Macroeconomics B Notes

Nicholas Umashev [*]

2019

Contents

Dynamic Optimization in Discrete Time

<u>Optimization Issues</u>: solving large problems (like that below is challenging using traditional methods and so we must rely on dynamic programming)

- <u>Example</u>: consider the problem

$$\max_{\{c_t, k_{t+1}\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \ln c_t$$

$$\text{subject to: } c_t + k_{t+1} \le A k_t^{\alpha} \theta_t \ \ \forall t, \quad k_0 \text{ given}, \quad \ln(\theta_t) \sim_{iid} N(0, \sigma^2)$$

  where c_t is consumption, k_t is capital, θ_t is a productivity shock
  - <u>Issue</u>: this problem involves maximization over an infinite sequence of controls, subjects to an infinite sequence of constraints, and where each period's optimal decisions depend on the history of shocks and past decisions - making the problem more difficult
  - <u>Solution</u>: we are looking for a solution of the form $\{c_t^*(\theta^t)\}_{t=0}^{\infty}$ and $\{k_t^*(\theta^t)\}_{t=0}^{\infty}$ where $\theta^t = \{\theta_0, \theta_1, \ldots, \theta_t\}$ is the history of shocks up to tS

<u>Dynamic Programming</u>: breaks up optimization problems into a series of small and more tractable problems where the smaller problems are casted in a recursive way and are solved for a time-invariant policy function. Here the solution to such problems (under certain regularity conditions) can be used to recover the solution to the original problem

- <u>Bellman's Optimality Principle</u>: an optimal policy has the property that whwatever the initial state and initial decisions are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision.
  - <u>Note</u>: based on this principle the solution to the dynamic programming problem is a valid solution to the original problem
- <u>Horizon Types</u>: can involve either finite horizon or infinite horizon problems
- <u>State Variables</u>: set of variables summarizing the state of the economy at each point in time
- <u>Control Variables</u>: the set of choice variables at each point in time
- <u>Value Function</u>: the optimal value of the original problem, given the states
- <u>Policy Function</u>: the optimal value for the control variables, given the states

<u>Primer on Dynamic Programming Concepts</u>: any optimization problem has some objective (such as minimizing cost) with the mathematical function that describes this objective being called the objective function.

Dynamic programming breaks a multi-period planning problem into smaller steps at different points in

time and therefore it requires keeping track of how the decision situation is evolving over time. The information about the current situation that is needed to make a correct decision is called the "the state". For example, to decide how much to consume and spend at each point in time, people would need to know their initial wealth and therfore wealth would be one of their state variables.

The variables chosen at any given point in time are often called the "control" variables. For example, given their current wealth, people might decide how much to consume now. Choosing the control variables now may be equivalent to choosing the next state, since the next state is affected by the current control (in addition to other factors). For example, today's wealth (the state) and consumption (the control) might exactly determine tomorrow's wealth (the new state).

The dynamic programming approach describes the optimal plan by finding a rule that tells what the control should be, given any possible value of the state. For example, if consumption (c) depends only on wealth (w) then we would seek a rule $c(w)$ that gives consumption as a function of wealth. Such a rule, determining the controls as a function of the states, is called a "policy function".

Finally, the optimal decision rule is the one that achieves the best possible value of the objective. For example, if someone chooses consumption, given wealth, in order to maximize happiness (represented by a utility function), then each level of wealth with be associated with some highest possible level of happines $H(w)$. The best possible value of the objective, written as a function of the state, is called the value function.

Therefore, a dynamic optimization problem in discrete time can be stated in a recursive, step-by-step, form known as backward induction by writing down the relationship between the value function in one period and the value function in the next period. The relationship between these two value functions is called the "Bellman Equation". In this approach, the optimal policy in the last time period is specified in advance as a function of the state variable's value at that time, and the resulting optimal value of the objective function is thus expressed in terms of that value of the state variable.

Next, the next-to-last period's optimization involves maximizing the sum of that period's period-specific objective function and the optimal value of the future objective function, giving that period's optimal policy contingent upon the value of the state variable as of the next-to-last period decision.

This logic continues recursively back in time, until the first period decision rule is derived, as a function of the initial state variable value, by optimizing the sum of the first-period-specific objective function and the value of the second period's value function, which gives the value for all the future periods. Thus, each period's decision is made by explicitly acknowledging that all future decisions will be optimally made.

<u>Finite Horizon Problem</u>: consider a decision-maker solving the deterministic problem

$$\max_{\{u_t\}_{t=0}^{T-1}} \sum_{t=0}^{T-1} \beta^t r(x_t, u_t) + W(x_T)$$

$$\text{subject to: } x_{t+1} = g(x_t, u_t) \; \forall t, \quad x_0 \text{ is given}$$

Here we assume a convex set, to ensure a maximum exists, and that the state space is discrete (ie $\mathscr{X} = \{x^1, x^2, \ldots, x^n\}$). Our variables include:

$$\beta \in (0, 1) : \text{ the discount factor}$$

$$x \in \mathscr{X} : \text{ the state variable}$$

$$u \in \mathscr{U} : \text{ the control variable}$$

$$r(x_t, u_t) : \text{ concave return function}$$

$$g(x_t, u_t) : \text{ the law of motion for the state}$$

$$W(x_T) : \text{ the terminal condition at } x_t$$

- <u>Recursive Transformation</u>: based on the original problem being time separable we can write it as a series of recursive functions. The original problem asks us to find a sequence of T controls (i.e. find $\{u_t^*\}_{t=0}^{T-1}$) while the recursive formulation asks us to solve T problems with one control each (i.e. find the series of $U_t^*(x)$ over the set t).

In other words, we transform the original problem of finding an infinite sequence of controls that maximizes our function for the problem of finding the optimal value function $V(x)$ and a function h that solves the continuum of maximum problems (ie one maximum problem for each value of x). Here h is a time-invariant policy function that maps the state x_t into the control u_t such that we can iteratively use $u_t = h(x_t)$ and $x_{t+1} = g(x_t, u_t)$ to find the sequence $\{u_s\}_{s=0}^{\infty}$ We make the following transformation of the Objective Function to a Bellman Function

$$\max_{\{u_t\}_{t=0}^{T-1}} \sum_{t=0}^{T-1} \beta^t r(x_t, u_t) + W(x_T)$$

$$\text{subject to: } x_{t+1} = g(x_t, u_t) \; \forall t, \quad x_0 \text{ is given}$$

$$\Downarrow$$

$$V_t(x) = \max_u \{r(x, u) + \beta V_{t+1}(g(x, u))\}$$

- <u>Value Function $V_t(x)$</u>: is the greatest feasible payoff from time t forward if the state at time t is x_t.

Here the value function is

$$V_t(x_t) = \max_{\{u_s\}_{s=t}^{T-1}} \sum_{s=t}^{T-1} \beta^s r(x_s, u_s) + W(x_T)$$

subject to : $x_{s+1} = g(x_s, u_s) \ \forall s, \quad x_t$ is given

- ○ Bellman Condition: given the structure of the problem, the value function satisfies the Bellman Equation. Here we want to solve for $V(x)$ and $h(x)$ together which are linked by the Bellman Equation:

$$V_t(x) = \max_u \{r(x, u) + \beta V_{t+1}(g(x, u))\}$$

  with the terminal condition $V_T(x) = W(x)$.
  - ∗ Intuition: the value function at time t is going to be the maximum utility given next period's value function, today's return, and the terminal condition. Note that the Bellman Equation describes the value function, at a given point in time, as a function of itself in another point in time. Note that the maximizer of the Bellman equation is a policy function $h(x)$ (also notated as $U_t(x)$) that satisfies

$$V(x) = r[x, h(x)] + \beta V \{g[(x, h(x)]\}$$

- • Policy Function $U_t(x)$: the policy function describes the utility that, given the state, is the maximum of the value function. In other words, the maximizer of the RHS of the bellman equation is a policy function $h(x)$. Given $V_t(x)$, we define the policy function by

$$U_t(x) = \arg \max_u \{r(x, u) + \beta V_{t+1}(g(x, u))\}$$

- • Assumptions: in order to solve the Bellman Equation we need to impose the following assumptions
  - ○ Assuumption 1: the Bellman Equation has a unique strictly concave solution
  - ○ Assumption 2: the solution is approached in the limit as $j \to \infty$ by iterations on

$$V_{j+1}(x) = \max_u \{r(x, u) + \beta V_j(\tilde{x})\}$$

  Subject to $\tilde{x} = g(x, u)$, x given, and starting from any bounded and continuous initial V_0. Note that $\tilde{x}$ is tomorrows x
  - ○ Assumption 3: there is a unique and time-invariant optimal policy of the form $u_t = h(x_t)$, where h is chosen to maximize the RHS of the Bellman equation
  - ○ Assumption 4: off the corners, the limiting value function V is differentiable

- <u>Solution</u>: the Bellman Equation and the Policy Function are the building blocks for constructing the optimal solution
  - ○ <u>Step 1 - Backward Induction</u>: compute $\{V_t(x)\}_{t=0}^{T}$ and $\{U_t(x)\}_{t=0}^{T-1}$ by backward induction using the Bellman Equation
    - ∗ <u>Part 1</u>: compute $V_t(x) = W(x) \; \forall x \in \mathfrak{X}$
    - ∗ <u>Part 2</u>: for all $t = T - 1, T - 2, \ldots, 0$ compute $V_t(x) = \max_u \{r(x, u) + \beta V_{t+1}(g(x, u))\}$ and $U_t(x) = \arg\max_u \{r(x, u) + \beta V_{t+1}(g(x, u))\}$ for all $x \in \mathscr{X}$
    - ∗ <u>Example</u>

$$\boxed{t = T}$$

$$\forall x^i, \; V_T(x^i) = W(x^i) \to \text{ final condition}$$

$$\boxed{t = T - 1}$$

$$\forall x^i, \; V_{T-1}(x^i) = \max_u \left\{ r(x^i, u) + \beta V_T(x^i) \right\}$$

$$U_{T-1}(x^i) = \arg\max_u \left\{ r(x^i, u) + \beta V_T(x^i) \right\}$$

$$\ldots$$

repeat these steps until we reach $\boxed{t = 0}$

Note that because the state variable x is discrete, we know how many times we need to iterate backwards until $t = 0$

  - ○ <u>Step 2 - Iterating Forward</u>: given x_0 compute $\{u_t^*\}_{t=1}^{T-1}$ and $\{x_t^*\}_{t=1}^{T}$ by iterating forward on

$$\{U_t(x)\}_{t=0}^{T-1}$$

- ∗ <u>Method</u>: we iterate forward using x

$$u_0^* = U_0(x_0^*), \;\; x_1^* = g(x_0^*, u_0^*),$$

$$u_1^* = U_1(x_1^*), \;\; x_2^* = g(x_1^*, u_1^*),$$

$$\ldots$$

$$u_t^* = U_t(x_t^*), \;\; x_{t+1}^* = g(x_t^*, u_t^*),$$

$$\ldots$$

$$u_{T-1}^* = U_{T-1}(x_{T-1}^*), \;\; x_T^* = g(x_{T-1}^*, u_{T-1}^*),$$

- ∗ <u>Intuition</u>: the law of motion will indicate what the future value of x_t is given the control variable u, allowing us to derive the solution. Note that the control variable u is able to be

6

calculated due to having previously found the value of u via backward induction in step 1

<u>Infinite Horizon Problem</u>: consider the problem

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

$$\text{subject to}: \ x_{t+1} = g(x_t, u_t), \quad x_0 \text{is given}$$

We want to compute the solution $\{u_t^*, x_{t+1}^*\}_{t=0}^{\infty}$

- <u>Assumptions</u>: in addition to the previous finite horizon problem assumptions we assume that $r(\cdot)$ is bounded, also note that now the bellman equation is a functional equation in V
- <u>No Backward Induction</u>: since there is no terminal condition to start with (ie V is time-invariant) we cannot use backward induction
- <u>Conditions for a Solution</u>: the following four conditions must hold for the infinite horizon problem to be solved
    - <u>Condition 1</u>: the Bellman Equation has a unique strictly concave solution V^*
    - <u>Condition 2</u>: let V^0 be any bounded and continuous VF and let V^j satisfy

    $$V^j(x) = \max_u \left\{ r(x, u) + \beta V^{j-1}(g(x, u)) \right\} \quad \text{for } j \geq 1$$

    Then $V^* = \lim_{j \to \infty} V^j$
    - <u>Condition 3</u>:There is a unique and time-invariant policy function U^* such that the solution to the original problem can be obtained by iterating on $u_t^* = U^*(x_t)$
    - <u>Condition 4</u>: off the corners, the limiting value function V^* is differentiable with

    $$V^{*'}(x) = r_1(x, U(x)) + \beta V^{*'}(g(x, U(x)))g_1(x, U(x))$$

    Note that the second condition implies that if you have any function V_0 that is bounded and continuous then if you iterate on j you will eventually find the solution V^* (ie a steady state). The third condition implies that from finding V^* you can use u_* to then backwards iterate to find U^*
- <u>Guess and Verify Method</u>: also known as the method of undetermined coefficients, this involves guessing and verifying a solution V to the Bellman Equation. This relies on the uniqueness of the solution to the Bellman Equation (condition 1) and can be solved analytically with pen and paper (though does not always work)
    - <u>Step 1</u>: guess that the true value function has a particular parametric form $V^G(x; A)$ where A is a vector of unknown parameters
    - <u>Step 2</u>: compute $\widehat{V}(x; A) = \max_u \{r(x, u) + \beta V^G(g(x, u); A)\}$

* Note: you can solve for the control variable that is implied by the policy function by using the FOC (provided in condition 4). This then allows you to solve for the optimized $\widehat{V}(x; A)$
  - Step 3: if there exists parameters A^* such that $\widehat{V}(x; A^*) = V^G(x; A^*)$ then $V^G(x; A^*)$ is the solution
    * Note: compare the derived optimized value function $\widehat{V}(x; A)$ from step 2 to $V^G(x; A)$
- Value Function Iteration Method: we proceed by constructing a sequence of value functions and associated policy functions. The sequence is created by iterating on the following equation, starting from $V_0 = 0$, and continuing until V_j has converged:

$$V_{j+1}(x) = \max_u \{r(x, u) + \beta V_j(\widetilde{x})\}$$

$$\text{subject to} \quad \widetilde{x} = g(x, u), \ x \ \text{given}$$

This method always works because by solution method condition 2 we have $V^* = \lim_{j \to \infty} V^j$, however this method can be very slow particularly for problems with more than one state
  - Step 1: start from any bounded and continuous guess V^0
  - Step 2: iterate on $V^j(x) = \max_u\{r(x, u) + \beta V^{j-1}(g(x, u))\}$ until convergence (ie until $||V^j - V^{j-1}|| < \varepsilon$)
- Policy Function Iteration Method
  - Step 1: pick a feasible policy function $U^0(x)$ (or $u = h_0(x)$) and compute the value associated with operating forever with that policy

$$V_{h_j}(x) = \sum_{t=0}^{\infty} \beta^t r[x_t, h_j(x_t)]$$

where $x_{t+1} = g[x_t, h_j(x_t)]$ with $j = 0$
  - Step 2: generate a new policy $u = h_{j+1}(x)$ that solves the two period problem

$$\max_u \{r(x, u) + \beta V_{h_j}[g(x, u)]\}$$

for each x
  - Step 3: iterate over j until convergence, repeating steps 1 and 2
- Example: consider a social planner solving

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

$$\text{subject to:} \ c_t + k_{t+1} = Ak_t^\alpha, \quad k_0 \text{ is given}$$

where $\alpha \in (0, 1)$ and $\beta \in (0, 1)$

○ Simplification: rearranging the constraint and substituting c_t into the objective function we have

$$\max_{\{k_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \ln(Ak_t^{\alpha} - k_{t+1})$$

$$\text{subject to}: \; k_0 \text{ is given}$$

Here the State Variable is $t \to k^t$ (i.e. today's capital) and the control variable is $t \to k_{t+1}$ (i.e. tomorrow's capital). Note that indexes have been dropped since it is an infinite horizon problem with k being the state and k' being the control. The Bellman Equation given this simplication is

$$V(k) = \max_{k'}\{\ln(Ak^{\alpha} - k') + \beta V(k')\}$$

○ Guess and Verify Solution: here we guess the functional form $V^G(k; E, F) = E + F \ln(k)$ where E and F are unknown constants. We then plug in $V^G(x; E, F)$ into the RHS of the Bellman equation yielding

$$\widehat{V}(k; E, F) = \max_{k}\left\{ \ln(Ak^{\alpha} - k') + \beta V^G(k'; E, F) \right\} = \max_{k}\left\{ \ln(Ak^{\alpha} - k') + \beta(E + F \ln(k')) \right\}$$

Taking first order conditions and log-linearizing results in the optimal policy function $k'(k) = \frac{\beta F}{1+\beta F}Ak^{\alpha}$. Substituting the optimal policy function into the objective function yields

$$\widehat{V}(k; E, F) = \underbrace{\left[\ln\left(\frac{A}{1 + \beta F}\right) + \beta F \ln\left(\frac{\beta F}{1 + \beta F}A\right) + \beta E \right]}_{E'} + \underbrace{\alpha(1 + \beta F)\ln(k)}_{F'}$$

$$\Updownarrow$$

$$\widehat{V}(k; E', F') = E' + F' \ln(k)$$

When grouping the constants for $\widehat{V}(k; E', F')$ it has the same functional form as $V^G(k; E, F)$. Finally, we pin down the unknown coefficients by solving the system of equations

$$E^* = \left[\ln\left(\frac{A}{1 + \beta F}\right) + \beta F \ln\left(\frac{\beta F}{1 + \beta F}A\right) + \beta E \right]$$

$$F^* = \alpha(1 + \beta F)\ln(k)$$

Note that the equation for E^* is derived from collecting constant terms and F^* is derived from collecting the terms containing our dependent variable. This leaves us with the final solution $V^* = E^* + F^* \ln(k)$ and the optimal policy function $k'^*(k) = \frac{\beta F^*}{1+\beta F^*}Ak^{\alpha}$

  * Solution Write Out: here the Bellman Equation is $V(k) = \max_{k'}\left\{\ln(Ak^{\alpha} - k') + \beta V(k')\right\}$
    ⊛ Step 1: my guess is $V^G(k) = F \ln(k) + E$ where F and E are our unknown constants

9

⊛ <u>Step 2</u>: using my guess we have

$$\widehat{V}(k) = \max_{k'} \left\{ \ln(Ak^\alpha - k') + \beta(Ak^\alpha - k') \right\} = \max_{k'} \left\{ \ln(Ak^\alpha - k') + \beta(E + F\ln(k')) \right\}$$

Taking first order conditions we have the following

$$-\frac{1}{Ak^\alpha - k'} + \beta F \frac{1}{k'} = 0$$

$$\Downarrow$$

$$k' = \frac{\beta F}{1 + \beta F} Ak^\alpha \qquad (\star)$$

Plugging $\star$ into the objective function, we obtain the optimized value function $\widehat{V}(k)$

$$\widehat{V}(k; E, F) = \underbrace{\left[\ln\left(\frac{A}{1 + \beta F}\right) + \beta F \ln\left(\frac{\beta F}{1 + \beta F} A\right) + \beta E \right]}_{E'} + \underbrace{\alpha(1 + \beta F)}_{F'} \ln(k)$$

⊛ <u>Step 3</u>: to evaluate the parameters (E^*, F^*) we solve the system of equations implied by the functional form

$$\widehat{V}(k) = E' + F' \ln(k)$$

$$V^G(k) = E + F \ln(k)$$

This yields

$$E' = E \Rightarrow E = \ln\left(\frac{A}{1 + \beta F}\right) + \beta F \ln\left(\frac{\beta F}{1 + \beta F} A\right) + \beta E$$

$$F' = F \Rightarrow F = \alpha(1 + \beta F)$$

Note that by $F = \alpha(1 + \beta F)$ and $F' = F$ we have $F^* = \frac{\alpha}{1 - \alpha\beta}$ which gives us the policy function with one unknown $k'^*(k) = \frac{\beta F^*}{1 + \beta F^*} Ak^\alpha$

<u>Stochastic Problems</u>: consider the stochastic problem

$$\max_{\{u_t\}_{t=0}^\infty} \mathbb{E}_0 \sum_{t=0}^\infty r(x_t, u_t, \varepsilon_t)$$

$$\text{subject to}: \ x_{t+1} = g(x_t. U_t, \varepsilon_{t+1}), \quad \text{given } x_t$$

- <u>iid Shocks</u>: ε_t is a sequence of iid shocks with the cumulative distribution function $\Pr[\varepsilon_t \leq e] = F(e)$
  - <u>Timing</u>: ε_{t+1} is realized at $t + 1$ after u_t has been chosen, therefore x_t is a sufficient statistic at t

○ iid Assumption: this is important for recursive formulation for if shocks were persistent instead then the state vector should also incorporate past shock realizations

- Bellman Equation: the Bellman Equation for this problem is

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

where $\mathbb{E}[V(g(x, u, \varepsilon))|x] = \int V(g(x, u, \varepsilon))dF(\varepsilon)$ and $V(x)$ is the optimal value of the problem starting from x at $t = 0$

- Solution Methods: the previous solution methods still apply and the solution $V(x)$ of the Bellman Equation can be computed by iterating on

$$V_{j+1}(x) = \max_u \left\{ r(x, u) + \beta \mathbb{E}\left[V_j[g(x, u, \varepsilon)|x]\right] \right\}$$

starting from any bounded continuous initial V_0. The first order necessary condition for the problem on the RHS of the Bellman Equation is

$$r_2(x, u) + \beta \mathbb{E}\left\{ V'[g(x, u, \varepsilon)]g_2(x, u, \varepsilon)|x \right\} = 0$$

which we obtain simply by differentiating the Bellman Equation and passing the differentiation operation under the $\mathbb{E}$ (an integration) operator. Off corners, the value function satisfies

$$V'(x) = r_1[x, h(x)] + r_2[x, h(x)]h'(x) + \beta \mathbb{E}\left\{ V'\{g[x, h(x), \varepsilon]\}\{g_1[x, h(x), \varepsilon] + g_2[x, h(x), \varepsilon]h'(x)\}|x \right\}$$

When the states and controls can be defined in such a way that x does not appear in the transition equation, the formula for V' becomes

$$V'(x) = r_1[x, h(x)]$$

Substituting this formula into the first order necessary condition for the problem gives the stochastic Euler equation

$$r_2(x, u) + \beta \mathbb{E}[r_1(\tilde{x}, \tilde{u})g_2(x, u, \varepsilon)|x] = 0$$

where tildes over x and u denote next period values.

