision has a negligible effect on prices that individual agents do not take into account. But by choosing a consumption policy for each agent in the economy, the planner can affect prices in the right way. In this sense, it is as if the planner has an additional instrument for redistributing income across states which is not available in the decentralized market.

We follow Davila, Hong, Krusell, and Rios-Rull (DHKR) in the exposition of the recursive problem. Let's start from the competitive equilibrium and let $a' = g^*(a,\varepsilon)$ be the decision rule of the agent. The necessary FOC of the agent in the steady state with invariant distribution λ^* is

$$u_c\left[R(\lambda^*)a + w(\lambda^*)\varepsilon - g^*(a,\varepsilon)\right] \geq \beta R(\lambda^*)\sum_{\varepsilon'\in E} u_c\left[R(\lambda^*)g^*(a,\varepsilon) + w(\lambda^*)\varepsilon' - g^*(g^*(a,\varepsilon),\varepsilon')\right]\pi(\varepsilon',\varepsilon),$$

which we can compactly rewrite as

$$u_c \geq \beta R(\lambda^*)\sum_{\varepsilon'\in E} u_c'\pi(\varepsilon',\varepsilon). \tag{1}$$

The problem of the planner who maximizes social welfare by choosing a saving policy $g(a,\varepsilon)$ (i.e., a saving level a' for every point in the state space) is

$$\begin{aligned}
\Omega(\lambda) &= \max_{g(a,\varepsilon)\in A}\int_{A\times E} u\left[R(\lambda)a + w(\lambda)\varepsilon - g(a,\varepsilon)\right]d\lambda + \beta\Omega(\lambda') \\
& \quad s.t. \\
R(\lambda) &= F_K(K,H) \text{ and } w(\lambda) = F_H(K,H) \\
H &= \int_{A\times E}\varepsilon d\lambda \\
K &= \int_{A\times E} a d\lambda \\
\lambda'(\mathcal{A}\times\mathcal{E}) &= \int_{A\times E} 1_{\{g(a,\varepsilon)\in\mathcal{A}\}}\pi(\varepsilon'\in\mathcal{E},\varepsilon)\,d\lambda(a,\varepsilon)
\end{aligned}$$

It is easy to see that the necessary FOC for the planner who chooses the level a' for a particular pair (a,ε) is:

$$u_c \geq \beta R(\lambda')\sum_{\varepsilon'\in E} u_c'\pi(\varepsilon',\varepsilon) + \beta\int_{A\times E}(a'F_{KK}' + \varepsilon' F_{HK}')\,u_c' d\lambda'.$$

Compared to the competitive equilibrium Euler equation (1), we have an extra term which comes from the fact that the planner internalizes the effects that individual savings have on prices, so it "takes derivatives" also with respect to equilibrium prices that are, in turn, equal to marginal productivities of the factors of production.

In particular, the term in parenthesis under the integral captures the effect of an additional unit of savings on next-period individual labor income (ε') and on next-period individual capital income (a') of all agents through next-period price changes, F_{KH}' and F_{KK}'. More savings increase the capital stock, thus raise the marginal product of labor

F'_H, wages, and labor income, while decrease the marginal product of capital F'_K, the interest rate, and capital income. This effect is averaged across all agents through weights equal to their marginal utility of consumption. So, poor agents receive more weight.

This extra term can be either positive or negative since $F'_{HK} > 0$ but $F'_{KK} < 0$. Note that, in the representative agent case, this term is zero because if F is CRS, then F_K and F_H are homogenous of degree zero.

We can use the CRS assumption on F by rewriting the planner's Euler equation as

$$u_c \geq \beta R(\lambda') \sum_{\varepsilon' \in E} u'_c \pi(\varepsilon', \varepsilon) + \beta F'_{KK} K' \int_{A \times E} \left(\frac{a'}{K'} - \frac{\varepsilon'}{H} \right) u'_c d\lambda'.$$

This expression clarifies that if income of the poor agents (those with u'_c large) is labor-intensive, then the extra term will be positive since $F'_{KK} < 0$ and, for the poor agents who have the highest weight, $\frac{a'}{K'} < \frac{\varepsilon'}{H}$. The opposite is true if income of the poor is capital-intensive. Therefore the factor composition of income of the poor agents is key in determining the constrained-efficiency properties of this economy. In the first case, arguably the more plausible, the planner wants agents to save more than in the decentralized equilibrium, and hence equilibrium allocations display *under-accumulation* relative to the constrained optimum. This is a surprising result.

The intuition comes from the fact that the planner always wants to redistribute from rich to poor. If the poor have mostly labor income, then the way to redistribute is to increase equilibrium wages by inducing agents to save more than in equilibrium. Larger individual savings increase the aggregate capital stock and increase wages. Another way to understand this result is that in this economy wealth inequality is a symptom of inefficiency because it is generated by uninsurable shocks. Higher capital stock reduces the return on saving and reduces the wealth inequality induced by the missing markets.

Quantitatively, DHKR calibrate the model to the US economy (and the match wealth inequality using the strategy in Diaz-Gimenez, Castaneda, and Rios-Rull) and find that the constrained efficient capital stock is a staggering 3.5 times higher than the laissez-faire economy capital stock.

1 Transitional Dynamics in Bewley Models

So far, we focused on stationary equilibria of economies with heterogeneous agents and incomplete markets. However, certain policy questions that can be asked in this class of models are better addressed by computing the entire transition path across steady states. For example, Here we study the welfare effect of a tax reform, e.g., a permanent rise in the labor income tax rate from τ^* to τ^{**} with the tax revenues being used to finance lump sum government transfers ϕ.

Steady-state comparison– One way to evaluate the tax reform is computing a stationary RCE for these two levels of the tax rate and compare aggregate variables and welfare between the two steady-states. However, this approach is not fully satisfactory. It can only be used to assess whether a household would prefer to live in the stationary equilibrium of an economy with tax rate equal to τ^* or in the stationary equilibrium of an economy with tax rate equal to τ^{**}. For example, it is useful to answer the question whether an individual would prefer to live, forever, in a country with a given tax rate τ^* or in another country with tax rate τ^{**}.

Transition– A more interesting and relevant policy question is: consider a household living in the stationary equilibrium of an economy with initial tax rate τ^*. What is the welfare change (gain or loss) of the tax reform? To answer this question properly, one needs to compute the whole transition: the new policy will change the individual consumption/saving and, possibly, labor supply decision, hence aggregate prices and will induce dynamics away from the current steady-state towards the new one (assuming the system has stable dynamics). See Domeij and Heathcote (2003) for an example that applies these techniques.

How do we attack this problem? Since the transition is characterized by a sequence of aggregate prices and quantities, we need to modify appropriately the definition of recursive competitive equilibrium.

1.1 Definition of Equilibrium with Transition

First, let's define the household problem at time t still in recursive form

$$
\begin{aligned}
v_t(a,\varepsilon) &= \max_{c_t,a_{t+1}} \left\{ u(c_t(a,\varepsilon)) + \beta \sum_{\varepsilon_{t+1}\in E} v_{t+1}(a_{t+1}(a,\varepsilon),\varepsilon_{t+1})\pi(\varepsilon_{t+1},\varepsilon) \right\} \qquad (1) \\
& \quad s.t. \\
c_t(a,\varepsilon) + a_{t+1}(a,\varepsilon) &= (1+r_t)\, a + w_t\,(1-\tau_t)\,\varepsilon + \phi_t \\
a_{t+1}(a,\varepsilon) &\geq -b
\end{aligned}
$$

Note now that value functions and policies are also a function of time since policies (τ_t, ϕ_t) and, hence, aggregate prices (r_t, w_t) are time-varying. It is important to understand that the transitional dynamics induced by the tax reform are *deterministic*. Since we know the exact future path for taxes, we know there will be a deterministic path for prices and for the distribution λ_t. Therefore, we do not need to keep track of the distribution as an additional state, as time is a sufficient statistic. We'll see that in presence of aggregate uncertainty, instead, the distribution λ becomes an aggregate state.

Let's denote the initial stationary distribution with λ^*. Given an initial distribution λ^*, and a sequence of tax rates $\{\tau_t\}_{t=0}^{\infty}$, a *recursive competitive equilibrium* is a sequence of value functions $\{v_t\}_{t=0}^{\infty}$ and decision rules for households $\{c_t, a_{t+1}\}_{t=0}^{\infty}$, firm choices $\{H_t, K_t\}_{t=0}^{\infty}$, prices $\{w_t, r_t\}_{t=0}^{\infty}$, government transfers $\{\phi_t\}_{t=0}^{\infty}$ and distributions $\{\lambda_t\}_{t=0}^{\infty}$ such that, for all t:

- given prices $\{r_t, w_t\}$ and policies $\{\tau_t, \phi_t\}$, the decision rules $a_{t+1}(a,\varepsilon)$ and $c_t(a,\varepsilon)$ solve the household's problem (1) and $v_t(a,\varepsilon)$ is the associated value function,

- given prices $\{r_t, w_t\}$, the firm chooses optimally its capital K_t and its labor H_t, i.e. $r_t + \delta = F_K(K_t, H_t)$ and $w_t = F_H(K_t, H_t)$,

- the labor market clears: $H_t = \int_{A\times E} \varepsilon d\lambda_t = H$,

- the asset market clears: $K_{t+1} = \int_{A\times E} a_{t+1}(a,\varepsilon) d\lambda_t$,

- the goods market clears: $\int_{A\times E} c_t(a,\varepsilon) d\lambda_t + K_{t+1} - (1-\delta)\, K_t = F(K_t, H_t)$,

- the government budget constraint is balanced: $\phi_t = \tau_t w_t H$,

- for all $(\mathcal{A} \times \mathcal{E}) \in \mathcal{S}$, the probability measure λ_{t+1} satisfies

$$\lambda_{t+1}(\mathcal{A} \times \mathcal{E}) = \int_{A \times E} Q_t((a, \varepsilon), \mathcal{A} \times \mathcal{E})\, d\lambda_t,$$

  where Q_t is the transition function defined as

$$Q_t((a, \varepsilon), \mathcal{A} \times \mathcal{E}) = I_{\{a_{t+1}(a,\varepsilon) \in \mathcal{A}\}} \sum_{\varepsilon_{t+1} \in \mathcal{E}} \pi(\varepsilon_{t+1}, \varepsilon). \tag{2}$$

1.2 Numerical Computation of the Transition Path

The economy at time $t = 0$ is in steady-state with stationary distribution $\lambda_0 = \lambda^*$ over assets and individual productivities and tax rate τ^*. At the end of period $t = 0$, the government makes the surprise announcement that from $t = 1$ onward the tax policy will change to $\tau^{**} > \tau^*$ and that the additional revenues will augment the lump-sum transfer ϕ_t. Hence, the relevant tax sequence needed to compute the equilibrium is

$$\tau_t = \begin{cases} \tau^*, & \text{for } t = 0 \\ \tau^{**}, & \text{for } t \geq 1. \end{cases}$$

Next, we will assume that after T periods, with T arbitrarily large but finite, the economy will settle to the final steady-state. This assumption is helpful because it allows us to guess a finite sequence of aggregate capital stocks and use backward induction for the solution of the household problem.

To compute the equilibrium follows these steps:

1. Fix T (say $T = 200$).

2. Compute the initial steady state objects $\{v^*, c^*, a^*, K^*\}$ corresponding to $\tau = \tau^*$ and the final steady state objects $\{v^{**}, c^{**}, a^{**}, K^{**}\}$ corresponding to $\tau = \tau^{**}$.

3. Guess a sequence of aggregate capital stocks $\left\{\hat{K}_t\right\}_{t=1}^{T}$ of length T such that $\hat{K}_1 = K^*$ (capital at time 1 is predetermined at time $t = 0$ which is a steady-state) and $\hat{K}_T = K^{**}$. For example, you can make a guess based on the representative agent equivalent of your economy. Note that $H_t = H$ (i.e. constant) for every t in absence of endogenous labor supply. Hence, it is easy to determine, for each t,

$$\begin{aligned} \hat{w}_t &= F_H(\hat{K}_t, H), \\ \hat{r}_t &= F_K(\hat{K}_t, H), \\ \hat{\phi}_t &= \tau_t \hat{w}_t H, \end{aligned}$$

which are all the elements we need in the budget constraint of the household to solve the household problem at time t.[1]

4. Since we know that $c_T(a,\varepsilon) = c^{**}(a,\varepsilon)$, we can solve the household problem *by backward induction* and derive $\{\hat{c}_t(a,\varepsilon)\}_{t=1}^{T-1}$ from (1) and the associated policy functions $\{\hat{a}_{t+1}(a,\varepsilon)\}_{t=1}^{T-1}$ by iterating over the Euler equation

$$u_c\left(R_t a + (1-\tau_t) w_t \varepsilon - a_{t+1}(a,\varepsilon)\right) \geq \beta R_{t+1} \sum_{\varepsilon'} u_c\left(R_{t+1} a_{t+1}(a,\varepsilon) + (1-\tau_{t+1}) w_{t+1}\varepsilon_{t+1} - a_{t+2}(a_{t+}\right.$$

and note that the function $a_{t+2}(\cdot)$ is always known.

5. Given the policy functions, we can reconstruct the sequence of transition functions $\left\{\hat{Q}_t\right\}_{t=1}^{T}$ and, since we know that $\lambda_0 = \lambda^*$, we can recover the whole sequence of measures $\left\{\hat{\lambda}_t(a,\varepsilon)\right\}_{t=1}^{T}$ and calculate

$$\hat{A}_{t+1} = \int_{A\times E} \hat{a}_{t+1}(a,\varepsilon) d\hat{\lambda}_t.$$

To compute this integral, we use simulation techniques. We simulate histories of length T for N workers (say $N = 10,000$) starting from the steady-state distribution at $t = 1$ (distribution that we also obtain by simulation). Note that when computing the optimal consumption and saving choices of each of the N individuals in our sample at time t, we must use the time t decision rules computed in the previous step.

