

Public Economics Lecture Notes

Matteo Paradisi

Contents

| | | |
|----------|---|-----------|
| 1 | Section 1-2: Uncompensated and Compensated Elasticities; Static and Dynamic Labor Supply | 4 |
| 1.1 | Uncompensated Elasticity and the Utility Maximization Problem | 4 |
| 1.2 | Substitution Elasticity and the Expenditure Minimization Problem | 6 |
| 1.3 | Relating Walrasian and Hicksian Demand: The Slutsky Equation | 6 |
| 1.4 | Static Labor Supply Choice | 7 |
| 1.5 | Dynamic Labor Supply | 10 |
| 2 | Section 2: Introduction to Optimal Income Taxation | 12 |
| 2.1 | The Income Taxation Problem | 12 |
| 2.2 | Taxation in a Model With No Behavioral Responses | 12 |
| 2.3 | Towards the Mirrlees Optimal Income Tax Model | 13 |
| 2.4 | Optimal Linear Tax Rate | 13 |
| 2.5 | Optimal Top Income Taxation | 15 |
| 3 | Section 3-4: Mirrlees Taxation | 17 |
| 3.1 | The Model Setup | 17 |
| 3.2 | Optimal Income Tax | 19 |
| 3.3 | Diamond ABC Formula | 20 |
| 3.4 | Optimal Taxes With Income Effects | 21 |
| 3.5 | Pareto Efficient Taxes | 23 |
| 3.6 | A Test of the Pareto Optimality of the Tax Schedule | 23 |
| 4 | Section 5: Optimal Taxation with Income Effects and Bunching | 28 |
| 4.1 | Optimal Taxes with Income Effects | 28 |
| 4.2 | Bunching Estimator | 31 |
| 5 | Section 6: Optimal Income Transfers | 34 |
| 5.1 | Optimal Income Transfers in a Formal Model | 34 |
| 5.2 | Optimal Tax/Transfer with Extensive Margin Only | 35 |
| 5.3 | Optimal Tax/Transfer with Intensive Margin Responses | 36 |
| 5.4 | Optimal Tax/Transfer with Intensive and Extensive Margin Responses | 37 |
| 6 | Section 7: Optimal Top Income Taxation | 38 |
| 6.1 | Trickle Down: A Model With Endogenous Wages | 38 |
| 6.2 | Taxation in the Roy Model and Rent-Seeking | 39 |
| 6.3 | Wage Bargaining and Tax Avoidance | 40 |
| 7 | Section 8: Optimal Minimum Wage and Introduction to Capital Taxation | 44 |
| 7.1 | Optimal Minimum Wage | 44 |

| | | |
|-----------|--|-----------|
| 8 | Section 9: Linear Capital Taxation | 49 |
| 8.1 | A Two-Period Model | 49 |
| 8.2 | Infinite Horizon Model - Chamley (1986) | 50 |
| 8.3 | Infinite Horizon - Judd (1985) | 54 |
| 9 | Section 10: Education Policies and Simpler Theory of Capital Taxation | 57 |
| 9.1 | Education Policies | 57 |
| 9.2 | A Simpler Theory of Capital Taxation | 59 |
| 10 | Section 11: Non-Linear Capital Taxation | 62 |
| 10.1 | Non-Linear Capital Taxation: Two-Periods Model | 62 |
| 10.2 | Infinite Horizon Model | 65 |

1 Section 1-2: Uncompensated and Compensated Elasticities; Static and Dynamic Labor Supply

In this section, we will briefly review the concepts of substitution (compensated) elasticity and uncompensated elasticity. Compensated and uncompensated labor elasticities play a key role in studies of optimal income taxation. In the second part of the section we will study the context of labor supply choices in a static and dynamic framework.

1.1 Uncompensated Elasticity and the Utility Maximization Problem

The utility maximization problem: We start by defining the concept of *Walrasian demand* in a standard *utility maximization problem (UMP)*. Suppose the agent chooses a bundle of consumption goods x_1, \dots, x_N with prices p_1, \dots, p_N and her endowment is denoted by w . The optimal consumption bundle solves the following:

$$\max_{x_1, \dots, x_N} u(x_1, \dots, x_N)$$

s.t.

$$\sum_{i=1}^N p_i x_i \leq w$$

We solve the problem using a Lagrangian approach and we get the following optimality condition (if an interior optimum exists) for every good i :

$$u_i(\mathbf{x}^*) - \lambda^* p_i = 0$$

Solving this equation for λ^* and doing the same for good j yields:

$$\frac{u_i(\mathbf{x}^*)}{u_j(\mathbf{x}^*)} = \frac{p_i}{p_j}$$

This is an important condition in economics and it equates the relative price of two goods to the *marginal rate of substitution (MRS)* between them. The MRS measures the amount of good j that the consumer must be given to compensate the utility loss from a one-unit marginal reduction in her consumption of good i . Graphically, the price ratio is the slope of the budget constraint, while the ratio of marginal utilities represents the slope of the indifference curve.¹

We call the solution to the utility maximization problem *Walrasian* or *Marshallian demand* and we represent it as a function $\mathbf{x}(\mathbf{p}, w)$ of the price vector and the endowment. The Walrasian demand has the following two properties:

- *homogeneity of degree zero*: $x_i(\alpha \mathbf{p}, \alpha w) = x_i(\mathbf{p}, w)$
- *Walras Law*: for every $\mathbf{p} \gg 0$ and $w > 0$ we have $\mathbf{p} \cdot \mathbf{x}(\mathbf{p}, w) = w$

¹Notice that in a two goods economy by differentiating the indifference curve $u(x_1, x_2(x_1)) = k$ wrt x_1 you get:

$$u_1 + u_2 \frac{dx_2}{dx_1} = 0$$

which delivers

$$\frac{dx_2}{dx_1} = -\frac{u_1}{u_2}$$

which shows that the ratio of marginal utilities is the slope of the indifference curve at a point (x_1, x_2) .

We define *uncompensated elasticity* as the percentage change in the consumption of good i when we raise the price p_k . Using the Walrasian demand we can write the uncompensated elasticity as:

$$\varepsilon_{i,p_k}^u = \frac{\partial x_i(\mathbf{p}, w)}{\partial p_k} \frac{p_k}{x_i(\mathbf{p}, w)}$$

Elasticities can also be defined using logarithms such that:

$$\varepsilon_{i,p_k}^u = \frac{\partial \log x_i(\mathbf{p}, w)}{\partial \log p_k}$$

Indirect utility: We introduce the concept of indirect utility that will be useful throughout the class. It also helps interpreting the role of the Lagrange multiplier. The indirect utility is the utility that the agent achieves when consuming the optimal bundle $\mathbf{x}(\mathbf{p}, w)$. It can be obtained by plugging the Walrasian demand into the utility function:

$$v(\mathbf{p}, w) = u(\mathbf{x}(\mathbf{p}, w))$$

The indirect utility has the following properties:

- *homogeneity of degree zero*: since the Walrasian demand is homogeneous of degree zero, it follows that the indirect utility will inherit this property
- $\partial v(\mathbf{p}, w) / \partial w > 0$ and $\partial v(\mathbf{p}, w) / \partial p_k \leq 0$

Roy's Identity and the multiplier interpretation: Using the indirect utility function, the value of the problem can be written as follows at the optimum:

$$v(\mathbf{p}, w) = u(\mathbf{x}^*(\mathbf{p}, w)) + \lambda^*(w - \mathbf{p} \cdot \mathbf{x}^*(\mathbf{p}, w))$$

Applying the Envelope theorem, we can study how the indirect utility responds to changes in the agent's wealth:

$$\frac{\partial v(\mathbf{p}, w)}{\partial w} = \lambda^*$$

The value of the Lagrange multiplier at the optimum is the shadow value of the constraint. Specifically, it is the increase in the value of the objective function resulting from a slight relaxation of the constraint achieved by giving an extra dollar of endowment to the agent. This interpretation of the Lagrangian multiplier is particularly important in the study of optimal Ramsey taxes and transfers. You will see more about it in the second part of the PF sequence.

The Envelope theorem also implies that:

$$\frac{\partial v(\mathbf{p}, w)}{\partial p_i} = -\lambda^* x_i(\mathbf{p}, w)$$

Using the two conditions together we have:

$$-\frac{\frac{\partial v(\mathbf{p}, w)}{\partial p_i}}{\frac{\partial v(\mathbf{p}, w)}{\partial w}} = x_i(\mathbf{p}, w)$$

This equation is known as the *Roy's Identity* and it derives the Walrasian demand from the indirect utility function.

1.2 Substitution Elasticity and the Expenditure Minimization Problem

In this section we aim to isolate the substitution effect of a change in price. An increase in the price of good i typically generates two effects:

- *substitution effect*: the relative price of x_i increases, therefore the consumer substitutes away from this good towards other goods,
- *income effect*: the consumer's purchasing power has decreased, therefore she needs to reoptimize her entire bundle. This reduces even more the consumption of good i .

We define *substitution or compensated elasticity* as the percentage change in the demand for a good in response to a change in a price that ignores the income effect. In order to get at this new concept, we focus on a problem that is “dual” to the utility maximization problem: the *expenditure minimization problem (EMP)*. The consumer solves:

$$\min_{x_1, \dots, x_N} \sum_{i=1}^N p_i x_i$$

s.t.

$$u(x_1, \dots, x_N) \geq \bar{u}$$

The problem asks to solve for the consumption bundle that minimizes the amount spent to achieve utility level \bar{u} . The solution delivers two important functions: the *expenditure function* $e(\mathbf{p}, \bar{u})$, which measures the total expenditure needed to achieve utility \bar{u} under the price vector \mathbf{p} , and the *Hicksian* (or *compensated*) *demand* $\mathbf{h}(\mathbf{p}, \bar{u})$, which is the demand vector that solves the minimization problem.

The Walrasian and Hicksian demands answer two different but related problems. The following two statements establish a relationship between the two concepts:

1. If \mathbf{x}^* is optimal in the UMP when wealth is w , then \mathbf{x}^* is optimal in the EMP when $\bar{u} = u(\mathbf{x}^*)$. Moreover, $e(\mathbf{p}, \bar{u}) = w$.
2. If \mathbf{x}^* is optimal in the EMP when \bar{u} is the required level of utility, then \mathbf{x}^* is optimal in the UMP when $w = \mathbf{p} \cdot \mathbf{x}^*$. Moreover, $\bar{u} = u(\mathbf{x}^*)$.

The Hicksian demand allows us to isolate the pure substitution effect in response to a price change. We call it compensated since it is derived following the idea that, after a price change, the consumer will be given enough wealth (the “compensation”) to maintain the same utility level she experienced before the price change. Suppose that under the price vector \mathbf{p} the consumer demands a bundle \mathbf{x} such that $\mathbf{p} \cdot \mathbf{x} = w$. When the price vector is \mathbf{p}' , the consumer solves the new expenditure minimization problem and switches to \mathbf{x}' such that $u(\mathbf{x}) = u(\mathbf{x}')$ and $\mathbf{p}' \cdot \mathbf{x}' = w'$. The change $\Delta w = w' - w$ is the compensation that the agent receives to be as well off in utility terms after the price change as she was before. Thanks to the compensation there is no income effect coming from the reduction in the agent's purchasing power.