○ Integral Valuation: the integral in expectation can be approximated by quadrature, if closed form solutions are infeasible

Problem Example 1: consider the following infinite horizon problem's Bellman Equation:

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

Let V^* be the equations solution, show that

$$V^{*'}(x) = r_1(x, U^*(x)) + \beta V^{*'}(g(x, U(x)))g_1(x, U(x))$$

- Solution: the first order condition for the Bellman Equation is $0 = \frac{\partial}{\partial u} = r_2(x,u) + \beta V'(g(x,u)) \cdot g_2(x,u)$.
  Let $U(x)$ be the optimal policy function that solves V^* then we have:

$$V^*(x) = r(x, U(x)) + \beta V^*(g(x, U(x)))$$

. If $U(x)$ is differentiable then we have the following first order condition:

$$\frac{\partial}{\partial x} = 0 \rightarrow V'^*(x) = r_2(x, U(x)) \cdot U'(x) + r_1(x, U(x)) + \beta V'^*(g(x, U(x))) \cdot \{g_1(x, U(x)) + g_2(x, U(x)) \cdot U'(x)\}$$

$$V'^*(x) = r_1(x, U(x)) + \beta V'^*(g(x, U(x)))g_1(x, U(x)) + U'(x)\{r_2(x, U(x)) + \beta V'^*(g(x, U(x)))g_2(x, U(x))\}$$

Due to our first order condition from before we have:

$$V^*(x) = r_1(x, U(x)) + \beta V'^*(g(x, U(x)))g_1(x, U(x))$$

Problem Example 2: consider the Brock-Mirman stochastic growth model:

$$\max_{\{c_t, k_{t+1}\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \ln(c_t)$$

$$\text{subject to}$$

$$c_t + k_{t++1} = A k_t^\alpha \theta_t$$

$$k_0 \text{ is given}$$

where $\{\theta_t\}$ is an i.i.d sequence with $n(\theta_t) \sim N(0, \sigma^2)$, $\sigma > 0$, $A > 0$, $\alpha \in (0,1)$, and $\beta \in (0,1)$. It is assumed that the shock θ_t is realised before k_{t+1} is chosen. Solve this problem analytically by applying policy function iteration

- Solution Step 1: write the problem recursively
  - $V_t = \max_{k_{t+1}} \{\ln(A k_t^\alpha \theta_t - k_{t+1}) + \beta \mathbb{E} V_{t+1}\}$ which can be written as $J(k, \theta) = \max_{k'} \left\{\ln(A k^\alpha \theta - k') + \beta \mathbb{E}_0[J(k', \theta'}\right.$
- Solution Step 2: guess a policy function of the form $k_{t+1} = h_0(A k_t^\alpha \theta_t)$ for some constant $h_0 \in (0,1)$

and compute the value from following this policy function forever, i.e.

$$J_0(k_0, \theta_0) = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \ln(Ak_t^\alpha \theta_t - h_0 Ak_t^\alpha \theta_t)$$

○ Our guess is $k_{t+1} = h_0(Ak_t^\alpha \theta_t)$. We want to make this a function of k_0 and θ_0 only. We can reduce the policy function as follows:

$$
\begin{aligned}
J_0(k_0, \theta_0) &= \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \ln(Ak_t^\alpha \theta_t - h_0 Ak_t^\alpha \theta_t) \\
&= \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \ln(A(1-h)k_t^\alpha \theta_t) \\
&= \tfrac{1}{1-\beta} \ln(A(1-h)) + \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \ln(\theta_t) + \sum_{t=0}^{\infty} \beta^t \ln(k_t^\alpha) \\
&= \tfrac{\ln(A(1-h))}{1-\beta} + \ln(\theta_0) + \mathbb{E}_0 \sum_{t=1}^{\infty} \ln(\theta_t)\beta^t + \alpha \ln(k_0) + \sum_{t=1}^{\infty} \ln(k_t^\alpha)\beta^t \\
&= \tfrac{\ln(A(1-h))}{1-\beta} + \alpha \ln(k_0) + \ln(\theta_0) + \sum_{t=1}^{\infty} \ln(k_t)\beta^t \alpha
\end{aligned}
$$

Using the fact that $k_{t+1} = h_0(Ak_t^\alpha \theta_t)$ we can derive an equation for all k_t as a function of the parameters:

$$k_1 = h_0(Ak_0^\alpha \theta_0)$$

$$
\begin{aligned}
k_1 : \ln(k_1) &= \ln(k_1) - \ln(h_0 Ak_0^\alpha \theta_0) \\
&= \ln(h_0 A) + \ln(\theta_0) + \alpha \ln(k_0) \\
k_2 : \ln(k_2) &= \ln(h_0 Ak_1^\alpha \theta_1) \\
&= \ln(h_0 A) + \alpha \ln(k_1) + \ln(\theta_1) \\
&= \ln(h_0 A) + \alpha \{\ln(h_0 A) + \ln(\theta_0) + \alpha \ln(k_0)\} + \ln(\theta_1) \\
&= (1 - \alpha) \ln(h_0 A) + \alpha \ln(\theta_0) + \alpha^2 \ln(k_0) + \ln(\theta_1) \\
k_3 : \ln(k_3) &= \ln(h_0 Ak_2^\alpha \theta_2) \\
&= \ln(h_0 A) + \ln(\theta_2) + \alpha \ln(k_2) \\
&= \ln(h_0 A) + \ln(\theta_2) + \alpha \{(1 - \alpha) \ln(h_0 A) + \alpha \ln(\theta_0) + \alpha^2 \ln(k_0) + \ln(\theta_1)\} \\
&= (1 - \alpha - \alpha^2) \ln(h_0 A) + \alpha^3 \ln(k_0) + \ln(\theta_2) + \alpha \ln(\theta_1) + \alpha^2 \ln(\theta_0) k_4 : \ln(k_4) \qquad = \ln(h_0 Ak_3^\alpha \\
&= \ln(h_0 A) + \ln(\theta_3) + \alpha \ln(k_3) \\
&= \ln(h_0 A) + \ln(\theta_3) + \alpha \{(1 - \alpha - \alpha^2) \ln(h_0 A) + \alpha^3 \ln(k_0) + \ln(\theta_2) + \alpha \ln(\theta_1) + \alpha^2 \ln(\theta_0)\} \\
&= (1 - \alpha - \alpha^2 - \alpha^3) \ln(h_0 A) + \alpha^4 \ln(k_0) + \ln(\theta_3) + \alpha \ln(\theta_2) + \alpha^2 \ln(\theta_1) + \alpha^3 \ln(\theta_0)
\end{aligned}
$$

Based on the above, we have:

$$\sum_{t=1}^{\infty} \ln(k_t) = \sum_{t=1}^{\infty} \left[\tfrac{1-\alpha^t}{1-\alpha} \ln(h_0 A) + \alpha^t \ln(k_0) + \ln(\theta_{t-1}) + \alpha \ln(\theta_{t-2}) + \alpha^2 \ln(\theta_{t-3}) + \cdots + \alpha^{t-1} \ln(\theta_0) \right]$$

Plugging this into $J_0(k_0, \theta_0)$ we have:

$$J_0(k_0, \theta_0) = \tfrac{1}{1-\beta} \ln(A(1-h)) + \alpha \ln(k_0) + \ln(\theta_0) + \sum_{t=1}^{\infty} \left(\frac{1-\alpha^t}{1-\alpha} \ln(h_0 A) + \alpha^t \ln(k_0) + \ln(\theta_{t-1}) + \alpha \ln(\theta_{t-2}) + \alpha^2 \ln \right.$$

$$= \tfrac{1}{1-\beta} \ln(A(1-h)) + \alpha \ln(k_0) + \ln(\theta_0) + \sum_{t=1}^{\infty} \left\{ \tfrac{1-\alpha^t}{1-\alpha} \ln(h_0 A) \right\} + \sum_{t=1}^{\infty} \left\{ \alpha^t \ln(k_0) \right\} + \sum_{t=1}^{\infty} \left\{ \ln(\theta_{t-1}) + \alpha \ln(\theta) \right.$$

Since θ is i.i.d we have that $E_0 \theta_t = 0 \ \ \forall t > 0$ and therefore:

$$J_0(k_0, \theta_0) = \overbrace{\tfrac{1}{1-\beta} \ln(A(1-h)) + \sum_{t=1}^{\infty} \beta_t \left\{ \tfrac{1-\alpha^t}{1-\alpha} \ln(hA) \right\}}^{H} + \alpha \sum_{t=0}^{\infty} (\beta\alpha)^t \left\{ \ln(k_0) \right\} + (1 + \sum_{t=1}^{\infty} (\beta\alpha)^t) \ln(\theta_0)$$

$$J_0(k_0, \theta_0) = H + \tfrac{\alpha}{1-\beta\alpha} \ln(k_0) + \sum_{t=0}^{\infty} (\beta\alpha)^t \ln(\theta_0)$$

$$J_0(k_0, \theta_0) = H + \tfrac{\alpha}{1-\beta\alpha} \ln(k_0) + \tfrac{1}{1-\beta\alpha} \ln(\theta_0)$$

- <u>Solution Step 3</u>: update your guess for the policy function by solving $\max_{k'} \left\{ \ln(Ak^\alpha \theta - k') + \beta \mathbb{E}[j_0(k', \theta')|k'] \right\}$
  - based on our solution to step 2, we can rewrite the above equation as

$$\max_{k'} \left\{ \ln(Ak^\alpha \theta - k') + \beta \mathbb{E}[H_0 + \tfrac{\alpha}{1-\beta\alpha} \ln(k') + \tfrac{1}{1-\beta\alpha} \ln(\theta')] \right\}$$

Taking first order conditions yields:

$$0 = \tfrac{\partial}{\partial k'} = \frac{-1}{Ak^\alpha \theta - k'} + \frac{\beta\alpha}{(1-\beta\alpha)k'}$$

$$\therefore \frac{1}{Ak^\alpha \theta - k'} = \frac{\beta\alpha}{(1-\beta\alpha)k'}$$

$$Ak^\alpha \theta = \frac{(1-\beta\alpha)k'}{\beta\alpha} + k'$$

$$Ak^\alpha \theta \beta\alpha = (1-\beta\alpha)k' + k'\beta\alpha$$

$$k' = Ak^\alpha \theta \beta\alpha$$

- <u>Solution Step 4</u>: compute the value from following the new policy function forever
  - In this case $h_1 = \beta\alpha$, which has the same form as our initial guess. Since this policy function is the same as our initial policy function, the value of operating with this policy function will be of

14

the same form as that in step 2:

$$J_0(k_0, \theta_0) = H_1 + \frac{\alpha}{1-\beta\alpha} \ln(k_0) + \frac{1}{1-\beta\alpha} \ln(\theta_0)$$

- <u>Solution Step 5</u>: iterate until the policy function converges. How many iterations were necessary?

   ○ we have that

$$\max_{k'} \left\{ \ln(Ak^\alpha\theta - k') + \beta\mathbb{E}[H_1 + \frac{\alpha}{1-\beta\alpha} \ln(k') + \frac{1}{1-\beta\alpha} \ln(\theta')] \right\}$$

   The first order conditions for this function is:

$$0 = \frac{\partial}{\partial k'} = \frac{-1}{Ak^\alpha\theta - k'} + \frac{\beta\alpha}{(1 - \beta\alpha)k'}$$

   This yields the same policy function as before so no more iterations are necessary

- <u>Summary</u>: in summary we executed the following procedure

   ○ <u>(1)</u> Make a guess on the policy function

   ○ <u>(2)</u> Compute the value of operating with this function forever by making it time invariant and only dependent on our state variables

   ○ <u>(3)</u> In the value function at $t = 0$, substitute V_1 for the value obtained in step 2. Take FOC of the value function and solve for the control variables to derive a new policy function

   ○ <u>(4)</u> Repeat from step 2, iterating until the derived policy functions converge

Dynamic Optimisation in Continuous Time

Continuous vs Discrete Time: in discrete time the intervals between periods are $\Delta > 0$ (e.g. the interval between t and $t+1$ is 1), in continuous time the intervals between periods are $\Delta \to 0$

- Continuous Variables Issue: while state variables are often naturally continuous, when using computational methods we can only evaluation the value function on a finite number of points. Since continuous variables are not countable this becomes an issue
  - Solution: we can address this problem via using (1) discretization of the state space, or, (2) polynomial approximation to the value function
- Derivatives with Respect to Time: note that $\cdot a(t) = a_t - a_{t-1}$. Therefore, $\frac{\cdot a(t)}{a(t)} = \frac{a_t - a_{t-1}}{a_t}$ which is equivalent to $\frac{\partial ln(a)}{\partial t}$. This is as $\int \frac{\partial a}{a} = ln(a)$ and therefore if were multiply $\frac{\cdot a(t)}{a(t)}$ by an integral and derivative we have

$$\int \frac{\partial a(t)}{a(t)} \cdot \partial t = ln(a(t)) \cdot \partial t = \frac{\partial ln(a)}{\partial t}$$

Discretization Method: consider a problem with continuous state and control variables $\mathfrak{x} \in \mathbb{R}$, discretization just replaces $\mathfrak{x}$ and $\mathfrak{u}$ by the finite grids $\widehat{\mathfrak{x}} = \{x^1, \ldots, x^n\}$ and $\widehat{\mathfrak{u}} = \{u^1, \ldots, u^n\}$. Here we discretize for a vector of unknown capital levels $k = [0, k_1, \ldots, \overline{k}]$, we need to decide how big of a grid should we create when discretizing

- Value Function Impact: now the value function becomes a finite list of numbers, $V = [V^1, \ldots, V^n]^T$
- Advantage: the maximization step is much simpler than under the original bellman equation, which is a key advantage of discretization methods
- Disadvantage: there is a "curse of dimensionality" in muiltidimensional state spaces where we can have either N points for a one-dimensional state space vs N^k points for a k-dimensional state space, which increases the number of states.
  - Grid Decision: requires some a-priori information about the state space, which is sometimes difficult to obtain (e.g. upper and lower bounds). Typically, you choose a steady state level of capital and then select the grid up to that steady state
- Optimal Growth Illustration: suppose $V(k) = \max_{k'} \{\ln(Ak^\alpha - k') + \beta V(k')\}$ where in this case $\mathfrak{x} = \mathfrak{u}$
  - Discretization: if we discretize $\mathfrak{x}$ then the Bellman equation becomes $V_i = \max_j \{\ln(Ak_i^\alpha - k_j) + \beta V_j\}$ for all $i = 1, \ldots, n$ where i indexes the different possible current levels of capital and j indexes the different possible levels of future capital
  - Value Function Iteration Illustration: suppose we applied value function iteration and therefore iterate on the mapping $V_i^s = \max \{\ln(Ak_i^\alpha - k_1) + \beta V_1^{s-1}, \ldots, \ln(Ak_i^\alpha - k_n) + \beta V_n^{s-1}\}$ for all $i = 1, \ldots, n$ where s indexes the iteration step

* <u>Step 1</u>: start with a guess

$$v_{n\times 1}^0 = \begin{bmatrix} v_1^0 \\ v_2^0 \\ \ldots \\ v_n^0 \end{bmatrix}, \quad \text{where} \quad v_i^0 \equiv v^0(k_i)$$

Note that our guess on V^0 has a dimension of $N \times 1$ which is the vector of all values of v_0 given different levels of capital. Therefore you have $N \times 1$ guesses

* <u>Step 2</u>: find V^1 by solving $V_i^1 = \max_j \left\{ \ln(Ak_i^\alpha - k_j) + \beta V_j^0 \right\}$ where

$$V_i^1 = \max \begin{bmatrix} \ln(Ak_i^\alpha - k_i) + \beta V_1^0 \\ \ln(Ak_i^\alpha - k_2) + \beta V_2^0 \\ \ldots \\ \ln(Ak_i^\alpha - k_n) + \beta V_n^0 \end{bmatrix} \quad \forall i, \quad \text{which reduces to} \quad V_{n\times 1}^1 = \begin{bmatrix} V_1^1 \\ V_2^1 \\ \ldots \\ V_n^1 \end{bmatrix}$$

  ⊛ <u>Note</u>: V_j^1 is the maximum element contained in the $N \times 1$ vector of $v_{n\times 1}^1$'s. Note that this involves just picking the maximum possible element in the v_i^1 vector containing every possible level of capital for today (ie period i)

* <u>Step 3</u>: check for convergence, setting the tolerance level to the euclidean norm
 ⊛ <u>Tolerance Level(ε)</u>: $||v^1 - v^0|| < \varepsilon$ where $\varepsilon = \sqrt{\sum_{i=1}^n (v_i^1 - v_i^0)}$
* <u>Step 4</u>: compute v^s by solving $v_i^s = \max \left\{ \ln(Ak_i^\alpha - k_j) + \beta v_j^{s-1} \right\} \quad \forall i$
* <u>Step 5</u>: check convergence $||v^s - v^{s-1}|| < \varepsilon$, if convergence holds then stop. Else go back to step 4 and iterate again

<u>Ordinary Differential Equations (ODEs)</u>: a "differential" equation" is one where the unknown is a function (instead of a variable) and the equation includes one or more of the derivatives of the function, an "ordinary differential equation" equation is one for which the unknown is a function of only one variable (typically time). The order of an ODE is that of the highest derivative - i.e. if the highest derivative is an ODE of order n then it is an nth-order ODE. The goal when solving an ODE is to find the behaviour of of the equation with respect to the unknown (i.e. how it evolves with time)

* <u>Partial ODEs</u>: where the unknown is a function of more than one variable
* <u>First-Order ODE Form</u>: $\dot{x}(t) = F(t, x(t))$, where $\dot{x}(t) \equiv dx(t)/dt$ and $t \in [t_a, t_b]$
  ○ <u>Unknown</u>: here the unknown is a function $x(t)$ with $x : [t_a, t_b] \to \mathbb{R}$ and can be found (without uniqueness) by integrating and combining with a initial condition
  ○ <u>Uniqueness</u>: the solution is not unique with the form having infinitely many solutions indexed by an integrating constant C. However, generally the constant C can be uniquely determined by

requiring the solution to pass through a given point on the $x(t)$-plane, providing us with a general and particular solution

    ○ Example: $\dot{x}(t) = 2t$ with $t \in [0, \infty)$ where the dot on top of $x(t)$ represents the derivative of $x(t)$ with respect to time

        * Solution without Initial Condition: the general solution is $x(t) = t^2 + C$

        * Solution with Initial Condition: suppose we impose the initial condition $x(0) = 1$, therefore $C = 1$ and the particular solution is $x(t)t^2 + 1$
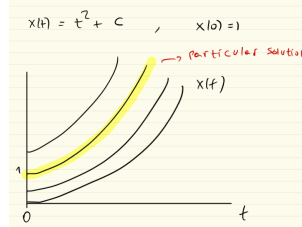

- Finite-Difference Methods for Solution: here we approximate derivates using finite-differences to approximate the solution to the ODE. This involves finding $\dot{x}(t) = F(t, x(t))$, $x(t_a) = x_a$, and $t \in [t_a, t_b]$. There are a wide range of methods depending on how the derivatives are approximated

    ○ Euler's Method: given the equation implied by the ODE

        * Step 1: specify a grid for t, i.e. $t_0 = t_a < t_1 < t_2 < \cdots < t_N = t_b$

        * Step 2: we can approximate the ODE using the difference equation

$$\frac{x(t_{i+1}) - x(t_i)}{t_{i+1} - t_i} = F(t_i, x(t_i))$$

        iterating over $i = 0, \ldots, N - 1$, where $x(t_0) = x_a$ is fixed by the initial condition

            ⊛ Note: the difference equation is obtained by rearranging the equation in step 1, we know all elements on the RHS of the difference equation which provides us with an explicit solution for $x(t_1)$

    ○ Euler's Method Illustration:

        * Step 1: discretize $[t_a, t_b]$

        * Step 2: iterate forward on $x(t_{i_1}) = x(t_i) + (t_{i+1} - t_1)F(t_i, x(t_i))$ $\forall i$

            ⊛ i=0: $x(t_1) = x(t_0) + (t_1 - t_0)F(t_0, x(t_0))$

            ⊛ i=1: ... (etc)

    ○ Euler's Method Numerical vs Analytical Solution Equivalency: $\dot{x}(t) = 2t$, $x(0) = 1$, $t \in [0, 1000]$
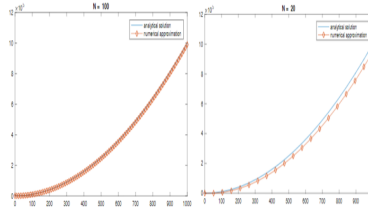
- ODE Systems: consider the system of two equations in two unknowns

$$\dot{x}(t) = f(t, x(t), y(t))$$
$$\dot{y}(t) = g(t, x(t), y(t))$$

where $t \in [t_a, t_b]$. The general solution typically depends on two arbitrary constants, say A and B, and can be written as

$$x(t) = \phi_1(t; A, B)$$
$$y(t) = \phi_2(t; A, B)$$

Here the two arbitrary constants A and B can be pinned down via two conditions on the solution which can causes two types of problems to emerge. In other words, the function must be solved such that either or both the initial and boundary conditions hold. Note that, like single-equation ODEs, exact solutions ODE systems can only be obtained under special cases

- Initial Value Problem(IVP): where you have conditions on the initial time (a) values - here $x(t_a) = x_a$, $y(t_a) = y_a$ can be solved numerically by applying Euler's Methods
  * Step 1: discretize the domain
  * Step 2: iterate forward starting from $x(t_a) = x_a$ and $y(t_a) = y_a$
- Boundary Value Problem (BVP): where you have a condition on the value at a time period (b) that different from the initial time value (a). Here $x(t_a) = x_a$ and $y(t_b) = y_b$, which can be solved by applying a shooting algorithm
  * Step 1: make an initial guess for $y(t_a)$ called y_a^G
  * Step 2: solve the ODE system by applying Euler's method given $x(t_a|y_a) = x_a$ and $y(t_a) = y_a^G$
  * Step 3: if $y(t_b)$ is 'close enough' to y_b then stop, else update y_a^G and go back to step 2

19
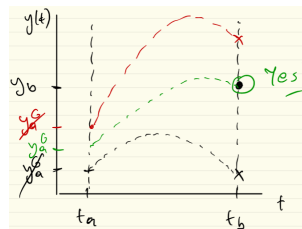
⊛ Note: the solution must cause the value of $x(t_a)$ to equal the a-priori $x(a)$ at time t_a and the value of $y(t_b)$ to equal the a-priori $y(b)$ at time t_b

