1 Economies with Idiosyncratic Risk and Incomplete Markets: Stationary Equilibrium

We next study the general equilibrium of an economy in which agents face idiosyncratic risk and in which market are incomplete. 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. Consider an individual subject to exogenous income shocks. These shocks are not fully insurable because of the lack of a complete set of Arrow-Debreu contingent claims. We assume that there is only a risk-free asset (i.e. with fixed rate of return) in which the individual can save/borrow, and that the individual faces an exogenously set borrowing (liquidity) constraint. There are two reasons for savings: intertemporal substitution and precautionary motive, the latter as a self-insurance strategy to hedge against low earnings in the future. Individuals who bear a long sequence of bad shocks will have low wealth and will be close to the constraint, individuals with long realizations of good shocks will have high wealth. 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. We will show that the steady state of this economy will be characterized by an interest rate r, such that $\beta(1+r) < 1$ (recall with complete markets $\beta(1+r) = 1$), and higher aggregate savings rate (and therefore stock of capital) than the economy with complete markets we have studied in the beginning of this course.

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 $\pi(\varepsilon', \varepsilon) = \Pr\left(\varepsilon_{t+1} = \varepsilon' \mid \varepsilon_t = \varepsilon\right)$. 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 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}).

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.

[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.

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 $\delta \in (0,1)$ .The firm rents capital from the households at price r_t and thus solves

$$\max_{K_t,H_t} F(K_t, H_t) - (r_t + \delta) K_t - w_t H_t$$

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 through the concept of *Recursive Competitive Equilibrium* (RCE). Most of the requirement of this RCE definition will be standard (agents optimize, markets clear). However, in the stationary equilibrium of this economy we require the distribution of agents across states to be invariant.[2] The 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. The aggregate state of the economy is the distribution of agents across states, i.e. $\lambda(a, \varepsilon)$. 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 state space S be the Cartesian product $A \times E$ and B_S its Borel sigma-algebra (smallest $\sigma-$algebra (collection of subsets of S closed under

[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.

complementaton and union) containing all open and closed sets). The space (S, B_s) is a measurable space, and for any set $\mathcal{S} = \mathcal{A} \times \mathcal{E} \in B_s$, $\lambda(\mathcal{S})$ 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 $Q\left((a,\varepsilon), \mathcal{A} \times \mathcal{E}\right)$ as the probability that an individual with current state (a,ε) transits to the set $\mathcal{A} \times \mathcal{E}$ next period, formally $Q : S \times B_s \to [0,1]$, and

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

where $I_{\{\cdot\}}$ is the indicator function, and $a'(a,\varepsilon)$ is the optimal saving policy. We will search for a stationary distribution of agents λ. Stationarity requires that the distribution self-replicates over time, i.e., the measure of agents in $\mathcal{A} \times \mathcal{E}$ is constant over time. Letting

$$(T\lambda)(\mathcal{A} \times \mathcal{E}) = \int_{A \times E} Q\left((a,\varepsilon), \mathcal{A} \times \mathcal{E}\right) d\lambda(a,\varepsilon)$$

the stationary distribution satisfies

$$\lambda^*(\mathcal{A} \times \mathcal{E}) = (T\lambda^*)(\mathcal{A} \times \mathcal{E}) \tag{2}$$

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

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

where, for clarity, we have made explicit the dependence of prices from the distribution of agents (although, strictly speaking, it is *redundant* in a stationary environment and it can be omitted since it's just a 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 $c : S \to \mathbb{R}_+$; policies for the firm 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: $H = \int_{A\times E} \varepsilon d\lambda^* (a,\varepsilon)$,

- the asset market clears: $K = \int_{A\times E} a'(a,\varepsilon) d\lambda^* (a,\varepsilon)$,

- the goods market clears:[3] $\int_{A\times E} c(a,\varepsilon) d\lambda^* (a,\varepsilon) + \delta K = F(K,H)$,

- for all $(\mathcal{A}\times\mathcal{E}) \in B_s$, the invariant probability measure λ^* satisfies

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

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

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 (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.

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

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

[3] This condition is redundant by Walras law.

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]

This discussion can be summarized as follows: $A(-1) + K(-1) < 0$ and $\lim_{r\to\infty} A(r) + K(r) > 0$.Thus, if $A(r)$ is continuous, the excess demand function will be equal to 0 for some r, guaranteeing existence.