We call the elasticity of the Hicksian demand function *compensated elasticity* and it reads:

$$\varepsilon_{i,p_k}^c = \frac{\partial h_i(\mathbf{p}, \bar{u})}{\partial p_k} \frac{p_k}{h_i(\mathbf{p}, \bar{u})}$$

1.3 Relating Walrasian and Hicksian Demand: The Slutsky Equation

We now establish a relationship between the Walrasian and the Hicksian demand elasticities. We know that $u(x_i(\mathbf{p}, w)) = \bar{u}$ and $e(\mathbf{p}, \bar{u}) = w$. Start from the following identity:

$$x_i(\mathbf{p}, e(\mathbf{p}, \bar{u})) = h_i(\mathbf{p}, \bar{u})$$

and differentiate both sides wrt p_k to get:

$$\begin{aligned}
\frac{\partial h_i(\mathbf{p}, \bar{u})}{\partial p_k} &= \frac{\partial x_i(\mathbf{p}, e(\mathbf{p}, \bar{u}))}{\partial p_k} + \frac{\partial x_i(\mathbf{p}, e(\mathbf{p}, \bar{u}))}{\partial e(\mathbf{p}, \bar{u})} \frac{\partial e(\mathbf{p}, \bar{u})}{\partial p_k} \\
&= \frac{\partial x_i(\mathbf{p}, w)}{\partial p_k} + \frac{\partial x_i(\mathbf{p}, w)}{\partial w} h_k(\mathbf{p}, \bar{u}) \\
&= \frac{\partial x_i(\mathbf{p}, w)}{\partial p_k} + \frac{\partial x_i(\mathbf{p}, w)}{\partial w} x_k(\mathbf{p}, e(\mathbf{p}, \bar{u})) \\
&= \frac{\partial x_i(\mathbf{p}, w)}{\partial p_k} + \frac{\partial x_i(\mathbf{p}, w)}{\partial w} x_k(\mathbf{p}, w)
\end{aligned}$$

Rearranging, we derive the following relation:

$$\underbrace{\frac{\partial x_i(\mathbf{p}, w)}{\partial p_k}}_{\text{uncompensated change}} = \underbrace{\frac{\partial h_i(\mathbf{p}, \bar{u})}{\partial p_k}}_{\text{substitution effect}} - \underbrace{\frac{\partial x_i(\mathbf{p}, w)}{\partial w} x_k(\mathbf{p}, w)}_{\text{income effect}}$$

we have thus decomposed the uncompensated change into income and substitution effect. Notice also how the income effect is the product of two terms: $\frac{\partial x_i(\mathbf{p}, w)}{\partial w}$ is the response of the Walrasian demand for good i to a change in wealth; $x_k(\mathbf{p}, w)$ is the *mechanical effect* of an increase in p_k on the agent's purchasing power: an agent whose demand for k was $x_k(\mathbf{p}, w)$ experiences a mechanical reduction of her purchasing power amounting to $x_k(\mathbf{p}, w)$ when p_k increases by 1. J. R.

1.4 Static Labor Supply Choice

In this paragraph we study a simple framework of labor supply choice and we derive uncompensated labor elasticities. Assume an agent derives utility from consumption, but disutility from labor. Her preferences are represented by the utility function $u(c, n)$ where $\partial u / \partial c > 0$ and $\partial u / \partial n < 0$. The agent has I amount of wealth and earns salary w . We normalize the price of consumption to 1.² The utility maximization problem now is:

$$\max_{c, n} u(c, n)$$

s.t.

$$c = wn + I$$

Taking FOCs and rearranging we get the following:

$$-\frac{u_n}{u_c} = w$$

This condition is similar to the one we derived above. It equates the cost of leisure w to the marginal rate of substitution between labor and consumption. Dividing the marginal disutility of labor by the marginal utility of consumption we get the marginal utility cost of labor in consumption units. The condition therefore equates the marginal utility cost of labor to the salary.

We now want to study the labor supply response to a change in salary. Suppose that the wage increases. Since the consumer gets paid more for every hour she works, she will tend to work more (which implies that she will consume less leisure). This is the *substitution effect*. However, since the agent earns more for every hour of work, she gets paid more for the amount of hours she were already

²Notice that we can normalize the price of consumption in a two goods economy and interpret salary w as the relative price of leisure over consumption.

working. Since the consumer is wealthier, if leisure is a normal good, she will tend to work less and consume more leisure. This is the *income effect*. Notice that, even if the cost of leisure has increased, the income and substitution effects do not go in the same direction unlike in standard consumer problems where an increase in the price of good i generates a negative income and substitution effect for good i . The reason is that this is an endowment economy where we think about leisure l as the difference between total time endowment T and labor. We have $l = T - n$. In this setup the agent is a net seller of leisure and therefore the income effect is positive for leisure when the salary increases.

Now we get a little more formal and we study analytically the response of labor supply to changes in the wage rate. Totally differentiating the optimality condition wrt w we get:

$$\frac{\partial n}{\partial w} = -\frac{u_c + n(u_{nc} + wu_{cc})}{w^2u_{cc} + 2wu_{nc} + u_{nn}}$$

Notice that the denominator of the expression is the second order condition of the problem and can therefore be signed. If we assume the problem is concave (in order to get an interior solution), the denominator is negative. This implies that:

$$\frac{\partial n}{\partial w} \propto u_c + n(u_{nc} + wu_{cc})$$

This expression captures the intuition provided above. The first term is the substitution effect, which is always positive and proportional to the marginal utility of consumption: the extent to which the consumer substitutes labor and consumption depends on how attractive consumption is. The second term measures the income effect. It depends on the cross-derivative of consumption and labor and the concavity of the utility function in consumption. The cross-derivative measures how changes in consumption affect the labor disutility. Faster decreasing marginal returns to consumption imply lower incentive to consume when the agent becomes wealthier (remember that $u_{cc} < 0$). The income effect is scaled by n , which is the mechanical effect on endowment of a one unit increase of w .

Example: We now study a functional form for preferences that is particularly convenient for the study of optimal tax problems. Suppose the agent has the following utility:

$$u(c, n) = c - \frac{n^{1+\frac{1}{\varepsilon}}}{1 + \frac{1}{\varepsilon}}$$

This is a quasi-linear utility function whose property is to rule out income effects. We will come back to this point later.

The optimality condition reads:

$$n^{\frac{1}{\varepsilon}} = w$$

Taking logs we get:

$$\frac{1}{\varepsilon} \log n = \log w$$

Since $\varepsilon_{n,w}^u = \partial \log n / \partial \log w$ we can write:

$$\varepsilon_{n,w}^u = \partial \log n / \partial \log w = \varepsilon$$

Therefore, this utility function has a constant elasticity of labor supply. Also, given the absence of income effects, we know that $\varepsilon_{n,w}^u = \varepsilon_{n,w}^c$.

Compensated Labor Supply Elasticity: We can derive the compensated response of labor supply by using the Slutsky equation. We already know the uncompensated response to wage changes and we therefore need to find $\partial n / \partial I$. Totally differentiating the FOC wrt I we get:

$$\frac{\partial n}{\partial I} = -\frac{u_{nc} + w u_{cc}}{w^2 u_{cc} + 2w u_{nc} + u_{nn}}$$

The Slutsky equation is the following:

$$\frac{\partial n}{\partial w} = \frac{\partial n^c}{\partial w} + \frac{\partial n}{\partial I} n$$

Notice that the sign of the income effect is flipped since w is the price of leisure, while we are studying the response of labor. We therefore conclude:

$$\frac{\partial n^c}{\partial w} = -\frac{u_c}{w^2 u_{cc} + 2w u_{nc} + u_{nn}}$$

By comparing the compensated and uncompensated response we clearly see why quasi-linear preferences imply no income effect: they are separable and linear in consumption. Therefore, $u_{nc} = 0$ and $u_{cc} = 0$.

λ -constant Elasticity: We introduce a concept that will be useful later in the analysis of intertemporal elasticities. The first order conditions for the static labor supply model solved with a Lagrangian approach are:

$$\begin{aligned} u_c &= \lambda \\ u_n &= -\lambda w \end{aligned}$$

Define the λ -constant or Frisch elasticity the elasticity that is computed assuming λ does not change. Totally differentiating we get:

$$\begin{bmatrix} u_{cc} & u_{cn} \\ u_{nc} & u_{nn} \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial c^F}{\partial w} \\ \frac{\partial n^F}{\partial w} \end{bmatrix} = \begin{bmatrix} 0 \\ -\lambda \end{bmatrix}$$

By inverting the 2×2 matrix we can solve the system. The solutions are:

$$\begin{aligned} \begin{bmatrix} \frac{\partial c^F}{\partial w} \\ \frac{\partial n^F}{\partial w} \end{bmatrix} &= \begin{bmatrix} u_{cc} & u_{cn} \\ u_{nc} & u_{nn} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ -\lambda \end{bmatrix} \\ &= \frac{1}{u_{cc} u_{nn} - u_{cn}^2} \begin{bmatrix} u_{nn} & -u_{cn} \\ -u_{nc} & u_{cc} \end{bmatrix} \begin{bmatrix} 0 \\ -\lambda \end{bmatrix} \\ &= \begin{bmatrix} \frac{\lambda u_{cn}}{u_{cc} u_{nn} - u_{cn}^2} \\ \frac{-\lambda u_{cc}}{u_{cc} u_{nn} - u_{cn}^2} \end{bmatrix} \end{aligned}$$

A Comparison Among Elasticities: We now draw a comparison among the three elasticities presented so far. We already know that $\varepsilon_{l,w}^c \geq \varepsilon_{l,w}^u$ since the income effect is negative on labor supply. We therefore need to compare compensated and λ -constant elasticity. We will prove that $\varepsilon_{l,w}^F \geq \varepsilon_{l,w}^c$. Start by writing the following:

$$\begin{aligned}
\frac{1}{\varepsilon_{l,w}^c} - \frac{1}{\varepsilon_{l,w}^F} &= -\frac{w^2 u_{cc} + 2w u_{nc} + u_{nn}}{u_c} + \frac{u_{cc} u_{nn} - u_{cn}^2}{u_c u_{cc}} \\
&= \frac{1}{u_c} \left(-w^2 u_{cc} - 2w u_{nc} - \frac{u_{cn}^2}{u_{cc}} \right)
\end{aligned}$$

The definition of λ -constant elasticity implies that $u_{nn} \geq \frac{u_{cn}^2}{u_{cc}}$. It follows that:

$$\begin{aligned}
-w^2 u_{cc} - 2w u_{nc} - \frac{u_{cn}^2}{u_{cc}} &\geq -w^2 u_{cc} - 2w u_{nc} - u_{nn} \\
&= -SOC \\
&\geq 0
\end{aligned}$$

Where the last inequality uses the fact that the second order condition must be negative. Hence, we established that $\frac{1}{\varepsilon_{l,w}^c} - \frac{1}{\varepsilon_{l,w}^F} \geq 0$, which implies $\varepsilon_{l,w}^F \geq \varepsilon_{l,w}^c$. Keeping the marginal utility of consumption constant implies that there are no income effects: higher wealth given the same amount of hours of work does not change preferences towards consumption. Thus, the λ -constant elasticity is at least as big as the compensated one.

We therefore conclude that the following is always true:

$$\varepsilon_{l,w}^F \geq \varepsilon_{l,w}^c \geq \varepsilon_{l,w}^u$$

1.5 Dynamic Labor Supply

In the previous paragraph we studied the static labor supply choice. Now we will switch to a dynamic setting that allows us to study labor supply responses to over time changes in salaries. Agents make labor supply decisions in view of their lifetime. Current labor supply depends on current and future wages and income. Compared to static labor supply models, the substitution effect is similar, but the income effect differs since the agent faces a lifetime budget constraint. MaCurdy (1981) provides a useful framework to study labor supply elasticities over the lifecycle. In order to achieve our goal, we need to separate exogenous static changes (such as the ones studied above) from evolutionary changes, due to shifts in the life-cycle wage profile. In this analysis we need to distinguish between expected and unexpected wage changes. While expected changes do not lead to wealth effects, permanent unexpected changes generate strong wealth effects.

We distinguish among three dimensions of labor supply:

1. the pure *lifecycle dimension*. Usually, wages have a hump-shaped pattern over the lifecycle. Agents adjust the hours of work in response to the different salaries they observe along their lifetime.
2. the *macro dimension*. Hours of work vary over the business cycle following unexpected shocks.
3. the *idiosyncratic dimension*. A person may have temporarily higher wages in some period.

In order to isolate the labor supply response to expected changes in wage we need to rule out wealth effects. We will employ the concept of Frisch elasticity, which allows us to keep the marginal utility of consumption constant.

We study intertemporal labor supply in the same framework as before, but we introduce the time dimension. Preferences are now:

$$\sum_{s=t}^{+\infty} \beta^s u(c_s, n_s)$$

The consumer faces the following period-by-period budget constraint:

$$A_{t+1} = (1 + r_t) (A_t + y_t + w_t n_t - c_t)$$

The Bellman equation for the problem is:

$$V(A_t) = \max_{c_t, n_t} u(c_t, n_t) + \beta V((1 + r_t)(A_t + y_t + w_t n_t - c_t))$$

The FOCs for the problem read:

$$\begin{aligned} \lambda_t &= \beta (1 + r_t) V'_{t+1}(A_{t+1}) \\ u_n(c_t, n_t) &= w_t \lambda_t \end{aligned}$$

By envelope $\lambda_t = V'(A_t)$. Therefore, the conditions become:

$$\begin{aligned} \lambda_t &= \beta (1 + r_t) V'_{t+1}(\lambda_{t+1}) \\ u_n(c_t, n_t) &= \lambda_t w_t \end{aligned}$$

Since $\lambda_t = u_c(c_t, n_t)$ the static labor supply choice is the same as in the previous paragraph:

$$-\frac{u_n(c_t, n_t)}{u_c(c_t, n_t)} = w_t$$

Rearranging the budget constraint we have:

$$c_t = w_t n_t + y_t + \left[\frac{A_{t+1}}{1 + r_t} - A_t \right]$$

Notice that the problem is identical to the previous one where income $I_t = y_t + \left[\frac{A_{t+1}}{1 + r_t} - A_t \right]$.