- Higher-Order ODEs: techniques for solving ODE systems also apply to analysing higher-order Equations. Here we can express the second order ODE equation as a first order ODE system
  - Example: consider the second-order ODE

$$\ddot{x}(t) = f(t, \dot{(t)}, x(t))$$

where $\ddot{x}(t) \equiv \partial^2 x(t)/\partial t^2$. Define the new variables $y = \dot{x}$ and you are left with the two-equation ODE system

$$\dot{y}(t) = f(t, y(t), x(t))$$
$$\dot{x}(t) = y(t)$$

The Maximum Principle: the typical continuous time optimization problem is

$$\max_{x(t), y(t)} \int_0^{t_1} f(t, x(t), y(t)) \quad dt$$

subject to

$$\dot{x}(t) = g\big(t, x(t), y(t)\big), \quad x(t) \in \mathscr{X} \ \forall t, \quad y(t) \in \mathscr{Y} \ \forall t, \quad x(0) = x_0$$

There are two main issues with conventional solution methods to this problem; (1) we are choosing over infinitely dimensional objects such as the function $x : [0, t_1] \to \mathscr{X}$, (2) the constaints include a differential equation rather than a set of equalities/inequalities. In order to overcome these issues and find a solution, we apply the maximum principle theorem

- Notation and Assumptions: x is the state variable, y is the control variable, $\mathscr{X} \subset \mathbb{R}$ and $\mathscr{Y} \subset \mathbb{R}$ are nonempty and convex, f and g are continuously differentiable in their arguments, we define a Hamiltonian, and for simplicity we assume $t_1 < \infty$
  - Hamiltonian: we define the hamiltonian

$$H(t, x(t), y(t), \mu(t)) \equiv f(t, x(t), y(t)) + \mu(t)g(t, x(t), y(t))$$

where $\mu(t)$ is a continuously differentiable function called the costate variable

- Maximum Principle Theorem: suppose that the aforementioned continuous time problem has an interior continuous solution $(\widehat{x}(t), \widehat{y}(t))$. Then there exists a continuously differentiable function $\mu(t)$ such

that

$$H_y(t, \widehat{x}(t), \widehat{y}(t), \mu(t)) = 0 \quad \forall t \in [0, t_1]$$

$$\dot{\mu}(t) = -H_x(t, \widehat{x}(t), \widehat{y}(t), \mu(t)) \quad \forall t \in [0, t_1]$$

$$\mu(t_1) = 0$$

These operate as the necessary conditions for the Hamiltonian

- ○ <u>ODE Usage</u>: by $H_y(t, \widehat{x}(t), \widehat{y}(t), \mu(t)) = 0$ we can solve for y as a function of x thereby writting $y = Y(t, x\mu)$. Plugging $y = Y(t, x\mu)$ into $\dot{\mu}(t) = -H_x(t, \widehat{x}(t), \widehat{y}(t), \mu(t))$ and $\dot{x}(t) = g\big(t, x(t), y(t)\big)$ and combining the law of motion we are left with

$$\dot{x} = g\big(t, x, Y(t, x, \mu)\big)$$

$$\dot{\mu} = -H_x\big(t, x, Y(t, x, \mu), \mu\big)$$

$$x(0) = x_0, \quad \mu(t_1) = 0$$

which is an ODE system in x and μ that can be solved numerically by applying the shooting algorithm

- ○ <u>Sufficient Condition Requirements</u>: the maximum principle only provides necessary conditions for an optimum, sufficient conditions for a maximum rely on concavity properties of the objective and on convexity of the feasible set
- ○ <u>Technique Generalization</u>: the techniques generalize to multidimensional problems (where x and y are vectors), infinite horizon problems (ie $t_1 = \infty$), and to including terminal conditions on the final state (e.g. $x(t_1) = x_1$)
- ○ <u>Index</u>: the variable t may represent time or any other index

Heuristic Proof of The Maximum Principle: to solve problem (x) it would be natural to start trying "some sort" of lagrangian optimization such as constructing the function

$$\mathscr{L} = \int_0^{t_1} \{F(t, x(t), y(t)) + \mu(t)[g(t, x(t), y(t)) - \dot{x}(t)]\}$$

where $\mu(t)$ is a continuously differentiable function from the hamiltonian. Supposing that $(\widehat{x}(t), \widehat{y}(t))$ is an interior solution to the lagrange problem, then $(\widehat{x}, \widehat{y})$ should maximize $\mathscr{L}$.

- • <u>Lemma</u>: the function $\mathscr{L}$ can be written as

$$\mathscr{L} = \int_0^{t_1} \{f(t, x(t), y(t)) + \mu(t)g(t, x(t), y(t)) + \dot{\mu}(t)x(t)\} \quad dt - \mu(t_1)x(t_1) + \mu(0)x(0)$$

○ <u>Proof</u>: note that

$$\int_0^{t_1} \frac{\partial(\mu(t)x(t))}{\partial t} \; dt = \int_0^{t_1} \dot{\mu}(t)x(t) \; dt + \int_0^{t_1} \mu(t)\dot{x}(t) \; dt$$

If we integrate both sides over $[0, t_1]$ then we have

$$\mu(t_1)x(t_1) - \mu(0)x(0) = \int_0^{t_1} \dot{\mu}(t)x(t) \; dt + \int_0^{t_1} \mu(t)\dot{x}(t) \; dt$$

This can be rearranged as

$$\mu(t)\dot{x}(t) \; dt = \mu(t_1)x(t_1) - \mu(0)x(0) - \int_0^{t_1} \dot{\mu}(t)x(t) \; dt$$

Plugging this into our lagrangian yields the above lemma

- <u>Proof of Function Features</u>: consider a variation/perturbation around the optimal path $(\widehat{x}(t), \widehat{y}(t))$. Specifically, take an arbitrary function $P_y(t)$, let $\varepsilon \in \mathbb{R}$, and define $y(t, \varepsilon) = \widehat{y}(t) + \varepsilon P_y(t)$. Here $y(t, \varepsilon)$ is the solution plus a deviation from the solution $(\widehat{y}(t) + \varepsilon P_y(t))$ Graphically
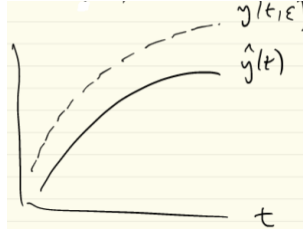
Similarly, let $x(t, \varepsilon) = \widehat{x}(t) + \varepsilon P_x(t)$ where $P_x(t)$ denotes the corresponding perturbation to $\widehat{x}(t)$ when $\widehat{y}(t)$ is varied according to $P_y(t)$. Note that the sum of the perturbations is the variation around the optimum. Define the 'perturbed' lagrangian function as

$$\mathscr{L}(\varepsilon) = \int_0^{t_1} \{f(t, x(t, \varepsilon), y(t, \varepsilon)) + \mu(t)g(t, x(t, \varepsilon), y(t, \varepsilon)) + \dot{\mu}(t)x(t, \varepsilon)\} \; dt - \mu(t_1)x(t_1, \varepsilon) + \mu(0)x_0$$

Recall that $\mathscr{L}$ is maximized at $\widehat{x}, \widehat{y}$, where the marginal gain from increasing ε (ie the perturbation) is 0, and therefore we must have

$$\frac{\partial \mathscr{L}(\varepsilon)}{\partial \varepsilon}\big|(\varepsilon, x, y) = (0, \widehat{x}, \widehat{y}) = 0$$

Using the definitions of $x(t, \varepsilon)$, $y(t, \varepsilon)$, and $\mathcal{L}$ the above equation can be expressed as:

$$0 = \int_0^{t_1} \left\{ \overbrace{f_x(t, \widehat{x}(t), \widehat{y}(t)) + \mu(t)g_x(t, \widehat{x}(t), \widehat{y}(t))}^{H_x(t,\widehat{x}(t),\widehat{y}(t),\mu(t))} + \dot{\mu}(t) \right\} P_x(t) \ dt$$
$$+ \int_0^{t_1} \underbrace{\left\{ f_y(t, \widehat{x}(t), \widehat{y}(t)) + \mu(t)g_y(t, \widehat{x}(t), \widehat{y}(t)) \right\}}_{H_y(\cdot)} P_y(t) \ dt - \mu(t_1)P_x(t_1)$$

○ <u>Note</u>: this expression can only hold for all feasible perturbation paths (P_x, P_y) if each integrand vanishes and $\mu(t_1) = 0$ such that $P_y(t)$ and $P_x(t)$ equal 0, which are the conditions of the maximum principle

Enforcement Frictions I - One-Sided

<u>Overview</u>: the analysis of contractual arrangements and enforcement frictions aims to rectify conflicts of interest that arise from violations of two assumptions - (1) both parties were perfectly committed to the rules of the exchange, (2) both parties in the exchange shared the same information. Dropping either of these assumptions creates conflicts of interest

- <u>Assumption 1 Violation</u>: here an agent who is not committed to the rules of the exchange would deviate from such rules when she finds it more convenient
- <u>Assumption 2 Violation</u>: here an agent who possesses private information would manipulate such information to her advantage

<u>Contract Theory</u>: the study of the economic aspects of contract design where a contract is a set of rules that facilitates cooperation among individuals with conflicting objectives. The goal of contract theory is twofold with the normative goal of how to draw up better contracts and the positive goal of analysing why real life contracts have various forms and designs. We can arbitrarily classify contract theory models by two classifications - (1) according to the type of private information, (2) according to the duration of the contractual relationship

- <u>Type of Private Information</u>: here we assume that the uninformed party designs the contract and address two types of information issues (adverse selection and moral hazard)
    - ○ <u>Adverse Selection</u>: asymmetric information about the characteristics of the informed party (e.g. life insurance where the insured's state of health is unknown by the insurer)
    - ○ <u>Moral Hazard</u>: asymmetric information about the actions of the informed party (e.g. CEO compensation where the CEO's actions are imperfectly observed by the Board of Directors)
- <u>Duration of Contractual Relationship</u>: here we address two types of contract periods (static and dynamic)
    - ○ <u>Static Contracts</u>: relationship lasts for one period
    - ○ <u>Dynamic Contracts</u>: relationship lasts for more than one period (gives rise to enforcement/commitment issues)

<u>Principal-Agent Paradigm</u>: since contracts involve two parties we need to address how the parties bargain over the terms of the exchange. We bypass the difficulties inherent to the bargaining process by allocating all bargaining power to one of the parties (the Principal) who will offer a take-it or leave-it contract to the Agent. The Principal-Agent Paradigm is the norm in contract theory

- <u>w.l.o.g</u>: the set of Pareto optimal contracts can be characterised by maximizing the utility of one agent subject to a given level of utility of the other agent. This outcome is a result of the Principal-Agent

analogy

<u>Enforcement Frictions</u>: occur when agents are free to walk away from the contract at any time, note that for most models of enforcement frictions we assume that all information is public

- <u>One-Sided Limited Commitment</u>: occurs when the principal is committed but the agent is not
    - <u>Applications</u>: long term labour contracts (principal is the firm and agent is the worker), international lending (principal is the lending country and agent is the borrowing country), firm dynamics (principal is the bank, agent is the entrepreneur seeking to finance a project)

<u>Basic Model</u>: the economy lasts for $T = \infty$ periods and consists of a village with a large number of households. Household preferences are

$$U = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_t)$$

where $\beta \in (0, 1)$ and u is concave and has the usual properties.

Each HH receives a stochastic endowment stream $\{y_t\}_{t=0}^{\infty}$ where $y_t \in \{\overline{y}_1, \overline{y}_2, \ldots, \overline{y}_S\}$ and $\overline{y}_{s+1} > \overline{y}_s$. Note that for each $t \geq 0$, y_t is independently and identically distributed according to the discrete probability distribution $\Pr(y_t = \overline{y}_s) = \pi_s$ where s is the set from which HHs draw endowments from. At time $t \geq 1$ the household has received a history of endowments $h_t = (y_t, y_{t-1}, \ldots, y_0)$

- <u>Assumptions</u>: the endowment processes are iid both across time and across households, overline represents the realizations of the endowments, due to concavity HHs value insurance against endowment realizations, y_t is non-storable and therefore self-insruance is precluded
- <u>No Contractual Issue Solution</u>: if there were a competitive equilibrium with complete markets, at $t = 0$ HHs would trade history and date contigent claims before the realization of endowments. Since all HHs are ex ante identical, each HH would end up consuming the per capita endowment in every period and their its lifetime utility would be

$$v_{pool} = \sum_{t=0}^{\infty} \beta^t u\Big(\sum_{s=1}^{S} \Pi_s \overline{y}_s\Big) = \frac{1}{1-\beta} u\Big(\sum_{s=1}^{S} \Pi_s \overline{y}_s\Big)$$

    HHs would thus insure away all of the risk associated with their individual endowment process. However, due to limited commitment and other contractual issues, this allocation is unattainable

<u>One-Sided Limited Commitment Contract Issue</u>: in this case, the HH can only trade with the moneylender and can walk away from the contract. If the HH walks away then the HH consumers their endowment and lives in autarky evermore (i.e. the HH cannot return from autarky). Note that only the HH can break the

contract, while the moneylender is fully committed to keeping their promise. Therefore, HHs must be induced not to walk away from the contract via structuring the contract in such a way that it is "self-enforcing" in that HHs prefer to conform to it.

This involves a contract between the moneylender and the HH as a sequence of history dependent functions $\{c_t\}_{t=0}^{\infty}$ with $c_t = f_t(y^t)$. The contract specifies that in each period the HH contributes her endowment realization y_t to the moneylender in exchange for consumption c_t.

From this arrangement, the moneylender earns an ex ante expected present value

$$P = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t (y_t - c_t)$$

where $(y_t - c_t)$ is the net transfer from HH at time t. The moneylender seeks to maximize the expected present value of profits from making contracts.

Given the HHs preferences in the basic model, the contract assigns the HH an expected present value of

$$v = \mathbb{E} \sum_{t=0}^{\infty} \beta^t u(c_t)$$

If the HH walks away from the contract, the ex ante value associated with consuming the endowment stream in autarky is

$$v_{aut} = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(y_t) = \frac{1}{1-\beta} \sum_{s=1}^{S} \Pi_s u(\overline{y}_s)$$

where the continuation utility under autarky is the weighted sum of the possible realizations from the endowment (which due to iid we can remove the dependency of y on time). Note that at time t the HH can guarantee itself a present value of utility of $u(y_t) + \beta v_{aut}$ by consuming its own endowment and therefore the moneylender's contract must offer the HH at least this utility at every possible history and every date (this forms the basis of the participation constraint).

Here the optimal contract can be analysed using recursive tools, this relies on the inclusion of a participation constraint and promise constraint to entice the HH to stay in the contract

- Assumptions: information is symmetric and therefore the moneylender can observe y_t, the HHs cannot borrow or lend with one another and can only trade with the moneylender, the money lender is risk-neutral and can borrow or lend at the risk free rate $R = \beta^{-1}$, only the money lender has access to the

risk-free loan market outside the village, R is exogeneous (i.e. this is a partial equilibrium model)

- <u>Notation</u>: our variables include

$\beta \in (0,1)$: the discount factor

c_t : the consumption that the moneylender gives to the HH in exchange for the endowment y_t under the contract

y_t : the endowment that the HH receives in time period t

y^t : refers to the endowment history up until time t (i.e. the sum of prior endowments)

$\overline{y_s}$: the observed endowment that is realized at time t

W_t : the sum of the money lender's net gains from each previous period's contract up until time t

w_s : refers to the moneylender's utility transfer to the HH the contract given a particular future state s from time t o

$\sum_{s=1}^{S} \Pi_s w_s$: the money lender's expected utility transfer to the HH from time $t \to \infty$

v_{aut} : the sum of future endowments (as utility) that the HH receives if living in autarky

v_t : the discounted sum of future gains (as utility) that the HH receives under the contract from time t onwards

s : refers to the different possible states at at a given time period

Here unspecified t refers to present time.

- HH "Forward-Looking" State Variable: note that for the HH, under the contract, the state variable is the promised expected discounted future value (i.e. the sum of all future earnings from the contract). In otherwords, the HH's state variable v_t is a "forward-looking" variable where $v_t = \mathbb{E}_0 \sum_{j=0}^{\infty} \beta^j u(c_{t+j})$. This "forward-looking" variable summarizes a stream of future utilities

- Moneylender "Backward-Looking" State Variable: note that for the moneylender, under the contract, the state variable is the total net gain $(y - c)$ at time t. In otherwords, the moneylender's state variable W_t is a "backward-looking" variable where $W_t = \sum_{t=0}^{t=T}(y_t - c_t)$ (i.e. W_t encodes all history dependence in the contract)

  - <u>Logic</u>: to see how this works, consider the following way of representing a contract $\{f_t\}$ in terms of a state variable x_t:

  $$c_t = g(x_t)$$
  $$x_{t+1} = \mu(x_t, y_t)$$

  Here g and μ are time-invariant functions. Notice that by iterating the $\mu(\cdot)$ function t times starting from (x_0, y_0), one obtains

  $$x_t = m_t(x_0; y_{t-1}, \dots, y_0), \quad t \geq 1$$

This allows one to summarize large dimensional shocks and rewards into a single variable and thus x_t summarizes histories of endowments

- Participation Contraint: since autarky is the outside option of the agents, the optimal contract must satisfy the participation constaints where the value of the contract is worth more than the value of autarky and is therefore "self-enforcing" (otherwise the HH will default). Since the money lender arrives at period y before y_t is realized and having previously made promised value v_t our participation constraint is:

$$u(\overbrace{f_t(y^t)}^{c_t}) + \underbrace{\beta\mathbb{E}_t \sum_{j=1}^{\infty} \beta^{j-1} u(\overbrace{f_{t+j}(y^{t+j})}^{c_{t+j}})}_{w_t(y^t):\text{ continuation utility at t}} \geq u(y_t) + \beta v_{aut}$$

for all t, y^t

We can rewrite our participation constraints as

$$u(f_t(y^t)) + \beta w_s(y^t) \geq u(y_t) + \beta v_{aut}$$

for all t, y^t where

$$w_s(y^t) \equiv \mathbb{E}_t \sum_{j=1}^{\infty} \beta^{j-1} u(f_{t+j}(y^{t+j}))$$

is the continuation utility of the agent at time t, given endowment history y^t

- Promise Keeping Constraint: the Promise Keeping Constraint states that the amount the HH gets in expectation must be greater than v, in other words the moneylender must be able to fulfil the contract such that

$$\sum_{s=1}^{S} \Pi_s[u(c_s) + \beta w_s] \geq v$$

- Moneylender Problem: each period the money lender chooses how much to give to the HH and how much that they will promise to transfer in the future (which is encoded in w_s) in order to make the contract "self-enforcing". Here the moneylender chooses c_s and a promised value starting in the next time period v. Therefore, the the money lender solves

$$P(v) = \max_{\{c_s, w_s\}} \sum_{s=1}^{S} \Pi_s[\overline{y}_s - c_s + \beta P(w_s)]$$

subject to

$$\sum_{s=1}^{S} \Pi_s[u(c_s) + \beta w_s] \geq v \rightarrow \text{ Promise Keeping Constraint}$$

$$u(c_s) + \beta w_s \geq u(\bar{y}_s) + \beta v_{aut}, \quad s = 1, \ldots, S \rightarrow \text{ Participation Constraint}$$

Here $w_s \in [v_{aut}, \bar{v}]$ where $\bar{v}$ is a large number. w_s is the *future* (i.e. next period) realized transfer under the contract established at time t given state possible s, given that $y = \bar{y}_s$ this period. w_s therefore represents the promised value which the HH will enter next period). We treat promised utility v as a state variable for the HH. Note that P must be a decreasing function of v because the higher the consumption stream of the villager, the lower must be the profits of the moneylender

- Characterisation: it can be shown that the constraint set is convex and that the value function $P(v)$ is stictly concave (as well as differentiable). Given this we have the following lemmas that essentially state that agents have incentives to walk away when income realisations are large enough - wherein the participation constraint will bind for a $\bar{y}_s$ that is significantly high.
  - * Lemma 1: for each v, there exists a threshold $\bar{y}(v)$ such that the PC binds binds if and only if $\bar{y}_s \geq \bar{y}(v)$
    - ⊛ Logic: in other words, for every possible "discounted sum of future gains", there exists a thesold endowment level such that the PC will bind if the actual endowment level is greater or equal to the figurative endowment level. This makes sense, since if the HH receives a particularly high endowment this period then the moneylender will offer distribute y_t with a lower proportion of $u(c_t)$ (though still a higher level of $u(c_t)$) and a higher proportion c_{t+j} to make the contract "self-enforcing"
  - * Lemma 2: the threshold $\bar{y}(v)$ is increasing in v
    - ⊛ Logic: this makes sense since a higher v occurs when c_{s+j} increases. c_{t+j} increases when $\bar{y}(v)$ is reached, with the moneylender offering a lower proportion (relative to y_s) of $u(c_t)$ as part of the contract (though $u(c_t)$ will still be higher than when receiving $y_1 < y_s$)
  - * Proof: we write the lagrangian as

$$\mathscr{L} = \sum_{s=1}^{S} \Pi_s[\bar{y}_s - c_s + \beta P(w_s)]$$

$$+ \varepsilon \left\{ \sum_{s=1}^{S} \Pi_s[u(c_s) + \beta w_s - v] \right\}$$

$$+ \sum_{s=1}^{S} \eta_s \left\{ u(c_s) + \beta w_s - u(\bar{y}_s) - \beta v_{aut} \right\}$$

where ε and η are the lagrangian multipliers on the Promise Keeping Constraint and Participation Contraint respectively. Taking first order conditions and applying the envelope

theorem yields the following:

$$\frac{\partial \mathscr{L}}{\partial C_s} = 0 \rightarrow -\Pi_s + \varepsilon\Pi_s u^{'}(C_s) + \eta_s u^{'}(C_s) = 0 \qquad (1)$$

$$\frac{\partial \mathscr{L}}{\partial w_s} = 0 \rightarrow \eta_s + \varepsilon\Pi_s = -\Pi_s P^{'}(w_s) \qquad (2)$$

$$\text{Envelope Theorem}: \quad \rightarrow P'(v) = -\varepsilon < 0 \qquad (3)$$

Rearranging equations (1) and (2) for $\varepsilon\Pi_s + \eta_s$ and solving them simultaneously yields

$$u^{'}(C_s) = -P^{'}(w_s)^{-1} \qquad (4)$$

It is clear to see that the LHS of this equation falls with C_s due to concavity of $u(\cdot)$ and the RHS of this equation falls with w_s due to concavity of $P(\cdot)$. This implies that C_s and w_s are positively related. In otherwords, as our lemma states, a higher C_s results in a higher w_s Combining (3) and (2) yields the law of motion for continuation Utility

$$P^{'}(w_s) = P^{'}(v) - \frac{\eta_s}{\Pi_s} \qquad (5)$$

<u>Lemma 1 Proof</u>

Take as given that $\overline{y}_s$ subject to PC is not binding (i.e. $\eta_s = 0$), therefore by (5) we have

$$P^{'}(w_s) = P^{'}(v) \Rightarrow w_s = v \quad \text{by concavity}$$

Using this fact into (4) we can write c as a function of only v and so have
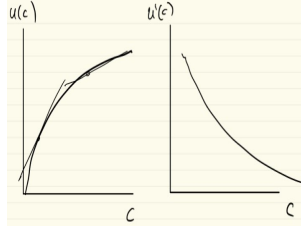
$$u^{'}(C_s) = -P^{'}(v)^{-1}$$

Which can be rewritten as $C_s = g_1(v), \quad g_1^{'} > 0$. This allows us to rewrite the PC as

$$u(C_s) + \beta w_s > u(\overline{y}_S) + \beta v_{aut}$$

$$\Downarrow$$

$$u(g_1(v)) + \beta v > u(\overline{y}) + \beta v_{aut} \equiv \overline{y}(v)$$

$$\Downarrow$$

$$\underbrace{u^{-1}(u(g_1(v))_\beta(v - v_{aut}))}_{\equiv \overline{y}(v)} > \overline{y}_s$$

Thus, given the definition of $\overline{y}(v)$, the PC is not binding if and only if $\overline{y}_s < \overline{y}(v)$

Lemma 2 Proof

Given the definition of $\bar{y}(v)$ from the proof for Lemma 1, this is straightforward


- Properties of PC Binding vs Not Binding Suppose PC is not binding, then $w_s = v$ and $C_s = g_1(v)$ where $g_1' > 0$. Therefore, if PC is not binding then w_s and C_s remain constant with consumption in state s depending on promised utility v but not on the endowment in state s

  Suppose PC is binding, then $w_s = l(\bar{y}_s) > v$ and $C_s = g_2(\bar{y}_s) \in (g_1(v), \bar{y}_s)$ where $l', g_2' \geq 0$.