6. Check market clearing in the asset market in every period t, i.e. check if the guess of equilibrium capital stocks $\left\{\hat{K}_t\right\}_{t=1}^{T}$ is consistent with aggregate wealth $\left\{\hat{A}_t\right\}_{t=1}^{T}$ that households would accumulate when facing the sequence of prices induced by the proposed sequence of aggregate capital. In other words, choose a convergence criterion η and check whether

$$\max_{1\leq t\leq T} |\hat{A}_t - \hat{K}_t| < \eta. \tag{3}$$

Note that if $|\hat{A}_T - K^{**}| < \eta$ is satisfied, we have implicitly also checked that T is large enough.

[1] In this step, one can equally guess a path for the interest rate or for wages. Then, from the FOC's of the firm, one would recover aggregate capital at each t.

7. If inequality (3) is not satisfied at every t, we need a new guess of the capital stock, for example

$$\hat{K}_t^{new} = 0.5\left(\hat{K}_t^{old} + \hat{A}_t\right),$$

and go back to step 3. with this new guess.

1.3 Results on the Monotonicity of Transitional Dynamics

Consider first an economy with heterogeneity in individual endowments and a borrowing constraint, but no idiosyncratic risk. Then, the Euler equation for every agent in the economy is

$$u'(c_t) \geq \beta R_{t+1} u'(c_{t+1}),$$

where there is no expectation on the RHS because of the lack of risk. Let $R_{t+1} = 1 + f'(K_{t+1} - \delta) \equiv F'(K_{t+1})$. Assume that the aggregate capital stock is below the steady-state level, so that

$$\beta R_{t+1} = \beta F'(K_{t+1}) > \beta F'(K^*) = \beta R^* = 1,$$

since we know that in this economy in steady-state the borrowing constraints do not bind and therefore $\beta R^* = 1$. Then, we have

$$u'(c_t) \geq \beta R_{t+1} u'(c_{t+1}) > u'(c_{t+1})$$

which implies that $c_{t+1} > c_t$. Consumption grows for every agent in the economy, hence aggregate consumption grows. One can show that also the capital stock is increasing monotonically towards K^*. For more details, see Hernandez (1991).

Consider now the neoclassical growth model with idiosyncratic risk (and borrowing constraints). Assume $u' > 0, u'' < 0, u''' > 0$ and the Inada condition $\lim_{c\to\infty} u'(c) = 0$. Huggett (1997) proves the following theorem.

Theorem: If $\beta F'(K_t) > 1$ and $K_t > 0$, then $K_{t+1} > K_t > K_{t-1}$.

Recall that the steady state of this model features $\beta R^* = \beta F'(K^*) < 1$, so the Theorem says that, when the capital stock is sufficiently below the steady-state, the convergence will be monotonic. The proof is by contradiction. Assume that $\{K_t\}$ is decreasing (by Lemma 2 in Hugget's paper it turns out that it's not restrictive). Then, the sequence $\{\beta R_\tau\}_{\tau=t}^{\infty}$ will be increasing and always above 1 from date t onward. The Euler equation for the individual is

$$u'\left(c_t\left(a_t,\varepsilon_t\right)\right)\geq\beta R_{t+1}E_t\left[u'\left(c_{t+1}\left(g_t\left(a_t,\varepsilon_t\right),\varepsilon_{t+1}\right)\right)\right].$$

Integrate both sides of the expression wrt λ_t and obtain

$$\int_{A\times E}u'\left(c_t\left(a_t,\varepsilon_t\right)\right)d\lambda_t\geq\beta R_{t+1}\int_{A\times E}E_t\left[u'\left(c_{t+1}\left(g_t\left(a_t,\varepsilon_t\right),\varepsilon_{t+1}\right)\right)\right]d\lambda_t.$$

From Theorem 8.3 in SLP, we know that for a continuous function h:

$$\int\left(Th\right)\left(z\right)\lambda\left(dz\right)=\int h\left(z'\right)\left(T^{*}\lambda\right)\left(dz'\right)$$

which implies

$$\int_{A\times E}E_t\left[u'\left(c_{t+1}\left(g_t\left(a_t,\varepsilon_t\right),\varepsilon_{t+1}\right)\right)\right]d\lambda_t=\int_{A\times E}u'\left(c_{t+1}\left(a_{t+1},\varepsilon_{t+1}\right)\right)d\lambda_{t+1}$$

since both integrals express the expected value of h next period: it does not matter the order in which the integration is performed. Put it differently, the LHS integrates wrt ε_{t+1} through the E operator and then integrates over all the pairs (a,ε) the function $g_t\left(a,\varepsilon\right)$ which is equivalent to integrating over a_{t+1}. So, both integrals are with respect to $\left(a_{t+1},\varepsilon_{t+1}\right)$. Therefore we obtain

$$\int_{A\times E}u'\left(c_t\left(a_t,\varepsilon_t\right)\right)d\lambda_t\geq\beta R_{t+1}\int_{A\times E}u'\left(c_{t+1}\left(a_{t+1},\varepsilon_{t+1}\right)\right)d\lambda_{t+1}$$

And repeating this argument, we get

$$\int_{A\times E}u'\left(c_t\left(a_t,\varepsilon_t\right)\right)d\lambda_t\geq\beta^{n}\left(R_{t+1}\cdot R_{t+2}\cdot\ldots\cdot R_{t+n}\right)\int_{A\times E}\left[u'\left(c_{t+n}\left(a_{t+n},\varepsilon_{t+n}\right)\right)\right]d\lambda_{t+n}$$

The first term of the RHS goes to infinity because $\beta R_\tau>1$ for every $\tau>t$. Therefore, $\int_{A\times E}\left[u'\left(c_{t+n}\left(a_{t+n},\varepsilon_{t+n}\right)\right)\right]d\lambda_{t+n}$ must converge to zero because the LHS is finite. Given any $\eta>0$, let n such that

$$\eta>\int_{A\times E}\left[u'\left(c_{t+n}\left(a_{t+n},\varepsilon_{t+n}\right)\right)\right]d\lambda_{t+n}>u'\left(\int_{A\times E}c_{t+n}\left(a_{t+n},\varepsilon_{t+n}\right)d\lambda_{t+n}\right)=u'\left(C_{t+n}\right)$$

where the second inequality requires $u'''>0$ and Jensen's inequality. This second inequality implies aggregate consumption C_{t+n} goes to infinity by the Inada condition, which violates the resource constraint, since the aggregate capital stock is assumed to be a decreasing sequence. This cannot be an equilibrium, and we found a contradiction.

1.4 Computing the Welfare Change from the Tax Reform

The crucial question to ask, from a policy perspective, is: how much agents gain/lose from the tax reform? In the first steady-state, an agent with initial individual state (a, ε) has expected lifetime utility associated with the stationary decision rule $c^*(a, \varepsilon)$ given by

$$v^*(a, \varepsilon) = E_0 \left[\sum_{t=0}^{\infty} \beta^t u \left(c^* \left(a_t, \varepsilon_t \right) \right) | \left(a_0 = a, \varepsilon_0 = \varepsilon \right) \right], \tag{4}$$

where the conditional expectation E_0 is taken over histories of the shocks conditional on a time-zero realization of the shock equal to ε (as made clear by the second equality) and conditional to an agent's wealth level equal to a. Note that, we can also define $v^*(a, \varepsilon)$ as the fixed point of

$$v^*_{n+1}(a, \varepsilon) = u \left(c^* (a, \varepsilon) \right) + \beta \sum_{\varepsilon'} v^*_n \left(a'(a, \varepsilon), \varepsilon' \right) \pi \left(\varepsilon', \varepsilon \right), \tag{5}$$

where the subscript n denotes the iteration, which is a contraction mapping. These two expressions gives us two ways to compute v^*. First, inspired by (4), by simulating S histories of an agent with initial conditions (a, ε), computing discounted utility for every history and the averaging across the S simulations. Second, by iterating over the contraction mapping (5). We set a grid over (a, ε), we guess a matrix v^*_n over the grid and use the steady-state decision rules to compute v^*_{n+1}, and continue until convergence. To evaluate the function v^*_n outside grid point, we need standard interpolation methods.

Let's turn to the transition. Define the expected discounted utility of an agent with initial state (a, ε) at date $t = 0$ going through the transition as

$$\tilde{v}_0(a, \varepsilon) = E_0 \left[\sum_{t=0}^{\infty} \beta^t u \left(\tilde{c}_t \left(a_t, \varepsilon_t \right) \right) | \left(a_0 = a, \varepsilon_0 = \varepsilon \right) \right],$$

where $\tilde{v}$ differs from v^* because it is computed using the sequence of decision rules $\{\tilde{c}_t, \tilde{a}_{t+1}\}_{t=0}^{\infty}$ along the transition path.

How do we compute this value function? We can always do it by brute force, by simulation. But, it is more efficient to do it by backward induction. We know that, for an agent going through the transition:

$$\begin{aligned}
\tilde{v}_0(a,\varepsilon) &= u(c^*(a,\varepsilon)) + \beta \sum_{\varepsilon'} \tilde{v}_1(a^*(a,\varepsilon),\varepsilon')\,\pi(\varepsilon',\varepsilon) \\
&\quad \ldots \\
\tilde{v}_t(a,\varepsilon) &= u(\tilde{c}_t(a,\varepsilon)) + \beta \sum_{\varepsilon'} \tilde{v}_{t+1}(\tilde{a}_{t+1}(a,\varepsilon),\varepsilon')\,\pi(\varepsilon',\varepsilon) \\
&\quad \ldots \\
\tilde{v}_{T-1}(a,\varepsilon) &= u(\tilde{c}_{T-1}(a,\varepsilon)) + \beta \sum_{\varepsilon'} v^{**}(a^{**}(a,\varepsilon),\varepsilon')\,\pi(\varepsilon',\varepsilon)
\end{aligned}$$

and using this recursion, we can construct $\tilde{v}_0(a,\varepsilon)$ by backward induction. Note that at every date t we don't need to iterate, we can construct $\tilde{v}_t$ from $\tilde{v}_{t+1}$ directly, given that we know $\tilde{v}_{t+1}$ and the decision rules.

Conditional welfare change– The first question we can ask is: how much would an agent with initial state (a,ε) gain, in percentage terms of lifetime consumption, if he went through the transition induced by the policy reform, compared to the no-reform scenario where he lives in the initial steady-state forever? So, welfare changes are expressed in terms of *consumption-equivalent variation*. Precisely, we ask: "how much do we need to change consumption of the agent in every state in the stationary equilibrium so that he'd be indifferent between living through the tax reform and living in the pre-reform economy?".

The answer to this question is a function $\omega(a,\varepsilon)$ that solves the equation

$$E_0 \sum_{t=0}^{\infty} \beta^t u([1+\omega(a,\varepsilon)]\,c_t^*) = E_0 \sum_{t=0}^{\infty} \beta^t u(\tilde{c}_t).$$

For the case of power-utility this calculation is really easy to make. When $u(c) = c^{1-\sigma}$, we can exploit the homogeneity of the utility function and the equation above becomes

$$\begin{aligned}
[1+\omega(a,\varepsilon)]^{1-\sigma}\, v^*(a,\varepsilon) &= \tilde{v}_0(a,\varepsilon), \\
\omega(a,\varepsilon) &= \left[\frac{\tilde{v}_0(a,\varepsilon)}{v^*(a,\varepsilon)}\right]^{\frac{1}{1-\sigma}} - 1. \qquad (6)
\end{aligned}$$

This welfare change is called *conditional welfare change*, because it is computed for an individual that is in a particular state (a,ε). Thus, we can compute the welfare change for the rich household, the poor household, the productive household, the unproductive household, etc... Moreover, we can compute the entire distribution of welfare changes and study whether the reform would be politically feasible, e.g. whether the majority of agents have positive welfare gains, hence they would support the reform.

Utilitarian social welfare change– The second type of welfare calculation is based on a Benthamian social welfare function that puts equal weight to every household in the initial steady-state (which is also the initial period of the reform), i.e. it uses the weighting criterion $\lambda^*(a,\varepsilon)$. The solution to this welfare calculation, for the power utility case is *one number only*, ω^U that solves

$$\omega^U = \left[\frac{\int_{A\times E} \tilde{v}_0(a,\varepsilon)\, d\lambda^*}{\int_{A\times E} v^*(a,\varepsilon)\, d\lambda^*}\right]^{\frac{1}{1-\sigma}} - 1.$$

So, ω^U computes the welfare change for "society", where every agent in society is given equal weight: some will lose, some will gain and we average across everyone with equal weights.

An alternative interpretation of this welfare criterion is that of an *ex-ante* welfare gain, or welfare gain *under the veil of ignorance*. In other words, $\int_{A\times E} \tilde{v}_0(a,\varepsilon)\, d\lambda^*(a,\varepsilon)$ represents the expected discounted utility of a newborn agent who is dropped at random in the first steady-state without knowing at which point in the distribution she will be, i.e. under the veil of ignorance.

1.5 A Welfare Change Decomposition

Sources of changes in social welfare– The typical utilitarian equal-weight social welfare function is:

$$\int_{A\times E} v(a,\varepsilon)\, d\lambda = \int_{A\times E} E_0 \sum_{t=0}^{\infty} \beta^t u(c_t)\, d\lambda. \tag{7}$$

Welfare can increase for three reasons:

1. If average consumption $E(c_t)$ increases, since utility is monotone. If the tax reform generates additional resources, at least one individual can be made better off, while the others receive the same utility. We can call this effect, the *level effect*. For example, by reallocating resources more efficiently a tax reform can increase average consumption.