In order to assess the Frisch elasticity, we need to compute the labor responses to changes in w when we keep the λ constant avoiding any wealth effect. The Frisch demands are defined as follows:

$$\begin{aligned} c_t &= c_t^F(w_t, \lambda_t) \\ n_t &= n_t^F(w_t, \lambda_t) \end{aligned}$$

Since the model is identical to the static labor supply choice and we already derived the Frisch elasticity for the latter, we can write:

$$\varepsilon_{n_t, w_t}^F = \frac{\partial n_t^F}{\partial w_t} \frac{w_t}{n_t} = \frac{-\lambda u_{cc}(c_t, n_t)}{u_{cc}(c_t, n_t) u_{nn}(c_t, n_t) - u_{cn}^2(c_t, n_t)} \frac{w_t}{n_t}$$

References

- MaCurdy, T. E.. *An Empirical Model of Labor Supply in a Life-Cycle Setting*, Journal of Political Economy, 1981, vol. 89, issue 6, pages 1059-85
- Mas-Colell, A., M. Whinston, and J. R. Green. *Microeconomic Theory*. New York: Oxford University Press, 1995.
- Miller, N. H. *Notes on Microeconomic Theory*

2 Section 2: Introduction to Optimal Income Taxation

In this section we will introduce the problem of optimal income taxation. We will set up the government problem and derive optimal taxes. We will study optimal linear tax rate, optimal top tax rate and the revenue maximizing tax rate.

2.1 The Income Taxation Problem

Our goal for most of this class is to derive the properties of optimal taxes in different context. We will define the tax in a flexible way using the mathematical object $T(z)$, where z is the income reported by the agent. The tax $T(z)$ generates the *retention function* $R(z) = z - T(z)$. $R(z)$ measures how much the agent can retain out of total income z . We denote transfers to income z with $-T(z)$ so that the transfer $-T(0)$ to non-working individuals is the intercept of the retention function.

If $T(z)$ is differentiable, $T'(z)$ represents the *marginal tax rate*. It measures how much the agent gets taxed out of one additional dollar of income.

In order to study the extensive margin decision between working and remaining unemployed, we need to know the *participation tax rate* $\tau_p = \frac{T(z)-T(0)}{z}$. It is the fraction of income that an agent pays in taxes when she moves from 0 income to z .

2.2 Taxation in a Model With No Behavioral Responses

We start with a simple version of an optimal income taxation problem that ignores the labor supply response to taxation. Suppose the agent has utility $u(c)$ such that $u'(c) > 0$ and $u''(c) \leq 0$. Labor does not enter the utility function and it is supplied inelastically. The agent consumes everything that is left after taxes so that $c = z - T(z)$. The economy is populated by several agents and their income is distributed according to $h(z)$ with support $[0, \infty]$.

We study the problem of a government, whose goal is to maximize the total utility of the economy. Every agent in the economy is equally weighted such that:

$$\int_0^\infty u(z - T(z)) h(z) dz$$

We call this type of social welfare function *utilitarian*. The government targets a level of revenues E and its budget constraint is:

$$\int_0^\infty T(z) h(z) dz \geq E$$

The Lagrangian for the problem reads:

$$L = [u(z - T(z)) + \lambda T(z)] h(z)$$

Where λ is constant across individuals and measures the value of government revenues in equilibrium. The optimal choice of $T(z)$ delivers the following first order condition:

$$\frac{\partial L}{\partial T(z)} = [-u'(z - T(z)) + \lambda] h(z) = 0$$

Rearranging:

$$u'(z - T(z)) = \lambda$$

Notice that since λ is constant and all agents have the same preferences, the equilibrium condition implies that consumption is equalized across all individuals. This is a direct consequence of the utilitarian social welfare function and the concavity of the utility. Suppose that we taxed a rich individual who would otherwise have a high level of consumption to redistribute to a poor who would otherwise have low consumption. The marginal utility gain of the poor would be higher than the marginal utility loss of the rich if the utility has decreasing marginal returns (implied by the concavity of the utility function). This implies that until all consumption levels are equalized across the economy the government can increase social welfare through “redistribution” from rich to poor individuals. Since every agent has the same weight in the government social welfare function, the optimal policy will treat all individuals equally. There is no gain for the government from guaranteeing a higher level of consumption to a particular group of individuals.

Taxes will serve the purpose of collecting the revenues needed to meet the requirement E . Each individual consumes $c = \bar{z} - E$, where $\bar{z} = \int_0^\infty zh(z) dz$ is the average income. Therefore, we have a 100% marginal tax rate above $\tilde{z} = \bar{z} - E$.

2.3 Towards the Mirrlees Optimal Income Tax Model

The main limitation of the model presented in the previous paragraph is the absence of behavioral responses. Agents were not allowed to respond to fiscal incentives and adjust the labor supply according to the tax schedule. We showed that an extreme case of 100% marginal tax rate can be optimal without causing a loss of revenues due to lower labor supply. We now relax the assumption of inelastic labor supply and study a more flexible model.

Suppose the agent has preferences over consumption and labor represented by the utility function $u(c, l)$. Each agent earns income wl when supplying l hours of labor and consumes $c = wl - T(wl)$ after taxes. Individuals are heterogeneous in the salary w that we will interpret as a measure of ability. Salaries are distributed according to $f(w)$.

Changes in taxes have labor supply effects that depends on the characteristics of the change. A lump-sum change in the level of taxes at a given income changes labor supply through an income effect. On the other hand, a shift in the marginal tax rate causes a distortion in the labor supply through a substitution effect.

Social Welfare Functions: The general problem in Mirrlees (1971) assumes that individual welfare is aggregated through a social welfare function $G(\cdot)$. We typically assume that $G(\cdot)$ is concave in order to represent redistributive preferences of the government. We define the following a social marginal welfare weight:

$$g_i = \frac{G'(u^i) u_c^i}{\lambda}$$

It measures the government marginal utility from giving a dollar to individual i . The expression is scaled by the marginal value of revenues to the government (λ), that converts the marginal utility in money metric. The concavity of the utility implies that g_i is decreasing in z_i . The social welfare effect of giving \$1 to all the individuals in the economy is therefore $\int_i g_i$.

2.4 Optimal Linear Tax Rate

In this paragraph we study the optimal income tax when we restrict the instruments that the government can use to tax income. We focus on *linear taxes* τ . The revenues of the tax are rebated through lump-sum transfers. The individual therefore consumes:

$$c_i = (1 - \tau) w_i l_i + \tau Z$$

where Z represents the total income level in equilibrium and therefore τZ is the total tax revenue from the tax.

The government sets the linear tax to maximize the following:

$$\int_i G[u_i((1-\tau)w_i l_i + \tau Z, l_i)]$$

Notice that we do not have any government budget constraint since the entire revenue is rebated. Applying the Envelope theorem we get:

$$\begin{aligned} \int_i G'(u_i) u'_i \left[-w_i l_i + Z - \tau \frac{dZ}{d(1-\tau)} \right] &= 0 \\ \int_i G'(u_i) u'_i \left[-z_i + Z - \frac{\tau}{(1-\tau)} Z \varepsilon_{z,1-\tau} \right] &= 0 \end{aligned}$$

Where the second line exploits the definition of uncompensated elasticity. Unlike z_i , we implicitly differentiate Z since the individual does not maximize over Z , but takes the transfer as given. In other words the agent does not internalize the effect of her labor supply choice on aggregate revenues and transfers. This is why the Envelope theorem does not apply to Z .

The two terms in the expression above are central in the optimal taxation literature:

- $Z - z_i$ is the mechanical effect of the tax. Suppose we keep labor supply unchanged, an increase in τ generates a drop in income of z_i and a mechanical increase in transfers of Z due to higher revenues.
- $\frac{\tau}{(1-\tau)} Z \varepsilon_{z,1-\tau}$ is the behavioral effect of the tax. If we allow individuals to adjust their labor supplies we have to take into account the fiscal externality on revenues: when people work less the government collects lower revenues.

We could expect to see in the formula the utility consequence of a change in labor supply. However, any welfare effect related to the behavioral response of the individual is excluded. The reason is that although the agent changes the labor supply, if the tax change is small enough we can neglect the utility effect invoking the envelope theorem. Remember that the logic of the envelope theorem is that after we shift a parameter (the tax in this case) the agent is moving to a new bundle on the same indifference curve.

Rearranging the optimality condition we find:

$$\begin{aligned} Z \int_i g_i - \int_i g_i z_i &= \frac{\tau}{(1-\tau)} Z \varepsilon_{z,1-\tau} \int_i g_i \\ 1 - \frac{\int_i g_i z_i}{Z \int_i g_i} &= \frac{\tau}{(1-\tau)} \varepsilon_{z,1-\tau} \end{aligned}$$

We define $\bar{g} = \frac{\int_i g_i z_i}{Z \int_i g_i}$ and rewrite the condition above to get the optimal tax rate:

$$\tau^* = \frac{1 - \bar{g}}{1 - \bar{g} + \varepsilon_{z,1-\tau}}$$

The optimal tax is decreasing in $\varepsilon_{z,1-\tau}$ and \bar{g} . When income is very elastic to taxes, the government will tax less to avoid negative effects on revenues and transfers coming from distortions to the labor supply. This is the efficiency part of the formula. On the other hand, \bar{g} is a measure of inequality in the economy. It is low when income is extremely polarized. Therefore, the government increases taxes at the optimum when inequality is high. This is the equity part of the formula.

2.5 Optimal Top Income Taxation

We now derive taxes as in Saez (2001). Instead of specifying a model, we consider the different effects of a tax change and derive the tax by imposing that their sum is zero in equilibrium. Suppose the government wants to optimally set a constant marginal tax rate τ above an income threshold z^* . The average income above z^* is denoted by $z(1 - \tau)$ and it depends on the tax rate in place. The uncompensated elasticity of z for top earners is constant and denoted by $\varepsilon_{z,1-\tau}$.

When tax τ is raised we have no effects on individuals with income below z^* , while all income above z^* are affected by the change. We will study three different effects of the tax.

Mechanical Effect Suppose labor supply was inelastic, when τ increases we would see a mechanical increase in revenues of the following form:

$$dM = d\tau (z - z^*)$$

The mechanical effect is proportional to the difference between the average income above z^* and z^* . It measures the mechanical increase in revenues that is generated by the tax change.

Behavioral Effect Top earners react to the tax increase by adjusting their labor supply. The behavioral response triggers a fiscal externality and a reduction in revenues. The behavioral effect is:

$$\begin{aligned} dB &= \tau dz = -\tau \frac{dz}{d(1-\tau)} d\tau \\ &= -\frac{\tau}{1-\tau} \frac{1-\tau}{z} \frac{dz}{d(1-\tau)} d\tau \\ &= -\frac{\tau}{1-\tau} \varepsilon_{z,1-\tau} z d\tau \end{aligned}$$

It is proportional to the elasticity of labor supply since the more elastic is labor the higher is the revenue loss.

Welfare Effect Denote with \bar{g} the (assumed) constant social marginal welfare weight for earners above z^* . The tax change mechanically raises revenues on top income individuals generating the following welfare effect:

$$dW = d\tau \bar{g} (z - z^*)$$

We also showed that the tax increase triggers a behavioral response. The reason it is not included in the welfare effect is that if the tax change is small people reoptimize at the margin and their utility level is unaffected. Again, this is an Envelope theorem argument.

Optimal Tax In equilibrium the three effects must sum to zero. If they did not the government would have margin to adjust the tax rate and achieve a higher social welfare. We therefore have:

$$dM + dB + dW = d\tau \left[(1 - \bar{g}) [z - z^*] - \varepsilon_{z,1-\tau} \frac{\tau}{1-\tau} z \right] = 0$$

Rearranging:

$$\tau^* = \frac{1 - \bar{g}}{1 - \bar{g} + a \varepsilon_{z,1-\tau}}$$

with $a = \frac{z}{z - z^*}$ measuring the thinness of the right tail in the income distribution. The optimal tax is decreasing in the social marginal welfare weight of top earners \bar{g} : the more the government cares

about top income individuals, the less they will be taxed. As we could expect, the optimal tax is also decreasing in the elasticity of labor supply. Higher elasticity implies larger efficiency costs. Finally, τ^* decreases in a . The shape of the income distribution matters: the government sets lower top income taxes when earners above z^* are mostly concentrated around z^* . If instead there is a thicker tail, the top income tax is higher.