  Therefore, if PC is binding then w_s and C_s must increase to induce the agent not to walk away from the contract - wherein w_s and c_s only depend on current income realisations (and not on the history of shocks given by v)

  ○ Consumption under Optimal Contract: graphically, given v, consumption under the optimal contract is


  Full insurance against low $\bar{y}_s$.

  $cov(c_s, \bar{y}_s) > 0$ for large $\bar{y}_s$.
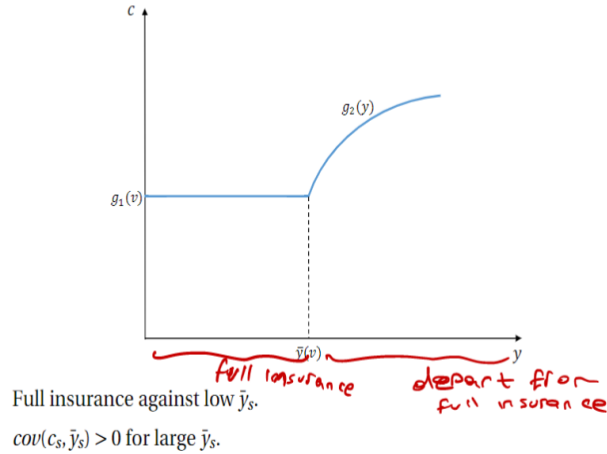
  ○ Long Run Dynamics: c_t increases over time until $\bar{y}_S$ is realized. All agents eventually become fully insured (no fluctuations in consumptions) when their highest income realization $\bar{y}_S$ is realized and therefore C eventually becomes constant.
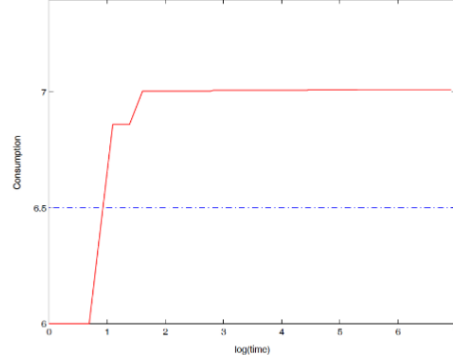
In graph: $c_t \to g_1(\bar{v})$, where $\bar{v}$ is given by:

$$u(g_1(\bar{v})) + \beta\bar{v} = u(\bar{y}_S) + \beta v_{aut}.$$

With many HHs, all agents eventually get to consume $g_1(\overline{v})$: $\lim_{t\to\infty} \Pr(c_t = g_1(\overline{v})) = 1$. Overall, we have "fanning in" of the cross-sectional distribution of consumption and full risk sharing in the long run (ie t=500)
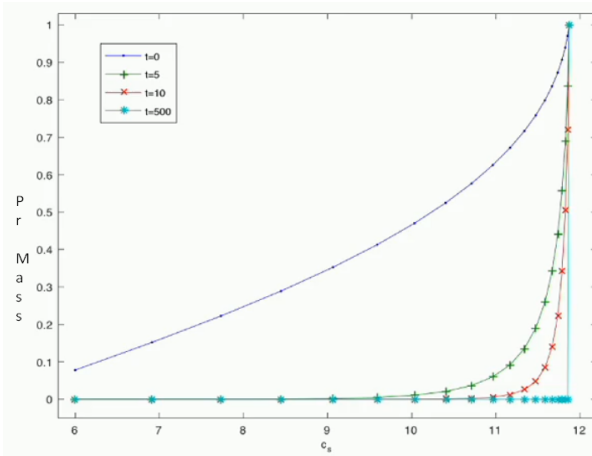
○ Not Binding Proof: shown in the proof of the lemma

○ Binding Proof: assume that PC is binding

* Step 1: show that w_s and c_s are only functions of $\overline{y}_s$

  By FOC (4) we have $u^1(C_s) = -P'(w_s)^{-1}$ and given the fact that PC is binding we have $u(C_s) + \beta w_s = u(\bar{y}_s) + \beta v_{aut}$. Note that these two equations form a system of 2 equations in 2 unknowns (C_s, w_s). Since promised v does not show up in this system we can write its solution as $w_s = l(\overline{y}_s)$ and $C_s = g_2(\overline{y}_s)$

* Step 2: show that $g_2' > 0$ and $l' > 0$

  By $u(C_s) + \beta w_s = u(\bar{y}_s) + \beta v_{aut}$ we know that $[u(C_s) + \beta w_s]$ is non-decreasing in $\overline{y}_s$. Also, by $u^1(C_s) = -P'(w_s)^{-1}$ (as shown previously), C_s is non-decreasing in w_s and hence C_s should be non-decreasing in $\overline{y}_s$

* Step 3: show that $w_s > v$

  By $P'(w_s) = P'(v) - \frac{\eta_s}{\Pi_s}$, if the PC binds then $\eta_s > 0$ and so $P'(w_s) < P'(v) \Rightarrow w_s > v$ via

32

concavity

* <u>Step 4</u>: show that $C_s \in (g_1(v), \overline{y}_s)$

  A binding PC implies that $u(C_s) + \beta w_s = u(\overline{y}_s) + \beta v_{aut}$. Therefore, $c_s < \overline{y}_s$. Because $w_s > v$ we have

$$-P'(w_s)^{-1} < -P'(v)^{-1}$$

$$\Downarrow$$

$$u'(C_s) < u'(g_1(v)) \quad \text{by (4)}$$

$$\Downarrow$$

$$C_s > g_1(v)$$

• <u>Connection to Labour Contracts</u>: early work on one-sided limited commitment models aimed at analysing the characteristics of labour contracts. In our model this can be mapped as C being the wage offered within a firm, y shock being wages offered in the spot market, and the enforcement friction being labour mobility. Since workers are risk-averse, they value insurance against fluctuations in the spot market for wages (i.e. implicit insurance premium in contractual wages). Firms find contractual arrangement beneficial because in the short run they can pay wages below those in the spot market (due to the implicity insurance premium), which is embedded by $C_s < \overline{y}_s$ when PC binds in our model

Enforcement Contracts II - Two-Sided

Overview: two agents "share risk" if they use state-contingent transfers to increase the expected utility of both, by reducing the risk of at least one. In the absence of contractual frictions there should be full risk sharing with all idiosyncratic risk eliminated and only aggregate risk affecting individuals' consumption. However, despite this we find that, even after conditioning for aggregate income, individual income is positively correlated with current and lagged individual income (i.e. risk is not shared across time). This inconsistency has three causes.

- Cause 1 - Exogeneously Incomeplete Markets: occurs when there are more states than assets and therefore transferring wealth across certain is restricted. For example, in Bewley-Hugget-Aiyagari models agents can only use a single bond to insure against idiosyncratic labour income shocks. However, this involves assuming that some asset markets are shut down - such as through arbitray assumptions about which assets are available (conclusions are sensitive to these assumptions)
- Cause 2 - Private Information Frictions: occurs when consumption is correlated with individual income to induce truth-telling or adequate efforce. This involves two major drawbacks:
  - Drawback 1 - Imperfect Risk Sharing: holds event in environments where informational asymmetries are small (e.g. village economies or within families)
  - Drawback 2 - Longrun Implications of Private Infomration: these models are counterfactual
- Cause 3 - Two-sided Limited Commitment: occurs when there is an optimal insurance arrangement between two agents who lack commitment. This differs from one-sided limited commitment (where agents can default but principals cannot) in that both the principal and the agent can default. The key results of modelling two-sided limited commitment are that incomeplete risk sharing can be perpetual and the amount of risk sharing endogenously emanates from limited enforcement frictions

Kocherlakota Model of Two-sided Limited Commitment: involves two infinitely-lived agents subject to endowment shocks where the state of the world is determined by realization of $\theta_t \in \{1, \ldots, S\}$, which is iid with $\Pr(\theta_t = s) = \pi_s > 0$. In other words, the endowment profile of the agents, (y_t^1, y_t^2) is determined by realization of θ_t. The aggregate endowment is $Y_t = y_t^1 + y_t^2 \geq 0$ and the joint distribution of endowments is symmetric such that $\Pr(y, y^{'}) = \Pr(y', y)$ [1]. Given identical preferences, both agents have the following objective function

$$\mathbb{E}_0 \sum_{t=1}^{\infty} \beta_{t-1} u(c_t)$$

where $\beta \in (0,1)$, u is increasing, strictly concave, and differentiable. An allocation $\mathbf{c} = \left\{ c_t^j \right\}_{t=1, j=1}^{\infty, 2}$ is

---

[1] a special case occurs where there is no aggregate uncertainty (i.e. $Y_t = 1$), $\theta_t \in [0, 1]$, $y_t^1 = \theta_t$, $y_t^2 = 1 - \theta_t$, and the distribution of θ_t is symmetric around $1/2$

feasible if and only if

$$c_t^1 + c_t^2 \leq Y_t$$

for all $t = 1, 2, \ldots, \infty$

- <u>Assumptions</u>: there is no outside lender (agents can only contract with other agents)
- <u>Contract</u>: agents sign a contract to insure against endowment shocks and each agent can walk away from the contract at time t. Agents proceed with the following steps per period
  - <u>Step 1</u>: θ_t is realized and therefore (y_t^1, y_t^2) is observed by both agents
  - <u>Step 2</u>: each individual j simultaneously transfers $TR_t^j = y_t^j - c_t^j$ to the other
- <u>Autarky</u>: if an agent defaults, they live in autark forever with utility

$$v_{aut} = \sum_{s=1}^{S} \pi_s \left[u(y_s^j) + \beta V_{aut} \right]$$

  Note that the symmetry of distribution of (y_t^1, y_t^2) guarantees that $v_{aut}^1 = v_{aut}^2$

- <u>Participation Constraints</u> in order to induce participation and prevent an agent from defaulting the following participation constraint must hold

$$u(c_t^j) + \mathbb{E}_t \sum_{\tau=1}^{\infty} \beta^\tau u(c_{t+\tau}^j) \geq u(y_t^j) + \beta V_{aut}, \quad j = 1, 2$$

  for all dates and states. Note that an allocation $\mathbf{c}$ is sustainable if it is feasible and satisfies participation constraints

- <u>Efficient Allocations</u>: a sustainable allocation $\mathbf{c}$ is efficient if there is no other sustainable allocation providing both individuals with at least as much utility and one individual with more (i.e. an efficient allocation is Pareto Optimal)
- <u>Contract Design with a Principal and an Agent</u>: efficient allocations can be found by solving the planning problem
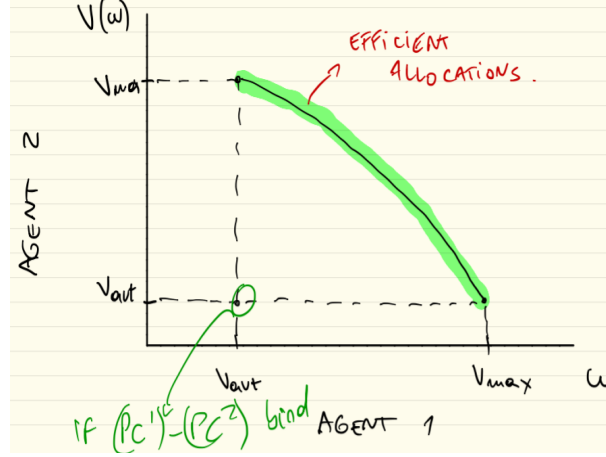
$$V(w) = \max_{c_s, w_s} \sum_{s=1}^{S} \pi_s \left[u(Y_s - c_s) + \beta V(w_s) \right]$$

$$\text{s.t:} \quad \sum_{s=1}^{S} \pi_s [u(c_s) + \beta w_s] = w \qquad (PK_c)$$

$$u(c_s) + \beta w_s \geq u(y_s^1) + \beta V_{aut}, \quad \forall s \qquad (PC^1)$$

$$u(Y_s - c_s) + \beta V(w_s) \geq u(Y_s - y_s^1) + \beta V_{aut}, \quad \forall s \qquad (PC^2)$$

  where $w \in [V_{aut}, V_{max}]$ Note that (c_s, w_s) are the consumption and continuation utility allocations of Agent 1 given state s, $(Y_s - c_s, V(w_s))$ are the consumption and continuation utility allocations of Agent 2 given state s, and $V(w)$ represents the Pareto frontier (traced out by changing w).

Here the planner enters period t having promised agent 1 a certain amount of ex ante utility u_0 and given this promise seeeks to maximize the amount of ex ante utility agent 2 receives. The planner determines how much consumption to give to or take from agent 1 and how much future utility to promise agent 1, contingent on each state of the world. The planner must take into account the participation constraints that capture his inability to force the agents to give up consumption beyond threatening them with future autarky.

- <u>First Best Allocation</u>: an allocation is first best if $c_t^1 + c_t^2 = Y_t$ for all dates and states and $u'(c_t^1)/u'(c_t^2)$ is also constant over all dates and states. Therefore, the individual's consumption is a time and state invariant function of the aggregate endowment. Therefore, we solve:

$$V(w) = \max_{c_s, w_s} \sum_{s=1}^{S} \pi_s [u(Y_s - c_s) + \beta V(w_s)]$$

$$\text{s.t:} \ \sum_{s=1}^{S} \pi_s [u(c_s) + \beta w_s] = w$$

  * <u>Proof</u>: writing the Lagrangian we have

$$\sum_{s=1}^{S} \pi_s [u(Y_s - c_s) + \beta V(w_s)] - \lambda \left\{ \sum_{s=1}^{S} \pi_s [u(c_s) + \beta w_s] - w \right\}$$

  The FOCs of the First Best Allocation includes:

$$\frac{\partial \lambda}{\partial c} = 0 \Rightarrow -\Pi_s u'(Y_s - c_s) + \lambda \Pi_s u'(c_s) = 0$$

Rearranging this we have

$$\frac{u'(Y_s - c_s)}{u'(c_s)} = \lambda$$

This implies full risk sharing with individual consumption only function with average income

○ <u>Solution</u>: the solution will be driven by the pattern of binding PCs where the possible state of the world can be divided into four groups

　* <u>State 1</u>: in which PC_1 binds

　* <u>State 2</u>: in which PC_2 binds

　* <u>State 3</u>: in which neither PC binds

　* <u>State 4</u>: in which both PCs bind

Since state 4 never holds if the allocation is efficient only state 1-3 are possibilities

　* <u>Proof</u>: suppose that PC_1 is binding, then $u(c_s) + \beta w_s = u(y'_s) + \beta v_{aut}$. This implies that $c_s \leq y_s^1$ since $w_s \geq v_{aut}$ Suppose that PC_2 is binding, then $u(Y_s - c_s) + \beta v(w_s) = u(y_s^2) + \beta v_{aut}$. This implies the following:

$$u(Y_s - c_s) - u(y_s^2) = \beta \overbrace{[v_{aut} - v(w_s)]}^{\leq 0}$$
$$u(y_s - c_s) \leq u(y_s^2)$$
$$u(Y_s - c_s) \leq u(Y_s - Y_s^1)$$
$$Y_s - c_s \leq Y_s - Y_s^1$$

Therefore $c_s \geq y_s^1$.

This means that if both participation constraints bind we have $PC^1 \rightarrow c_s \leq y_s^1$ and $PC^2 \rightarrow c_s \geq y_s^1$, therefore $c_s = y_s^1$. This implies that $w_s = v_{aut}$ and $v(w_s) = v_{aut}$, thus both PCs binding is not an efficient solution

○ <u>Efficient Allocations</u>: it can be shown that V is differentiable, decreasing, and concave. This implies the following evolution of promised utility given our states:

　* <u>State 1 Evolution</u>: if PC_1 binds then $w_s > w$

　* <u>State 2 Evolution</u>: if PC_2 binds then $w_s < w$

　* <u>State 3 Evolution</u>: if neither PC binds then $w_s = w$

This means that, unlike with one-sided commitment, promised values can fall since agents will alter their promised values with eachother in order to achieve insurance under the contract

* Proof: we have the following objective function

$$v(w) = \max_{c_s, w_s} \sum_s \Pi_s [u(y_s - c_s) + \beta v(w_s)]$$

$$\text{s.t:} \ \ \eta \sum_s \Pi_s [u(c_s) + \beta w_s] = w$$

$$\mu_s^1 u(c_s) + \beta w_s \geq u(y_s') + \beta v_{aut}, \quad \forall s$$

$$\mu_s^2 u(Y_s - c_s) + \beta v(w_s) \geq u(y_s^2) + \beta v_{aut}, \quad \forall s$$

Note that η, μ_s^2, and μ_s^1 are arbitrarily attached multipliers. Taking first order conditions and applying the envelope theorem yields the following:

$$\frac{\partial v}{\partial w_s} = 0 \rightarrow \Pi_s \beta v'(w_s) + \eta \Pi_s \beta + \mu_s^1 \beta + \mu_s^2 \beta V'(w_s) = 0 \tag{1}$$

$$\frac{\partial v}{\partial c_s} = 0 \rightarrow -\Pi_s u'(Y_s - c_s) + \eta \Pi_s u'(c_s) + \mu_s^1 u'(c_s) - \mu_s^2 u'(y_s - c_s) = 0 \tag{2}$$

Envelope Theorem $\rightarrow v'(w) = -\eta \tag{3}$

We need to consider the 3 states: (1) when PC_1 binds and therefore $\mu_s^1 > 0$, $\mu_s^2 = 0$, (2) when PC_2 binds and therefore $\mu_s^1 = 0$, $\mu_s^2 > 0$, (3) no PC binds and therefore ($\mu_s^1 = \mu_s^2 = 0$)

⊛ State 1 - PC_1 Binds: by (1) and (3) we have

$$\Pi_s v'(w_s)\underline{\beta} - v'(w)\Pi_s \underline{\beta} + \mu_s^1 \underline{\beta} = 0$$

$$\Updownarrow$$

$$v'(w_s) = v'(w) - \frac{\mu_s^1}{\Pi_s}$$

Therefore $v'(w_s) < v'(w)$ and thus, by concavity, $w_s > w$

⊛ State 2 - PC_2 Binds: by rearranging equations (1) and (3) we get

$$v'(w_s) = \frac{\Pi_s}{\Pi_s + \mu_s^2} v'(w)$$

$$\Downarrow$$

$$\frac{v'(w_s)}{v'(w)} = \frac{\Pi_s}{\Pi_s + \mu_s^2} < 1$$

$$\Downarrow$$

$$v'(w_s) > v'(w)$$

Therefore, $w_s < w$

38

⊛ <u>State 2 - No PC Binds</u>: by (1) and (3) we have that $v^{'}(w_s) = v^{'}(w)$. Therefore, $w_s = w$

Analogously to what we showed in the one-sided limited commitment model, it follows that:

  * PC_1 binds when y_s^1 is high

  * PC_2 binds when y_s^2 is high

  * w_s is a non-decreasing function of c_s

Combined with the evolution of promised utilitys this implies that $Cov(c_t, y_t|Y_t) \geq 0$ and therefore we have imperfect risk sharing. Given the correlation result with imperfect risk sharing, lagged individual income occurs since in an efficient allocation

$$Cov(c_t^j, y_{t-k}^j|Y_{t-k}) \geq 0$$

for $k \geq 0$ and $j = 1, 2$

○ <u>Long-Run Dynamics</u>: suppose the first best allocation is not sustainable. Then as $t \to \infty$, $\Pr(w_t|w_0)$ converges in distribution to the same non-degenerate limiting distribution for all w_0. As a result the first best is typically not sustainable whenever β is small enouggh. This result strengthens the previous result on imperfect risk sharing in two dimensions:

  * <u>Dimension 1</u>: it holds in the long run, as the limiting distribution is non-degenerate

  * <u>Dimension 2</u>: the same distribution (and hence correlation) is obtained regardless of initial promised values

<u>Graphical Intuition for Perpetual Imperfect Risk Sharing</u>

One Sided Limited Commitment
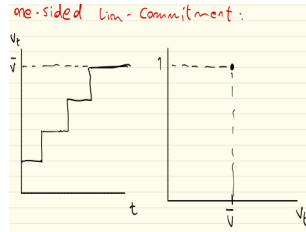
Two Sided Limited Commitment

Mirrlees Taxation Model with Two Types

<u>Variable Overview</u>: let y denote before-tax income and let $T(y)$ denote the income tax schedule. The optimal income tax system $T^*(y)$ is the one that maximizes social welfare subject to a revenue requirement. $T^*(y)$ can be found by either the Ramsey approach of Mirrlees approach

<u>Ramsey Approach</u>: makes parametric assumptions on the tax function $T(y)$ (i.e. a progressive tax function such as $T(y) = y - \lambda y^{1-\tau}$) and precludes non-distortionary lump sum taxes (i.e. $T(y) = constant$ is not allowed)

- <u>Advantages</u>: tractability
- <u>Disadvantages</u>: the set of possible tax instruments is exogeneous, by precluding lump-sum taxation the taxes are distortionary by assumption, the optimal tax function is arbitrarily restricted

<u>Mirrlees Approach</u>: feasible tax instruments are endogeneised given informational frictions by featuring workers who are traditionally heterogeneous across labour productivity/skill (which are private information). This leads to a trade off between redistribution and work incentives where the right balance in this trade-off will shape tax distortions and hence the optimal tax function $T^*(y)$

- <u>Contractual Friction</u>: applies tools from contract theory and adverse selection where private information acts as a contractual frictions
- <u>Static Environments</u>: involves either a discrete number of different worker skill types or a continuous number of different worker skill types

<u>Mirrlees Model with Two Types</u>: involves a static economy with a single consumption good and a continuum of workers of one of two skill types. No ad-hoc structure on tax design ($T^*(y)$) is imposed and the only contraints on $T^*(y)$ arise endogenously due to informational frictions. Workers earn money on the basis of their skills and their effort, so there is a tradeoff between effort and taxes the planner applies

- <u>Assumptions</u>: θ and n are private information, y is observable, aggregate output is linear in effective effort (implies that in a competitive equilibrium with a perfect labor market that y equals before tax income), workers have no income effects due to quasi-linear preferences (there are only substitution effects of after-tax income)
- <u>Heterogeneity of Workers</u>: workers are heterogeneous across skills/labor productivity where $\theta \in \Theta$ and $F : \Theta \to [0,1]$ denotes the cumulative distribution function of θ
- <u>Worker Preferences and Transformation</u>: workers have quasi-linear preferences over consumption c and

effort n:

$$U(c, n) = c - v(n)$$

where $v', v'' > 0$ (i.e. v is convex). Workers transform effort into output y (or effective labour) using the technology function $y = \theta \cdot n$. Note that n is substituted by y/θ due to n not being observable, allowing us to write the contract in terms of observables

- Key Tradeoff: there is a tradeoff between equity and efficieny. This is as confiscatory taxation improves upon the distribution of the 'social pie' but also reduces production and thus the size of the pie
- Allocation (the contract): an allocation is an assignment of c and y given θ:

$$\{c(\theta), y(\theta)\}_{\theta \in \Theta} \quad \Rightarrow \quad \text{the contract}$$

This allocation is feasible if and only if

$$\int c(\theta) dF(\theta) + G \leq \int y(\theta) dF(\theta)$$

where the LHS is aggregate consumption plus exogenous government expenditure, the RHS is aggregate output.