2. If the uncertainty/volatility of each individual consumption path $\{c_t\}_{t=0}^{\infty}$ is reduced, since agents are risk averse. We can call this the *uncertainty effect*. For example, by

redistributing from the lucky to the unlucky, the tax reform can provide additional insurance.

3. If inequality across individuals at any point in time is reduced, since the value function is concave. This is easily seen by just applying Jensen's inequality to the left hand side of (7). We can call this effect, the *egalitarian effect*. Note the key difference between 2) and 3): even when there is no uncertainty in consumption sequences, a policy that redistributes initial wealth more equally across agents would achieve a welfare gain, under this social welfare function. This makes the social welfare function not a fully desirable criterion when studying the welfare implication of a policy reform because it *mixes concern for risk/uncertainty with concern for interpersonal equality*.

Conditional welfare is preferable– For this reason, conditional welfare is a somewhat more satisfactory welfare criterion because only the level and uncertainty effect play a role. We now show that, starting from the conditional welfare criterion, the welfare change $\omega(a,\varepsilon)$ can be decomposed additively into 1 and 2. This is a useful result because, when evaluating the welfare implication of a policy reform, one can compute separately the welfare change due to the fact that the policy is 1) generating more/less aggregate consumption, and 2) increasing/decreasing consumption insurance.

Decomposition– For simplicity, we focus on a welfare comparison between steady-states, but the methodology can be extended to transitions. Let the ex-ante welfare change between economy A (e.g. the initial steady state) and economy B (e.g. the final steady-state) be

$$E_0 \sum_{t=0}^{\infty} \beta^t u\left((1+\omega)\, c_t^A\right) = E_0 \sum_{t=0}^{\infty} \beta^t u\left(c_t^B\right), \tag{8}$$

where E_0 is the expectation taken at date $t = 0$ conditional on an initial value for the pair (a,ε). To simplify the notation, we have omitted the dependence of ω from the pair (a,ε).

Let C^j denote the average consumption in economy $j = A, B$, i.e.

$$C^j = \int c^j(a,\varepsilon)\, d\lambda^j, \text{ with } j = A, B.$$

Then the *welfare gain of increased consumption levels* between A and B ω^{lev} is defined by

$$\left(1+\omega^{lev}\right) C^A \equiv C^B. \tag{9}$$

Next, let the certainty equivalent consumption bundle be defined by $\bar{C}^j$ that solves

$$E_0 \sum_{t=0}^{\infty} \beta^t u\left(c_t^j\right) = \sum_{t=0}^{\infty} \beta^t u\left(\bar{C}^j\right). \tag{10}$$

Then, we can define the cost of uncertainty p^j as

$$\sum_{t=0}^{\infty} \beta^t u\left(\left(1-p^j\right) C^j\right) \equiv \sum_{t=0}^{\infty} \beta^t u\left(\bar{C}^j\right), \tag{11}$$

which is the fraction of average consumption that an individual in economy j would be willing to give up to avoid all the risk associated to productivity fluctuations.

Then, the *welfare gain of reduced uncertainty* between economy A and economy B is

$$\omega^{unc} \equiv \frac{1-p^B}{1-p^A} - 1. \tag{12}$$

We are now ready to state:

Proposition (Floden, 2001): Assume that $u(xc)$ is "homogenous" in the sense that $u(xc) = g(x)\, u(c)$, then

$$1+\omega = \left(1+\omega^{unc}\right)\left(1+\omega^{lev}\right) \Rightarrow \omega \simeq \omega^{unc} + \omega^{lev}.$$

Proof: The total welfare change is given by that value for ω that solves (8). Consider the expected utility in economy B, i.e. the R.H.S. of equation (8):

$$E_0 \sum_{t=0}^{\infty} \beta^t u\left(c_t^B\right) = \sum_{t=0}^{\infty} \beta^t u\left(\bar{C}^B\right) = \sum_{t=0}^{\infty} \beta^t u\left(\left(1-p^B\right) C^B\right) = g\left(1-p^B\right) \sum_{t=0}^{\infty} \beta^t u\left(C^B\right),$$

where the first equality follows from (10), the second from (11) and the third from the homogeneity assumption. Then,

$$g\left(1-p^B\right) \sum_{t=0}^{\infty} \beta^t u\left(C^B\right) = g\left(1-p^B\right) \sum_{t=0}^{\infty} \beta^t u\left(\left(1+\omega^{lev}\right) C^A\right),$$

where the equality follows from definition (9). Next,

$$\begin{aligned}
g\left(1-p^{B}\right)\sum_{t=0}^{\infty}\beta^{t}u\left(\left(1+\omega^{lev}\right)C^{A}\right) &= g\left(1-p^{B}\right)g\left(1+\omega^{lev}\right)\sum_{t=0}^{\infty}\beta^{t}u\left(C^{A}\right) \\
&= g\left(\frac{1-p^{B}}{1-p^{A}}\right)g\left(1+\omega^{lev}\right)g\left(1-p^{A}\right)\sum_{t=0}^{\infty}\beta^{t}u\left(C^{A}\right) \\
&= g\left(\frac{1-p^{B}}{1-p^{A}}\right)g\left(1+\omega^{lev}\right)\sum_{t=0}^{\infty}\beta^{t}u\left(\left(1-p^{A}\right)C^{A}\right)
\end{aligned}$$

The line above follows from the homogeneity assumption. Using definition (11),

$$g\left(\frac{1-p^{B}}{1-p^{A}}\right)g\left(1+\omega^{lev}\right)\sum_{t=0}^{\infty}\beta^{t}u\left(\left(1-p^{A}\right)C^{A}\right)=g\left(\frac{1-p^{B}}{1-p^{A}}\right)g\left(1+\omega^{lev}\right)\sum_{t=0}^{\infty}\beta^{t}u\left(\bar{C}^{A}\right).$$

From the definition of certainty equivalent consumption

$$g\left(\frac{1-p^{B}}{1-p^{A}}\right)g\left(1+\omega^{lev}\right)\sum_{t=0}^{\infty}\beta^{t}u\left(\bar{C}^{A}\right)=g\left(\frac{1-p^{B}}{1-p^{A}}\right)g\left(1+\omega^{lev}\right)E_{0}\sum_{t=0}^{\infty}\beta^{t}u\left(c_{t}^{A}\right).$$

And, finally, from homogeneity

$$g\left(\frac{1-p^{B}}{1-p^{A}}\right)g\left(1+\omega^{lev}\right)E_{0}\sum_{t=0}^{\infty}\beta^{t}u\left(c_{t}^{A}\right)=E_{0}\sum_{t=0}^{\infty}\beta^{t}u\left(\left(1+\omega^{lev}\right)\left(1+\omega^{unc}\right)c_{t}^{A}\right).\ \textbf{QED}$$

Floden (2001) contains a more general proof of additive decomposition for preferences which also depend on leisure, and for the notion of ex-ante welfare which also includes a third component, what we called the egalitarian effect.

References

[1] Floden, Martin (2001), The Effectiveness of Government Debt and Transfers as Insurance, *Journal of Monetary Economics*

[2] Domeij, David and Jonathan Heathcote (2003), On the Distributional Effects of Reducing Capital Taxes, *International Economic Review*

1 Bewley Economies with Aggregate Uncertainty

So far we have assumed away aggregate fluctuations (i.e., business cycles) in our description of the incomplete-markets economies with uninsurable idiosyncratic risk à la Bewley-Aiyagari. In this section, the objective is to combine aggregate and idiosyncratic risk into an equilibrium model.

The good news is that we can use the recursive language to do this. The bad news is that solving exactly for the equilibrium allocations of this economy is *impossible*. The reason is that the measure of agents across states becomes an aggregate state of the economy, since households need to know it to forecast future prices. Future prices are a function of the future capital stock that, in turn, is the aggregation of individual saving decisions under the equilibrium distribution. However, a distribution is an infinitely dimensional object: how do we keep track of such a monster? The answer is that we will approximate the exact equilibrium and argue that the approximation is very good for standard parameterizations.

Aggregate and Idiosyncratic Risk– We introduce aggregate fluctuations through an aggregate productivity shock z that shifts the production function, i.e.

$$Y = zF(K, H),$$

and assume that the aggregate shock can take only two values, $z \in Z = \{z_b, z_g\}$ with $z_b < z_g$. To keep things simple, we also assume only two values for the individual productivity shock, $\varepsilon \in E = \{\varepsilon_b, \varepsilon_g\}$ with $\varepsilon_b < \varepsilon_g$. For example, if $\varepsilon_b = 0$, then it's as if the worker is unemployed for a period.

Let

$$\pi(z', \varepsilon' | z, \varepsilon) = \Pr(z_{t+1} = z', \varepsilon_{t+1} = \varepsilon' | z_t = z, \varepsilon_t = \varepsilon)$$

be the Markov chain that describes the joint evolution of the exogenous shocks. This notation allows the transition probabilities for ε to depend on z (the dependence of z on ε is also allowed in principle but does not make sense!). For example, one should expect that

$$\pi(z_g, \varepsilon_g | z_b, \varepsilon_b) > \pi(z_b, \varepsilon_g | z_b, \varepsilon_b), \text{ and } \pi(z_b, \varepsilon_b | z_g, \varepsilon_g) > \pi(z_g, \varepsilon_b | z_g, \varepsilon_g)$$

i.e. finding a job is easier if the economy is exiting from a recession, and losing a job is more likely when the economy is entering a recession.

State variables– The two individual states are $(a, \varepsilon) \in S$ and the two aggregate states are $(z, \lambda) \in Z \times \Lambda$ where $\lambda(a, \varepsilon)$ is the measure of households across states. The individual states are directly budget relevant, whereas the aggregate states are needed to compute and forecast prices.[1]

Household Problem– The household problem can be written in recursive form as:

$$\begin{aligned}
v(a, \varepsilon; z, \lambda) &= \max_{c, a'} \left\{ u(c) + \beta \sum_{\varepsilon' \in E, z' \in Z} v(a', \varepsilon'; z', \lambda') \pi(z', \varepsilon' | z, \varepsilon) \right\} \qquad (1) \\
& \quad s.t. \\
c + a' &= w(z, K(\lambda)) \varepsilon + R(z, K(\lambda)) a \\
a' &\geq 0 \\
\lambda' &= \Psi(z, \lambda, z')
\end{aligned}$$

where $\Psi(z, \lambda, z')$ is the law of motion of the endogenous aggregate state, and depends on z'. This dependence is inherited from π since the fraction of agents with $\varepsilon' = \varepsilon_b$ and $\varepsilon' = \varepsilon_g$ next period, given that the current aggregate productivity level is z, depends on z'.

The key complication is that the value function v depends on λ which is a distribution. Where is this dependence coming from? To solve their problem, households need to compute current prices and, most importantly, forecast prices next period. Prices depend on aggregate capital, and aggregate capital this period and next period, K and K', depends on how assets are distributed in the population, through λ, because in equilibrium

$$K = \int_{A \times E} a d\lambda \text{ and } K' = \int_{A \times E} a' d\lambda$$

Let's be more specific about why agents need to know λ. Consider the Euler equation associated to the problem above. Let the saving policy be denoted by g, and drop the depndence of K and K' on λ to ease the notation:

$$\begin{gathered}
u_c(R(z, K) a + w(z, K) \varepsilon - g(a, \varepsilon; z, K)) \geq \\
\beta E\left[R(z', K') u_c(R(z', K') g(a, \varepsilon; z, K) + w(z', K') \varepsilon' - g(g(a, \varepsilon; z, K), \varepsilon'; z', K'))\right].
\end{gathered}$$

It is clear that to solve for g, households need to forecast prices next period, and next period prices depend on K' which, in turn, depends on λ. Since λ is a state variable,

[1] Note that, although it seems reasonable that λ is enough to complete the description of the state (i.e. a recursive equilibrium exists), this is not obvious at all and there are counterexamples of economies for which one needs to keep track of a longer history of distributions.

agents need to know its equilibrium law of motion Ψ which is a complicated mapping of distributions into distributions. A note: prices depend on the K/H ratio, not just on K, but the dynamics of H can be perfectly forecasted, conditional on z', through π because labor supply is exogenous. H could be time varying, but we know how to forecats it.

A **Recursive Competitive Equilibrium** for this economy is a value function v; decision rules for the household a', and c; choice functions firm H and K; pricing functions r and w; and, a law of motion Ψ such that:

- given the pricing functions $r(z,K)$ and $w(z,K)$ and the law of motion Ψ, the decision rules a' and c solve the household's problem (1) and v is the associated value function,

- given the pricing functions $r(z,K)$ and $w(z,K)$, the firm chooses optimally its capital K and its labor H, i.e.

$$\begin{aligned} r(z,K)+\delta &= zF_K(K,H), \\ w(z,K) &= zF_H(K,H), \end{aligned} \tag{2}$$

- the labor market clears: $H = \int_{A\times E} \varepsilon d\lambda$,

- the asset market clears: $K = \int_{A\times E} a d\lambda$,

- the goods market clears:

$$\int_{A\times E} c(a,\varepsilon;z,\lambda)d\lambda + \int_{A\times E} a'(a,\varepsilon;z,\lambda)d\lambda = zF(K,H) + (1-\delta)K,$$

- For every pair (z,z'), the aggregate law of motion Ψ is generated by the exogenous Markov chain π and the policy function a' as follows:

$$\lambda'(\mathcal{A}\times\mathcal{E}) = \Psi_{(\mathcal{A}\times\mathcal{E})}(z,\lambda,z') = \int_{A\times E} Q_{z,z'}((a,\varepsilon),\mathcal{A}\times\mathcal{E})\,d\lambda, \tag{3}$$

  where $Q_{z,z'}$ is the transition function between two periods where the aggregate shock goes from z to z' and is defined by

$$Q_{z,z'}((a,\varepsilon),\mathcal{A}\times\mathcal{E}) = I_{\{g(a,\varepsilon;z,\lambda)\in\mathcal{A}\}} \sum_{\varepsilon'\in\mathcal{E}} \pi_\varepsilon(\varepsilon'|z,\varepsilon,z'), \tag{4}$$

  where I is the indicator function, $g(a,\varepsilon;z,\lambda)$ is the optimal saving policy, and $\pi_\varepsilon(\varepsilon'|z,\varepsilon,z')$ is the conditional transition probability for ε which can be easily derived from π.