References

Piketty, Thomas and Emmanuel Saez “Optimal Labor Income Taxation,” Handbook of Public Economics, Volume 5, Amsterdam: Elsevier-North Holland, 2013. (web)

3 Section 3-4: Mirrlees Taxation

In this section we will solve the Mirrlees tax problem. We will and derive optimal taxes introducing the concept of wedges and study the model with and without income effects.

3.1 The Model Setup

Suppose the agent has preferences over consumption and labor represented by the utility function $u(c, l)$ that we assume separable and quasi-linear such that $u(c, l) = c - v(l)$. We assume that $v'(l) > 0$ and $v''(l) \geq 0$. Each agent earns income $z = nl$ when supplying l hours of labor and consumes $c = nl - T(nl)$ after taxes. Individuals are heterogeneous in the salary n that represents their type and we will interpret as a measure of ability. Salaries are distributed according to $f(n)$, with $n \in [\underline{n}, \bar{n}]$. Individual welfare is aggregated through a social welfare function $G(\cdot)$, that we assume differentiable and concave.

Revelation Principle Throughout all of the tax problems that we study we will assume that the government cannot observe the labor choice of the agent and her type. Income is the only observed choice that the government can target. We solve the model using a revelation mechanism. Our goal is to define an optimal tax schedule that delivers an allocation $z(n), c(n)$ to each agent n . The *Revelation Principle* claims that if an allocation can be implemented through some mechanism, then it can also be implemented through a direct truthful mechanism where the agent reveals her information about n .

We imagine that each agent reports to the government her type n' and that allocations are a function of n' such that we can write $c(n'), l(n'), z(n')$ and $u(n')$. By revelation principle, the government cannot do better than defining functions $c(n), z(n)$ such that the agent finds optimal to reveal her true productivity:

$$c(n) - v\left(\frac{z(n)}{n}\right) \geq c(n') - v\left(\frac{z(n')}{n}\right)$$

for every n and n' where n is the true type of the agent. Notice that since n is continuous we have an infinity of constraints. In order to reduce the dimensionality of the problem, we assume that the marginal rate of substitution between consumption and before-tax income is decreasing in n :

$$-MRS_{cz} = \frac{v'(z(n)/n)}{nu'(c(n))} \text{ decreases in } n$$

This is the so called *single-crossing condition* (or *Spence-Mirrlees condition*). Single-crossing and incentive compatibility imply the monotonicity of allocations (i.e. $c(n), z(n)$ are increasing in n). If monotonicity and single-crossing are satisfied, we can replace the incentive constraint with the first-order necessary conditions of the agent that provide a local incentive condition. Under monotonicity and single-crossing the local conditions are also sufficient. While solving these problems we will only impose local incentive constraints and ignore the monotonicity of allocations, which is then verified ex-post.

Incentive Compatibility We reduce the dimensionality of the problem by taking a *first order approach* that replaces the infinity of constraints for each individual with a local condition relying on the optimal revelation choice. When reporting, the individual of type n solves the following problem:

$$\max_{n'} c(n') - v\left(\frac{z(n')}{n}\right)$$

the first order necessary condition for this problem is:

$$c'(n') - \frac{z'(n')}{n} v' \left(\frac{z(n')}{n} \right) = 0$$

If the government wants the agent to reveal her true type, it must be:

$$c'(n) = \frac{z'(n)}{n} v' \left(\frac{z(n)}{n} \right)$$

Under the concavity assumption on the preferences, this is a global incentive constraint condition. Suppose we study local utility changes by totally differentiating the utility wrt n , we get:

$$\frac{du(n)}{dn} = \left(c'(n) - \frac{z'(n)}{n} v' \left(\frac{z(n)}{n} \right) \right) + \frac{z(n)}{n^2} v' \left(\frac{z(n)}{n} \right)$$

Notice that the term in the first bracket is the first order condition of the agent. We can thus write $du(n)/dn = (z(n)/n^2) v'(z(n)/n)$. This equation pins down the slope of the utility assigned to the agent at the optimum. By convexity of $v(\cdot)$, the slope is always positive: the government assigns higher utility to higher types at the optimum. Higher types have a lower marginal disutility of labor for a given level of hours worked and they get *informational rents* in the equilibrium.

Labor Supply and Labor Wedge The individual solves the following optimization problem:

$$\max_z z - T(z) - v \left(\frac{z}{n} \right)$$

The first order condition is:

$$T'(z) = 1 - \frac{v'(l)}{n}$$

The second term on the right-hand-side of the equation is the marginal rate of substitution between consumption and income and we can always write that $T'(z) = 1 - MRS(n)$. When the agent is not distorted, the MRS is equal to 1 implying $T'(z) = 0$. We can interpret $T'(z)$ as a *wedge* on the optimal labor supply: whenever it is different from zero, labor supply is distorted. Wedges are a central concept in the optimal taxation literature and we will encounter them throughout the class.

From the optimality condition, we can derive the elasticity of labor wrt the net of tax wage. Rewrite the optimality condition as:

$$v' \left(\frac{z}{n} \right) = (1 - T'(z)) n$$

Totally differentiating wrt $(1 - T'(z)) n$, we have:

$$\frac{dl}{d(1 - T'(z)) n} v''(l) = 1$$

Which implies the following elasticity:

$$\varepsilon = \frac{dl}{d(1 - T'(z)) n} \frac{(1 - T'(z)) n}{l} = \frac{v'(z)}{lv''(z)}$$

Resource Constraint Suppose the government has an exogenous revenue requirement E . The revenues collected through taxation must be at least equal to E . Using the agent's budget constraint we can write the tax levied on a single individual as $T(z(n)) = z(n) - c(n)$. Summing over all the individuals in the economy we get:

$$\int_{\underline{n}}^{\bar{n}} c(n) f(n) dn \geq \int_{\underline{n}}^{\bar{n}} z(n) f(n) dn - E$$

This is the resource constraint for this economy. Notice that unlike incentive constraint, this constraint is unique.

3.2 Optimal Income Tax

We now solve the constrained maximization problem using optimal control theory. Instead of having taxes as a choice variable, we assume that the government chooses an allocation for each agent. Given the individual's budget constraint, this is equivalent to choosing a tax level. The government problem is:

$$\begin{aligned} & \max_{c(n), u(n), z(n)} \int_{\underline{n}}^{\bar{n}} G(u(n)) f(n) \\ \text{s.t.} & \frac{du(n)}{dn} = \frac{z(n)}{n^2} v' \left(\frac{z(n)}{n} \right) \\ & \int_{\underline{n}}^{\bar{n}} c(n) f(n) dn \geq \int_{\underline{n}}^{\bar{n}} z(n) f(n) dn - E \end{aligned}$$

We solve the problem with a Hamiltonian where we interpret n as the continuous variable and choose $u(n)$ as state variable and $z(n)$ as control. The incentive constraint becomes the law of motion of the state variable: it measures how utility changes across types in equilibrium. In order to setup the Hamiltonian, we need to replace consumption in the resource constraint with a function of state and control variables. Using the definition of indirect utility, we can write $c(n) = u(n) + v(z(n)/n)$. We replace this condition into the resource constraint and setup the following Hamiltonian:

$$H = \left[G(u(n)) + \lambda \left(z(n) - u(n) - v \left(\frac{z(n)}{n} \right) \right) \right] f(n) + \mu(n) \frac{z(n)}{n^2} v' \left(\frac{z(n)}{n} \right)$$

$\mu(n)$ denotes the multiplier on the incentive constraint of type n and λ is the multiplier on the resource constraint.

The first order conditions of the optimal control problem are:

$$\frac{\partial H}{\partial z(n)} = \lambda \left[1 - \frac{v'(z(n)/n)}{n} \right] f(n) + \frac{\mu(n)}{n^2} \left[v' \left(\frac{z(n)}{n} \right) + \frac{z(n)}{n} v'' \left(\frac{z(n)}{n} \right) \right] = 0 \quad (1)$$

$$\frac{\partial H}{\partial u(n)} = [G'(u(n)) - \lambda] f(n) = -\mu'(n) \quad (2)$$

The transversality (boundary) conditions read:

$$\mu(\underline{n}) = \mu(\bar{n}) = 0$$

The Hamiltonian solution requires $\mu(\bar{n}) u(\bar{n}) = 0$. However, if we want to provide positive utility to type \bar{n} we must require $\mu(\bar{n}) = 0$. At the same time, since at the optimum the incentive constraints will be binding downwards, we require $\mu(\underline{n}) = 0$. As it is standard in this kind of problems the lowest type does not want to imitate any other agent in equilibrium implying that her incentive constraint is slack, while everyone else is indifferent between her allocation and the allocation provided to the immediately lower type.

If we integrate equation (2) over the entire type space and use transversality conditions we find:

$$\lambda = \int_{\underline{n}}^{\bar{n}} G'(u(n)) f(n) dn$$

This is an expression for the marginal value of public funds to the government. It states that the value of public funds depends on the marginal social welfare gains across the entire type space and

it is equal to the welfare effect of transferring \$1 to every individual in the economy. In other words, public funds are more valuable the higher are the social welfare gains achievable in the economy.

We can also integrate equation (2) to find the value of $\mu(n)$:

$$-\mu(n) = \int_n^{\bar{n}} [\lambda - G'(u(m))] f(m) dm \quad (3)$$

Using the definition of labor elasticity, we rearrange the following:

$$\left[v' \left(\frac{z(n)}{n} \right) + \frac{z(n)}{n} v'' \left(\frac{z(n)}{n} \right) \right] = v' \left(\frac{z(n)}{n} \right) \left[1 + \frac{1}{\epsilon} \right]$$

Exploiting the definition of the tax wedge, we simplify equation (1) to get:

$$\lambda T'(z(n)) = \frac{\mu(n)}{f(n)} (1 - T'(z(n))) \left(1 + \frac{1}{\epsilon} \right)$$

Using the expression for μ derived in equation (3) we get:

$$\frac{T'(z(n))}{1 - T'(z(n))} = \left(\frac{1 + \epsilon}{\epsilon} \right) \frac{\int_n^{\bar{n}} [1 - g(m)] f(m) dm}{nf(n)} \quad (4)$$

The optimal tax is decreasing in the elasticity of labor supply. We define $g(n) = \frac{G'(u(n))}{\lambda}$ the relative social welfare weight of individual n such that $\int_n^{\bar{n}} g(n) f(n) dn = 1$. Remember that λ aggregates the social welfare weights across the entire economy. Thus, a higher $g(n)$ means that the government cares relatively more about individual n and will tax her less.

A Rawlsian government would have $g(n) = 0$ for any $n > \underline{n}$ and the formula would reduce to:

$$\frac{T'(z(n))}{1 - T'(z(n))} = \left(\frac{1 + \epsilon}{\epsilon} \right) \frac{1 - F(n)}{nf(n)}$$

The second part of the expression captures the ratio of the mass above type n and the density at n . It is a measure of thickness and the lower it is the higher marginal tax rate will be.

3.3 Diamond ABC Formula

In this paragraph we derive a tax formula presented in Diamond (1998). We change our assumption about welfare weights and assume that they are distributed according to a function $\psi(n)$ with cdf $\Psi(n)$. The government objective function becomes:

$$\int_n^{\bar{n}} u(n) \psi(n) dn$$

By assumption $\int_n^{\bar{n}} \psi(n) dn = 1$ implies $\lambda = 1$. First order conditions can be derived exactly as before. We therefore have:

$$-\mu'(n) = \psi(n) - \lambda f(n)$$

and after integration:

$$\begin{aligned} -\mu(n) &= \int_n^{\bar{n}} (f(n) - \psi(n)) dn \\ &= \Psi(n) - F(n) \end{aligned}$$

Using the expression above the tax formula reads:

$$\frac{T'(z(n))}{1 - T'(z(n))} = \left(\frac{1 + \epsilon}{\epsilon} \right) \frac{\Psi(n) - F(n)}{nf(n)}$$

To write the ABC formula we divide and multiply by $1 - F(n)$ to get:

$$\frac{T'(z(n))}{1 - T'(z(n))} = \underbrace{\left(\frac{1 + \epsilon}{\epsilon} \right)}_{A(n)} \underbrace{\frac{\Psi(n) - F(n)}{1 - F(n)}}_{B(n)} \underbrace{\frac{1 - F(n)}{nf(n)}}_{C(n)}$$

$A(n)$ captures the standard elasticity and efficiency argument. $B(n)$ measures the desire for redistribution: if the sum of weights below n is high relative to the mass above n , the government will tax more. Finally, $C(n)$ measures the thickness of the right tail of the distribution. A thicker tail will be associated to higher tax rates.