- Social Welfare Function: the government aims to allocate c and y across θ to maximize the SWF:

$$SWF = \int \Psi\big(c(\theta) - v(\frac{y(\theta)}{\theta})\big) dF(\theta)$$

where $\Psi' > 0$ and $\Psi'' < 0$ (ie is increasing and concave). The SWF encapsulates the normative criterion followed by the government and given the concavity of Ψ displays a preference for redistribution across θs

- Aim: to design an income tax system that maximizes social welfare. More precisely, let $T(y)$ be the income tax schedule so that agents face the budget constraint $c \leq y - T(y)$. In this case, we are seeking a function $T^*(y)$ that maximizes the SWF in a competitive equilibrium and focus on characterising optimal marginal tax rates $T^{*'}(y)$ across the income distribution
- Solution for Optimal Taxes: we find $T^*(y)$ by following two steps
  - Step 1 - Centralization (Contract Design): characterise the allocations for $\{c^*, y^*\}$ that maximise the SWF subject to informational constraints
    - First Best: Start by analyzing the case in which the government can observe θ. At this 'first

best', the government solves:

$$\max_{\{c(\theta),y(\theta)\}_{\theta\in\Theta}} \int \Psi\Big(c(\theta) - v(\frac{y(\theta)}{\theta})\Big)dF(\theta)$$

subject to

$$\int c(\theta)dF(\theta) + G \leq \int y(\theta)dF(\theta) \tag{RC}$$

Assume that the set of possible types (i.e. two) is given by $\Theta = \{\theta_L, \theta_H\}$ with $\theta_L < \theta_H$ and let $\pi_i \in (0,1)$ denote the probability of being type $theta_i$. Therefore, with two types, the first best becomes:

$$\max_{c_L,y_L,c_H,y_H} \sum_{i\in\{L,H\}} \pi_i \Psi\Big(c_i - v(\frac{y_i}{\theta_i})\Big)$$

subject to

$$\sum_{i\in\{L,H\}} \pi_i c_i + G \leq \sum_{i\in\{L,H\}} \pi_i y_i \tag{RC}$$

It can be shown that at the first best:

⊛ <u>Property 1</u>: everyone gets the same utility, $c_L^{FB} - v(\frac{y_L^{FB}}{\theta_L}) = c_H^{FB} - v(\frac{y_H^{FB}}{\theta_H})$

⊛ <u>Property 2</u>: high types work more, $y_H^{FB} > y_L^{FB}$

<u>Issue</u>: the first best cannot be implemented under private information since high types have incentives to misreport their type

⊛ <u>Proof</u>: using $\theta_H > \theta_L$ and $y_H^{FB} > y_L^{FB}$ we get:

$$c_L^{FB} - v(\frac{y_L^{FB}}{\theta_H}) > c_L^{FB} - v(\frac{y_L^{FB}}{\theta_L})$$
$$= c_H^{FB} - v(\frac{y_H^{FB}}{\theta_H^{FB}})$$

which follows property 1 and implies that if the principal offered the First Best contract, both types of agents would choose (c_L^{FB}, y_L^{FB}). Therefore, there will not be equal utility between types.

∗ <u>Second Best</u>: the principal can generally do better than the first best by screening types. This is done via offering a contract through which different types get different allocations - wherein, under the Revelation Principle, asking agents directly about their types can occur without loss of generality.

With the Revelation Principle we add the Incentive Compatibility Contraints (IC_H and IC_L), which requires that individuals should be better-off by revealing their true types. Therefore,

the second best problem becomes:

$$\max_{c_L,y_L,c_H,y_H} \sum_{i\in\{L,H\}} \pi_i \Psi\left(c_i - v(\frac{y_i}{\theta_i})\right)$$

subject to

$$\sum_{i\in\{L,H\}} \pi_i c_i + G \leq \sum_{i\in\{L,H\}} \pi_i y_i \tag{RC}$$

$$c_H - v(\frac{y_H}{\theta_H}) \geq c_L - v(\frac{y_L}{\theta_H}) \tag{IC_H}$$

$$c_L - v(\frac{y_L}{\theta_L}) \geq c_H - v(\frac{y_H}{\theta_L}) \tag{IC_L}$$

This is a direct mechanism since allocations are indexed by θ-types. It can be shown that only IC_H will be binding at the optimum.

⊛ <u>Logic</u>: this is a reflection of the equity and efficiency tradeoff. Since the redistribution from θ_H to θ_L is desired, θ_H might mimic θ_L and work less (thus decreasing the size of the pie). The binding of IC_H means that θ_H should be provided with incentives not to shirk. The ICs effectively are just constraints that a types utility from correctly reporting must be greater than or equal to their utility from misreporting

Since only IC_H is binding, the second best problem becomes:

$$\max_{c_L,y_L,c_H,y_H} \sum_{i\in\{L,H\}} \pi_i \Psi\left(c_i - v(\frac{y_i}{\theta_i})\right)$$

subject to

$$\sum_{i\in\{L,H\}} \pi_i c_i + G \leq \sum_{i\in\{L,H\}} \pi_i y_i \tag{λ}$$

$$c_H - v(\frac{y_H}{\theta_H}) \geq c_L - v(\frac{y_L}{\theta_H}) \tag{μ}$$

where $\lambda, \mu > 0$ and denote the multipliers on RC and IC_H respectively.

⊛ <u>First Order Conditions</u>: taking FOCs on the second best problem with IC_H binding yields:

$$0 = \frac{\partial U}{\partial c(\theta_L)} = \pi(\theta_L)\Psi'[c(\theta_L) - v(\frac{y(\theta_L)}{\theta_L})] - \lambda\pi(\theta_L) - \mu \tag{1}$$

$$0 = \frac{\partial U}{\partial c(\theta_H)} = \pi(\theta_H)\Psi'[c(\theta_H) - v(\frac{y(\theta_H)}{\theta_H})] - \lambda\pi(\theta_H) - \mu \tag{2}$$

$$0 = \frac{\partial U}{\partial y(\theta_L)} = -\theta(\lambda_L)\Psi'[c(\lambda_L) - v(\frac{y(\theta_L)}{\theta_L})]v'(\frac{y(\theta_L)}{\theta_L})\frac{1}{L} + \lambda\pi(\theta_L) + \mu v'(\frac{y(\theta_L)}{\theta_H})\frac{1}{\theta_H} \tag{3}$$

$$0 = \frac{\partial U}{\partial y(\theta_H)} = -\theta(\lambda_H)\Psi'[c(\lambda_H) - v(\frac{y(\theta_H)}{\theta_H})]v'(\frac{y(\theta_H)}{\theta_H})\frac{1}{H} + \lambda\pi(\theta_H) + \mu v'(\frac{y(\theta_H)}{\theta_H})\frac{1}{\theta_H}$$

$$\tag{4}$$

○ <u>Step 2 - Decentralisation</u>: back out the $T^*(y)$ that implement $\{c^*, y^*\}$ as a competitive equilibrium

  * <u>Framework</u>: in a decentralised economy an agent of type θ solves:

$$\max_{c,y} c - v(\frac{y}{\theta})$$

$$\text{subject to}$$

$$c = y - T(y)$$

This implies that T' satisfies:

$$1 - T'(y) = v'(\frac{y}{\theta})\frac{1}{\theta} \tag{5}$$

This, together with the FOCs, can be used to characterise optimal marginal taxes $T^{*'}(y)$ for y_L and y_H.

  * <u>High Skilled Taxation</u>: recall that

$$1 - T^{*'}(y(\theta_H)) = v'(\frac{y(\theta_H)}{\theta_H})\frac{1}{\theta_H}$$

By equation (2) we have that:

$$\pi(\theta_H)\Psi'(\mathscr{U}(\theta_H)) + \mu = \lambda\pi(\theta_H) \tag{2'}$$

$$\text{where:} \quad \mathscr{U}(\theta_H) = c(\theta_H) - v(\frac{y(\theta_H)}{\theta_H})$$

By equation (4) we have that:

$$v'(\frac{y(\theta_H)}{\theta_H})\frac{1}{\theta_H}\left[\pi(\theta_H)\Psi'(\mathscr{U}(\theta_H)) + \mu\right] = \lambda\pi(\theta_H) \tag{4'}$$

By (2') and (4') we have that:

$$v'(\frac{y(\theta_H)}{\theta_H})\frac{1}{\theta_H} = 1 \tag{*}$$

By (5) we have

$$T'(y(\theta_H)) = 0 \tag{**}$$

implying that we have zero marginal taxation at the top.

* <u>Low Skilled Taxation</u>: by (1) we have:

$$\lambda \pi(\theta_L) = \pi(\theta_L)\Psi^{'}(\mathscr{U}(\theta_L)) - \mu \tag{1'}$$

By (3) we have:

$$\lambda \pi(\theta_L) = \pi(\theta_L)\Psi^{'}(\mathscr{U}(\theta_L))v^{'}(\frac{y(\theta_L)}{\theta_L})\frac{1}{\theta_L} - \mathscr{U}v^{'}(\frac{y(\theta_L)}{\theta_H})\frac{1}{\theta_H} \tag{3'}$$

Combining (1') and (3') yields:

$$\pi(\theta_L)\Psi^{'}(\mathscr{U}(\theta_L)) - \mu = \pi(\theta_L)\Psi^{'}(\mathscr{U}(\theta_L))v^{'}(\frac{y(\theta_L)}{\theta_L})\frac{1}{\theta_L} - \mu v^{'}(\frac{y(\theta_L)}{\theta_H})\frac{1}{\theta_H}$$

which is equivalent to

$$\pi(\theta_L)\Psi^{'}(\mathscr{U}(\theta_L))\underbrace{\left[1 - v^{'}(\frac{y(\theta_L)}{\theta_L})\frac{1}{\theta_L}\right]}_{\equiv T^{'}(y(\theta_L))} = \mu\left[1 - v^{'}(\frac{y(\theta_L)}{\theta_H})\frac{1}{\theta_H}\right]$$

Rearranging this and inserting $+/- \mu v^{'}(\frac{y(\theta_H)}{\theta_H})\frac{1}{\theta_H}$ we have:

$$T^{'}(y(\theta_L)) = \mu\left[1 - v^{'}(\frac{y(\theta_L)}{\theta_H})\frac{1}{\theta_H}\right]\frac{1}{\pi(\theta_L)\Psi^{'}(u(\theta_L))}$$

$$T^{'}(y(\theta_L)) = \mu\Big[\underbrace{1 - v^{'}(\frac{y(\theta_H)}{\theta_H})\frac{1}{\theta_H}}_{\equiv 0 \text{ by (**)}} + v^{'}(\frac{y(\theta_H)}{\theta_H})\frac{1}{\theta_H} - v^{'}(\frac{y(\theta_L)}{\theta_H})\frac{1}{\theta_H}\Big]\frac{1}{\pi(\theta_L)\Psi^{'}(u(\theta_L))}$$

$$T^{'}(y(\theta_L)) = \mu\Big[v^{'}(\frac{y(\theta_H)}{\theta_H})\frac{1}{\theta_H} - v^{'}(\frac{y(\theta_L)}{\theta_H})\frac{1}{\theta_H}\Big]\frac{1}{\pi(\theta_L)\Psi^{'}(u(\theta_L))} \tag{**}$$

Note that since $y(\theta_H) \geq y(\theta_L)$, ** must be positive. Therefore, $0 \leq T^{'}(y(\theta_L)) \leq 1$

∘ <u>Summary</u>: optimal marginal taxes satisfy (A): $T^{*'}(y_h) - 0$ and (B): $0 \leq T^{'}(y(\theta_L)) \leq 1$. Overall, high incomes are untaxed at the margin but low incomes are levied a positive marginal tax (optimal income tax is therefore regressive)

* <u>Proof</u>: (A) follows from combining (2), (4), and (5). (B) follows from combining (1), (3), (5), and that $y(\theta_H) \geq y(\theta_L)$

* <u>Intuition</u>: there should be no taxation at the top since the distribution of skills is bounded - therefore, a marginal tax on the highest type would only decrease work incenstives without yielding additional revenue. The tax should be regressive to prevent the high type from deviating, so the planner distorts the low type at the margin

Mirrless Taxation Model with a Continuum of Types

<u>Setup</u>: involves a static economy with a single consumption good and a continuum of workers of a continuum of skill types. No ad-hoc structure on tax design ($T^*(y)$) is imposed and the only contraints on $T^*(y)$ arise endogenously due to informational frictions.

<u>Assumptions</u>: θ and n are private information, y is observable, aggregate output is linear in effective effort (implies that in a competitive equilibrium with a perfect labor market that y equals before tax income)

<u>Heterogeneity of Workers</u>: workers are heterogeneous across skills/labor productivity where $\theta \in \Theta$ and $F :$ $\Theta \to [0,1]$ denotes the cumulative distribution function of θ. Here $\Theta = [\theta_{LB}, \theta^{UB}] \to$ continuum of types, wherein θ_{UB} can be finite or infinite (unbounded distribution)

<u>Worker Preferences and Transformation</u>: workers have quasi-linear preferences over consumption c and effort n:

$$U(c,n) = c - v(n)$$

where $v', v'' > 0$. Workers transform effort into output y (or effective labour) using the technology function $y = \theta \cdot n$. Note that n is substituted by y/θ, allowing us to write the contract in terms of observables

<u>Key Tradeoff</u>: there is a tradeoff between equity and efficieny. This is as confiscatory taxation improves upon the distribution of the 'social pie' but also reduces production and thus the size of the pie

<u>Allocation</u>: an allocation is an assignment of c and y given θ:

$$\{c(\theta), y(\theta)\}_{\theta \in \Theta} \quad \Rightarrow \quad \text{the contract}$$

This allocation is feasible if and only if

$$\int c(\theta)dF(\theta) + G \leq \int y(\theta)dF(\theta)$$

where the LHS is aggregate consumption plus exogenous government expenditure, the RHS is aggregate output.

<u>Social Welfare Function</u>: the government aims to allocate c and y across θ to maximize the SWF:

$$SWF = \int \Psi\big(c(\theta) - v(\frac{y(\theta)}{\theta})\big)dF(\theta)$$

where $\Psi' > 0$ and $\Psi'' < 0$. The SWF encapsulates the normative criterion followed by the government and given the concavity of Ψ displays a preference for redistribution across θs

<u>Aim</u>: to design an income tax system that maximizes social welfare. More precisely, let $T(y)$ be the income tax schedule so that agents face the budget constraint $c \leq y - T(y)$. In this case, we are seeking a function

$T^*(y)$ that maximizes the SWF in a competitive equilibrium and focus on characterising optimal marginal tax rates $T^{*'}(y)$ across the income distribution

Solution for Optimal Taxes: we find $T^*(y)$ by following two steps

Step 1 - Contract Design: characterise the allocations for $\{c^*, y^*\}$ that maximise the SWF subject to informational Constraints

- Planning Problem: the social planner solves:

$$\max_{\{c(\theta), y(\theta)\}} \int \Psi\left(c(\theta) - v(\frac{y(\theta)}{\theta})\right)$$

$$\text{subject to} \int c(\theta) dF(\theta) \leq \int y(\theta) dF(\theta) \tag{RC}$$

$$c(\theta) - v(\frac{y(\theta)}{\theta}) \geq c(\theta') - v(\frac{y(\theta')}{\theta}), \quad \forall(\theta, \theta') \tag{IC}$$

- Double Infinite IC Constraints: As we have a continuum number of types we have an infinite number of θs, since for each θ we also have an infinite number of constraints we therefore have a double infinite number of IC constraints.

- IC Constraints Simplification: due to there being a double infinite number of ICs, they must be simplified to characterise the solution. The double infinite series of ICs is simplified by the proposition that (IC) holds if and only if for all θ:

$$c'(\theta) - v'(\frac{y(\theta)}{\theta})\frac{y'(\theta)}{\theta} = 0 \tag{IC_1}$$

$$\text{and}$$

$$y'(\theta) \geq 0 \tag{IC_2}$$

which are the simplified IC constraints

  ○ Proof: the idea of the proof is to show that under (IC_1) and (IC_2) we have that $\mathscr{U}(\theta'; \theta) = c(\theta') - v(\frac{y(\theta')}{\theta})$ attains a global maximum at $\theta' = \theta$ for all θ. Let

$$\mathscr{U}(\theta'; \theta) = c(\theta') - v(\frac{y(\theta)}{\theta}) \tag{1}$$

  be the utility of agent θ from reporting as θ'. This involves proving lemma 1 and then lemma 2

  ○ First Lemma: *a local maximum of $\mathscr{U}(\theta'; \theta)$ at $\theta' = \theta$ is attained if and only if $\forall \theta \in \Theta$ (IC_1) and (IC_2) hold*

    * Lemma 1 Proof: The first order conditions for local maximum are

$$\frac{\partial \mathscr{U}(\theta'; \theta)}{\partial \theta'}|_{\theta' = \theta} = 0$$

More succinctly, by $\mathscr{U}_1(\theta;\theta) = 0$ and recalling that $\mathscr{U}(\theta';\theta) = c(\theta') - v(\frac{y(\theta')}{\theta})$ we have

$$\mathscr{U}_1(\theta';\theta) = \underbrace{c'(\theta') - v'(\frac{y(\theta')}{\theta})\frac{y'(\theta')}{\theta} = 0}_{IC_1}$$

The second order conditions for a local maximum at $\theta' = \theta$ are:

$$\frac{\partial^2 \mathscr{U}(\theta';\theta)}{\partial \theta'^2}\Big|_{\theta'=\theta} \leq 0$$

Which can be written as

$$\mathscr{U}(\theta;\theta) \leq 01) \tag{()}$$

Recall that we showed at a local maximum that $\mathscr{U}_1(\theta;\theta) = 0$ and therefore $\mathscr{U}_{11}(\theta;\theta) \leq 0$. Differentiating this with respect to θ we have

$$\mathscr{U}_{11}(\theta;\theta) + \mathscr{U}_{12}(\theta;\theta) = 0 \tag{2}$$

This is as the total derivative is equal to the sum of partial derivatives and the total second order derivative equals 0 at a critical value (e.g. a local minima). By (1) we must then have that $\mathscr{U}_{12}(\theta;\theta) \geq 0$
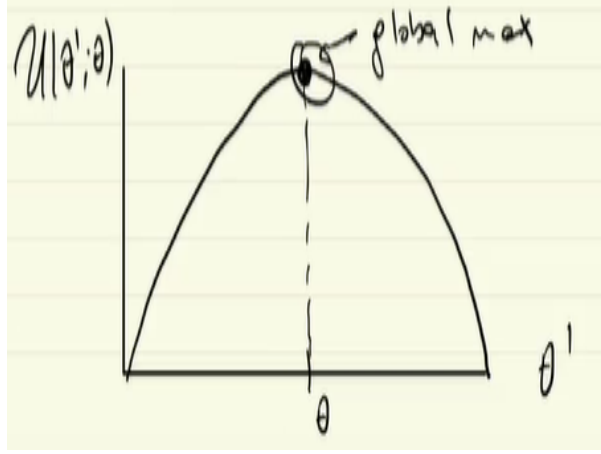
Expanding on the LHS of (2) and recalling that $\mathscr{U}_1(\theta';\theta) = c'(\theta') - v'(\frac{y(\theta')}{\theta'})\frac{y'(\theta')}{\theta} = 0$ we have

$$\underbrace{\mathscr{U}_{12}(\theta;\theta)}_{\geq 0 \text{ by } (2)} = v''(\frac{y(\theta)}{\theta})\frac{y(\theta)}{\theta^2}\frac{y'(\theta)}{\theta} + v'(\frac{y(\theta)}{\theta})\frac{y'(\theta)}{\theta^2}$$

$$= \underbrace{[v''(\frac{y(\theta)}{\theta})\frac{y(\theta)}{\theta^3} + v'(\frac{y(\theta)}{\theta})\frac{1}{\theta^2}]}_{\geq 0} y'(\theta)$$

This implies that $y'(\theta) \geq 0$ (IC_2)

○ <u>Second Lemma:</u> *a global maximum of $\mathscr{U}(\theta';\theta)$ at $\theta' = \theta$ is attained at $\theta' = \theta$ if and only if $\forall \theta \in \Theta$ (IC_1) and (IC_2) hold*

  ∗ <u>Lemma 2 Proof</u>: the idea is to show that the sign of $\mathscr{U}_1(\theta',\theta)$ is the same as $(\theta' - \theta)$

Here we show that the sign of $\mathscr{U}_1(\theta'; \theta)$ is the same as $(\theta - \theta')$ (which is greater or equal to zero). First note that

$$\mathscr{U}_1(\theta'; \theta) = c'(\theta') - v'(\frac{y(\theta')}{\theta'})\frac{y'(\theta')}{\theta} \tag{i}$$

By condition (IC_1) evaluated at θ' we have:

$$c'(\theta') = v'(\frac{y(\theta')}{\theta'})\frac{y'(\theta')}{\theta'} \tag{ii}$$

Combining (i) and (ii) yields:

$$\mathscr{U}_1(\theta'; \theta) = v'(\frac{y(\theta')}{\theta'})\frac{y'(\theta')}{\theta'} - v'(\frac{y(\theta')}{\theta})\frac{y'(\theta)}{\theta}$$
$$= \underbrace{[v'(\frac{y(\theta')}{\theta'})\frac{1}{\theta'} - v'(\frac{y(\theta')}{\theta})\frac{1}{\theta}]}_{\geq 0 \text{ by } (IC_2)} y'(\theta') \tag{[·]}$$

By (IC_2) and because $v'' \geq 0$ (i.e. since we have convexity in v), we get the following:

$$\text{sign}(\mathscr{U}(\theta', \theta)) = \text{sign}([·]) = \text{sign}(\theta - \theta')$$

This proposition has the advantage of reducing dimensionality (from double infinity to infinity constraints) since (IC_1) and (IC_2) are conditions for local maxima and therefore we actually work with a local version of (IC).