1.4 Continuity of $A(r)$

What follows is a sketch of the arguments used to prove continuity of $A(r)$.

1. 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).

2. We next need to establish continuity of λ_r^*in r. To do so we need to show that an invariant distribution λ_r^* exists and is unique.

To establish the existence and uniqueness of the invariant distribution we need to verify four properties: compactness of the state space and Q satisfying the 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 operator T defined above 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

[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.

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 $\pi(\varepsilon_H, \varepsilon_H) \geq \pi(\varepsilon_H, \varepsilon_L)$, 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

$$h(a, \varepsilon) = Tf = \sum_{\varepsilon' \in E} \int_A f(a', \varepsilon') Q((a, \varepsilon), da' \times \varepsilon')$$

  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:* Let $P^N(s, \mathcal{S})$ be the probability of transiting from s to $\mathcal{S}$ in N steps. MMC requires there to exist $s^* \in S$ and $\delta > 0$ and a N such that $P^N(s_{\max}, \{s : s \leq s^*\}) > \delta$ where $s_{\max} = (\bar{a}, \varepsilon_{\max})$ and $P^N(s_{\min}, \{s : s \geq s^*\}) > \delta$ where $s_{\min} = (-b, \varepsilon_{\min})$.

  Heuristically, 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.

[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.

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.

Clearly, we must have $r < \frac{1}{\beta} - 1$ in equilibrium. To see this, notice that at $r = \frac{1}{\beta} - 1$ we have $A(r) \to \infty$ while $K(r)$ is finite.

1.5 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 $c(a, \varepsilon; r^0)$.

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) One way 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.

7. Compare $K(r^0)$ with $A(r^0)$ to verify the asset market clearing condition. If $A(r^0) > (<) K(r^0)$, 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 $F_K(A(r^0), H) - \delta$ 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.

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

The model we study above is one originally studied by Aiyagari (1994). To evaluate the model's predictions, Aiyagari assigns values to model parameters as follows.

He interprets the model period as 1 year and chooses $\beta = 0.96$. The production function is Cobb-Douglas with K-share 0.36. $\delta = 0.08$. $U(c) = \frac{c^{1-\gamma}}{1-\gamma}$ with $\gamma = \{1, 3, 5\}$. The earnings process is a Markov chain with 7 states with a transition probability chosen according to the Tauchen method to mimic the moments of

$$\log(\varepsilon_t) = \rho \log(\varepsilon_{t-1}) + \sigma \left(1 - \rho^2\right)^{\frac{1}{2}} u_t, u_t \sim N(0, 1)$$

Aiyagari considers a number of values for $\sigma = \{0.2, 0.4\}$ and $\rho = 0, 0.3, 0.6, 0.9$. Estimates from PSID are usually around $\rho = .9$ and $\sigma = .2$.

Recall that when markets are complete, $r = 4.17\%$ (1/.96-1) and the savings rate is $s = \frac{\delta K}{f(K,1)} = 23.67\%$

The tables below show what these numbers are as we vary σ and ρ and γ

ρ	σ	γ	r	s
0	.2	1	4.17	23.67
0.9	**0.2**	**1**	**3.93**	**23.73**
0.9	0.4	1	3.30	25.47

ρ	σ	γ	r	s
0	.2	5	4.09	23.83
0.9	0.2	5	2.53	27.36
0.9	0.4	5	-0.35	37.63

The departures from complete markets are clearly largest when shocks are more volatile/persistent and when households have a strong motive to substitute intertemporaly. The entry that is closest to what has been argued reasonable for US is in bold typeface.

1.5.2 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, but in general it is quantitatively bigger.

1.6 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, 3) modify appropriately the income process.

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 (1997) 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.

1.7 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 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.

With leisure, the new equilibrium condition in the labor market becomes

$$H = \int_{A\times E} \varepsilon h(a, \varepsilon) d\lambda^*.$$

The government budget constraint (balanced in equilibrium) reads

$$T = \tau w H,$$

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).

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

$$\max_{\tau} W(\tau) = \int_{A\times E} u(c(a, \varepsilon; \tau), 1 - h(a, \varepsilon; \tau)) d\lambda^*(\tau),$$

with the constraint that the allocations are a competitive equilibrium (it is a Ramsey taxation problem).[6]

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 insurance is very good 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.