Notice that in the Rawlsian case $\Psi(n) = 1$ for every $n > \underline{n}$ and the formula converges to the one presented in the previous paragraph.

3.4 Optimal Taxes With Income Effects

We now relax the assumption of no income effects. Suppose the utility of the agent takes the form $\tilde{u}(c, l) = u(c) - v(l)$ where $u'(c) > 0$ and $u''(c) \leq 0$.

Elasticity of Labor Supply The optimality condition for the labor supply choice becomes:

$$\frac{v'(l)}{u'(c)} = (1 - T'(z))n$$

The uncompensated response of labor supply to the net of tax wage is:

$$\frac{\partial l^u}{\partial (1 - T'(z))n} = \frac{u'(c) + l(1 - T'(z))nu''(c)}{v''(l) - (1 - T'(z))^2 n^2 u''(c)}$$

implying the following uncompensated elasticity:

$$\varepsilon^u = \frac{u'(c)/l + \frac{v'(l)^2}{u'(c)^2} u''(c)}{v''(l) - \frac{v'(l)^2}{u'(c)^2} u''(c)}$$

The response of labor to income changes is given by:

$$\frac{\partial l}{\partial I} = \frac{(1 - T'(z))nu''(c)}{v''(l) - (1 - T'(z))^2 n^2 u''(c)}$$

Using the Slutsky equation (as we did in Section notes 1):

$$\begin{aligned} \frac{\partial l^c}{\partial (1 - T'(z))n} &= \frac{u'(c) + l(1 - T'(z))nu''(c)}{v''(l) - (1 - T'(z))^2 n^2 u''(c)} - \frac{l(1 - T'(z))nu''(c)}{v''(l) - (1 - T'(z))^2 n^2 u''(c)} \\ &= \frac{u'(c)}{v''(l) - (1 - T'(z))^2 n^2 u''(c)} \end{aligned}$$

Therefore:

$$\varepsilon^c = \frac{v'(l)/l}{v''(l) - (1 - T'(z))^2 n^2 u''(c)}$$

Optimal Tax Everything is similar to the previous case except for the fact that now we cannot replace the variable $c(n)$ in the resource constraint using the definition of indirect utility. We will define consumption as an expenditure function $\tilde{c}(\tilde{u}(n), z(n), n)$ and implicitly differentiate it wrt to $\tilde{u}(n)$ and $z(n)$. Start from the definition of indirect utility:

$$\tilde{u}(n) = u(\tilde{c}(n)) - v(z^*(n)/n)$$

It follows that the following two conditions will hold at the optimum:

$$d\tilde{u}(n) = u'(\tilde{c}(n)) d\tilde{c}(n)$$

$$0 = u'(\tilde{c}(n)) d\tilde{c}(n) - \frac{1}{n} v'(z^*(n)/n) dz^*(n)$$

Rearranging:

$$\frac{d\tilde{c}(n)}{d\tilde{u}(n)} = \frac{1}{u'(\tilde{c}(n))}$$

$$\frac{d\tilde{c}(n)}{dz^*(n)} = \frac{v'(z^*(n)/n)}{nu'(\tilde{c}(n))}$$

The Hamiltonian for the problem is:

$$H = [G(u(n)) + \lambda(z(n) - \tilde{c}(\tilde{u}(n), z(n), n))] f(n) + \mu(n) \frac{z(n)}{n^2} v'\left(\frac{z(n)}{n}\right)$$

and FOCs are:

$$\frac{\partial H}{\partial z(n)} = \lambda \left[1 - \frac{v'(z(n)/n)}{nu'(c(n))} \right] f(n) + \frac{\mu(n)}{n^2} \left[v'\left(\frac{z(n)}{n}\right) + \frac{z(n)}{n} v''\left(\frac{z(n)}{n}\right) \right] = 0$$

$$\frac{\partial H}{\partial u(n)} = \left[G'(u(n)) - \frac{\lambda}{u'(c^*(n))} \right] f(n) = -\mu'(n)$$

In order to find the equilibrium value of the multiplier, we can integrate the second FOC:

$$\mu(n) = \int_n^{\bar{n}} \left[G'(u(m)) - \frac{\lambda}{u'(c(m))} \right] f(m) dm$$

We exploit the definition of the two elasticities to write:

$$\begin{aligned} \left[v'\left(\frac{z(n)}{n}\right) + \frac{z(n)}{n} v''\left(\frac{z(n)}{n}\right) \right] &= v'\left(\frac{z(n)}{n}\right) \left[1 + \frac{z(n)}{n} \frac{v''\left(\frac{z(n)}{n}\right)}{v'\left(\frac{z(n)}{n}\right)} \right] \\ &= v'\left(\frac{z(n)}{n}\right) \left(\frac{1 + \varepsilon^u}{\varepsilon^c} \right) \end{aligned}$$

The optimal tax formula will then become:

$$\frac{T'(z(n))}{1 - T'(z(n))} = \left(\frac{1 + \varepsilon^u}{\varepsilon^c} \right) \frac{\eta(n)}{nf(n)} \quad (5)$$

where $\eta(n) = \frac{u'(c(n))\mu(n)}{\lambda}$.

3.5 Pareto Efficient Taxes

We now ask the question of whether a tax system $T_0(z)$ in place is Pareto-optimal, meaning that there exists no feasible adjustment in the tax schedule such that all individuals in the economy are weakly better off.

We can characterize the Pareto frontier of the previous problem by solving the following:

$$\begin{aligned} & \max \int_{\underline{n}}^{\bar{n}} u(n) \psi(n) dn \\ & \text{s.t.} \\ & u(c(n)) - h(z(n)/n) \geq u(c(n')) - h(z(n')/n) \quad \forall n, n' \\ & \int_{\underline{n}}^{\bar{n}} [z(n) - c(n)] f(n) dn \geq E \end{aligned}$$

By varying the social marginal welfare weights, we can trace out every point on the Pareto frontier. However, there might be points on the Pareto frontier that can be improved upon increasing the utility of all the agents in the economy.

Werning (2007) develops a test for the Pareto optimality of a tax schedule. The first important result of the paper is the following:

Proposition 1: A tax code fails to be constrained Pareto optimal *if and only if* there exists a feasible tax reform that (weakly) reduces taxes at all incomes.³

Proof: (*if*) suppose we weakly reduce taxes all over the entire economy, then every individual is at least as well off.

(*only if*) suppose there exists a Pareto improving feasible tax reform $T_1(z)$. Then we have:

$$\begin{aligned} U(z_1(n) - T_1(z_1(n)), z_1(n), n) & \geq U(z_0(n) - T_0(z_0(n)), z_0(n), n) \\ & \geq U(z_1(n) - T_0(z_1(n)), z_1(n), n) \end{aligned}$$

where the first inequality comes from the assumption of Pareto-improvement and the second from the assumption that under $T_0(z)$ the agent truthfully reveals her type and chooses $z_0(n)$. The chain of inequalities implies that $T_1(z_1(n)) \leq T_0(z_1(n))$ for every n .

Proposition 1 implies that since the resource constraint is satisfied and both tax systems raise revenues at least equal to E , a Pareto improvement can only occur through a tax reduction that does not generate a drop in revenues. This can be interpreted as a Laffer effect: although the government lowers taxes, the behavioral response (increase in labor supply) is strong enough to more than compensate the revenue loss.

3.6 A Test of the Pareto Optimality of the Tax Schedule

In order to implement the test, Werning takes a dual approach to the optimal taxation problem that we studied in the previous paragraphs. We rewrite the problem such that instead of maximizing the social welfare function, the government maximizes the resources to provide a minimum level $\bar{v}(n)$ of indirect utility to every agent in the economy. We write the problem as follows:

$$\max_{u(n), z(n)} \int_{\underline{n}}^{\bar{n}} (z(n) - \tilde{c}(v(n), z(n), n)) f(n) dn$$

³Feasible means that it satisfies the resource constraint.

s.t.

$$\frac{dv(n)}{dn} = U_n(\tilde{c}(v(n), z(n), n), z(n), n) \quad (6)$$

$$v(n) \geq \bar{v}(n) \quad \forall n \quad (7)$$

Notice that $\tilde{c}(v(n), z(n), n)$ is the expenditure function that we introduced to study optimal taxes with income effects. The problem would also have a monotonicity constraint that we relax for the moment, as we usually do. Notice that by changing the levels of $\bar{v}(n)$ we can characterize the entire Pareto frontier.⁴

We solve the problem with a Lagrangian by attaching multiplier $\psi(n) f(n)$ to (7) and $\mu(n)$ to the local incentive constraint:

$$\begin{aligned} L = & \int_{\underline{n}}^{\bar{n}} (z(n) - \tilde{c}(v(n), z(n), n)) f(n) dn + \int_{\underline{n}}^{\bar{n}} \psi(n) v(n) f(n) dn \\ & + \int_{\underline{n}}^{\bar{n}} \mu(n) v'(n) dn - \int_{\underline{n}}^{\bar{n}} \mu(n) U_n(\tilde{c}(v(n), z(n), n), z(n), n) dn \end{aligned}$$

Notice that this is identical to the Lagrangian that we would obtain in a classical optimal tax problem with welfare weights $\psi(n) f(n)$.⁵ We integrate $\int_{\underline{n}}^{\bar{n}} \mu(n) u'(n) dn$ by parts to obtain:

$$\int_{\underline{n}}^{\bar{n}} \mu(n) v'(n) dn = \mu(\bar{n}) v(\bar{n}) - \mu(\underline{n}) v(\underline{n}) - \int_{\underline{n}}^{\bar{n}} \mu'(n) v(n) dn$$

and rewrite the Lagrangian as follows:

$$\begin{aligned} L = & \int_{\underline{n}}^{\bar{n}} (z(n) - \tilde{c}(v(n), z(n), n)) f(n) dn + \int_{\underline{n}}^{\bar{n}} \psi(n) v(n) f(n) dn \\ & + \mu(\bar{n}) v(\bar{n}) - \mu(\underline{n}) v(\underline{n}) - \int_{\underline{n}}^{\bar{n}} \mu'(n) v(n) dn - \int_{\underline{n}}^{\bar{n}} \mu(n) U_n(\tilde{c}(v(n), z(n), n), z(n), n) dn \end{aligned}$$

The FOC wrt $z(n)$ is:

$$\left(1 - \frac{d\tilde{c}(v(n), z(n), n)}{dz(n)}\right) f(n) - \mu(n) \left[U_{nc}(n) \frac{d\tilde{c}(v(n), z(n), n)}{dz(n)} + U_{nz}(n)\right] = 0 \quad (8)$$

and the FOC wrt $v(n)$ is:

⁴The first order conditions for this problem will be sufficient. We can rewrite the problem in terms of $\tilde{u}(n) = u(c(n))$ and $\tilde{h}(n) = h(z(n)/n)$ so that the objective function becomes $n\tilde{h}^{-1}(\tilde{h}(n)) - u^{-1}(\tilde{u}(n))$ and is concave in $\tilde{u}(n)$ and $\tilde{h}(n)$ that become the new control and state variables.