- Simplified IC Qualities: (IC_2) requires that higher types work more. (IC_1) requires that higher types get higher utility at the optimum.

- $\circ$ <u>Proof</u>: let $\mathscr{U}'(\theta) \equiv U(\theta;\theta)$ be the utility for type θ at the optimum. Note that

$$\mathscr{U}'(\theta) = \underbrace{\left[c'(\theta) - v'(\frac{y(\theta)}{\theta})\frac{y'(\theta)}{\theta}\right]}_{=0 \ by \ (IC_1)} + \underbrace{v'(\frac{y(\theta)}{\theta})\frac{y(\theta)}{\theta^2}}_{\geq 0}$$

  Hence, $\mathscr{U}'(\theta) \geq 0$. Note that the above is derived by solving the simplified IC_1 and substituting it into the utility formula

- $\circ$ <u>Intuition</u>: $\mathscr{U}(\lambda)$ should increase with λ to incentivise high types to work at full potential. More technically, we need that the informational rent (embedded in $\mathscr{U}(\lambda)$) increases with λ. These results extend what we obtained in the 2-types case

- <u>Final Planning Problem</u> let $\mathscr{U} = c(\theta) - v(\frac{y(\theta)}{\theta})$ be the utility of agent θ at the optimum. WE can therefore rewrite the planning problem using $\mathscr{U}(\theta)$ as a state variable where:
  - $\circ$ <u>Consumption</u>: can be written as $c(\theta) = \mathscr{U}(\theta) + v(\frac{y(\theta)}{\theta})$
  - $\circ$ <u>IC_1</u>: is equivalent to $\mathscr{U}'(\theta) = v'(\frac{y(\theta)}{\theta})\frac{y(\theta)}{\theta^2}$

  Therefore, the planning problem can be rewritten as:

$$\max_{\{\mathscr{U}(\theta),y(\theta)\}_{\theta \in \Theta}} \int \Psi(\mathscr{U}(\theta))f(\theta)d\theta$$

subject to

$$\int \left(\overbrace{\mathscr{U}(\theta) + v(\frac{y(\theta)}{\theta})}^{\equiv c(\theta)} - y(\theta)\right)f(\theta)d\theta = 0 \tag{RC}$$

$$\mathscr{U}'(\theta) = v'(\frac{y(\theta)}{\theta})\frac{y(\theta)}{\theta^2}, \quad \forall\theta \tag{IC_1}$$

$$y'(\theta) \geq 0, \quad \forall\theta \tag{IC_2}$$

- <u>Relaxing the Planning Problem</u>: by applying optimal control techniques we can handle the inequality contraint (IC_2) and integral constraint (RC)
  - $\circ$ <u>Inequality Constraint (IC_2)</u>: to make progress we neglect the monotonicity condition by assuming it holds at the optimum - thus working with a 'relaxed' planning problem by ignoring (IC_2)s existence. This is equivalent to assuming that there is no bunching at the optimum (i.e. different types are offered a different contract)

    This is the so called 'first-order approach' to the planning problem because (IC_1), the condition we keep, is a first order condition for truth telling to be optimal while (IC_2) came from second order conditions. In practice, we typically verify (IC_2) ex-post via simulations
  - $\circ$ <u>Integral Constraint (RC)</u>: for applying the Maximum Principle, the constraint set should only include differential equations, not integral constraints. To do this, we define a new state variable

$z(\theta)$ such that $z(\theta_{LB}) = z(\theta_{UB}) = 0$ and $z'(\theta) = \underbrace{\big(\mathscr{U}(\theta) + v(\frac{y(\theta)}{\theta}) - y(\theta)\big)f(\theta)}_{\text{integrand of RC}}$. Clearly these

conditions are equivalent to the RC, therefore:

$$0 = z(\theta_{UB}) - z(\theta_{LB}) = \int_{\theta_{LB}}^{\theta_{UB}} z'(\theta)d\theta = \int_{\theta_{LB}}^{\theta_{UB}} \big(\mathscr{U}(\theta) + v\frac{y(\theta)}{\theta} - y(\theta)\big)f(\theta)d\theta$$

◦ <u>Relaxed Problem</u>: by handling (RC) and (IC_2), the relaxed problem becomes:

$$\max_{\{\mathscr{U}(\theta),y(\theta)\}_{\theta\in\Theta}} \int \Psi(\mathscr{U}(\theta))f(\theta)d\theta$$

subject to

$$z'(\theta) = \big(\mathscr{U}(\theta) + v(\frac{y(\theta)}{\theta}) - y(\theta)\big)f(\theta), \ \forall\theta \qquad\qquad (\lambda, \ RC)$$

$$\mathscr{U}'(\theta) = v'(\frac{y(\theta)}{\theta})\frac{y(\theta)}{\theta^2}, \ \forall\theta \qquad\qquad (\mu, \ IC_1)$$

$$z(\theta_{LB}) = z(\theta_{UB}) = 0$$

• Solving the Relaxed Problem: this can be done via applying the Maximum Principle with $y(\theta)$ as the control variable and $z(\theta), \mathscr{U}(\theta)$ as the state Variables. The Hamiltonian is:

$$H = \Psi(\mathscr{U}(\theta))f(\theta) + \lambda(\theta)\big[\mathscr{U}(\theta) + v(\frac{y(\theta)}{\theta}) - y(\theta)\big]f(\theta) + \mu(\theta)v'(\frac{y(\theta)}{\theta})\frac{y(\theta)}{\theta^2}$$

where $\lambda(\theta)$ and $\mu(\theta)$ are the costates associated with $z(\theta)$ and $\mathscr{U}(\theta)$. According to the Maximum Principle, the necessary consitions for an optimum are (dropping explicit dependencies on θ):

$$\frac{\partial H}{\partial y} = 0 \ \Rightarrow \ H_y = \lambda\big[v'(\frac{y}{\theta})\frac{1}{\theta} - 1\big]f + \mu\big[v''(\frac{y}{\theta})\frac{1}{\theta^3} + v'(\frac{y}{\theta})\frac{1}{\theta^2}\big] = 0$$

$$\frac{\partial H}{\partial z} = -\lambda' \Rightarrow \lambda' = 0$$

$$\frac{\partial H}{\partial\mathscr{U}} = -\mu' \Rightarrow \mu' = -\big[\Psi'(\mathscr{U}) + \lambda\big]f$$

Boundary conditions are (recall that $\mathscr{U}$ is free at the boundaries):

$$\mu(\theta_{LB}) = \mu(\theta_{UB}) = 0$$

◦ <u>Disutility of Effort</u>: the disutility for effort is isoelastic

$$v(n) = \frac{n^{1+\frac{1}{\varepsilon}}}{1+\frac{1}{\varepsilon}}$$

where $\varepsilon > 0$ is the Frisch elasticity of labour supply and measures the elasticity of effort to a

change in wages, holding constant the marginal utility of wealth

○ Average Marginal SW: let

$$D(\theta) \equiv \frac{1}{1 - F(\theta)} \int_{\theta_{LB}}^{\theta_{UB}} \Psi'(\mathscr{U}(\widetilde{\theta})) f(\widetilde{\theta}) d\widetilde{\theta}$$

be the average marginal social welfare weight on individuals with $\widetilde{\theta} \geq \theta$

* $D(\theta)$: summarises the redistributive tastes of the planner where $D(\theta) \geq 0$ and is decreasing in θ by the concavity of Ψ

* Note: $F(\theta)$ is the proportion of agents with skill type $\theta \leq \widetilde{\theta}$ and will increase as $\widetilde{\theta}$ increases. This implies a higher social weight on individuals with higher $\widetilde{\theta}$ and therefore that they should be taxed less (all else equal). $(1 - F(\theta)) = \int_{\theta_{UB} - \theta_{LB}} f(\theta) d\theta$

• Optimal Marginal Taxes: the optimal marginal taxes satisfy

$$\frac{T^{*'}(y^*(\theta))}{1 - T^{*'}(y^*(\theta))} = (1 + \frac{1}{\varepsilon})(1 - \frac{D(\theta)}{D(\theta_{LB})})\frac{1 - F(\theta)}{\theta f(\theta)}$$

This has three key determinants: (1) elasticity of labour supply, (2) redistributive tastes, (3) distribution of skills. Recall that $y' \geq 0$.

○ Proof Step 1 - Centralization: characterise the allocations for $\{c^*, y^*\}$ that maximise the SWF subject to informational constraints using the planning problem.

By the maximum principal we have:

$$\lambda[v'(\frac{y}{\theta})\frac{1}{\theta} - 1]f + \mu[v''(\frac{y}{\theta})\frac{y}{\theta^3} + v'(\frac{y}{\theta})\frac{1}{\theta^2}] = 0 \tag{1}$$

$$\lambda' = 0 \tag{2}$$

$$\mu' = -[\Psi'(\mathscr{U}) + \lambda]f \tag{3}$$

Recall that $\mathscr{U}$ is free at the boundaries so $\mu(\theta_{LB}) = \mu(\theta_{UB}) = 0$. Therefore, by integrating (3) over all θ we have:

$$\underbrace{\int_{\theta_{LB}}^{\theta_{UB}} \mu'(\widetilde{\theta}) d\widetilde{\theta}}_{\mu(\theta_{UB}) - \mu(\theta_{LB}) = 0} = -\int_{\theta_{LB}}^{\theta_{UB}} [\Psi'(\mathscr{U}(\widetilde{\theta})) + \lambda] f(\widetilde{\theta}) d\widetilde{\theta} \tag{4}$$

Based on (4), we have that:

$$\lambda' = -\int_{\theta_{LB}}^{\theta_{UB}} \Psi'(\mathscr{U}(\widetilde{\theta})) f(\widetilde{\theta}) d\widetilde{\theta}$$

$$\lambda' = -D(\theta_{LB}) \tag{5}$$

Since when $\widetilde{\theta} = \theta_{LB}$ we have that $(1 - F(\theta) = 1$, yielding the above result. Also, by integrating (3) between $[\theta, \theta_{UB}]$ we get:

$$\underbrace{\mu(\theta_{UB})}_{=0} - \mu(\theta) = -\int_{\theta}^{\theta_{UB}} [\Psi'((\widetilde{\theta})) + \lambda] f(\widetilde{\theta}) d\widetilde{\theta}$$

$$\mu(\theta) = \underbrace{\int_{\theta}^{\theta_{UB}} [\Psi'((\widetilde{\theta}))] f(\widetilde{\theta}) d\widetilde{\theta}}_{(1-F(\theta))D(\theta)} + \underbrace{\lambda}_{-D(\theta_{LB})} \underbrace{\int_{\theta}^{\theta_{UB}} f(\widetilde{\theta}) d\widetilde{\theta}}_{1-F(\theta)}$$

Therefore, we have that:

$$\mu(\theta) = (1 - F(\theta))(D(\theta) - D(\theta_{LB})) \tag{6}$$

Note that $\mu(\theta) \leq 0$ since $(D(\theta) - D(\theta_{LB})) \leq 0$ as $D(\theta)$ is decreasing in θ (based on the planner's taste for redistribution). Using the functional form for $v(\cdot)$ we have:

$$v(n) = \frac{n^{1+\frac{1}{\varepsilon}}}{1 + \frac{1}{\varepsilon}}$$

$$\therefore v'(n) = n^{\frac{1}{\varepsilon}} \tag{7}$$

$$\therefore v''(n) = \frac{1}{\varepsilon} n^{\frac{1}{\varepsilon}-1} \tag{7}$$

Using (5), (6), and (7) into (1), we get:

$$\underbrace{D(\theta_{LB})[(\frac{y}{\theta})^{\frac{1}{\varepsilon}} \frac{1}{\theta} - 1] f(\theta)}_{\lambda} = \underbrace{(1 - F(\theta))(D(\theta) - D(\theta_{LB}))}_{\mu(\theta)} [\frac{1}{\varepsilon}(\frac{y}{\theta})^{\frac{1}{\varepsilon}-1} \frac{y}{\theta^3} + (\frac{y}{\theta})^{\frac{1}{\varepsilon}} \frac{1}{\theta^2}]$$

$$\underbrace{(\frac{y}{\theta})^{\frac{1}{\varepsilon}} \frac{1}{\theta} - 1}_{-T'(y)} = \frac{1 - F(\theta)}{\theta f(\theta)} (\frac{D(\theta)}{D(\theta_{LB})} - 1)(\frac{1}{\varepsilon} + 1) \underbrace{(\frac{y}{\theta})^{\frac{1}{\varepsilon}} \frac{1}{\theta}}_{1-T'(y)} \tag{8}$$

(8) can be used in conjunction with the decentralization step to solve for the optimal tax formula

○ <u>Proof Step 2 - Decentralisation:</u> back out the $T^*(y)$ that implement $\{c^*, y^*\}$ as a competitive equilibrium using the individual problem.

The individual solves the following problem in the decentralization

$$\max_{y} \underbrace{y - T(y)}_{c} - \underbrace{(\frac{y}{\theta})^{1+\frac{1}{\varepsilon}} \frac{1}{1 + \frac{1}{\varepsilon}}}_{v(\frac{y}{\theta})}$$

The first order conditions to this problem are:

$$1 - T'(y) = (\frac{y}{\theta})^{\frac{1}{\varepsilon}} \frac{1}{\theta} \tag{9}$$

By combining (8) and (9) we derive the optimal tax formula

$$\frac{T^{*'}(\theta)}{1 - T^{*'}(\theta)} = (1 + \frac{1}{\varepsilon})(1 - \frac{D(\theta)}{D(\theta_{LB})})\frac{1 - F(\theta)}{\theta f(\theta)}$$

<u>Characteristics of $T^{*'}$</u>: (1) $T^{*'}(\theta)$ decreases with ε since large ε means that marginal taxes create large distortions, (2) $T^{*'}$ decreases with $D(\theta)$ since more social weight on individuals with $\widetilde{\theta} \geq \theta$ implies that the government taxes such individuals less, (3) $T^{*'}$ increases with $\frac{1 - F(\theta)}{\theta f(\theta)}$

- <u>Logic for (3)</u>: suppose you increase $T^{*'}$. Then, all else equal, the government collects more tax revenue on individuals above θ, who are $(1 - F(\theta))$ in number. This explains the positive dependence on $(1 - F(\theta))$. However, the resulting distortion is proportional to the mass of individuals at θ and to their productivity level θ. This explains the negative dependence on $\theta f(\theta)$

<u>Quantifying Optimal Taxes</u>: optimal marginal taxes satisfy $\frac{T^{*'}(\theta)}{1 - T^{*'}(\theta)} = (1 + \frac{1}{\varepsilon})(1 - \frac{D(\theta)}{D(\theta_{LB})})\frac{1 - F(\theta)}{\theta f(\theta)}$ with its key determinants being (1) elasticity of labour supply, (2) redistributive tastes, (3) distribution of skills

- <u>Elasticity of Labour Supply (ε)</u>: ε can be estimated from the data which suggests that the reasonable range among US males is $\varepsilon \in [0, 0.5]$ and for US females $\varepsilon \geq 0.5$
- <u>Redistributive Tastes $(D(\theta))$</u>: the shape of $D(\theta)$ is driven by the parameterisation of Ψ, typically the 'weighted utilitarian' SWF is used:

$$\Psi(\mathscr{U}(\theta)) = a(\theta)\mathscr{U}(\theta) \Rightarrow D(\theta) = \frac{1}{1 - F(\theta)} \int_{\theta}^{\theta_{UB}} a(\widetilde{\theta}) f(\widetilde{\theta}) d\widehat{\theta}$$

  where $a(\theta)$ decreases with θ. The Rawlsian $a(\theta_{LB}) = +\infty$ and Utilitarian $a(\theta) = 1$ for all θ
- <u>Skill Distribution $f(\theta)$</u>: given the issue of the distribution of skills being unobservable, there are two common approaches - (1) the Saez approach backs out distribution of skills from distribution of income, (2) the Mankiw approach proxies skills with wages
    - <u>Distribution of Skills</u>: if the distribution of skills is bounded then $\frac{1 - F(\theta_{UB})}{\theta_{UB} f(\theta_{UB})} = 0$ so that $T^{*'}(\theta_{UB}) = 0$ as in the two type model. However, if the distribution of skills is unbounded then this is not the case

<u>Saez Approach to Skill Distribution Analysis</u>: suppose that given the actual tax schedule $T(y)$ we observe an income distribution with cdf $Q(y)$. As long as $y'(\theta) \geq 0$, we can back out the skill distribution based on the relationship $F(\theta) = Q(y(\theta))$. This implies that $f(\theta) = q(y(\theta))y'(\theta)$, where $y'(\theta)$ on the RHS can be recovered from the utility maximization problem given the actual tax code

- Optimization Problem: given the actual tax code T, the agent solves

$$\max_{c,y} c - v(\frac{y}{\theta})$$

$$\text{Subject to: } c = y - T(y)$$

The first order conditions of this problem imply that

$$1 - T'(y) - v'(\frac{y}{\theta})\frac{1}{\theta} = 0$$

This implicity defines $y(\theta)$ and $y'(\theta)$. Therefore, we can identify the skill distribution from the distribution of observed incomes and the observed tax schedule

- Skill Distribution at the Top: both the Saez and Mankiw approach for estimating $f(\theta)$ sets $\theta_{UB} = \infty$ and usually assumes that skills follow a pareto distribution at the top (e.g. above the 95th percentile). Specifically, above some productivity level θ_p we have:

$$f(\theta) = \frac{\alpha(\theta_p)^\alpha}{\theta^{1+\alpha}} \text{ and } F(\theta) = 1 - (\frac{\theta_p}{\theta})^\alpha$$

where α measures the 'thinness' of the upper tail of F (i.e. low α implies a thick upper tail, therefore having distribution of top skills more spread out and more inequality within the top quantiles of skills). The below shows the pareto pdf for $\theta_p = 1$ (x is θ):
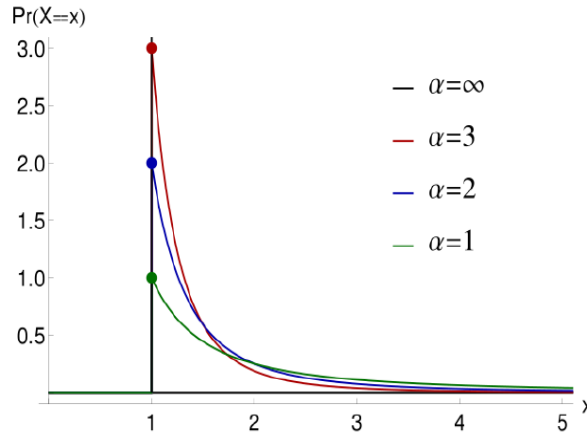
The reason for fitting a pareto distribution for skills at the top is because income distribution is well approximated with a pareto distribution. This suggests that the underlying skill distribution should have a "right pareto tail". If income is pareto distributed with tail index α_y, then

$$\lim_{y\to\infty} \frac{1 - Q(y)}{yq(y)} = \frac{1}{\alpha_y} \to \text{ constant}$$

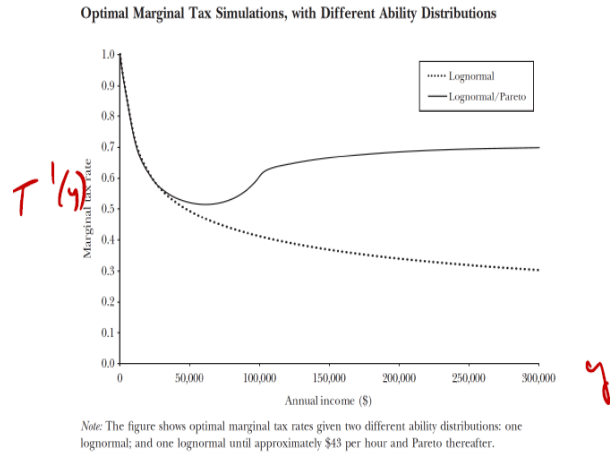In the data $\frac{1-Q(y)}{yq(y)}$ is extremely stable at the top as seen below
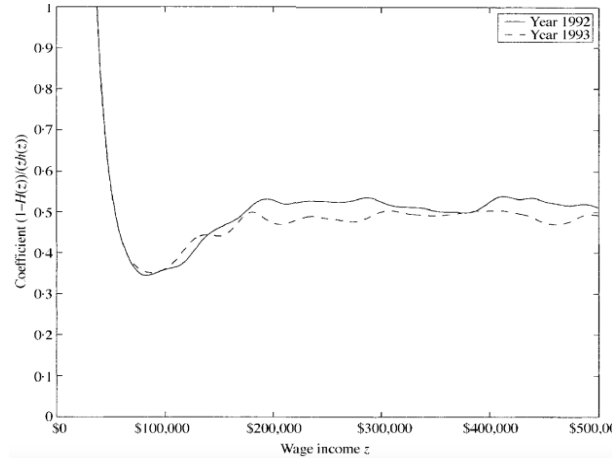
In summary, if the skill distribution is also pareto at the top then

$$\lim_{\theta \to \infty} \frac{1 - F(\theta)}{\theta f(\theta)} = \frac{1}{\alpha}$$

where α is backed out from α_y based on $a = (1 + \varepsilon)\alpha_y$ (this can be proven by manipulating the FOC of the agent's problem in decentralisation facing the linear top tax rate $\overline{\tau}$)

<u>Simulation of Optimal Taxes</u>: from simulating the model we can derive optimal taxes as being

Note: The figure shows optimal marginal tax rates given two different ability distributions: one lognormal; and one lognormal until approximately $43 per hour and Pareto thereafter.