1.1 Computation of an Approximate Equilibrium

The state space of the problem of the household is, technically, infinite-dimensional because it contains a distribution. The problem is to find an efficient way to compute the law of motion

$$\lambda' = \Psi\left(z, \lambda, z'\right).$$

Krusell and Smith (1998) contains the insight that, since we cannot work with an infinitely dimensional distribution, we need to approximate the distribution with a finite-dimensional object. Any distribution can be represented by its entire (in general, infinite) set of moments. Let $\bar{m}$ be a M dimensional vector of the first M moments (mean, variance, skewness, kurtosis,...) of the *wealth distribution*, i.e., the marginal of λ with respect to a. Our new state is exactly the vector $\bar{m} = \{m_1, m_2, ..., m_M\}$ with law of motion

$$\bar{m}' = \Psi_M\left(z, \bar{m}\right) = \begin{cases} m_1' = \psi_1\left(z, \bar{m}\right) \\ \quad ... \\ m_M' = \psi_M\left(z, \bar{m}\right) \end{cases}. \tag{5}$$

Note that we lost the dependence on z' since we are only interested in the wealth distribution. The law of motion for the marginal wealth distribution does not depend on z' since capital is pre-determined.

This method is based on the idea that households have *partial information* about λ. They don't know every detail about that measure, but only a set of moments, e.g. its mean, its variance, the Gini coefficient, the share held by the top 5% and so on. Hence, they use these M statistics to approximate the true distribution and form forecasts.

To make this approach operational, one needs to: 1) fix M and 2) specify a functional form for Ψ_M. Krusell and Smith (and this is their main finding) show that one obtains an excellent forecasting rule by simply setting $M = 1$ and by specifying a law of motion of the form:

$$\ln K' = b_z^0 + b_z^1 \ln K,$$

where only the first moment $m^1 = K$ would matter to predict the first moment next period.

The new partial-information problem of the agent becomes

$$
\begin{aligned}
v(a,\varepsilon;z,K) &= \max_{c,a'}\left\{u(c)+\beta\sum_{\varepsilon',z'}v(a',\varepsilon';z',K')\pi(z',\varepsilon'|z,\varepsilon)\right\} \qquad (6)\\
& \quad s.t.\\
c+a' &= w\left(z,K\right)\varepsilon+R\left(z,K\right)a\\
a' &\geq 0\\
\ln K' &= b_z^0+b_z^1\ln K.
\end{aligned}
$$

Note that this state space is definitely manageable: we collapsed an infinitely dimensional distribution λ into one variable, K. Now, we also know how to solve this problem. It will be a fixed-point algorithm over the law of motion (i.e., a function) for K: recall that in equilibrium the law of motion used by the agents has to be consistent with the aggregation of the optimal individual decisions ("aggregate consistency" of rational expectation equilibrium).[2]

1.1.1 Algorithm

The algorithm to solve this problem (and the associated equilibrium) is the following:

1. Guess the coefficients of the law of motion $\{b_z^0, b_z^1\}$

2. Solve the household problem and obtain the decision rules $a'\left(a,\varepsilon;z,K\right), c\left(a,\varepsilon;z,K\right).$ Note that with the law of motion for K in hand, we have all we need to solve for decision rules. For example, if we iterate on the Euler equation, then we need to compute the rule $a' = g\left(a,\varepsilon;z,K\right)$ that solves

$$
\begin{gathered}
u_c\left(R\left(z,K\right)a+w\left(z,K\right)\varepsilon-g\left(a,\varepsilon;z,K\right)\right)\geq\\
\beta E\left\{R\left(z',K'\right)u_c\left(R\left(z',K'\right)g\left(a,\varepsilon;z,K\right)+w\left(z',K'\right)\varepsilon'-g\left(a',\varepsilon';z',K'\right)\right)\right\}.
\end{gathered}
$$

   Thus we can use standard methods to obtain $g\left(\cdot\right).$

3. Simulate the economy for N individuals and T periods. For example, $N = 10,000$ and $T = 2,000$. Draw first a random sequence for the aggregate shocks. Next one

[2] Recall that to solve for the stationary equilibrium, since the distribution is time-invariant, we guess only one value for the capital stock (or the interest rate). To compute the equilibrium transitional dynamics, we guess a deterministic sequence of capital stocks. With aggregate uncertainty, we need to guess a law of motion for the aggregate capital stock.

for the individual productivity shocks for each $i = 1, ..N$, conditional on the time-path for the aggregate shocks. Use the decision rules to generate sequences of asset holdings $\{a_t^i\}_{t=1,i=1}^{T,N}$ and in each period compute the average capital stock

$$A_t = \frac{1}{N}\sum_{i=1}^{N} a_t^i.$$

4. Discard the first T^0 periods (e.g. $T^0 = 500$) to avoid dependence from the initial conditions. Using the remaining sequence, run the regression

$$\ln A_{t+1} = \beta_z^0 + \beta_z^1 \ln A_t \tag{7}$$

and estimate the coefficients (β_z^0, β_z^1). Note that this step requires running two regressions, one for each state z. Since the law of motion is time-invariant, we can separate the dates t in the sample where the state is z_b from those where the state is z_g and record (A_t, A_{t+1}) and run the regressions on these two separate samples.

5. If $(\beta_z^0, \beta_z^1) \neq (b_z^0, b_z^1)$, then try a new guess and go back to step 1. If the two pairs are equal for each $z \in \{z_g, z_b\}$, then it means that the approximate law of motion used by the agents is consistent with the one generated in equilibrium by aggregating individual choices. Notice that, once you reached the fixed point, you are sure that market clearing in the asset market holds: next period prices are determined by K' induced by the law of motion and this law of motion induces individual choices a' that aggregate into $A' = K'$.

6. Recall that this equilibrium computation is approximate: we still need to verify how good this approximation is to the fully rational-expectation equilibrium. For this purpose, compute a measure the fit of the regression in step 4), for example by using R^2. Next, try augmenting the state space with another moment, for example using $m^2 = E(a_i^2)$. Repeat steps 1)-5) until convergence. If the R^2 of the new equation (7) has improved significantly, keep adding moments until R^2 is large and does not respond to addition of new explanatory moments. Otherwise, stop: it means that additional moments do not add new useful information in forecasting prices.

1.2 A Near-Aggregation Result in the Krusell-Smith Economy

Krusell and Smith's main finding is that a law of motion based only on the mean, i.e.,

$$\ln K' = \begin{cases} 0.095 + 0.962 \ln K, & \text{for } z = z_g \\ 0.085 + 0.965 \ln K, & \text{for } z = z_b \end{cases}$$

delivers an $R^2 = 0.999998$ which means that the agents with this simple forecasting rule make very small errors, for example the maximal error in forecasting the interest rate 25 years into the future is around 0.1%. This result is called *near-aggregation* in the sense that in equilibrium, the evolution of aggregate quantities and prices does not depend on the distribution but, approximately, depends only on the aggregate shock and aggregate capital. Hence, it is almost like in a complete-markets economy, where aggregation of heterogenous individuals holds perfectly.

What is the intuition for the fact that keeping track of the mean of the distribution of assets is enough? Recall that if policy functions are linear, i.e.,

$$a'(a, \varepsilon, z, \lambda) = b_z^0 + b_z^1 a + b_z^2 \varepsilon,$$

then

$$K' = \int_{A\times E} a'(a, \varepsilon, z, \lambda) d\lambda = b_z^0 + b_z^1 K + b_z^2 H_z = \tilde{b}_z^0 + b_z^1 K$$

which would explain why the mean is a sufficient statistic. But saving functions are in general, not linear with uninsurable idiosyncratic shocks. They're exactly linear only with complete markets (recall the exact aggregation result of the Chatterjee economy with homothetic preferences?), where we showed that the distribution does not affect the dynamics of aggregate variables.

So, why do we get *near-aggregation* in practice? For three reasons. First, the saving functions $a'(a, \varepsilon, z, \lambda)$ for this class of problems usually display lots of curvature for low levels of ε and low levels of assets a, but beyond this region they're *almost linear*. Second, the agents with this high curvature are few and have low wealth, so they matter very little in determining aggregate wealth. What matters for the determination of the aggregate capital stock are the ones who hold a lot of capital, i.e., the rich, not the poor! Third, aggregate productivity shocks move the asset distribution only very slightly, and the mass of the distribution is always where the saving functions are linear.

But why do agents have linear saving functions, i.e. a constant marginal propensity to save out of wealth, for a very wide range of the asset space? After all, if they save for precautionary reasons (as they do in these economies) they should do so more when they hold few assets and less when they hold large assets, so the saving function should be nonlinear. The answer is that most of the consumers in this economy can smooth consumption very effectively through self-insurance, by cumulating a relatively small amount of wealth. Thus their saving behavior is guided mostly by their intertemporal motive rather than their insurance motive, like in complete markets.

To understand why the risk-free asset is such a good vehicle of self-insurance in the Krusell-Smith model, we discuss here two theoretical results in the literature that can help explain it.

First, Yaari (1976) analyzed the optimal consumption path of a perfectly impatient household ($\beta = 1$) with general concave preferences (hence with prudence) who lives for T periods, faces *iid* endowment shocks and saves and borrows at rate $r = 0$. Yaari shows that as $T \to \infty$, the optimal consumption plan converges to that of a consumer who eats a constant fraction of his wealth every period. In this sense, these households behave like certainty-equivalent consumers who are not concerned about future risk. Recall that an agent with quadratic preferences has consumption determined by

$$c_t = \frac{r}{1+r}\left[a_t + E_0 \sum_{j=0}^{\infty}\left(\frac{1}{1+r}\right)^j y_{t+j}\right] = \frac{r}{1+r}a_t + H\left(\mathbf{y},r\right),$$

where $\mathbf{y}$ is its whole future income sequence. His assets next period are determined by

$$\begin{aligned} a_{t+1} &= (1+r)\left(a_t - c_t + y_t\right) \\ &= (1+r)\left[a_t - \frac{r}{1+r}a_t - H\left(\mathbf{y},r\right)\right] + (1+r)\, y_t \\ &= a_t + \tilde{H}\left(\mathbf{y},r\right) \end{aligned}$$

and note that the coefficient on past wealth is exactly one, which is very close to the one computed by Krusell-Smith.

Second, Levine and Zame (2001) analyze an economy populated by infinitely-lived consumers with standard preferences satisfying $u''' > 0$ who face stationary individual endowment shocks (i.e., not random-walk) and trade a risk-free asset in zero net supply. They prove that, as $\beta \to 1$ and the individuals become perfectly patient, "market incompleteness will not matter" in the sense that the welfare of the optimal consumption plan in this economy tends to the welfare of a complete markets economy where every agent consumes her average endowment every period. In other words, a great deal of risk-sharing may take place even in absence of a complicated structure of financial markets.[3]

Finally, at this point it is not surprising that Krusell and Smith find that the cyclical properties of aggregates in their model economy (i.e. volatility of output, consumption, investment, cross-correlations, etc...) are very similar to those of the standard representative agent model.

References

[1] Krusell P., and A. Smith (1998);. "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, vol. 106(5), 867-896.

[2] Yaari, M. (1976); "A Law of Large Numbers in the Theory of Consumer's Choice under Uncertainty," *Journal of Economic Theory*, 12, 202-217.

[3] Levine, D., and W. Zame (2001); "Does Market Incompleteness Matter?," *Econometrica*, vol. 70(5), 1085-1839.

[3] Note however, that Levine and Zame show that their result holds in the presence of aggregate uncertainty only if markets are complete with respect to aggregate risk. In the Krusell-Smith economy there is no insurance against aggregate risk, but aggregate fluctuations are quantitatively small.

1 Micro and macro labor supply elasticity

The elasticity of labor supply is one of the crucial parameters in every macroeconomic model. For example, this elasticity determines the response of hours worked to changes in the tax rate and determines the degree of distortions tax introduce. This elasticity also determines how employment, and hence output, responds to fluctuations in productivity. therefore a key issue in macroeconomics is: how large is this elasticity?

In addition to being a very important issue, it is also well known to be quite controversial. In particular, there is a long-standing controversy driven by the fact that on the one hand, researchers who look at micro data typically estimate relatively small labor supply elasticities, while on the other hand, researchers who use representative agent models to study aggregate outcomes typically employ parameterizations that imply relatively large aggregate labor supply elasticities.

In this section, we first explain how labor economists arrived at the conclusion that the micro labor supply elasticity is small, and then we explain one way in which a small micro elasticity can be reconciled with a large aggregate elasticity.