⁵The problem would be:

$$\int_{\underline{n}}^{\bar{n}} v(n) \psi(n) f(n) dn$$

s.t.

$$\frac{dv(n)}{dn} = U_n(\tilde{c}(v(n), z(n), n), z(n), n)$$

$$\int_{\underline{n}}^{\bar{n}} [z(n) - \tilde{c}(v(n), z(n), n)] f(n) dn \geq E$$

$$-\frac{d\tilde{c}(v(n), z(n), n)}{dv(n)} f(n) - \mu'(n) - \mu(n) U_{nc}(n) \frac{d\tilde{c}(v(n), z(n), n)}{dv(n)} + \psi(n) f(n) = 0 \quad (9)$$

We know from the paragraph about optimal taxation with income effects that we can write:

$$\frac{d\tilde{c}(v(n), z(n), n)}{dz(n)} = -\frac{U_z(n)}{U_c(n)} = MRS(n)$$

Also, since $T'(n) = 1 - MRS(n)$ (see the discussion about wedges) we have that:

$$1 - \frac{d\tilde{c}(v(n), z(n), n)}{dz(n)} = 1 - MRS(n) = T'(n)$$

The term in square brackets in equation (8) can be written as follows:

$$\begin{aligned} U_{nc}(n) \frac{d\tilde{c}(v(n), z(n), n)}{dz(n)} + U_{nz}(n) &= -U_{nc}(n) \frac{U_z(n)}{U_c(n)} + U_{nz}(n) \\ &= U_c(n) \frac{U_{zn}(n) U_c(n) - U_{cn}(n) U_z(n)}{U_c(n)^2} \\ &= -U_c(n) \frac{\partial}{\partial n} \left[-\frac{U_z(n)}{U_c(n)} \right] \\ &= -U_c(n) \frac{\partial MRS(n)}{\partial n} \end{aligned}$$

Therefore, equation (8) becomes:

$$T'(n) f(n) = -\mu(n) U_c(n) \frac{\partial MRS(n)}{\partial n}$$

Using $MRS(n) = 1 - T'(n)$, we can rewrite the condition as:

$$\frac{T'(n)}{1 - T'(n)} f(n) = -\mu(n) U_c(n) \frac{\partial \log MRS(n)}{\partial n}$$

Now, we move to the second FOC. We know from before that:

$$\frac{d\tilde{c}(v(n), z(n), n)}{dv(n)} = \frac{1}{U_c(n)}$$

We can therefore rewrite (9) as:

$$-\frac{f(n)}{U_c(n)} - \mu'(n) - \mu(n) \frac{U_{nc}(n)}{U_c(n)} + \psi(n) f(n) = 0$$

Any Pareto Efficient allocation must satisfy (7) and provide at least utility $v(\bar{n})$ to every agent n . By the complementarity-slackness condition, this is equivalent to ask that $\psi(n) f(n) \geq 0$, which is that the multipliers associated to the constraints are never negative. We rewrite the FOC imposing the following inequality:

$$-U_c \mu'(n) - \mu(n) U_{nc}(n) \leq f(n) \quad (10)$$

If we change variables and define $\hat{\mu}(n) \equiv U_c(n) \mu(n)$, we have:

$$\hat{\mu}'(n) = U_c(n) \mu'(n) + \mu(n) [U_{cn}(n) + U_{cc}(n) c'(n) + U_{cz}(n) z'(n)]$$

Substituting into (10) we find:

$$-\hat{\mu}'(n) + \hat{\mu}(n) \frac{U_{cc}(n) c'(n) + U_{cz}(n)}{U_c(n)} \leq f(n)$$

The local incentive constraint of the agent (FOC for optimal reporting) implies that $c'(n)/z'(n) = -U_z(n)/U_c(n)$. It follows that:

$$\begin{aligned} \frac{U_{cc}(n) c'(n) + U_{cz}(n)}{U_c(n)} &= \frac{U_{cc}(n) \frac{c'(n)}{z'(n)} + U_{cz}(n)}{U_c(n)} z'(n) \\ &= \frac{-U_{cc}(n) U_z(n) + U_{cz}(n) U_c(n)}{U_c(n)^2} z'(n) \\ &= -\frac{\partial}{\partial c} \left[-\frac{U_z(n)}{U_c(n)} \right] z'(n) \\ &= -\frac{\partial MRS(n)}{\partial c} z'(n) \end{aligned}$$

We finally establish two conditions for Pareto efficiency:

$$\frac{T'(n)}{1 + T'(n)} f(n) = -\hat{\mu}(n) \frac{\partial \log MRS(n)}{\partial n} \quad (11)$$

$$-\hat{\mu}'(n) - \hat{\mu}(n) \frac{\partial MRS(n)}{\partial c} z'(n) \leq f(n) \quad (12)$$

Using the tax schedule in place, the preferences and the skill distribution we can derive the $\hat{\mu}$ from equation (11). We can then use equation (12) to test for the Pareto efficiency of the tax schedule.

Applying the Test and Interpreting the Conditions Suppose the agent has preferences $U(c, z, n) = c - \frac{1}{\gamma} (z/n)^\gamma$ with elasticity $\varepsilon = 1/(\gamma - 1)$. The FOCs of the dual problem read:

$$\mu'(n) = \psi(n) f(n) - f(n) \quad (13)$$

$$\frac{T'(n)}{1 - T'(n)} f(n) = \frac{\mu(n)}{n} \frac{1 + \varepsilon}{\varepsilon} \quad (14)$$

The tax schedule is Pareto-optimal if and only if $\psi(n) f(n) \geq 0$, which implies $-\mu'(n) \leq f(n)$. This inequality is the same as the one derived in (12) since $U_c(n) = 1$ and $\partial MRS(n)/\partial c = 0$. Equation (14) is the same as (11) when we notice that $\partial MRS(n)/\partial n = (1 + 1/\varepsilon)/n$.⁶

Suppose that the marginal tax rate is linear and equal to τ , when we put the two conditions together we get:

$$-\frac{\varepsilon}{1 + \varepsilon} \frac{\partial}{\partial n} \left[\frac{\tau}{1 - \tau} n f(n) \right] = -\mu'(n) \leq f(n) \quad \forall n$$

Taking the derivative wrt n the condition becomes:

⁶Notice that $MRS(n) = z(n)^{\gamma-1} n^{-\gamma}$
Therefore:

$$\frac{\partial \log MRS(n)}{\partial n} = \frac{\partial MRS(n)}{\partial n} \frac{1}{MRS(n)} = -\frac{\gamma}{n}$$

where $\gamma = 1 + 1/\varepsilon$.

$$-\frac{\varepsilon}{1+\varepsilon}\frac{\tau}{1-\tau}\left[-1-\frac{nf'(n)}{f(n)}\right]\leq 1 \quad \forall n \quad (15)$$

First, the condition in (15) shows that for any τ and ε there exists a set of $f(n)$ such that τ is Pareto efficient and a set of $f(n)$ such that it is not Pareto efficient. At the same time, for any ε and $f(n)$ we can find flat tax schedules τ that are efficient and set of τ s that are inefficient. It follows that it is crucial to know the distribution of skills. The test can also be written in terms of income distributions that are easier to infer from the data. Higher ε makes the condition harder to be satisfied: when individuals are reacting more to changes in taxes, a tax reduction is more likely to lead to a Pareto improvement. When taxes are locally lowered at some n , the individuals below n will tend to increase their labor supply and individuals above will reduce the labor supply. The term $nf'(n)/f(n)$ measures the elasticity of the skill distribution and captures how fast the skill distribution is decreasing at some n . Highly negative elasticity of skill distribution at n means that the distribution decreases fast and that the mass of individuals below n is significantly larger than above n , implying a local Laffer effect from the increase in labor supply of individuals below n . In other words, by locally decreasing taxes at n the government can increase revenues by incentivizing the labor supply of the large mass of individuals below n . For this reason, when the elasticity of the skill distribution is highly negative the test is harder to pass.⁷

References

- Diamond, Peter (1998), “Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates”, *American Economic Review*, 88, 83–95.
- Salani , Bernard (2002), “The Economics of Taxation”, The MIT Press
- Scheuer, Florian (2016), Lecture Notes
- Werning, Ivan (2007), “Pareto Efficient Income Taxation”, Working Paper

⁷In the second part of the sequence you will study how to run the same test in the inequality deflator framework using the concept of fiscal externality.

4 Section 5: Optimal Taxation with Income Effects and Bunching

In this section we will see the derivation of optimal income taxes proposed by Saez (2001). We will also introduce a way to estimate the elasticity of reported income that exploits the degree of bunching at the kinks of the existing tax schedule (Saez, 2010).⁸

4.1 Optimal Taxes with Income Effects

In this paragraph we derive optimal income taxes with income effects following the experiment in Saez (2001). Suppose individuals in the economy are heterogeneous in ability n and work l_n hours earning income z_n . We can write the labor supply as a Walrasian (uncompensated) demand such that $l_n(w_n, R_n)$ where $w_n = n(1 - T')$ is the net of tax wage that the agent earns in equilibrium and R_n is the virtual (non-labor) income. Define R_n under the assumption that the tax is linear and tangent to the tax schedule at z_n such that $c = (1 - T')nl_n + R_n$. Using the fact that $c = nl_n - T(nl_n)$ we can write $R_n = nl_n - T(nl_n) - nl_n(1 - T')$. Using the Walrasian demand we can define uncompensated elasticity and the income parameter as follows:

$$\zeta_n^u = \frac{\partial l_n}{\partial (1 - T')} \frac{(1 - T')}{l_n} \quad (16)$$

$$\eta_n = \frac{\partial l_n}{\partial R_n} (1 - T') \quad (17)$$

Using the definitions above and the Slutsky equation, we can write the compensated elasticity as:

$$\zeta_n^c = \zeta_n^u - \eta_n \quad (18)$$

We now derive a result that will be useful later. Totally differentiating l_n we get:

$$\dot{l}_n = \frac{\partial l_n}{\partial w_n} \left[1 - T' - n(n\dot{l}_n + l_n)T'' \right] + \frac{\partial l_n}{\partial R_n} (n\dot{l}_n + l_n) (nl_n T'')$$

rearranging

$$\dot{l}_n = \frac{w_n}{l_n} \frac{\partial l_n}{\partial w_n} \frac{l_n}{n} + \left[w_n \frac{\partial l_n}{\partial R_n} - \frac{w_n}{l_n} \frac{\partial l_n}{\partial w_n} \right] \frac{l_n T''}{(1 - T')} [l_n + n\dot{l}_n]$$

Notice that $\dot{z}_n = l_n + n\dot{l}_n$ and apply the definitions in (16) and (17) to get:

$$\dot{l}_n = \zeta_n^u \frac{l_n}{n} - \dot{z}_n \frac{l_n T''}{1 - T'} \zeta_n^c \quad (19)$$

Using (19) we can write:

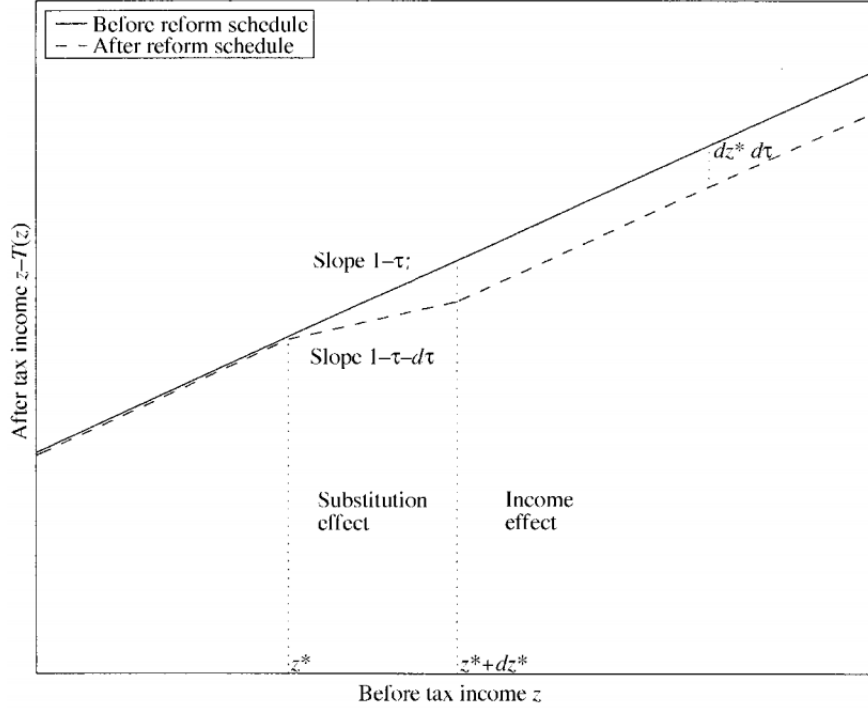
$$\frac{\dot{z}_n}{z_n} = \frac{l_n + n\dot{l}_n}{nl_n} = \frac{1 + \zeta_n^u}{n} - \dot{z}_n \frac{T''}{1 - T'} \zeta_n^c \quad (20)$$

In order to derive the optimal tax, we follow the experiment in Saez (2001). Suppose we introduce a perturbation around the optimal tax schedule such that we raise the marginal tax rate by $d\tau$ in a small interval $[z^*, z^* + dz^*]$.⁹ As we have already seen in previous Sections, the tax has three main effects: mechanical, welfare and behavioral. While the mechanical and welfare effects are similar to the ones we have previously studied, the behavioral effect now consists of two components: an elasticity effect for people in the interval $[z^*, z^* + dz^*]$ and an income effect for taxpayers above z^* .