[6] This welfare function maximizes also welfare from an ex-ante point of view, i.e. "under the veil of ignorance" if one assumes that the individual's initial assets and productivity shocks are drawn from the stationary distribution λ^*.

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. Often governments use capital income tax for redistribution, which is much more distortionary, or progressive labor income taxes. One feature of this latter type of taxation is that it reduces the variability of post-tax earnings (by taxing proportionately more high earnings and less low earnings), so progressivity could represent an additional source of insurance.

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.

1.8 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 indo unproductive debt. Note that the equilibrium condition in the asset market is now

$$K (r) + B = A (r) \Rightarrow K (r) = \tilde{A} (r) \equiv A (r) - B.$$

The aggregate supply shifts to the left in this economy as B increases (it's as if the effective borrowing constraint shifts), so the interest rate rises unambiguously.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.

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.

2 Constrained Efficiency

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 $R\beta = 1$. 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. An unconstrained planner can achieve Pareto efficiency (by effectively reintroducing the missing markets) because it has more freedom in allocating resources than is provided by the system of incomplete markets.

However, the important question is not so much whether a new economic structure can do better, but whether the market performs efficiently relative to the set of allocations achievable with the this same structure. 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) that agents face in the decentralized equilibrium.

We will find that the competitive equilibrium is constrained inefficient. The reason is the presence of a "pecuniary externality", i.e. each agents' decision has a small effect on prices that individual agents do not take into account. By choosing a consumption policy, the planner can also 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 competitive system.

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

$$u_c\left(R\left(\lambda^*\right)a + w\left(\lambda^*\right)\varepsilon - g^*\left(a,\varepsilon\right)\right) \geq \beta R\left(\lambda^*\right)\sum_{\varepsilon' \in E} u_c\left(R\left(\lambda^*\right)g^*\left(a,\varepsilon\right) + w\left(\lambda^*\right)\varepsilon' - g^*\left(g^*\left(a,\varepsilon\right),\varepsilon'\right)\right)\pi\left(\varepsilon',\varepsilon\right),$$

which we can compactly rewrite as

$$u_c \geq \beta R\left(\lambda^*\right)\sum_{\varepsilon' \in E} u_c' \pi\left(\varepsilon', \varepsilon\right).$$

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

$$\begin{aligned}
\Omega(\lambda) &= \max_{g(a,\varepsilon)\in A} \int_{A\times E} u\left(R(\lambda)a + w(\lambda)\varepsilon - g(a,\varepsilon)\right) d\lambda + \beta\Omega(\lambda') \\
& s.t. \\
R(\lambda) &= F_K(K,H) \text{ and } w(\lambda) = F_H(K,H) \\
H &= \int_{A\times E} \varepsilon d\lambda \\
K &= \int_{A\times E} a d\lambda \\
\lambda'(\mathcal{A}\times\mathcal{E}) &= \int_{A\times E} 1_{\{g(a,\varepsilon)\in\mathcal{A}\}}\pi(\varepsilon'\in\mathcal{E},\varepsilon)\, d\lambda(a,\varepsilon)
\end{aligned}$$

It is easy to see that the necessary FOC for the planner who chooses the level a' for a particular pair (a,ε) is:

$$u_c \geq \beta R(\lambda) \sum_{\varepsilon'\in E} u_c' \pi(\varepsilon',\varepsilon) + \beta \int_{A\times E} \left(\varepsilon' F_{HK}' + a' F_{KK}'\right) u_c' d\lambda'.$$

We have an extra term relative to the competitive equilibrium 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 prices.

In particular, the term in parenthesis under the integral captures the effect of an additional unit of savings on next-period individual labor income $(\varepsilon' F_{HK}')$ and capital income $(a' F_{KK}')$ of all agents through next-period price changes. More savings increase the capital stock, thus raise wages and labor income and decrease 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 it's zero because if F is CRS, then F_K and F_H are homogenous of degree zero. If income of the poor is labor-intensive (capital-intensive), then the extra term will be positive (negative). In the first case (arguably the more reasonable), the planner wants agents to save more than in the decentralized equilibrium, 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.

Quantitatively, DHKR calibrate the model to the US economy and find that the constrained efficient capital stock is a staggering 3.5 times higher than the laissez-faire economy capital stock.