1.1 The micro Frisch elasticity of labor supply

The Frisch elasticity of labor supply measures the percentage change in hours worked due to the percentage change in wages, holding constant the marginal utility of wealth (i.e., the multiplier on the budget constraint λ_t):

$$\varepsilon_t = \frac{dh_t/h_t}{dw_t/w_t}|_{\lambda_t}$$

It is also called the λ-constant elasticity, or intertemporal elasticity of labor supply. This elasticity measures how hours respond to wage changes abstracting from its effect on wealth. For example, consider a two-period model with no uncertainty, $r = 0$, zero initial wealth, and loose borrowing limits which do not bind. Suppose that $w_t = 1$ at $t = 1, 2$. Optimal hours worked would be equal in both periods. Now, change the path of wages so to keep the DPV of wages constant as follows: $w_1 = 0.5$ and $w_2 = 1.5$. The effect of this change in the time path of wages on hours worked is mediated precisely by the size of the Frisch elasticity.

In general, wage changes (i.e., if they not purely transitory) also have wealth effects on labor supply. The Frisch elasticity does not capture the total effect on hours from wage shocks. It captures the component due to intertemporal substitution effects, but not the one due to wealth effects.

1.1.1 General expression for Frisch elasticity

Let's derive the general expression for the Frisch elasticity. Households solve:

$$
\begin{aligned}
& \max_{\{c_t,h_t\}} E_0 \sum_{t=0}^{\infty} \beta^t u\left(c_t, h_t\right) \\
& s.t. \\
c_t + a_{t+1} &= Ra_t + w_t h_t \\
a_{t+1} &\geq -\underline{a}
\end{aligned}
$$

The Lagrangean for this problem is

$$
\mathcal{L} = E_0 \sum_{t=0}^{\infty} \beta^t \left\{ u\left(c_t, h_t\right) + \lambda_t \left[Ra_t + w_t h_t - c_t - a_{t+1}\right] + \phi_t \left[a_{t+1} + \underline{a}\right] \right\}
$$

where λ_t is the multiplier on the time t budget constraint and ϕ_t is the multiplier on the time t borrowing constraint. The first order conditions with respect to (c_t, h_t, a_{t+1}) yield

$$
\begin{aligned}
u_c &= \lambda_t && (1) \\
-u_h &= \lambda_t w_t && (2) \\
\lambda_t - \phi_t &= \beta R E_t\left[\lambda_{t+1}\right]. && (3)
\end{aligned}
$$

Differentiate the intratemporal FOC (2), keeping λ_t constant:

$$
\begin{aligned}
-u_{hh} dh_t - u_{hc} dc_t &= \lambda_t dw_t \\
-h_t u_{hh} \frac{dh_t}{h_t} - u_{hc} dc_t &= \left(\lambda_t w_t\right) \frac{dw_t}{w_t}
\end{aligned}
$$

Using the intratemporal FOC (2) again to substitute out $(\lambda_t w_t)$:

$$
\begin{aligned}
h_t u_{hh} \frac{dh_t}{h_t} + u_{hc} dc_t &= u_h \frac{dw_t}{w_t} \\
h_t u_{hh} \frac{dh_t}{h_t} + h_t u_{hc} \frac{dc_t}{dh_t} \frac{dh_t}{h_t} &= u_h \frac{dw_t}{w_t} && (4)
\end{aligned}
$$

Now, differentiating the FOC with respect to consumption (1) remembering to keep λ_t constant:

$$\begin{aligned} u_{cc}dc_t + u_{ch}dh_t &= 0 \\ \frac{dc_t}{dh_t} &= \frac{-u_{ch}}{u_{cc}} \end{aligned} \tag{5}$$

and using (5) into (4) one obtains:

$$h_t u_{hh}\frac{dh_t}{h_t} - h_t\frac{u_{hc}^2}{u_{cc}}\frac{dh_t}{h_t} = u_h\frac{dw_t}{w_t}$$

which gives the general expression for the Frisch elasticity :

$$\varepsilon_t \equiv \frac{dh_t/h_t}{dw_t/w_t}|_{\lambda_t} = \frac{u_h}{h_t u_{hh} - h_t\frac{u_{hc}^2}{u_{cc}}}. \tag{6}$$

Note that in this derivation we did not use the Euler equation, so this expression for the Frisch elasticity is independent on how agents behave intertemporally, or whether borrowing limits bind. It only requires agents to be on their intratemporal optimality condition. But it would not work for agents who are at a corner (e.g., optimally choose not to work). For example, for women, who tend to go in and out work much more than men, this derivation is problematic.

To understand the intuition for this expression, ignore the cross derivative,. Then, $\varepsilon = u_h/\left(h_t u_{hh}\right)$ which is akin to the inverse of the expression for risk aversion, i.e. akin to the expression for the IES with time-separable preferences. Indeed, the Frisch elasticity measures the willingness to substitute hours worked intertemporally.

1.1.2 Example: separable preferences

Consider the following utility function separable in consumption and hours worked:

$$u\left(c_t, h_t\right) = \frac{c_t^{1-\gamma}}{1-\gamma} - \alpha\frac{h^{1+1/\eta}}{1+1/\eta} \tag{7}$$

The Frisch elasticity under these preference specification is

$$\varepsilon_t = \frac{u_h}{h_t\left[u_{hh} - \frac{u_{hc}^2}{u_{cc}}\right]} = \frac{h_t^{1/\eta}}{h_t\left[\frac{h_t^{1/\eta-1}}{\eta} - 0\right]} = \eta,$$

hence it is a constant, independent of the level of hours. In general, this is not the case, and the Frisch elasticity depends on the level of hours worked.

Taking logs and first differences of equation (2) under the utility function (7) yields:

$$\Delta \ln h_t = \eta \Delta \ln w_t + \eta \Delta \ln \lambda_t.$$

One can estimate η through OLS only if it is possible to argue that the last term, which is unobservable (i.e. a residual in the OLS equation) is not correlated with wage growth. Unfortunately, in general it is: when $\Delta \ln w_t \uparrow \Rightarrow \Delta \ln c_t \uparrow \Rightarrow \Delta \ln \lambda_t \downarrow$ and hence $cov\left(\Delta \ln w_t, \Delta \ln \lambda_t\right) < 0$ which induces a downward bias in η. Several techniques can be used to estimate η without bias: 1) focusing on anticipated wage changes, 2) using some effective IV method, 3) focusing on wage changes that are clearly transitory in nature, 4) using consumption data to proxy λ_t. In case 1) and 3), the change in wage at date t has no impact on consumption, and hence on λ_t. Based on these techniques, labor economists have found estimates of η between 0 and 0.4.

Because of these results, some economists have maintained the view that large aggregate labor supply elasticities should be ruled out. Over time, we learned that this view is flawed. It is flawed for several reasons (see the survey by Keane and Rogerson, especially section 2.4). Here we'll pursue one particular direction to show that small micro elasticities can be reconciled with large aggregate elasticities. This direction takes seriously that there are adjustments along the extensive margin, population heterogeneity and aggregation. Most of the structural analyses based on micro data implicitly focus on adjustment along the intensive margin (choice of hours given employment). But adjustment along the extensive margin (work/not work) plays an important role: 2/3 of total fluctuations in aggregate hours are due to fluctuations in employment. In the next of the section we show that a model that, at the micro level, is consistent with this indivisibility in labor supply can, at the macro level, imply large aggregate elasticities.

1.2 The indivisible labor model with full insurance

We begin with a static model of labor supply adjustment along the extensive margin in complete markets. This model is due to Hansen (1985) and Rogerson (1988). A common view of labor supply is that individuals face an indivisibility in their choice of hours worked, i.e. at any date t, $h_{it} \in \{0, \overline{h}\}$. Then, the commodity set is nonconvex and the Welfare Theorem fails. To deal with this conceptual problem, economists often use lotteries. To be precise, one can introduce a new commodity: instead of choosing hours $h_{it} \in \{0, \overline{h}\}$, agents choose a probability of working $\pi_{it} \in [0, 1]$. If the outcome of the chosen lottery is good, they are employed with outcome $s_{it} = e$ and work $h_{it} = \bar{h}$ hours, if it is bad they remain unemployed ($s_{it} = n$) and do not work ($h_{it} = 0$).

1.2.1 Supply and demand of insurance

In what follows, we also assume that there are complete insurance markets. In particular, the market offers insurance claims that pay one unit of consumption contingent on the bad realization of the lottery, i.e. when the outcome is $s_{it} = u$, which occurs with probability $1 - \pi_{it}$. The insurance market is competitive, i.e. the insurance company asked to price a contract that pays q units of consumption if unemployment occurs with probability $1 - \pi$ makes profits

$$pq - (1 - \pi) q,$$

and therefore the zero profit condition implies: $p(\pi) = 1 - \pi$ which means that the the price per unit of consumption insured is $p(\pi)$ and the contract is actuarially fair.

Now, consider the problem of a household who has already chosen π_{it}. How much insurance would she purchase to insure her unemployment risk? Assume households have separable preferences in consumption and hours worked, i.e.

$$U(c, h) = u(c) - v(h).$$

The household who takes the price function $p(\pi)$ as given solves:

$$\begin{aligned}
&\max_{q_{it}} \pi_{it} \left[u\left(c_{it}^{e}\right) - v\left(\overline{h}\right)\right] + (1 - \pi_{it}) u\left(c_{it}^{u}\right) \\
&s.t. \\
c_{it}^{e} &= w_t \overline{h} - p(\pi_{it}) q_{it} \qquad \text{if employed} \\
c_{it}^{u} &= q_{it} - p(\pi_{it}) q_{it} \qquad \text{if unemployed}
\end{aligned}$$

The FOC with respect to q_{it} gives:

$$\pi_{it} u_c\left(c_{it}^{e}\right) p(\pi_{it}) = (1 - \pi_{it}) u_c\left(c_{it}^{u}\right) \left[1 - p(\pi_{it})\right].$$

Using the solution for the equilibrium price of insurance, we arrive at $u_c(c_{it}^e) = u_c(c_{it}^u)$ which implies $c_{it}^e = c_{it}^u$ from strict concavity of u. Therefore, the agent fully insures herself, $q_{it} = w_t\overline{h}$ and consumption of individual i at date t in every state is $c_{it} = \pi_{it} w_t \overline{h}$.

1.2.2 Choice of lottery

Now we are ready to determine how households choose their probability of working, i.e., their favorite lottery. This is also a static problem. Note that expected utility is

$$\pi_{it}\left[u(c_{it}^e) - v(\overline{h})\right] + (1-\pi_{it})\, u(c_{it}^u) = u(c_t) - \pi_{it} v(\bar{h})$$

where we used the full insurance feature of the consumption allocation. Therefore, the problem can be written as:

$$\begin{aligned} & \max_{\pi_{it}} u(c_{it}) - \pi_{it} v(\bar{h}) \\ & s.t. \\ c_{it} &= \pi_{it} w_t \overline{h} \end{aligned}$$

with solution

$$u_c(c_{it})\, w_t\overline{h} = v(\bar{h}) \rightarrow u_c\left(\pi_{it} w_t \overline{h}\right) w_t\overline{h} = v(\bar{h}) \rightarrow \pi_{it}^* = u_c^{-1}\left(\frac{v(\bar{h})}{w_t\overline{h}}\right)\frac{1}{w_t\overline{h}}$$

which implies that $\pi_{it}^* = \pi_t^*$ is the same across individuals. Therefore also consumption will be the same across all individuals at date t. Note that if preferences where heterogeneous in the disutility of labor, or if individuals differed in terms of initial wealth endowments, this would not be the case. You may ask: why does the agent bother about choosing a probability of working if, in every state, she gets the same consumption anyway? Because, by choosing optimally π, she can affect her level of consumption.

1.2.3 Aggregation

Because every agent makes the same consumption and lottery decision, it is clear that if we had a dynamic RBC model with capital accumulation and productivity shocks, we could define a representative agent —whose chosen allocations equal the aggregate quantities of the economy— solving the problem:

$$
\begin{aligned}
& \max_{C_t,H_t,K_{t+1}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[u\left(C_t\right) - AH_t\right] \qquad (8)\\
& s.t.\\
C_t + K_{t+1} &= (1-\delta) K_t + z_t F\left(K_t, H_t\right)\\
z_t &\sim F\left(z_{t-1}\right)
\end{aligned}
$$

where, without loss of generality, we have defined $A \equiv v\left(\bar{h}\right)/\bar{h}$ and we have defined aggregate hours $H_t \equiv \pi_t \bar{h}$. We have also used the equivalence between competitive equilibrium and social planner allocations, which we can do because, with lotteries and insurance contracts, the consumption set is convex and markets are complete. This is the dynamic model solved by Hansen (JME, 1985), a workhorse of RBC theory.[1]

What is the aggregate Frisch labor supply elasticity in this model? Recall that

$$\eta \equiv \frac{u_H}{Hu_{HH}} = \frac{A}{H \cdot 0} = +\infty.$$

The intuition for the infinite elasticity is that, with full insurance, everyone is ex-ante indifferent between working or not, so it is costless to move workers in and out of employment to respond to fluctuations in aggregate productivity and wages. Note though that, ex-post, the unemployed are better off because they get the same consumption and more leisure. So, ex-post, those who win the lottery would not like to work. Implicitly, commitment or perfect enforcement is assumed here, like in every complete market model.

1.2.4 Divisible vs indivisible labor

To further understand the implications of micro-indivisibility for aggregate labor supply, consider a version of problem (8) with log utility over consumption and where the aggregate labor supply elasticity is η. From the intratemporal FOC, we have

$$\frac{z_t}{C_t} F\left(K_t, H_t\right) = AH_t^{1/\eta}.$$

[1]This aggregation result does not depend on agents starting with the same (zero) wealth as in the static model we solved above. If agents differ by initial wealth and have isoelastic utility (and we still have complete markets), we can still define a representative agent even if consumption and lottery decisions differed among agents with different levels of initial wealth.