⁸The bunching paragraph is based on previous notes by Simon Jager.

⁹The experiment also assumes that $d\tau$ is second order relative to dz^* to avoid any bunching at kink.

Figure 1: Tax Reform Experiment



Mechanical and Welfare Effects Every taxpayer above z^* will pay an extra $dz^* d\tau$ of taxes, which for welfare purposes are valued according to the social marginal welfare weights $g(z)$. The net-of-welfare mechanical effect for pre-tax income z is $(1 - g(z)) dz^* d\tau$. Summing up over all incomes above z^* we get:

$$M = dz^* d\tau \int_{z^*}^{\infty} (1 - g(z)) h(z) dz$$

Social marginal welfare weights $g(z)$ represent the value for the government of giving one dollar to some level of income z . In particular, the government is indifferent between giving $1/g(z_1)$ dollars to individual 1 and $1/g(z_2)$ dollars to individual 2. The social marginal welfare weights are expressed in government money. Going back to the fully specified problem, we can interpret the weights as being normalized by λ , the multiplier on the resource constraint. The λ measures the value of transferring one dollar to every individual in the economy and captures the value of government funds. Higher λ means that the government can significantly raise welfare by transferring money to the individuals in the economy, a low λ on the other hand implies that the gain from transfers is low and public funds are less valuable.

Elasticity Effect The increase in the marginal tax rate has an effect on individuals' labor supply that is denoted by dz . The effect consists of two parts. First, there is the direct consequence of the increase in taxes that depends on the compensated elasticity of labor supply. Second, since the taxpayer changes her labor supply by dz shifting on the tax schedule, she will face an additional change in the tax. We write the change in the marginal tax rate induced by the shift dz as $dT' = T'' dz$. The behavioral response is proportional to the total tax change $d\tau + dT'$:

$$dz = \frac{dz^c}{d(1 - T')} (d\tau + dT') = -\zeta^c z^* \frac{d\tau + dT'}{1 - T'}$$

Rearranging:

$$dz = -\zeta^c z^* \frac{d\tau}{1 - T' + \zeta^c z^* T''} \quad (21)$$

Income Effect All individuals above z^* face a parallel shift of the tax schedule and pay additional taxes for $dz^* d\tau$. The mechanical increase in taxes paid has a direct income effect that depends on the income parameter $\eta = (dz/dR)(1 - T')$. Moreover, since the individual shifts along the tax schedule we must take into account a further change in the tax rate. The two effects combined are:

$$dz = -\zeta^c z \frac{T'' dz}{1 - T'} - \eta \frac{d\tau dz^*}{1 - T'}$$

rearranging

$$dz = -\eta \frac{d\tau dz^*}{1 - T' - \zeta^c T''}$$

In order to compute the total revenue effect we then need to sum over all taxpayers above z^* and account for the marginal tax rate T' :

$$- \int_{z^*}^{\infty} \eta \frac{d\tau dz^*}{1 - T' - \zeta^c T''} T' h(z) dz \quad (22)$$

Virtual and Actual Income Density Saez (2001) introduces the concept of virtual density in order to simplify the tax formulas. The virtual density is closely related to the virtual income in that it is the income density that would arise if the tax system was linear and tangent to the tax schedule $T(z)$ at every z . We denote the virtual density with $h^*(z)$. The mapping between virtual density and the type distribution $f(n)$ is given by the following:

$$h^*(z) \dot{z}^* = f(n)$$

where \dot{z}^* is the derivative of earnings wrt n when the linear tax schedule is in place. A similar relation holds for $h(z)$ and we have $h(z) \dot{z} = f(n)$. Using the definition in (27) and the fact that $T'' = 0$ when a linear tax schedule is in place we can write:

$$h(z_n) \frac{1 - T'}{1 - T' + z_n \zeta_n^c T''} \frac{1 + \zeta_n^u}{n} z_n = h^*(z_n) \frac{1 + \zeta_n^u}{n} z_n$$

It follows that:

$$\frac{h^*(z)}{1 - T'(z)} = \frac{h(z)}{1 - T'(z) + \zeta^c z T''(z)} \quad (23)$$

Optimal Income Taxes Starting from (21) and using equation (23) we can write the elasticity effect as follows:

$$E = -h(z) dz^* T' \zeta^c z^* \frac{d\tau}{1 - T' + \zeta^c z^* T''} = -\zeta^c z^* \frac{T'}{1 - T'} h^*(z) d\tau dz^* \quad (24)$$

Notice that in order to get the revenue effect that the elasticity effect should measure, we multiplied the expression in (21) by the marginal tax rate T' and by $h(z) dz^*$, which is the share of taxpayers affected by the tax reform. Using (23) we can write the income effect as follows:

$$I = -d\tau dz^* \int_{z^*}^{\infty} \eta \frac{T'}{1 - T'} h^*(z) dz \quad (25)$$

At the optimum, the sum of the three effects must be zero. We thus impose:

$$M + E + I = 0$$

and find

$$dz^* d\tau \int_{z^*}^{\infty} (1 - g(z)) h(z) dz - \zeta^c z^* \frac{T'}{1 - T'} h^*(z) d\tau dz^* - d\tau dz^* \int_{z^*}^{\infty} \eta \frac{T'}{1 - T'} h^*(z) dz = 0$$

Rearranging:

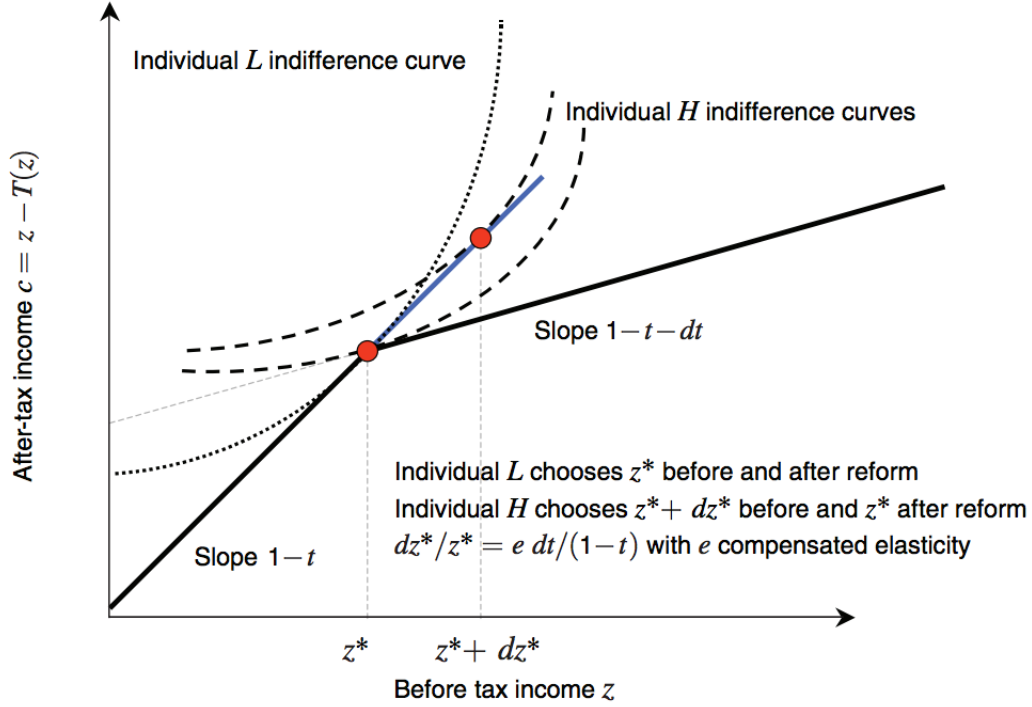
$$\begin{aligned} \frac{T'}{1 - T'} &= \frac{1}{\zeta^c} \frac{1}{h^*(z) z^*} \left[\int_{z^*}^{\infty} (1 - g(z)) h(z) dz - \int_{z^*}^{\infty} \eta \frac{T'}{1 - T'} h^*(z) dz \right] \\ &= \frac{1}{\zeta^c} \frac{1 - H(z^*)}{h^*(z) z^*} \left[\int_{z^*}^{\infty} \frac{h(z)}{1 - H(z^*)} (1 - g(z)) dz - \int_{z^*}^{\infty} \eta \frac{T'}{1 - T'} \frac{h^*(z)}{1 - H(z^*)} dz \right] \end{aligned} \quad (26)$$

As we have already seen in the previous Sections the formula consists of different terms. $1 - H(z^*)/h^*(z) z^*$ captures the shape of the income distribution and measures how many people are above z^* relative to how much income is accumulated at z^* (i.e. $z^* h(z^*)$). The former is proportional to the mechanical increase in revenues, while the second measures the total income that is distorted by the tax. The marginal tax is also decreasing in the compensated elasticity ζ^c following a classical efficiency argument, and increases the larger the income effect is (in absolute value): a stronger income effect means that the negative fiscal externality from higher taxes is reduced.

4.2 Bunching Estimator

In this paragraph we will study a way to derive income elasticity that was introduced by Saez (2010). The methodology exploits the degree of bunching at the kinks that characterize the tax schedule. Suppose that individual incomes z are distributed according to a smooth density distribution $h(z)$. There is a constant marginal tax rate t at income z^* and a reform introduces an increase in marginal taxes such that it becomes $t + dt$ for all incomes above z^* . The kink will induce people that were falling in the interval $[z^*, z^* + dz^*]$ before the reform to bunch at z^* . Denote with L an individual who is exactly indifferent between the pre and post-reform tax schedule and does not change her income in equilibrium. This individual's indifference curve has slope $1 - t$ at z^* . There is also an individual H who represents the highest pre-reform income bunching at z^* . The indifference curve of H has slope $1 - t - dt$ at z^* and is tangent to the slope of the retention function above z^* .

Figure 2: Bunching



The response of income to the tax reform for individual H is:

$$dz^* = \frac{dz}{d(1-t)} \Big|_{z=z^H} d(1-t) = e \frac{z^H}{1-t} d(1-t) = -e (z^* + dz^*) \frac{dt}{1-t}$$

Where e is the compensated elasticity of income. Notice that the dz^* is proportional to the ratio between the change in the tax rate and the net-of-tax rate $1-t$. It follows that, everything else being equal, a change in marginal tax rates from 0 percent to 10 percent should produce the same amount of bunching as a change from 90 percent to 91 percent. Rearranging:

$$dz^* \left(\frac{1-t+edt}{1-t} \right) = -e \frac{z^*}{1-t} dt$$

which implies:

$$dz^* = -e \frac{z^*}{1-t+edt} dt \quad (27)$$

It is not surprising that dz^* is increasing in the elasticity of income, implying that if income is more elastic more people will bunch.

Suppose the income distribution is locally continuous, the share of people bunching at the kink is:

$$s(z^*) = h(z^*) dz^*$$

Using the definition in (27):

$$\frac{s(z^*)}{h(z^*)} = -e \frac{z^*}{1-t+edt} dt$$

$$-e \left(\frac{s(z^*)}{h(z^*)} + z^* \right) dt = \frac{s(z^*)}{h(z^*)} (1-t)$$

$$e = - \frac{s(z^*)}{s(z^*) + z^* s'(z^*)} \frac{1-t}{dt} \approx - \frac{s(z^*)}{h(z^*)} \frac{1-t}{z^*} \frac{1}{dt}$$

where we assumed that $s(z^*)/z^* \approx 0$. The formula shows that using the share of people bunching at kink and the income distribution that would arise under the no reform scenario we can estimate the elasticity of labor supply.

References

Saez, E. "Using Elasticities to Derive Optimal Income Tax Rates", *Review of Economics Studies*, Vol. 68, 2001, 205-229.

Saez, E. "Do Taxpayers Bunch at Kink Points?", *AEJ: Economic Policy*, Vol. 2, 2010, 180-212

5 Section 6: Optimal Income Transfers

In this Section we study the optimal design of income transfers. We start from a formal model where we specify individuals' preferences and a government's social welfare function. We then take the approach by Saez (2002) to derive optimal transfers using an "experiment" where we introduce a perturbation around the optimal tax schedule for a generic "occupation" and derive a formula for the optimal tax.