- <u>Bottom and Top Distribution Results</u>: we have that the optimal T' at the bottom is very high, which deters high incomes from not working at full potential (counterpart of the $T'(\theta_L) > 0$ result in the two-type model). Moreover, $\frac{1-F(\theta)}{\theta f(\theta)}$ is very high at the bottom since (1) there are too many individual above the bottom making the numerator big, (2) the denominator being small because of few individual around the bottom and low θ. Thus distortions induced by high T' are not too large

- <u>Distribution Shape Results</u>: the optimal T' is U-shaped, mirroring the behaviour of ratio $\frac{1-F(\theta)}{\theta f(\theta)}$. T'

56

only becomes progressive at high incomes

- Qualitative Results: top tax rates can be quite large for high incomes. The 'zero taxation at the top' (see two type model) only applies to the very top income earner. With an unbounded skill distribution the level of top earnings is not known in advance. Hence, the 'zero taxation at the top' (and the implication that T' decreases at the top) has little empirical and policy relevance

Top Tax Formula: with our optimal tax formula $\frac{T^{*'}(\theta)}{1-T^{*'}(\theta)} = (1 + \frac{1}{\varepsilon})(1 - \frac{D(\theta)}{D(\theta_{LB})})\frac{1-F(\theta)}{\theta f(\theta)}$, as $\theta \to \infty$ we have that $D(\theta) \to 0$ and $\frac{1-F(\theta)}{\theta f(\theta)} \to \frac{1}{\alpha}$. Therefore, the optimal asymptotic top tax rate, known as the Diamond-Saez Top Tax Formula, is:

$$\overline{T}^{*'} = \frac{1}{1 + \alpha\frac{\varepsilon}{1+\varepsilon}}$$

- Implications: with low ε we have low distortions as $\overline{T}^{*'}$ increases. With low α we have thick right tails of skill distribution and therefore if $\overline{T}^{*'}$ increases few will be affected at the margin but many inframarginally
- Data: $\overline{T}^{*'}$ is a key policy instrument to achieve redistribution and accounts for a significant fraction of individual tax revenue (about 40 % in the US). Using a back-of-the-envelope calibration we have $\varepsilon = 0.25$ as a midrange estimate, $\alpha_y = 1.5 \to \alpha = 1.88$. This implies that $\overline{T}^{*'} = 73\%73$.

  Using a sensitivity analysis we have that:

	$\varepsilon = 0.1$	$\varepsilon = 0.25$	$\varepsilon = 0.5$
α	1.65	1.88	2.25
$\overline{T}^{*'}$	83%	73%	57%

which compares to the following statutory rates across time:
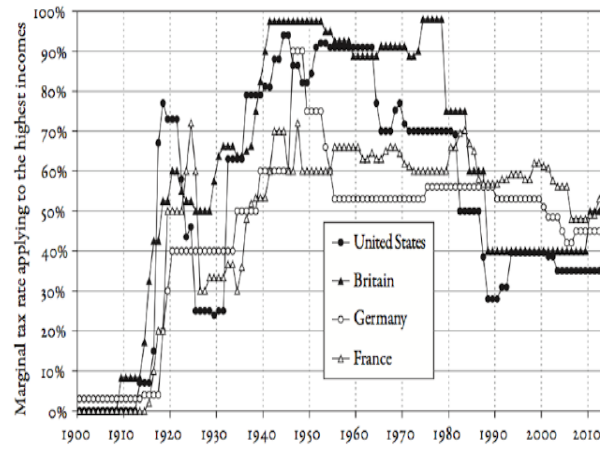
Optimal Dynamic Taxation

<u>Dynamic Mirrlees Approach</u>: extends the traditional mirrlees model to a dynamic one involving information friction that endogenise feasible tax instruments, a heterogeneous agents framework, and taxes on capital income <u>Baseline Model</u>: involves a discrete time economy lasting for $T \leq \infty$ periods and an economy populated by a continuum of ex-ante identical individuals

- <u>Expected Lifetime Utility</u>: the expected lifetime utility of individuals is

$$U_0 \equiv \mathbb{E}_0 \sum_{t-1}^{T} \beta^{t-1}(u(c_t) - v(n_t))$$

where $\beta \in (0,1)$, $c_t \geq 0$ is consumption, $n_t \geq 0$ is effort, v is strictly increasing and convex, u is strictly increasing and concave (we drop the quasilinearity assumption)

- <u>Skill</u>: individuals are heterogeneous with respect to skills/labour productivity where the skill of an agent at time t is given by $\theta_t \in \Theta \in \mathbb{R}_+$.

  Skill shocks are stochastic where $\theta^t = (\theta_1, \ldots, \theta_t) \in \Theta^t$ is the history of shocks up to t, $\pi_t(\theta^t)$ is the probability of history θ^t, and individuals learn their θ_t at the beginning of t with future realizations unknown.

  the set of skill shocks is discrete where

$$\Theta = \{\bar{\theta}_1, \bar{\theta}_2, \ldots, \bar{\theta}_N\}$$

  with $\bar{\theta}_1 > \bar{\theta}_2 > \cdots > \bar{\theta}_N$.

  Skill shocks are iid such that:

$$\pi_t(\theta_t | \theta^{t-1}) = \pi(\theta_t)$$

  , therefore previous history of skill shocks is not going to impact the probability of experience a skill shock of size θ_t and without loss of generality we assume $\mathbb{E}[\theta] = \sum_{\theta \in \Theta} \pi(\theta)\theta = 1$

- <u>Agent Output and Borrowing</u> Agents produce output at period t according to

$$y_t = \theta_t \cdot n_t$$

  where θ_t and n_t are private informations. Agents can borrow and lend at an exogenous gross R

- <u>Firm</u>: a representative firm produces output by operating a linear technology

$$F(Y_t, K_t) = Y_t + RK_t$$

  where $Y_t \equiv \sum_{\theta^t} \pi(\theta^t) \cdot y(\theta^t)$ is aggregate effective effort and K_t is aggregate savings

- <u>Aloocation</u>: an allocation $\{c, y\}$ is a sequence of $\{c_t, y_t\}_{t=0}^T$ with $(c_t, y_t) : \Theta^t \to \mathbb{R}_+^2$. An allocation $\{c, y\}$ is feasible if and only if

$$\underbrace{\sum_{t=1}^T \sum_{\theta^t \in \Theta^T} R^{1-t} \pi(\theta^t) c_t(\theta^t)}_{PV\,Consumption} \leq \underbrace{\sum_{t=1}^T \sum_{\theta^t \in \Theta^t} R^{1-t} \pi(\theta^t) y_t(\theta^t)}_{PV\,Output}$$

- <u>Social Welfare</u>: assume that a social planner evaluates $\{c, y\}$ according to a utilitarian social welfare function that maximizes the expected utility of the agents

$$SWF(c, y) = \sum_{t=1}^T \sum_{\theta^t \in \Theta^t} \beta^{t-1} \pi(\theta^t) \big(u(c_t(\theta^t)) - v(n_t(\theta^t)) \big)$$

i.e. $SWF(c, y) = U_0(c, y)$. Here the motive for redistribution from high to low types is captured by the concavity of u and due to no initial heterogeneity the planner's goal is to provide insurance against θ_t shocks

- <u>Incentive Compatibility</u>: we define a reporting strategy by $\sigma_t : \Theta^t \to \Theta$ and a history of reporting strategies by $\sigma^t = (\sigma_1, \ldots, \sigma_t)$. Given this, we have the following IC Contraints:

$$\sum_{t=1}^T \sum_{\theta^t \in \Theta} \beta^{t-1} \pi(\theta^t) \left(u(c_t(\theta^t)) - v\big(\frac{y_t(\theta^t)}{\theta_t}\big) \right) \geq \sum_{t=1}^T \sum_{\theta^t \in \Theta} \beta^{t-1} \pi(\theta^t) \left(u(c_t(\sigma^t(\theta^t))) - v\big(\frac{y_t(\sigma^t(\theta^t))}{\theta_t}\big) \right)$$

for all σ^T

- <u>Optimal Tax Design</u>: with the goal of the government designing a tax transfer system to provide insurance against θ_t shock, let $\mathcal{T}_t(\cdot)$ be a tax system so that agents' budget constraints are

$$c_t + k_{t+1} \leq y_t + R k_t - \mathcal{T}_t(\cdot)$$

We are seeking a function $\mathcal{T}_t^*(\cdot)$ that maximises ex-ante welfare U_0 in a competitive equilibrium. To find $\mathcal{T}_t^*(\cdot)$ we follow two steps

  ○ <u>Step 1</u>: find the allocations $\{c^*, y^*\}$ that maximize U_0 subject to informational frictions (applying dynamic contracts tools)

  ○ <u>Step 2</u>: back out a $\mathcal{T}_t^*(\cdot)$ that implements $\{c^*, y^*\}$ as a competitive equilibrium

One key difference with the static model is that $\mathcal{T}_t^*(\cdot)$ is generally not unique. For this reason, we focus on characterising the wedges (distortions) derived from the planning problem

- <u>Planning Problem</u>: the constrained optimal allocations $\{c^*, y^*\}$ solve

$$\max_{\{c,y\}} \mathbb{E}_0\left[\sum_{t=1}^{T} \beta^{t-1}\big(u(c_t(\theta^t)) - v(\tfrac{y_t(\theta^t)}{\theta^t})\big)\right]$$

$$\text{subject to} \quad \mathbb{E}_0\left[\sum_{t=1}^{T} R^{1-t}\big(c_t(\theta^t) - y_t(\theta^t)\big)\right] \leq 0 \qquad (RC)$$

$$\sum_{t=1}^{T}\sum_{\theta^t \in \Theta} \beta^{t-1}\pi(\theta^t)\Big(u(c_t(\theta^t)) - v(\tfrac{y_t(\theta^t)}{\theta_t})\Big) \geq \sum_{t=1}^{T}\sum_{\theta^t \in \Theta} \beta^{t-1}\pi(\theta^t)\Big(u(c_t(\sigma^t(\theta^t))) - v(\tfrac{y_t(\sigma^t(\theta^t))}{\theta_t})\Big) \ \forall \ \sigma^T IC)$$

$$(())$$

<u>Two Period Model</u>: focusing on a two period model $(T = 2)$ with constant type across time $(\theta_1 = \theta_2)$ and two possible types $\Theta = \{\theta_L, \theta_H\}$. We have ex-ante utility:

$$U_0 = \mathbb{E}_0\left[u(c_1) - v(\tfrac{y_1}{\theta}) + \beta\big(u(c_2) - v(\tfrac{y_2}{\theta})\big)\right]$$

We have the following resource constraints:

$$\sum_{\theta} \pi(\theta)[c_1(\theta) - y_1(\theta)] + K_2 = 0 \qquad (1)$$

$$\sum_{\theta} \pi(\theta)[c_2(\theta) - y_2(\theta)] = RK_2 \qquad (2)$$

Combining (1) and (2) yields:

$$\sum_{\theta} \pi(\theta)\left[c_1(\theta) - y_1(\theta) + \frac{1}{R}(c_2(\theta) - y_2(\theta))\right]$$

Overall, we have the following planning problem:

$$\max_{\{c,y\}} \sum_{\theta} \pi(\theta)\left[u(c_1) - v(\tfrac{y_1}{\theta}) + \beta\big(u(c_2) - v(\tfrac{y_2}{\theta})\big)\right]$$

$$\text{subject to}$$

$$\sum_{\theta} \pi(\theta)\left[c_1(\theta) - y_1(\theta) + \frac{1}{R}(c_2(\theta) - y_2(\theta))\right]$$

$$u(c_1(\theta)) - v(\tfrac{y_1(\theta)}{\theta}) + \beta\big(u(c_2(\theta)) - v(\tfrac{y_2(\theta)}{\theta})\big) \geq u(c_1(\theta')) - v(\tfrac{y_1(\theta')}{\theta}) + \beta\big(u(c_2(\theta')) - v(\tfrac{y_2(\theta')}{\theta})\big) \ \forall \ \theta, \theta'$$

With (IC_H) binding at the optimum we transform the planning problem to:

$$\max_{\{c,y\}} \sum_{\theta} \pi(\theta)\left[u(c_1) - v(\frac{y_1}{\theta}) + \beta\left(u(c_2) - v(\frac{y_2}{\theta})\right)\right]$$

subject to

$$\sum_{\theta} \pi(\theta)\left[c_1(\theta) - y_1(\theta) + \frac{1}{R}(c_2(\theta) - y_2(\theta))\right]$$

$$u(c_1(\theta_H)) - v(\frac{y_1(\theta_H)}{\theta_H}) + \beta\left(u(c_2(\theta_H)) - v(\frac{y_2(\theta_H)}{\theta_H})\right) \geq u(c_1(\theta_L)) - v(\frac{y_1(\theta_L)}{\theta_H}) + \beta\left(u(c_2(\theta_L)) - v(\frac{y_2(\theta_L)}{\theta_H})\right)$$

$$(IC_H)$$

- <u>First Best Allocation</u>: we have two key conditions for optimality
  - <u>Tntratemporal Condition</u>: the intratemporal efficiency condition must hold where $MRS_{l,c} = MRT_{l,c}$ such that

$$\frac{\frac{v'(y_t^{FB})}{\theta_t}}{u'(c_t^{FB})} = \theta_t$$

  - * <u>Logic</u>: if intratemporal efficiency does not hold then I can rearrange (c,l) and increase U_0
  - * <u>Alternative Form</u>: this can be alternatively written as

$$\underbrace{v'(l_t^{FB})}_{\text{Marginal Cost of Y Increase}} = \underbrace{\theta_t u'(c_t^{FB})}_{\text{Marginal Benefit of Y Increase}}$$

  - <u>Intertemporal Condition</u>: the intertemporal efficiency condition must hold where $MRS_{c_1,c_2} = MRT_{c_1,c_2}$ such that

$$\frac{u'(c_1^{FB})}{\beta u'(c_2^{FB})} = R$$

  - * <u>Alternative Form</u>: this can be alternatively written as

$$\underbrace{u'(c_1^{FB})}_{\text{Marginal Cost of } c_2 \text{ Increase}} = \underbrace{R\beta u'(c_2^{FB})}_{\text{Marginal Benefit of } c_2 \text{ Increase}}$$

The constrained optimal allocation also features two types of wedges/distortions. These can be interpreted as implicit marginal taxes and will shape $\mathscr{T}_t^*(\cdot)$ in the decentralization

  - <u>Labor Wedge at t</u>: this involves a distortion between MRS_{l_t,c_t} and MRT_{l_t,c_t} due to the implicit marginal tax on labor income at t. Note that

$$1 - \tau_t^y(\theta) \equiv \frac{v'(y_t^*(\theta)/\theta)/theta}{u'(c_t^*(\theta))}$$

We have for a two-period dynamic model with $|\Theta| = 2$ and constant types (i.e. $\theta_t = \theta$ for $t = 1, 2$) that for all t: $\tau_t^y(\theta_L) > 0$ and $\tau_t^y(\theta_H) = 0$

○ Savings Wedge at $t = 1$: this involves a distortion between MRS_{c_1,c_2} and MRT_{c_1,c_2} due to the implicit marginal tax on capital income. Note that

$$1 - \tau^s(\theta) \equiv \frac{u'(c_1^*(\theta))}{\beta R u'(c_2^*(\theta))}$$

We have for a two-period model with $|\Theta| = 2$ and constant types (i.e. $\theta_t = \theta$ for $t = 1, 2$) that $\tau^s(\theta_L) = \tau^s(\theta_H) = 0$

● Final Planner Problem: based on the wedges, when types are constant it is optimal not to distort savings decisions - a key assumption of this model. Thus our planning problem becomes:

$$\max_{\{c,y\}} \sum_{\theta_1,\theta_2} \pi(\theta_1)\pi(\theta_2)\left[u(c_1(\theta_1)) - v(\frac{y_1(\theta_1)}{\theta_1}) + \beta\left(u(c_2(\theta_2)) - v(\frac{y_2(\theta^2)}{\theta^2})\right)\right]$$

subject to

$$\sum_{\theta_1,\theta_2} \pi(\theta_1)\pi(\theta_2)\left[c_1(\theta_1) - y_1(\theta_1) + \frac{1}{R}(c_2(\theta^2) - y_2(\theta^2))\right] = 0$$

$$u(c_1(\theta_1)) - v(\frac{y_1(\theta_1)}{\theta_1}) + \beta\sum_{\theta_2} \pi(\theta_2)\left(u(c_2(\theta^2)) - v(\frac{y_2(\theta^2)}{\theta^2})\right) \geq u(c_1(\sigma_1(\theta_1))) - v(\frac{y_1(\sigma_1(\theta_1))}{\theta_1}) + \beta\sum_{\theta_2} \pi(\theta_2)\left(u(c_2(\sigma^2(\theta^2)))\right.$$

for all $(\theta_1, \sigma_1, \theta^2, \sigma^2)$.

○ Saving Wedge: the savings wedge with stochastic types is:

$$1 - \tau^s(\theta_1) \equiv \frac{u'(c_1^*(\theta_1))}{\beta R \mathbb{E}_1[u'(c_2^*(\theta_1,\theta_2))]}$$

○ Inverse Euler Equation: consider the two-period model with stochastic types. The constrained efficient allocation satisfies

$$\frac{1}{u'(c_1^*(\theta_1))} = \frac{1}{\beta R}\sum_{\theta_2}$$

This is essentially a euler equation for the reciprocal of the marginal utilities.

＊ Proof - Variational Approach with Utility Space: consider the following variation of the optimal solution: fix θ and let

$$u_1^\varepsilon(\theta_1) = u_1^*(\theta_1) - \varepsilon, \quad \varepsilon > 0$$
$$u_2^\varepsilon(\theta_1, \theta_2) = u_2^*(\theta_1, \theta_2) + \frac{\varepsilon}{\beta}, \quad \forall \theta_2$$

where $u_1^*(\theta_1)$ is consumption utility at optimum at $t = 1$ and $u_2^*(\theta_1, \theta_2)$ is consumption utility

at optimum at $t = 2$. It is easy to show that the perturbed allocation is such that (1) the objective function doesn't change and (2) the ICs are satisfied. Therefore, a necessary condition for the original allocation to be optimal is that resources spent in implementing the allocation are minimized at $\varepsilon = 0$. Let the resource cost of delivering the new allocation be:

$$\mathcal{L}^\varepsilon(\theta_1) = \underbrace{u^{-1}(u_1^\varepsilon(\theta_1))}_{c_1(\theta_1)} + \frac{1}{R}\sum_{\theta_2}\pi(\theta_2)\underbrace{u^{-1}(u_2^\varepsilon(\theta_1,\theta_2))}_{c_2^\varepsilon(\theta_1,\theta_2)}$$

If $u_1^*(\theta_1)$ and $u_2^*(\theta_1,\theta_2)$ are optimal, we must have that:

$$\frac{\partial \mathcal{L}^\varepsilon(\theta_1)}{\partial \varepsilon}\Big|_{\varepsilon=0} = 0 \tag{*}$$

Note that:

$$\mathcal{L}^\varepsilon(\theta_1) = u^{-1}(u_1^*(\theta_1) - \varepsilon) + \frac{1}{R}\sum_{\theta_2}\pi(\theta_2)u^{-1}(u_2^*(\theta_1,\theta_2) + \tfrac{\varepsilon}{\beta})$$

Then condition (*) boils down to:

$$\frac{\partial \mathcal{L}^\varepsilon(\theta_1)}{\partial \varepsilon}\Big|_{\varepsilon=0} = 0 \Leftrightarrow \frac{1}{u'(\underbrace{underbrace u^{-1}(u_1^*(\theta_1))}_{c_1^*(\theta_1)})} = \overbrace{\frac{1}{R\beta}\sum_{\theta_2}\pi(\theta_2)\frac{1}{u'(\underbrace{u^{-1}(u_2^*(\theta_1,\theta_2))}_{c_2^*(\theta_1,\theta_2)})}}^{\text{IEE}}$$

Use that: $[F^{-1}]'(x) = \frac{1}{f'(f^{-1}(x))}$

○ Savings Taxes: for the two period model with stochastic types, suppose that for some $\theta_1, c_2^*(\theta_1,\theta_2)$ is not independent of θ_2. Then we have that

$$\tau^s(\theta_1) > 0, \quad \forall \theta_1$$

* Proof: by the inverse Euler Equation we have that

$$\frac{1}{u'(c_1^*(\theta_1))} = \frac{1}{R\beta}\mathbb{E}_1\Big[\frac{1}{u'(c_2^*(\theta_1,\theta_2))}\Big]$$

Applying Jensen's inequality we have:

$$\frac{1}{u'(c_1^*(\theta_1))} \geq \frac{1}{R\beta\mathbb{E}_1[u'(c_2^*(\theta_1,\theta_2))]} \Leftrightarrow 1 \geq \frac{u'(c_1^*)}{\beta R u'(c_2^*)}$$

If $c_2^*(\theta_1,\theta^2)$ is not independent of θ_2 then the inequality is strict. Using this into the definition of $\tau^s(\theta_1)$ proves the result