Assuming a Cobb-Douglas aggregate production function, we obtain

$$\frac{z_t}{C_t} K_t^{\alpha} = A H_t^{(1/\eta + \alpha)}$$

and taking logs and first differences

$$\Delta \log H_t = \frac{1}{1/\eta + \alpha} \Delta \log z_t + \frac{\alpha}{1/\eta + \alpha} \Delta \log K_t - \frac{1}{1/\eta + \alpha} \Delta \log C_t$$

which shows that, in equilibrium (or equivalently, in the planner's solution), fluctuations in hours due to aggregate shocks are going to be much bigger if $\eta = \infty$.

1.3 Indivisible labor with imperfect insurance

The starting point is the paper by Chang and Kim (2006). Consider the neoclassical growth model with incomplete markets (as in Aiyagari) in its version with aggregate shocks (as in Krusell-Smith) with two important differences:

1. Individual labor supply is indivisible: individuals can work an indivisible amount of hours $h_t \in \{0, \bar{h}\}$, but there are no lotteries and no markets to insurance unemployment risk.

2. A household is composed by a couple (husband and wife) who pool income and maximize a household objective function where consumption is a public good for the couple (e.g., the house). The main role of this assumption is to add a source of consumption insurance beyond self-insurance through borrowing and saving.

1.3.1 A static example

Consider an economy where individuals can either work $h_i \in \{0, \bar{h}\}$ and where there is a distribution of wealth in the population (like in a typical incomplete markets model) which induces a distribution of reservation wages in the population. Call this distribution $\Phi(w^*)$, i.e., the fraction of the population who would work at wage w is $\Phi(w)$ since they have reservation wage below w. In other words, the aggregate labor supply curve is

$$H(w) = \int_0^w \bar{h} d\Phi(w^*) = \bar{h}\Phi(w)$$

and the aggregate elasticity at the prevailing wage w is

$$\eta(w) = \frac{dH/H}{dw/w} = \frac{\Phi'(w)/\Phi(w)}{1/w} = \frac{\Phi'(w)}{\Phi(w)}w$$

which depends on the concentration (the ratio of the density to the pdf) of workers at w. If the population is concentrated near w, then the elasticity is very large. Note that in Rogerson's model, for every worker, the reservation wage is the current wage since workers are ex-ante indifferent between working and not working, and hence this elasticity would be infinity.

1.3.2 The full-blown dynamic model

Demographics: the economy is populated by a measure one of infinitely lived households. A household is a couple of two individuals of gender $i \in (m, f)$ who pool income together and consume a good which is public within the household (e.g., a house).

Preferences: Intra-period utility over consumption and hours worked (h_m, h_f) are given by $u(c, h_m, h_f)$. Hours are indivisible, i.e. $h_i \in \{0, \bar{h}\}$. The employment status of an individual is

$$s_i = \begin{cases} e & \text{if } h_i = \bar{h} \\ n & \text{if } h_i = 0 \end{cases}$$

Therefore the household is in one of four possible employment states: $(e, e), (e, u), (u, e), (u, u)$. Couples discount the future at rate $\beta \in (0, 1)$.

Technology: The economy produces output through an aggregate CRS production technology

$$Y = zF(K, H)$$

where H is the aggregate of all hours-weighted efficiency units in the population. z is the aggregate productivity shock which follows a Markov chain.

Endowments: Household productivity is the pair $(\varepsilon_m, \varepsilon_f)$ which follows a joint Markov process independent of z. A household who chooses not to work can produce at home an amount b of the final good which can be exchanged in the market.

Markets: They are all competitive. The only asset traded is a claim to physical capital. Individuals can borrow up to an exogenously given limit $\underline{a}$.

We now turn to the household problem. The individual states are $(a, \varepsilon_m, \varepsilon_f)$ and the aggregate states are (z, μ), where μ is the measure of agents. The law of motion for the distribution is $\mu' = \Gamma(\mu, z)$.

Each period, the household observes its own asset holdings a, the pair of productivity $(\varepsilon_m, \varepsilon_f)$, aggregate states (z, μ) and makes its own labor supply decision $\left(s_m^*, s_f^*\right)$ by choosing

$$V^*(a, \varepsilon_m, \varepsilon_f; z, \mu) = \max\left\{V_{ee}(a, \varepsilon_m, \varepsilon_f; z, \mu), V_{en}(a, \varepsilon_m, \varepsilon_f; z, \mu), V_{ne}(a, \varepsilon_m, \varepsilon_f; z, \mu), V_{nn}(a, \varepsilon_m, \varepsilon_f; z, \mu)\right\}$$

Conditional on the employment decision, the household chooses consumption/saving. For example, the household problem for the (e, e) couple, in recursive formulation is:

$$\begin{aligned} V_{ee}(a, \varepsilon_m, \varepsilon_f; z, \mu) &= \max_{c, a'} \; u\left(c, \bar{h}, \bar{h}\right) + \beta E\left[V^*\left(a', \varepsilon_m', \varepsilon_f'; z', \mu'\right) | \varepsilon_m, \varepsilon_f, z, \mu\right] \\ & \quad s.t. \\ c + a' &= w\left(\varepsilon_m \bar{h} + \varepsilon_f \bar{h}\right) + Ra \\ a' &\geq -\underline{a} \end{aligned}$$

The household problem for the (n, n) couple in recursive formulation is:

$$\begin{aligned} V_{nn}(a, \varepsilon_m, \varepsilon_f; z, \mu) &= \max_{a'} \; u(c, 0, 0) + \beta E\left[V^*\left(a', \varepsilon_m', \varepsilon_f'; z', \mu'\right) | \varepsilon_m, \varepsilon_f, z, \mu\right] \\ & \quad s.t. \\ c + a' &= 2b + Ra \\ a' &\geq -\underline{a} \end{aligned}$$

The solution of the labor supply problem for spouse i is a policy function

$$s_i^*(a, \varepsilon_m, \varepsilon_f; z, \mu) \in \{e, n\}.$$

We can represent the labor supply decision also as a reservation productivity level (wage) $\varepsilon_m^*(a, \varepsilon_f; z, \mu)$ such that

$$s_m^* = \begin{cases} e & \text{if } \varepsilon \geq \varepsilon_m^*(a, \varepsilon_f; z, \mu) \\ n & \text{if } \varepsilon < \varepsilon_m^*(a, \varepsilon_f; z, \mu) \end{cases}$$

Intuitively, we have:

$$\frac{\partial \varepsilon_m^*(a, \varepsilon_f; z, \mu)}{\partial a} \geq 0, \frac{\partial \varepsilon_m^*(a, \varepsilon_f; z, \mu)}{\partial \varepsilon_f} \geq 0$$

where the first inequality descends from the fact that leisure is a normal good. The second will hold because there are gains from specialization: the spouse who "produces leisure" for the household must be the one with the lowest productivity.

We omit the definition of recursive competitive equilibrium with aggregate shocks, and the algorithm to compute this equilibrium (but you should work this out on your own).

Chang and Kim specify utility as:

$$u(c, h_m, h_f) = \log c - A_m \frac{h_m^{1+1/\eta}}{1+1/\eta} - A_f \frac{h_f^{1+1/\eta}}{1+1/\eta}$$

which is consistent with balanced growth (check). Recall η is the Frisch elasticity. In order to be consistent with the micro evidence, they set $\eta = 0.40$. This value implies a Frisch consistent with the micro estimates when it is estimated from an artificial panel of individuals generated from the model. The variation in hours in response to wages comes from the fact that they simulate the model at a quarterly frequency but they estimate this regression at an annual frequency (like most of the micro research does).

What is the aggregate Frisch elasticity implied by this model? One way to answer this question is to simulate the heterogeneous-agent incomplete-markets model, generate aggregate time series of $\{Y, C, I, H\}$ and compare it to a representative agent model with preferences

$$u(C, H) = \ln C - A\frac{H^{1+1/\gamma}}{1+1/\gamma}.$$

In terms of amplitude of aggregate fluctuations of $\{Y, C, I, H\}$, the heterogeneous agent incomplete-markets model reproduces those of a representative agent model with $\gamma = 2$. Therefore, the aggregate Frisch is 5 times as large as the micro Frisch in this model.

This result is reminiscent of Rogerson's result, but it is not as extreme. Why? Even though there are no contracts that can be explicitly traded to insure unemployment risk, in this economy self-insurance through borrowing/saving and though spousal labor supply provides good hedging against wage risk. So, we are somewhat close to the complete markets case, something that we already knew form Krusell and Smith.

1 Life-cycle economies

In this section, we introduce an explicit life-cycle dimension into the incomplete markets model. To close the model, we need to design an economy with generations that overlap so that the population structure is constant. We will use this model to study the evolution of inequality over the life cycle. Here are some basic facts we should keep in mind.

1. Wage inequality rises almost linearly. The variance of log wages rises by 0.30 point from age 25 to age 55.

2. Consumption inequality also rises over the life-cycle, but the increase is smaller, around 1/3 of the rise in wage inequality.

3. Hours inequality is fairly flat over the life cycle.

Before getting to the incomplete markets model, we will try to explain these life-cycle inequality patterns as a complete markets allocation.

1.1 Complete markets

We follow Storesletten, Telmer and Yaron, (2001). Consider the following economic environment. Time is discrete and indexed by $t = 0, 1, ...$ and the economy is stationary.

Demographics: The economy is populated by $J + 1$ overlapping cohorts of individuals. Each cohort is born with measure one and it is indexed by their birth date τ. Individuals indexed by i are born at age $j = 0$ and die for sure after reaching age J. They survive up to age j with probability φ^j. By definition, age j is equal to $t - \tau$, i.e., current date minus birth date.

Preferences: Intratemporal utility, as a function of consumption and hours, of an agent i is given by:

$$u(c, h) = \frac{c^{1-\gamma}}{1-\gamma} - A\frac{h^{1+\sigma}}{1+\sigma}.$$

Note that $1/\gamma$ is the intertemporal elasticity of substitution and $1/\sigma$ is the intertemporal labor supply elasticity, or Frisch labor supply elasticity.

Endowments: Individual productivity has a deterministic cohort- and age-specific component $\kappa_{j\tau}$ and a stochastic, idiosyncratic component $\varepsilon_{i,j,\tau}$ which could be, also, cohort specific. For example, the wage-age profile of some cohorts can start higher and be steeper than others. Or, the variance of the shocks $\varepsilon_{i,j,\tau}$ could be larger for some cohorts than for others. These two components (κ, ε) are multiplicative.

Technology: The production technology is linear in aggregate efficiency units of labor, $Y = zN$, where N are the aggregate efficiency-weighted hours worked, i.e., the sum over the population of $n_{i,j,\tau} = \kappa_{j\tau}\varepsilon_{i,j,\tau}h_{i,j,\tau}$.

We wish to use a planner's problem to characterize the complete markets allocations. Consider a planner with social welfare function putting positive weights only on individuals alive today only (i.e., the planner puts zero weight on all future cohorts). Its social welfare function is:

$$\max_{\{c_{ij\tau},h_{ij\tau}\}} \sum_{\tau=t-J}^{t} \sum_{j=t-\tau}^{J} \beta^{j-(t-\tau)}\varphi^{j} \int \lambda_{i\tau}\left[\frac{c_{ij\tau}^{1-\gamma}}{1-\gamma} - A\frac{h_{ij\tau}^{1+\sigma}}{1+\sigma}\right] d\mu_{\tau}$$

where $\lambda_{i\tau}$ is the weight put on individual i of cohort τ and μ_τ is the cross-sectional distribution (i.e., it integrates to one) of agents in cohort τ.

The aggregate resource constraint at every date t obeys the equation:

$$0 = \sum_{\tau=t-J}^{t} \varphi^{t-\tau} \int \left(z\kappa_{t-\tau,\tau}\varepsilon_{i,t-\tau,\tau}h_{i,t-\tau,\tau} - c_{i,t-\tau,\tau}\right) d\mu_{\tau}.$$

Denote by θ_t the Lagrange multiplier on the resource constraint at date t.

At date t, the planner's FOC with respect to consumption of individual i belonging to birth cohort τ yields:

$$\varphi^{t-\tau}\lambda_{i\tau}c_{i,t-\tau,\tau}^{-\gamma} = \varphi^{t-\tau}\theta_t \tag{1}$$

which establishes that the ratio of marginal utility of consumption between any two agents (i,τ) and (i',τ') is constant over time and equal to the ratio of planner weights, which is the definition of complete markets.

Taking logs of (1), using the notation $j = t - \tau$ and rearranging, we arrive at:

$$\begin{aligned}
\log \lambda_{i\tau} - \gamma \log c_{i,j,\tau} &= \log \theta_t \\
\log c_{i,j,\tau} &= \frac{1}{\gamma} \left[\log \lambda_{i\tau} - \log \theta_t \right].
\end{aligned}$$

The cross-sectional variance of log consumption across agents i conditional on age $j = t - \tau$, given by

$$var_j \left(\log c_{ij\tau} \right) = \frac{1}{\gamma^2} var \left(\log \lambda_{i\tau} \right)$$

is independent of age (albeit it could depend on cohort). Hence, we have reached three results:

1. Consumption inequality within a cohort can be positive only if the planner's weights differ across agents belonging to the same cohort. Remember, from the Negishi approach, that weights differ if initial endowments of wealth differ, for example.

2. Consumption inequality can vary across cohorts, if the distribution of planner weights differ across cohorts.

3. Consumption dispersion cannot rise over the life cycle under complete markets with separable preferences. Indeed, the discrepancy between the growth in consumption dispersion and the growth of wage dispersion can be interpreted as a metric of risk sharing in the economy. The larger this gap, the lower risk sharing.