* Jensen's Inequality: $E[\frac{1}{x}] \geq \frac{1}{E[x]}$

64

* Intuition behind $\tau^s(\theta_1) > 0$: with stochastic types, incentives to exert high effort should also be provided in period $t = 2$. When agents save, providing such inventives is more costly since higher savings creates a positive income effect which leads to reducing effort. This goes against the Ramsey approach, where taxing capital/savings is never optimal in the long run, and the Atkinson-Stiglitz result, where taxing capital/savings is never optimal in the presence of non-linear labour income taxes

○ Implementing the Savings Wedge: in an implementation with income taxes, consumers face the following budget constraints

$$c_1 + k_2 \leq y_1 - T_1(y_1)$$

$$c_2 \leq y_2 + Rk_2 - T_2(y_1, y_2, Rk_2)$$

Since tax implementations in the dynamic model are non-unique, different specifications for T_2 could implement the savings wedge τ^s

Finite Horizon Model: in the finite horizon version of the dynamic mirrlees model we have the following planning problem:

$$\max_{c,y} \mathbb{E}_0 \left[\sum_{t=1}^{T} \beta^{t-1} \big(u(c_t(\theta^t)) - v(\frac{y_t(\theta^t)}{\theta_t}) \big) \right]$$

subject to

$$\mathbb{E}_0 \big[\sum_{t=1}^{T} R^{1-t}(c_t(\theta^t) - y_t(\theta^t)) \big] \leq 0$$

$$\mathbb{E}_0 \big[\sum_{t=1}^{T} \beta^{t-1}(u(c_t(\theta^t)) - v(\frac{y_t(\theta^t)}{\theta})) \big] \geq \mathbb{E}_0 \big[\sum_{t=1}^{T} \beta^{t-1}(u(c_t(\sigma^t(\theta^t))) - v(\frac{y_t(\sigma^t(\theta^t))}{\theta})) \big] \quad \forall \sigma^t$$

Given history θ^{t-1}, we define continuation utility from reporting θ at t and reporting truthfully from $t + 1$ onwards by:

$$w_t(\theta^{t-1}, \theta) = \sum_{s=1}^{T-1} \sum_{\theta^{t+s}} \beta^{s-1} \pi(\theta^{t+s}) \big[u(c_{t+s}(\theta^{t-1}, \theta, \theta_{t+1}^{t+s})) - v(\frac{y_{t+s}(\theta^{t-1}, \theta, \theta_{t+1}^{t+1})}{\theta_{t+s}}) \big]$$

It can be shown that the original set of ICs is equivalent to:

$$u(c_t(\theta^{t-1}, \theta)) - v(\frac{y_t(\theta^{t-1}, \theta)}{\theta}) + \beta w_t(\theta^{t-1}, \theta) \geq u(c_t(\theta^{t-1}, \widehat{\theta})) - v(\frac{y_t(\theta^{t-1}, \widehat{\theta})}{\theta}) + \beta w_t(\theta^{t-1}, \widehat{\theta})$$

for all $t, \theta^{t-1}, \theta, \widehat{\theta}$. Therefore, the planning problem now becomes:

$$\max_{c,y} \mathbb{E}_0\Big[\sum_{t=1}^{T} \beta^{t-1}\big(u(c_t(\theta^t)) - v(\frac{y_t(\theta^t)}{\theta_t}))\big)\Big]$$

subject to

$$\mathbb{E}_0\Big[\sum_{t=1}^{T} R^{1-t}(c_t(\theta^t) - y_t(\theta^t))\Big] \leq 0$$

$$u(c_t(\theta^{t-1},\theta)) - v(\frac{y_t(\theta^{t-1},\theta)}{\theta}) + \beta w_t(\theta^{t-1},\theta) \geq u(c_t(\theta^{t-1},\widehat{\theta})) - v(\frac{y_t(\theta^{t-1},\widehat{\theta})}{\theta}) + \beta w_t(\theta^{t-1},\widehat{\theta})$$

$$w_t(\theta^t) = sum_\theta \pi(\theta)\Big[u(c_{t+1}(\theta^t,\theta)) - v(\frac{y_{t+1}(\theta^t,\theta)}{\theta}) + \beta w_{t+1}(\theta^t,\theta)\Big]$$

for all $t, \theta^{t-1}, \theta, \widehat{\theta}$. Since this problem has no recursive structure we can transofrm it to the dual problem of cost minimization given w_0:

$$\max_{c,y} -\mathbb{E}_0\Big[\sum_{t=1}^{T} R^{1-t}(c_t(\theta^t) - y_t(\theta^t))\Big] \text{subject to}$$

$$u(c_t(\theta^{t-1},\theta)) - v(\frac{y_t(\theta^{t-1},\theta)}{\theta}) + \beta w_t(\theta^{t-1},\theta) \geq u(c_t(\theta^{t-1},\widehat{\theta})) - v(\frac{y_t(\theta^{t-1},\widehat{\theta})}{\theta}) + \beta w_t(\theta^{t-1},\widehat{\theta})$$

$$w_t(\theta^t) = sum_\theta \pi(\theta)\Big[u(c_{t+1}(\theta^t,\theta)) - v(\frac{y_{t+1}(\theta^t,\theta)}{\theta}) + \beta w_{t+1}(\theta^t,\theta)\Big]$$

for all $t, \theta^{t-1}, \theta, \widehat{\theta}$. Thereby, we solve this problem by minimizing the cost of delivering w_0. This dual problem admits a nice recursive formulation:

$$K_t(x) = \max_{c(\theta),y(\theta),w(\theta)} \sum_\theta \pi(\theta)[y(\theta) - c(\theta) + R^{-1}K_{t+1}(w(\theta))]$$

subject to

$$x = \sum_\theta \pi(\theta)\Big[u(c(\theta)) - v(\frac{y(\theta)}{\theta}) + \beta w(\theta)\Big] \tag{PKC}$$

$$u(c(\theta)) - v(\frac{y(\theta)}{\theta}) + \beta w(\theta) \geq u(c(\widetilde{\theta})) - v(\frac{y(\widetilde{\theta})}{\theta}) + \beta w(\widetilde{\theta}), \quad \forall (\theta,\widetilde{\theta}) \tag{IC}$$

<u>Infinite Horizon</u>: extending the framework to $T = \infty$ we have

$$K(x) = \max_{c(\theta), y(\theta), w(\theta)} \sum_\theta \pi(\theta)[y(\theta) - c(\theta) + R^{-1}K(w(\theta))]$$

subject to

$$x = \sum_\theta \pi(\theta)\left[u(c(\theta)) - v(\frac{y(\theta)}{\theta}) + \beta w(\theta)\right] \tag{PKC}$$

$$u(c(\theta)) - v(\frac{y(\theta)}{\theta}) + \beta w(\theta) \geq u(c(\widetilde{\theta})) - v(\frac{y(\widetilde{\theta})}{\theta}) + \beta w(\widetilde{\theta}), \quad \forall (\theta, \widetilde{\theta}) \tag{IC}$$

Note that technical conditions should also be met such that solutions to the original problem and recursive formulation coincide. The key results of the two period model extend to the infinite horizon model, particularly the positive savings wedge and implementations discussed. We set $R = \beta^{-1}$ since this is the only configuration consistent with a steady state and otherwise aggregate consumption would not be constant across periods

- <u>Inequality Dynamics</u>: consider the special case with two types:

$$\Theta = \{\theta_L, \theta_H\}$$

The ICs for this problem are

$$u(c(\theta_H)) - v(\frac{y(\theta_H)}{\theta_H}) + \beta w(\theta_H) \geq u(c(\theta_L)) - v(\frac{y(\theta_L)}{\theta_H}) + \beta w(\theta_L)$$

$$u(c(\theta_L)) - v(\frac{y(\theta_L)}{\theta_L}) + \beta w(\theta_L) \geq u(c(\theta_H)) = v(\frac{y(\theta_H)}{\theta_L}) + \beta w(\theta_H)$$

Using that only (IC_H) is binding at the optimum we can show that at an interior solution

$$y(\theta_H) \geq y(\theta_L), \ c(\theta_H) \geq c(\theta_L), \ w(\theta_H) \geq w(\theta_L)$$

Here w serves as a new margin for incentive provision where to provide incentives the planner spreads out continuation utilities. Hence, providing incentives to reveal private information leads to inequality in lifetime utility promises and consumption. To analyse the optimal degree of inequality, let Ψ_t be the distribution of promised utilities at t. The optimal contract defines a mapping Ω such that

$$\Psi_{t+1} = \Omega(\Psi_t)$$

with Ψ_0 given. This gives us the law of motion for Ψ_t. The aim then is to find the steady state distribution of Ψ^* subject to $\Psi^* = \Omega(\Psi^*)$. Rearranging the planning problem's first order conditions

reveals that the optimal contract is such that

$$K^{'}(x) = \mathbb{E}[K^{'}(w(\theta))]$$

This implies that $\left\{K^{'}(w_t(\theta^t))\right\}$ is martingale and therefore has the following convergence property:

$$x_t \to RV$$

$$E_t[x_{t+1}] = x_t \to \text{ martingale property}$$

- Immiseration Result: suppose consumption utility u is unbounded below. Then

$$w_t(\theta^t) \to -\infty$$

as $t \to \infty$ with probability 1. Since this implies that $c_t \to 0$, in the long run almost everyone is trapped at misery. Note that this does not mean everyone's consumption converges to zero, there will still be some agents with strictly positive consumption but their measure vanishes at $t \to \infty$. In other words, inequality grows without bounds over time and a measure 0 of individuals consume the entire production.

  o Proof: important properties of the value function K are that: K is continuous, strictly concave, strictly decreasing, differentiable, that $\lim_{w \to -\infty} K(w) = \lim_{w \to -\infty} K^{'}(w) = 0$ and $\lim_{w \to \overline{w}} K(w) = \lim_{w \to \overline{w}} K^{'}(w) = \infty$.
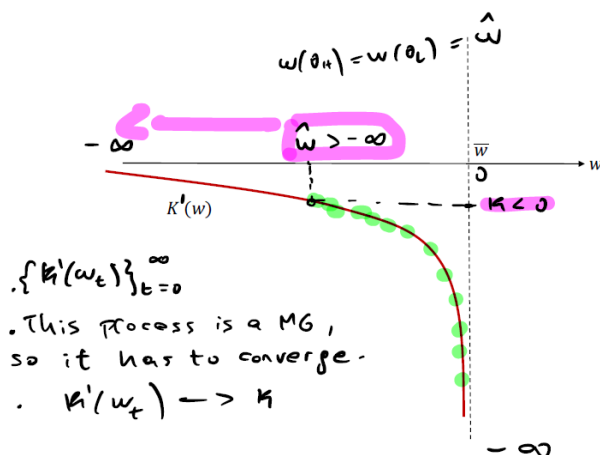
For being a MG, $K^{'}(w_t)$ converges to a random variable with probability 1. It is easy to see that $K^{'}(w_t)$ must converge to 0 since suppose that $K^{'}(w_t) \to k \leq 0$ (recall that K is decreasing) then $w_t \to \widehat{w} > -\infty$, where $\widehat{w}$ is defined by $K^{'}(\widehat{w}) = k$. This implies that, in the limit, continuation utilities are not spread out and so $w(\theta_L) = w(\theta_H)$ which is a contradition. Hence, $K^{'}(w_t)$ must converge to 0 so that $w_t \to -\infty$

○ <u>Intuition</u>: providing incentives is cheaper if the planner introduces a downward drift on consumption becaue utility is concave. Therefore, for poor individuals, even tiny increments in consumption would make them super happy.

○ <u>Issue</u>: immersiration is an issue and there is no well-defined answer regarding the optimal level of inequality (Ψ^* does not exist)

○ <u>Solving for Ψ^*</u>: many economists have provided restricted solutions for if we suppose that Ψ^* exists

  * <u>Lower Bound</u>: impose an exogenous lower bound on continuation utility $\underline{w} > -\infty$

  * <u>Commitment</u>: impose lack of commitment by the planner

  * <u>Philosophical Arguments</u>: impose political economy arguments to microfound $\underline{w} > -\infty$

  * <u>Discounting</u>: impose that the social discount factor ¡ private discount factor

Optimal Unemployment Insurance

Model: economy lasts for $T = \infty$ periods, populated by a worker (agent) and an unemployment insurance agency (principal). The worker can be employed or unemployed and the goal is to characterise the optimal unemployment insurance contract between an unemployment agency and a worker

- **Worker's Preferences**

$$U = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t [u(c_t) - a_t]$$

  where $c_t \geq 0$ is consumption, $u(\cdot)$ is strictly increasing, strictly concave, and bounded below, and $a_t \geq 0$ is search effort that determines the probability of finding a job at $t+1$

- **Employment**: employment is an absorbing state (one spell model) where all jobs are alike and pay wage $w > 0$ forever and the search effort is $a = 0$ if employed

- **Unemployment**: while unemployed the worker finds a job with probability $p(a)$ where $p'(a) > 0, p''(a) < 0, p(0) = 0$. During this time, the worker receives unemployment beneftis from the insurance agency

    - **Non-Storable Consumption Good**: the worker cannnot save or borrow. Unemployment benefits are the only source of consumption (c_t) during unemployment

- **Private Information Friction**: a_t is unobservable by the insurance agency, with the information friction coming from private action (a) rather than private types (making this a moral hazard rather than adverse selection problem)

    - **Note**: though this is a moral hazard instead of adverse selection problem, similar techniques (such as the Revelation Principle) still hold

- **Value of Employment V^e**: let V^e be the value of being employed for the worker. We assume that once a worker finds a job, she is beyond the reach of the agency (i.e. no further obligations such as taxes based on unemployment length). Given that jobs are permanent and pay w forever, we have

$$V^e = \frac{u(w)}{1 - \beta}$$

- **Value of Autarky V_{aut}**: let V_{aut} denote the value of being unemployed and not having acccess to unemployment insurance. V_{aut} is the solution to the Bellman Equation:

$$V_{aut} = \max_a \left\{ u(0) - a + \beta[p(a)V^e + (1 - p(a))V_{aut}] \right\}$$

Model with Full Information: suppose the UI agency can observe and control c and a. The agency designs a UI contract to give the unemployed worker expected utility $V \geq V_{aut}$. Let $C(V)$ be the expected discounted

cost of delivering $V \geq V_{aut}$, the Bellman Equation therefore becomes:

$$C(V) = \min_{c,a,V^u} \{c + \beta(1 - p(a))C(V^u)\}$$

$$\text{subject to}$$

$$u(c) - a + \beta \{p(a)V^e + (1 - p(a))V^u\} = V \quad \text{(PKC)}$$

Note that PKC is the promise keeping constraint, we assume the agency discounts future payoffs using β, c is the unemployment benefit, C is structky convex, and $V \geq V_{aut}, V^u \geq V_{aut}$. Let $\lambda > 0$ be the multiplier on the (PKC).

- Solution: the first order conditions are:

$$\frac{\partial}{\partial c} = 0 \rightarrow \frac{1}{u'(c)} = \lambda \quad (1)$$

$$\frac{\partial}{\partial a} = 0 \rightarrow C(V^u) = \lambda\Big[\frac{1}{\beta p'(a)} - (V_e - V^u)\Big] \geq 0 \quad (2)$$

$$\frac{\partial}{\partial V^u} = 0 \rightarrow C'(v^u) = \lambda$$

$$\text{Envelope Theorem} \rightarrow C'(V) = \lambda$$

By FOC (3) and (4) we have that $V^u = V$ using that C is convex. In sequential notation this is:

$$V_{t+1} = V_t \quad \forall t$$

Unemployment benefits can be written as:

$$C_t = c(V_t) \rightarrow c_t = c_{t+1} \; \forall t$$

$$a_t = a(V_t) \rightarrow a_t = a_{t+1} \; \forall t$$

Rearranging the optimality conditions we get the following proposition: *Under full information, c and a are held constant during the unemployment state.* In the absence of private information, unemployment benefits (c) are fully smoothed during the unemployment state. This implies that the replacement ratio c/w is also constant

- Private Information Issue: the full information solution is not achievable under private information. Note that by FOC (2), the socially optimal level of effort under full information a^* satisfies:

$$[\beta p'(a^*)]^{-1} > V^e - V^{u*}$$

71

However, if offered (c^*, v^{u*}) the worker could choose a according to:

$$a^{private} \in \arg\max_a \{u(c^*) - a + \beta[p(a)V^e + (1 - p(a))V^{u*}]\}$$

This leads to:

$$[\beta p'(a^{private})]^{-1} = V^e - V^{u*}$$

Since p is concave we have that $a^* > a^{private}$. This is as the worker does not take into account the cost of the UI scheme and therefore chooses a level of effort below the socially optimal one.

Model with Asymmetric Information: as effort is not observed, we must ensure that the agent optimally chooses the level of effort a prescribed by the contract. Therefore, we have that:

$$a \in \arg\max_{\widehat{a}} \{u(c) - \widehat{a} + \beta[p(\widehat{a})V^e + (1 - p(\widehat{a}))V^u]\}$$

The necessary and sufficient condition is that a satisfies

$$\beta p'(a)(V^e - V^u) = 1$$

Note that the optimal level of effort is $a > 0$ if and only if $V^e - V^u$. Therefore, we have the following planning problem:

$$C(V) = \min_{c,a,V^u} \{c + \beta(1 - p(a))C(V^u)\}$$

$$\text{subject to}$$

$$u(c) - a + \beta \{p(a)V^e + (1 - p(a))V^u\} = V \tag{PKC}$$

$$\beta p'(a)(V^e - V^u) = 1 \tag{IC}$$

Where we assume that $C(V)$ is strictly convex.

- Solution: let $\lambda > 0$ be the multiplier on the (PKC) and μ be the multiplier on the (IC). We have the following first order conditions:

$$\frac{\partial}{\partial c} = 0 \rightarrow \frac{1}{u'(c)} = \lambda \tag{1}$$

$$\frac{\partial}{\partial a} = 0 \rightarrow C(v^u) = -\mu \frac{p''(a)}{p'(a)}(V^e - V^u)$$

$$\frac{\partial}{\partial V^u} = 0 \rightarrow C'(V^u) = \lambda - \mu \frac{p'(a)}{1 - p(a)}$$

Rearranging the first order conditions we get the proposition: *under private information - (1) unemployment benefits decrease with the duration of unemployment, (2) search effort rises with the duration of unemployment.* The planner knows that $p'(a) > 0$. Exploiting $p'(a) > 0$, the principal introduces duration dependence to provide incentives to search. Using the 'carrot' we will have high beenfits and low effort early in the unemployment spell. Using the 'stick' we will have low benefits and high effort later in the spell.

- Numerical Example: $\beta = 0.99$ with a weekly frequency, $u(c) = \frac{c^{1-\sigma}}{1-\sigma}$, $\sigma = 0.5$, $p(a) = 1 - \exp(-ra)$ with $r = 3 \times 10^{-1}$, $w = 100$
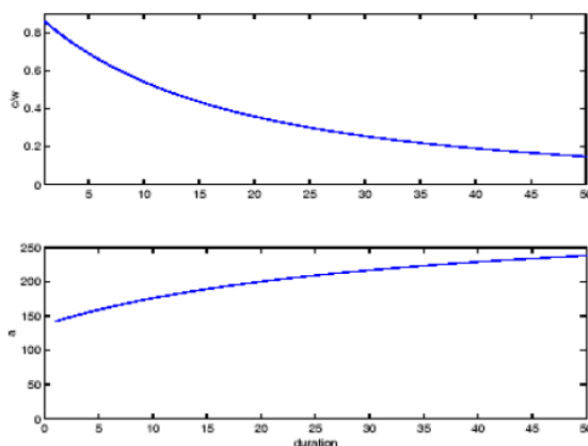
- Model vs. Data: in most OECD countries, unemployment benefits decrease with the duration of Unemployment

Model with Employment Taxes: if we allow the unemployment agency to tax workers once employed we have the key result that taxes typically increase with the duration of the unemployment spell and that the cost savings from switching to this scheme can be large. We have the same environment as before but now the UI agency can use a permanent tax on labor income τ

- Value of Employment: this is now:

$$V^e = \frac{u(w - \tau)}{1 - \beta}$$

We use V^e instead of τ as the control variable for convenience. Therefore, we write tax as a function of V^e as:

$$\tau = w - u^{-1}((1 - \beta)V^e)$$

- Continuation Cost: for the agency in the event of employment::

$$W(V^e) = \frac{-w + u^{-1}((1 - \beta)V^e)}{1 - \beta}$$

- Planning Problem: this becomes:

$$C(V) = \min_{c,a,V^u,V^e} \left\{ c + \beta(1 - p(a))C(V^u) + \beta p(a)W(V^e) \right\}$$

$$\text{subect to}$$

$$u(c) - a + \beta \left\{ (1 - p(a))V^u + p(a)V^e \right\} = V \tag{PKC}$$

$$\beta p'(a)(V^e - V^u) = 1 \tag{IC}$$

From this we have the following propositions:

- ○ Proposition 1: the wage tax τ is not independent of the unemployment history
- ○ Proposition 2: τ increases with unemployment duration under either of the following two conditions - (1) $-p''(a)/p'(a)^2$ is increasing in a, (2) $[-p''(a)(1 - p(a))p(a)]/p'(a)^3$ is increasing in a

To provide incentives, the contract punishes workers for continued unemployment by reducing their claims for future consumption. To minimize costs, consumption should be reduced in all possible states of nature (i.e. unemployment and employment). The latter explains why taxes can increase with unemployment duration (proposition 1). But we need to make sure that the negative incentive effects are not too strong (proposition 2)

- Numerical Example: from Hopenhayn-Nicolini (1997) we have the following:

| | OPTIMAL WITH TAX | | | OPTIMAL WITH NO TAX: |
WEEKS OF UNEMPLOYMENT	Replacement Ratio (1)	C Emp (2)	Tax (%) (3)	REPLACEMENT RATIO (4)
1	99.0	100.5	−.5	85.8
2	98.9	100.4	−.4	80.8
3	98.8	100.3	−.3	76.3
4	98.7	100.2	−.2	72.1
5	98.6	100.1	−.1	68.2
6	98.5	100.0	.0	64.7
7	98.4	99.9	.1	61.4
8	98.3	99.8	.2	58.4
12	97.9	99.4	.6	48.2
16	97.5	99.0	1.0	40.5
26	96.5	98.0	2.0	27.7
52	94.0	95.5	4.5	13.4

Here the gains from using the employment tax can be large where agency's savings can increase by about 30