From the intratemporal FOC for an agent i of cohort τ at date t is:

$$\lambda_{i\tau} A h_{i,t-\tau,\tau}^{\sigma} = \theta_t z \kappa_{t-\tau,\tau} \varepsilon_{i,t-\tau,\tau}. \tag{2}$$

Rearranging (2), taking logs, and using $j = t - \tau$ we reach:

$$\log h_{i,j,\tau} = \frac{1}{\sigma} \log \left(\frac{z\theta_t}{A} \right) - \frac{1}{\sigma} \log \left(\lambda_{i\tau} \right) + \frac{1}{\sigma} \log \varepsilon_{i,j,\tau} + \frac{1}{\sigma} \log \kappa_{j\tau}.$$

The cross-sectional variance of log hours across agents i for cohort τ conditional on age j is:

$$var_j \left(\log h_{ij\tau} \right) = \frac{1}{\sigma^2} var \left(\log \lambda_{i\tau} \right) + \frac{1}{\sigma^2} var_j \left(\log \varepsilon_{ij\tau} \right)$$

which shows that in complete markets hours inequality grows over the life cycle if the unconditional variance of the shock grows over the life cycle. Empirically, as discussed earlier, the variance of hourly wages increases linearly over the life-cycle which implies that $\varepsilon_{ij\tau}$ has a permanent component, e.g.,

$$\log \varepsilon_{i,j,\tau} = \log \varepsilon_{i,j-1,\tau} + v_{i,j,\tau}$$

which implies

$$var_j \left(\ln \varepsilon_{ij\tau} \right) = j \cdot var \left(v_{ij\tau} \right).$$

If wage inequality grows steeply over the life cycle, then hours inequality must grow too, unless σ is very large, i.e., labor supply elasticity is small. But if the Frisch elasticity is close to zero, then the variance of hours worked would be close to zero too, while in the data is positive and quite large (even net of measurement error).

We conclude that the full insurance model has counterfactual implications about the path of inequality over the life cycle. Before examining the incomplete markets model, let's take a look at the other extreme benchmark, autarky.

1.2 Autarky

In autarky, i.e. no insurance and no storage, consumption for every individual is equal to its earnings, or:

$$c_{i,j,\tau} = \kappa_{j\tau} \varepsilon_{i,j,\tau} h_{ij\tau} \tag{3}$$

and hours worked satisfy the standard intratemporal first-order condition:

$$c_{i,j,\tau}^{-\gamma} \kappa_{j\tau} \varepsilon_{i,j,\tau} = A h_{ij\tau}^{\sigma}. \tag{4}$$

Combining these two equations, we obtain

$$\begin{aligned} c_{i,j,\tau}^{-\gamma} \kappa_{j\tau} \varepsilon_{i,j,\tau} &= A c_{i,j,\tau}^{\sigma} \left(\kappa_{j\tau} \varepsilon_{i,j,\tau} \right)^{-\sigma} \\ c_{i,j,\tau} &= \left(\frac{1}{A} \right)^{\frac{1}{\sigma+\gamma}} \left(\kappa_{j\tau} \varepsilon_{i,j,\tau} \right)^{\frac{1+\sigma}{\sigma+\gamma}} \end{aligned}$$

which implies that in the cross section

$$var_j(\log c_{ij\tau}) = \left(\frac{1+\sigma}{\sigma+\gamma}\right)^2 var_j(\log \varepsilon_{ij\tau})$$

and consumption dispersion grows over the life cycle, indeed, potentially even more than wage dispersion, if $\gamma < 1$. For consumption inequality to grow slower than wage inequality, γ must be higher than one.

We now turn to hours inequality. Using (3) into the first-order condition (4):

$$\begin{aligned} (\kappa_{j\tau}\varepsilon_{i,j,\tau}h_{ij\tau})^{-\gamma}\kappa_{j\tau}\varepsilon_{i,j,\tau} &= Ah_{ij\tau}^{\sigma} \\ h_{ij\tau} &= \left(\frac{1}{A}\right)^{\frac{1}{\sigma+\gamma}} (\kappa_{j\tau}\varepsilon_{i,j,\tau})^{\frac{1-\gamma}{\sigma+\gamma}} \end{aligned}$$

which implies that, in the cross-section,

$$var_j(\ln h_{ij\tau}) = \left(\frac{1-\gamma}{\sigma+\gamma}\right)^2 var_j(\ln \varepsilon_{ij\tau}).$$

Therefore, depending how far γ is from 1, hours can rise, fall or be flat. If γ is sufficiently higher than 1, then consumption inequality will grow less than wage inequality and hours inequality can also be relatively flat.

Clearly autarky is an extreme. The key question is: does a plausibly calibrated incomplete-markets model (with a financial market structure in between autarky and complete markets) generate the observed increase in consumption inequality?

1.3 Lifecycle incomplete markets economy

This section is based on Storesletten, Telmer and Yaron (2004). Consider the following economy. Time is discrete and indexed by $t = 0, 1, ...$ and the economy is stationary.

Demographics: The economy is populated by overlapping cohorts of individuals. Each cohort of newborn agents has measure one. Individuals of a cohort are born at age $j = 0$ and die for sure upon reaching age J. They survive up to age j with probability φ^j. It is also useful to denote the conditional survival probability between age j and age $j+1$ with $\varphi_j = \varphi^j/\varphi^{j-1}$.

Preferences: Intra-period utility is given by $u(c_j)$ with $u' > 0, u'' < 0$. We abstract from labor supply.

Endowments: Individual productivity endowments are the sum (in log) of a deterministic age component, plus an innate ability component, plus two stochastic components:

$$\begin{aligned}\log y_{ij} &= \kappa_j + \varepsilon_{ij} \\ \varepsilon_{ij} &= \rho\varepsilon_{i,j-1} + \omega_{ij}\end{aligned}$$

where ε_{ij} is iid over time and η_{ij} follows an AR(1). Let's discretize all pieces and let $\pi(\varepsilon', \varepsilon)$ be the relevant conditional distribution of the shosck. Agents go through two phases of the life-cycle: work and retirement. Assume that

$$\varepsilon_{ij} = 0 \text{ for } j \geq J^{ret}$$

where age J^{ret} denotes mandatory retirement. New cohorts of agents are born with zero initial wealth.

Technology: Output is produced through the aggregate production function:

$$C + \delta K = Y = F(K, N)$$

Government: It taxes labor income at rate τ and finances a pay-as-you-go social security system. The social security system pays retirement benefits which depend on average lifetime earnings $\bar{y}^{ret}$, where

$$\bar{y}_i^{ret} = \frac{1}{J^{ret}} \sum_{j=0}^{J_{ret}-1} y_{ij},$$

based on a given formula given by the function $P(\bar{y}^{ret})$, with $P' > 0$ and $P'' < 0$.

The government expropriates accidental bequests of agents who die at ages $j < J$ and redistributes them across all living agents equally through a lump sum transfer ϕ.

Markets: Only one-period non state contingent bonds are traded (a mix of claims to physical capital and private IOUs). Workers can borrow up to $-\underline{a}$ and retirees cannot borrow. Asset and goods market are competitive.

1.3.1 Household problem

The problem of the household during working age can be written in recursive formulation as:

$$
\begin{aligned}
V_j\left(\varepsilon_j, a_j, \bar{y}_j\right) &= \max_{\{c_j, a_{j+1}\}} u\left(c_j\right) + \beta \varphi_{j+1} E_j\left[V_{j+1}\left(\varepsilon_{j+1}, a_{j+1}, \bar{y}_{j+1}\right)\right] \\
& \quad s.t. \\
c_j + a_{j+1} &= R a_j + (1-\tau) w \exp\left(\kappa_j + \varepsilon_j\right) + \phi \\
a_{j+1} &\geq -\underline{a} \\
\bar{y}_{j+1} &= \bar{y}_j + \frac{w \exp\left(\kappa_{j+1} + \varepsilon_{j+1}\right)}{J^{ret}}
\end{aligned}
$$

where here we abuse a bit notation and we set, implicitly, the continuation value

$$
V_{J^{ret}}\left(0, 0, 0, a_{J^{ret}}, \bar{y}_{J^{ret}}\right) = \tilde{V}_{J^{ret}}\left(a_{J^{ret}}, \bar{y}^{ret}\right).
$$

During retirement, the household problem becomes

$$
\begin{aligned}
\tilde{V}_j\left(a_j, \bar{y}^{ret}\right) &= \max_{\{c_j, a_{j+1}\}} u\left(c_j\right) + \beta \varphi_{j+1} \tilde{V}_{j+1}\left(a_{j+1}, \bar{y}^{ret}\right) \\
& \quad s.t. \\
c_j + a_{j+1} &= R a_j + P\left(\bar{y}^{ret}\right) + \phi \\
a_{j+1} &\geq 0
\end{aligned}
$$

Also, note that since there is no uncertainty for the retirees, there is no expectation in the Bellman equation.

1.3.2 Stationary equilibrium

Let $s \equiv \left(\varepsilon_j, a_j, \bar{y}\right)$ be the vector of states for the worker and $\tilde{s} \equiv \left(a_j, \bar{y}^{ret}\right)$ the vector of states for the retiree. A stationary equilibrium is a collection of: 1) decision rules $\left\{c_j(s), a_{j+1}(s)\right\}$ for workers and $\left\{\tilde{c}_j(\tilde{s}), \tilde{a}_{j+1}(\tilde{s})\right\}$ for retirees, 2) value functions $\left\{V_j(s), \tilde{V}_j(\tilde{s})\right\}$, 3) prices $\{w, R\}$, 4) aggregate quantities $\{K, N\}$, 5) tax rate τ (e.g., given the lump-sum transfer ϕ), and 6) stationary measures $\left\{\mu_j, \tilde{\mu}_j\right\}$ such that:

- The decision rules are the solution to the household problem and satisfy the associated value functions.

- Input prices equal marginal products of capital and labor.

- The labor market clears

$$N = \sum_{j=0}^{J^{ret}-1} \varphi^j \int \exp\left(\kappa_j + \varepsilon_j\right) d\mu_j.$$

- The capital market clears:

$$K = \sum_{j=0}^{J^{ret}-1} \varphi^j \int a_{j+1}(s)\, d\mu_j + \sum_{j=J^{ret}}^{J} \varphi^j \int \tilde{a}_{j+1}(\tilde{s})\, d\tilde{\mu}_j.$$

- The government budget is balanced

$$\tau w N = \sum_{j=J^{ret}}^{J} \varphi^j \int P\left(\bar{y}^{ret}\right) d\tilde{\mu}_j$$

and the government rebates all the assets of the deceased

$$\sum_{j=0}^{J^{ret}-1} \varphi^j \left(1-\varphi_{j+1}\right) \int a_{j+1}(s)\, d\mu_j + \sum_{j=J^{ret}}^{J} \varphi^j \left(1-\varphi_{j+1}\right) \int \tilde{a}_{j+1}(\tilde{s})\, d\tilde{\mu}_j = \phi \sum_{j=0}^{J} \varphi^j$$

- The goods market clears

$$\sum_{j=0}^{J^{ret}-1} \varphi^j \int c_j(s)\, d\mu_j + \sum_{j=J^{ret}}^{J} \varphi^j \int \tilde{c}_j(\tilde{s})\, d\tilde{\mu}_j + \delta K = F(K,N)$$

- The distributions $\left\{\mu_j, \tilde{\mu}_j\right\}$ are stationary. For example, for $j < J^{ret}$:

$$\mu_{j+1} = \int_S Q_j\left(s, \mathcal{S}\right) d\mu_j$$

where

$$Q_j\left(\left(\varepsilon, a, \bar{y}\right), \mathcal{E} \times \mathcal{A} \times \bar{\mathcal{Y}}\right) = I_{\{a_{j+1}(s) \in \mathcal{A}\}} \cdot \sum_{\varepsilon_{j+1} \in \mathcal{E}} \pi_{\varepsilon}\left(\varepsilon_{j+1}, \varepsilon_j\right) \cdot I_{\left\{\bar{y}_j + \frac{w \exp\left(\kappa_{j+1} + \varepsilon_{j+1}\right)}{J^{ret}} \in \bar{\mathcal{Y}}\right\}}$$

1.3.3 Solution method

We first guess $\{R^0, \phi^0\}$. From R^0 we obtain the age w^0, the capital stock K^0 and the aggregate labor input N^0 (exogenous). We don't need to guess τ^0 as well because the distribution of $\bar{y}_j$ across agents is exogenous, so from the balanced budget of the government, given our guesses, we can recover τ^0.

Once we have all we need in the agent's budget constraint, we can solve the household problem. We do it backward, starting from the last period in retirement which is a static problem, all the way to age $j = 0$. Then we use the $j + 1$ decision rule in the Euler equation for age j (like in the transitional dynamics).

The simulation step is the same as always, with the caveat that we need to respect the demographic structure. From the aggregation of wealth among agents alive at date t, we obtain A^0 (and the implied wealth of the deceased) which we need to compare to K^0 and ϕ^0.

1.3.4 What matters for the rise of consumption inequality over the life cycle?

Storesletten, Telmer and Yaron (2004) show that this model does a good job in matching the lifecycle profile of consumption inequality, given wage inequality as an exogenous input. In particular, the key determinants of risk sharing in this economy are:

- **Financial wealth:** the amount of financial wealth held in the economy is key because the larger is financial wealth, the smaller is the impact of earnings shocks on consumption. The intuition is that agents consume out of human wealth and financial wealth and, the larger is the latter relative to the former, the less earnings shocks impact consumption. Usually β is set so that $K/Y = 3.5$ to reproduce this same ratio for the US economy. For example, reducing β such that K/Y is 1.5 increases the rise of consumption inequality over the life cycle substantially.

- **Social security**: Social security redistributes across generations, but also within generations because P is concave, so it provides some additional insurance. More social insurance means less growth in consumption inequality.

1.4 Aggregate shocks in life cycle economies

This section is based on Krueger and Kubler (2003). Suppose we add to the model an aggregate productivity shock, i.e.

$$Y = zF(K, N)$$

with $z \in Z$ stochastic which follows the Markov chain $\pi(z', z)$.

Let's get rid of all idiosyncratic uncertainty in ε so that we have one type of agent for each age. Then, the full aggregate state of the economy is z together with the $(J+1)$-dimensional wealth vector $\bar{a} = (a_0, a_1, ..., a_J)$. Note that now $\bar{y}$ is a deterministic function of age, as labor income is deterministic.

The Euler equation for an agent of age j is

$$u_c\left(w(z,K)\kappa_j + R(z,K)a_j - a_{j+1}(a_j; z, \bar{a})\right) \geq$$
$$\beta \sum_{z' \in Z} u_c\left(w(z',K')\kappa_{j+1} + R(z',K')a_{j+1}(a_j; z, \bar{a}) - a_{j+2}\left(a_{j+1}(a_j; z, \bar{a}); z', \bar{a}'\right)\right)\pi(z', z)$$

with

$$\begin{aligned} K &= \sum_{j=0}^{J} a_j \\ K' &= \sum_{j=0}^{J} a_{j+1}(a_j, z, \bar{a}) \\ \bar{a}' &= \begin{bmatrix} a_0 \\ a_1(a_0, z, \bar{a}) \\ \dots \\ a_J(a_{J-1}, z, \bar{a}) \end{bmatrix} = \Gamma(\bar{a}, z) \end{aligned}$$

and so we have all the pieces to compute the Euler equation.

Can we use the Krusell-Smith approach implemented as in the infinite horizon model? I.e., can we instead of keeping track of the entire J dimensional vector $\bar{a}$ just keep track of average capital K and use an approximate law of motion for K only of the type

$$\ln K' = b_0^z + b_1^z \ln K$$

as in the original Krusell-Smith economy? Recall that, in that economy with infinite horizon, most of the agents have the same marginal propensity to save and consume, i.e., the saving decision rules are linear for a wide range. In a life-cycle economy, instead, because of the finite horizon (and no bequest motive), elderly agents have much lower propensity to save than young agents who have to save to finance their retirement. Therefore, keeping track of the mean of the asset distribution alone leads to large forecasting errors, i.e. in some cases R^2 can be as low as 0.66 and quasi aggregation fails badly. The key is the enormous heterogeneity in saving rates across agents of different ages.

This calls for a simplification: instead of keeping track of a $J+1$ dimensional vector, since we know that wealth is hump shaped by age, we can approximate the wealth distribution by, for example, a cubic polynomial of age, i.e., 4 parameters instead of $J+1$.

1 Industry equilibrium

We develop an equilibrium model of an industry with many plants. The model is based on Hopenhayn (1992), and Hopenhayn and Rogerson (1993). The building blocks of the model are as follows:

- The industry is competitive and produces a homogeneous good. The industry is "small", so it takes the wage and the interest rate as given. The equilibrium of the industry determines the price and quantity of the good and the amount of labor hired in the industry.
- Instead of modelling a household sector explicitly, we'll assume that there is an exogenously given demand function for the good produced in the industry.
- Plants in the industry operate a decreasing returns to scale technology and are subject to productivity shocks. They hire labor as their only input. They pay a fixed operating cost.
- Every period incumbent firms choose whether to exit the market. There is free entry into the industry, subject to paying a fixed entry cost.

Jovanovic (1982) is another seminal paper on industry equilibrium which characterizes the evolution of an industry where costs are random and different among firms. Firms all have the same prior belief about costs when they enter the industry, and learn over time about their own true cost.

Applications of the model: One can use the model to study, for example:

- *Plant turnover* (in equilibrium some plants die, others enter) and *job turnover* from the growing firms (those with good shocks) to shrinking firms (those with bad shocks).
- The size distribution of firms: in the model, the large firms are those who had long sequences of good shocks.
- Effects of policies [e.g. a wage subsidy, a profit tax, a firing tax] on the size distribution, plant turnover, average profits, average productivity, etc...

- The diffusion of technologies across plants. If, in the model, the technology diffuses from young/small plants to old/large plants then we can replicate "S-shaped" diffusion curves.

- The role of financing constraints for firm's growth.

- Trade dynamics and export decisions of firms.

1.1 The Economy

Household sector: We represent the household sector simply through a demand function $D(p)$ where p is the price of the good produced by the plants in the industry and $D' < 0$.

Plant-level production technology: Each plant produces the homogenous good with technology

$$y = zf(n),$$

where $f' > 0$, $f'' < 0$ and $f(0) = 0$. The term $z \in Z$ denotes the plant-level idiosyncratic productivity shock which follows the continuous process $\Gamma(dz', z) = \Pr\{z_{t+1} \in dz' | z_t = z\}$.

Problem of an incumbent plant: Incumbent plants (i.e., those producing in the current period) incur in the per-period fixed cost ϕ in order to operate (e.g., they hire one unit of managerial time every period at cost ϕ). They hire labor at the wage ω.

Let the profits of the plant be denoted by $\pi(z)$. A plant that takes $\{p, \omega\}$ as given, solves:

$$\pi(z) = \max_n \{pzf(n) - \omega n - \phi\}. \tag{1}$$

Note that the profit function is increasing in z. Note also that $\omega = pzf_n(n)$. Competitive labor markets means that in equilibrium all plants face the same wage, so the more productive plants hire more labor. For example, suppose $f(n) = n^{\alpha}$. Then

$$pz\alpha n^{\alpha-1} = \omega \Rightarrow n(z;p) = \left(\frac{zp\alpha}{\omega}\right)^{\frac{1}{1-\alpha}}.$$

Therefore, take two plants with productivity z_i and z_j. The ratio between their size will be directly proportional to the productivity ratio, i.e.

$$\frac{n(z_j;p)}{n(z_i;p)} = \left(\frac{z_j}{z_i}\right)^{\frac{1}{1-\alpha}}.$$

So, once the decreasing return parameter α is known, by observing the size distribution one can fully recover the productivity distribution.

Exit decision: At the beginning of every period, before realizing the productivity shock z, the plant decides to exit. The value of exit is normalized to zero, i.e. the plant has no "scrap value". The value of an incumbent plant is therefore:

$$v(z;p) = \pi(z;p) + \frac{1}{1+r} \max \left\{ \int_Z v(z';p) \Gamma(dz', z), 0 \right\} \tag{2}$$

where the first argument of the max operator is the expected continuation value of the incumbent firm. Since profits are increasing in z and Γ is monotone, the value function v is also increasing in z. Let $\chi(z;p) = \{0, 1\}$ be the exit decision. The exit decision involves a reservation rule. There exists a threshold productivity level $z^*(p)$ such that, for all $z < z^*(p)$, the firm will decide to exit. In equilibrium, incumbent firms may incur in negative profits temporarily and keep operating, if the shock is mean reverting.

Problem of an entrant plant: An entrant plant must pay the fixed cost κ for one period (e.g., hire the manager to set-up the plant) and then it draws its initial productivity z' from the distribution $G(dz')$. The value of an entrant plant is, therefore,

$$v^e(p) = -\kappa + \frac{1}{1+r} \int_Z v(z';p) G(dz'). \tag{3}$$

The entry decision is simple: a plant should enter as long as $v^e(p) \geq 0$. Free entry of plants will guarantee that, in equilibrium, $v^e(p) = 0$.[1] Let m denote the measure of entrant plants, i.e. those preparing to produce but not yet producing, hence not yet hiring labor.

1.2 Equilibrium

A *stationary recursive competitive equilibrium* for this industry is a list of: plants' decision rules $\{n, \chi\}$, value functions $\{v, v^e\}$, price $\{p\}$, an invariant measure of incumbent firms λ, and a measure of entrant plants m, such that:

- Given p, $n(z;p)$ solves the static hiring decision (1).

[1] Depending on the parameter values, in particular κ and ϕ, there may be equilibria where $v^e(p) < 0$, so there is no entry and no exit.

- Given p, $\chi(z;p)$ solves the exit decision (2) of the incumbent firm, and $v(z,p)$ is the associated value function

- Free entry of firms implies that $v^e(p) = 0$, i.e., at the equilibrium price p

$$\kappa = \frac{1}{1+r} \int_Z v(z';p)\, G(dz')$$

- The good market clears

$$D(p) = \int_Z z f(n(z;p))\, \lambda(dz;p) \tag{4}$$

- Let dz' be a generic set of the Borel sigma algebra on the state space Z. Then, the invariant measure of incumbent plants solves

$$\lambda(dz';p) = \int_Z \Gamma(dz',z)\,[1-\chi(z)]\,\lambda(dz;p) + mG(dz'). \tag{5}$$

  Note that the measure λ is linearly homogenous in m. To see this, suppose that Z is a discrete set with I values, then G is a vector and λ is a vector that satisfies

$$\underset{(I\times 1)}{\lambda} = \underset{(I\times I)}{\Gamma} \otimes \underset{(I\times I)}{X} \cdot \underset{(I\times 1)}{\lambda} + m \underset{(I\times 1)}{G}$$

  where Γ is the transition matrix for z, X is a matriz where each line is a collection of either 0 or 1 depending on the value of $1-\chi(z)$, and the symbol $\otimes$ denotes the element by element product. The above linear system of equations has solution

$$\lambda = m\left[\left(I - \tilde{\Gamma}\right)^{-1} G\right]$$

  where $\tilde{\Gamma} = \Gamma \otimes X$. This equation shows clearly that λ is homogeneous in m, the number of entrants.

1.3 Solution method

The key difference with respect to the computation of the equilibrium in Bewley models with heterogeneous households is that in models of industry equilibrium, we also need to determine the *number of plants,* which is endogenous. A fixed-point algorithm to solve for the equilibrium is as follows.

1. Guess a price p_0 and solve the static hiring problem (1) for the hiring decision $n(z, p_0)$. Let

$$\pi(z; p_0) = p_0 z f(n(z; p_0)) - \omega n(z; p_0) - \phi$$

2. Given the profit function $\pi(z; p_0)$, one can look for the fixed point of the Bellman equation

$$v(z; p_0) = \pi(z; p_0) + \frac{1}{1+r} \max\left\{ \int_Z v(z'; p_0) \Gamma(dz', z), 0 \right\}$$

and from this step, obtain $\{\chi(z; p_0), v(z; p_0)\}$. Note that $v(z; p_0)$ is increasing in p_0 since the flow profits are increasing in p_0.

3. Verify if the price p_0 satisfies the free-entry condition

$$\kappa = \frac{1}{1+r} \max\left\{ \int_Z v(z'; p_0) G(dz'), 0 \right\}$$

If, for example, the equation above holds with the $>$ sign, then guess a new value $p_1 > p_0$. Go back to step 1) with the new guess. Continue until we find a price that satisfies the free entry condition. Now that we solved for p, all we need to determine is the pair (λ, m).

4. Exploit the linear homogeneity of λ in m. Guess a value of entrant plants m_0 and from (5) compute, either by simulation or using the transition function, the invariant measure λ.

5. Compute total sales of the industry, the RHS of equation (4) and verify that, at the equilibrium price found in step 3), total sales equal the exogenously given aggregate demand. If, for example, total sales are below demand, update your guess of entrants to $m_1 > m_0$ and go back to step 4). However, now step 4) is much simpler because we don't need to recompute λ, we just rescale the invariant measure already obtained.

1.4 An industry with firing costs

Hopenhayn and Rogerson (1993) study the impact of firing restrictions on the average productivity of the industry. Suppose, for example, that the government imposes *firing costs* that can be summarized by the function

$$g(n, n_0) = \begin{cases} \chi(n_0 - n) & \text{if } n < n_0 \\ 0 & \text{if } n \geq n_0 \end{cases}$$

i.e., the government imposes a severance payment of size χ to the firm for every workers who is laid off.

The key novelty, with respect to the previous model, is that the employment choice is dynamic and past employment n_0 is a *state variable*. Plants need to keep track of their past employment to calculate the firing cost $g(\cdot)$ associated to their employment decision.

The hiring problem becomes similar to the investment problem (with or without adjustment costs) and the value function of the incumbent is:

$$v(z, n_0; p) = \max_n \left\langle pzf(n) - \omega n - \phi - g(n, n_0) + \frac{1}{1+r} \max\left\{ \int_Z v(z', n; p) \Gamma(dz', z), -g(0, n) \right\} \right\rangle$$

Results: A firing cost will reduce labor reallocation from the low-productivity firms (which should be shrinking by shedding workers) towards the high-productivity firms (which should be expanding by hiring workers). It also prevents inefficient firms from exiting because of the large exit cost associated to firing the entire workforce. Overall, it is easy to see that this policy reduces average productivity of the industry and labor turnover. It is a source of misallocation.

References

[1] Hopenhayn, Hugo (1992); Entry, Exit and Firm Dynamics in Long Run Industry equilibrium, *Econometrica*

[2] Hopenhayn, Hugo ad Richard Rogerson (1993); Job Turnover and Policy Evaluation: A General Equilibrium Analysis, *Journal of Political Economy*

[3] Jovanovic, Boyan (1982); Selection and Evolution of Industry, *Econometrica*