5.1 Optimal Income Transfers in a Formal Model

We introduce in this paragraph a model of discrete choices where we will derive optimal taxes. Suppose agents choose an occupation i among a set of occupations $\{1, 2, \dots, I\}$ and earn income w_i at occupation i . Each individual is indexed by $m \in M$ being a multidimensional set of measure one. The measure of individuals on M is denoted by $dv(m)$. The agents maximize $u^m(c_i, i)$ differentiable in consumption. Individual consumption after taxes is $c_i = w_i - T_i$. A tax schedule defines a vector (c_0, \dots, c_I) such that the set M will be partitioned in subsets M_1, M_2, \dots, M_I . Denote with $h_i(c_0, c_1, \dots, c_I)$ the fraction of individuals choosing occupation i such that $\sum_i h_i = 1$. h_i is differentiable under the assumption that tastes for work captured by $u^m(\cdot)$ are regularly distributed. We define the elasticity of participation for occupation i as follows:

$$\eta_i = \frac{c_i - c_0}{h_i} \frac{\partial h_i}{\partial (c_i - c_0)} \quad (28)$$

Suppose the government weights individual utilities through linear welfare weights μ^m and that the social welfare function is:

$$W = \int_M \mu^m u^m(w_{i^*} - T_{i^*}, i^*) dv(m) \quad (29)$$

The government has some revenue requirement H such that the budget can be written as:

$$\sum_i h_i T_i = H \quad (30)$$

We solve the problem with a Lagrangian where we attach multiplier λ to the government constraint. The FOC wrt T_i reads:

$$-\int_{M_i} \mu^m \frac{\partial u^m(c_{i^*}, i^*)}{\partial c_i} dv(m) + \lambda \left[h_i - \sum_{j=0}^I T_j \frac{\partial h_j}{\partial c_i} \right] = 0 \quad (31)$$

For the usual envelope argument equation (31) ignores the welfare effect of a change in c_i . A social marginal welfare weight is:

$$g_i = \frac{1}{\lambda h_i} \int_{M_i} \mu^m \frac{\partial u^m(c_{i^*}, i^*)}{\partial c_i} dv(m) \quad (32)$$

Using the definition of g_i we can rewrite (31) as:

$$(1 - g_i) h_i = \sum_{j=0}^I T_j \frac{\partial h_j}{\partial c_i} \quad (33)$$

This formula is very similar to the one you will see in the spring studying Ramsey taxation.¹⁰ Take

¹⁰The formula implies that the following is true for every i :

$$\sum_{j=0}^I \frac{T_j}{h_i (1 - g_i)} \frac{\partial h_j}{\partial c_i} = 1$$

a benchmark case of no income effects such that $h_j(c_0, \dots, c_I) = h_j(c_0 + R, \dots, c_I + R)$, the formula implies that $(1 - g_i) h_i = 0$. Summing over all i is:

$$\sum_i h_i g_i = \sum_i h_i = 1 \quad (34)$$

5.2 Optimal Tax/Transfer with Extensive Margin Only

Suppose each individual only chooses between some occupation i and being unemployed. This can be rationalized by a utility function where $u^m(c_j, j) = -\infty$ for any $j \neq i$. The assumption implies that $\partial h_i / \partial c_i + \partial h_0 / \partial c_i = 0$ and we can rewrite (33) as:

$$(1 - g_i) h_i = T_i \frac{\partial h_i}{\partial c_i} + T_0 \frac{\partial h_0}{\partial c_0} = (T_i - T_0) \frac{\partial h_i}{\partial c_i}$$

using the definition of the elasticity of participation:

$$\frac{T_i - T_0}{c_i - c_0} = \frac{1}{\eta_i} (1 - g_i) \quad (35)$$

The level of taxation at occupation i decreases in the elasticity of taxation for the usual efficiency argument.

Redistributive preferences imply $g_0 \geq g_1 \geq \dots \geq g_I$. Suppose there are no income effects, we know from (34) that the weighted average of the g_i s is 1 and therefore there is a i^* such that $g_j \leq 1$ for $j \leq i^*$ and $g_j > 1$ for $j > i^*$. This implies that $T_i - T_0 \leq 0$ for $i \leq i^*$, meaning that the government is providing a higher transfer to workers with low income relative to unemployed. Therefore, we established that it is optimal for the government to implement negative marginal tax rates at the bottom of the income distribution.

If the government was Rawlsian, we could have that g_0 only is higher than 1. When this is the case, the tax schedule does not display negative marginal tax rates and we have a classical negative income tax. On the other hand, a utilitarian government would have constant g_i s such that the budget constraint is satisfied. We therefore have two cases. First, if every individual can pay H , the government will charge a constant lump-sum tax equal to H to every taxpayer and $g_i = 1$ for every i . Second, if low incomes cannot afford the tax the government will only impose the tax on higher income setting their social marginal welfare weights below 1 and having positive marginal tax rates throughout occupations.

Tax Experiment The same formula for optimal taxes can be derived through the following experiment. Suppose taxes increase by dT_i for occupation i . The mechanical increase in tax revenues is $h_i dT_i$ and it will be valued $(1 - g_i) h_i dT_i$ by the government taking into account the welfare effect of the change. The government must also account for the fiscal externality generated by the behavioral response of agents in occupation i . Using the elasticity of participation, the share of people leaving occupation i is:

$$dh_i = -\eta_i \frac{h_i}{c_i - c_0} dT_i$$

Each worker leaving occupation i generates a loss in revenues equal to $T_i - T_0$. The total behavioral effect of the tax increase is:

$$dh_i (T_i - T_0) = -\eta_i h_i \frac{T_i - T_0}{c_i - c_0} dT_i$$

We can interpret the lhs as an index of how much labor supply is discouraged. The formula holds for every i and implies that discouragement is equalized across all occupations.

Summing the mechanical and behavioral effects at the optimum we get:

$$(1 - g_i) h_i dT_i - \eta_i h_i \frac{T_i - T_0}{c_i - c_0} dT_i = 0$$

Rearranging we can derive (35). The decomposition of the formula in mechanical and behavioral effects provides further intuition for why marginal tax rates can be negative at the optimum. For very low incomes the mechanical effect of providing an extra dollar is positive ($g_i > 1$) and at the same time a decrease in taxes at i provides incentives for unemployed workers to enter the labor force. The sum of the two effects is unambiguously positive.

5.3 Optimal Tax/Transfer with Intensive Margin Responses

Suppose that agents' preferences are such that they can only work in two adjacent occupations and that we can write the share of workers in occupation i as $h_i(c_{i+1} - c_i, c_i - c_{i-1})$ when we assume there are no income effects.¹¹ The behavioral elasticity is defined as follows:

$$\zeta_i = \frac{c_i - c_{i-1}}{h_i} \frac{\partial h_i}{\partial (c_i - c_{i-1})} \quad (36)$$

Equation (33) becomes:

$$(1 - g_i) h_i = -T_{i+1} \frac{\partial h_{i+1}}{\partial (c_{i+1} - c_i)} - T_i \frac{\partial h_i}{\partial (c_{i+1} - c_i)} + T_i \frac{\partial h_i}{\partial (c_i - c_{i-1})} + T_{i-1} \frac{\partial h_{i-1}}{\partial (c_i - c_{i-1})}$$

By assumption on agent's preferences $\partial h_{i+1} / \partial (c_{i+1} - c_i) = -\partial h_i / \partial (c_{i+1} - c_i)$ and rearranging we find:

$$(1 - g_i) h_i = -(T_{i+1} - T_i) \frac{\partial h_{i+1}}{\partial (c_{i+1} - c_i)} + (T_i - T_{i-1}) \frac{\partial h_i}{\partial (c_i - c_{i-1})}$$

Summing over $i, i+1, \dots, I$ and using the definition in (36) we can derive the optimal tax formula:

$$\frac{T_i - T_{i-1}}{c_i - c_{i-1}} = \frac{1}{\zeta_i} \left[\frac{(1 - g_i) h_i + (1 - g_{i+1}) h_{i+1} + \dots + (1 - g_I) h_I}{h_i} \right] \quad (37)$$

Non-increasing social marginal welfare weights imply that $(1 - g_i) h_i + (1 - g_{i+1}) h_{i+1} + \dots + (1 - g_I) h_I \geq 0$ for any $i > 0$. Thus, the tax T_i is increasing in i and it is not optimal to set negative marginal tax rates. Using (34), (37) and computing the formula for the tax rate at the bottom of the income distribution we get:

$$\frac{T_1 - T_0}{c_1 - c_0} = \frac{1}{\zeta_1} \left[\frac{(g_0 - 1) h_0}{h_1} \right] \quad (38)$$

A higher social marginal welfare weight g_0 implies a higher tax rate at the bottom. The reason is that if the government cares more about the unemployed individual it should set the lump-sum transfer $-T_0$ as large as possible by imposing large phasing-out tax rates at the bottom. Negative marginal tax rates at the bottom can still occur for $g_0 < 1$, but this would imply that the unemployed worker has a lower welfare weight than the average taxpayer in the economy, meaning that the government has unusual redistributive tastes.

¹¹We can write the share of people working in occupation i as $h_i(c_i, c_{i+1})$. No income effects imply $h(c_0, c_1, \dots, c_I) = h(c_0 + R, c_1 + R, \dots, c_I + R)$. It follows that $h(c_i, c_{i+1}) = h(c_i + c_i - c_{i-1}, c_i + c_{i+1} - c_i) = h(c_i - c_{i-1}, c_{i+1} - c_i)$.

Tax Experiment The formula in (37) can be derived through an experiment where taxes increase by dT for any occupation $i, i+1, \dots, i+I$. This change decreases $c_i - c_{i-1}$ by dT and leaves any other difference unaltered. The mechanical increase in revenues is $[h_i + h_{i+1} + \dots + h_I] dT$ and net-of-welfare it is valued $[h_i(1 - g_i) + h_{i+1}(1 - g_{i+1}) + \dots + h_I(1 - g_I)] dT$. The behavioral effect of the tax change arise from individuals in occupation i only when we assume income effects away. The impact on revenues is $dh_i = -h_i \zeta_i dT / (c_i - c_{i-1})$ and it must be scaled by the loss in revenues $T_i - T_{i-1}$ generated by each worker switching to occupation $i-1$. Summing the two impacts:

$$[h_i(1 - g_i) + h_{i+1}(1 - g_{i+1}) + \dots + h_I(1 - g_I)] dT - h_i \zeta_i (T_i - T_{i-1}) dT / (c_i - c_{i-1}) = 0$$

Rearranging we get the formula in (37). The mechanical and behavioral effects help providing intuition for why negative marginal tax rates are not optimal with intensive margin only. Suppose the government raised taxes at i when there is a negative marginal tax rate in the interval $[i-1, i]$. Individuals would respond by shifting their labor supply to $i-1$ and, given the higher tax rate, would pay more taxes. At the same time the tax change would mechanically increase revenues. Therefore, the government could always improve welfare by increasing taxes as long as the marginal tax rate is negative.

5.4 Optimal Tax/Transfer with Intensive and Extensive Margin Responses

We present for the sake of simplicity only the tax experiment derivation of the formula. Suppose taxes are raised by dT for everyone in occupation $i, i+1, \dots, i_I$. The mechanical effect is the same as the one observed in the previous paragraph. However, we have to add the participation effect of an increase in the tax for all the occupations above i . The share of people who become unemployed leaving a generic occupation i is $-h_i \eta_i dT / (c_i - c_0)$, generating a revenue loss equal to $-h_i \eta_i (T_i - T_0) dT / (c_i - c_0)$. Summing this effect over every occupation $j \geq i$ and setting the sum of behavioral and mechanical effects equal to 0, we can derive the following formula:

$$\frac{T_i - T_{i-1}}{c_i - c_{i-1}} = \frac{1}{\zeta_i h_i} \sum_{j=i}^I h_j \left[1 - g_j - \eta_j \frac{T_j - T_0}{c_j - c_0} \right] \quad (39)$$

When a tax is lowered in the pure extensive margin model, labor supply unambiguously increases. On the other hand, if a tax is decreased in a pure intensive margin model individuals will have incentives to lower their labor supply. The formula shows how to optimally trade-off the two effects.

Notice that (39) can be rewritten as (37) where we employ augmented social welfare weights $\hat{g}_i = g_i + \eta_i (T_i - T_0) / (c_i - c_0)$. When the participation elasticity is high enough, the augmented welfare weights are not necessarily decreasing in w_i if g_i s are. This explains why an earning income tax credit could be optimal in a mixed model.

References

- E. Saez "Optimal Income Transfer Programs: Intensive Versus Extensive Labor Supply Responses" Quarterly Journal of Economics, 117, 2002, 1039-1073
- E. Saez, P. Diamond. "The Case for a Progressive Tax: From Basic Research to Policy Recommendations", Journal of Economic Perspectives 25(4), Fall 2011, 165-190

6 Section 7: Optimal Top Income Taxation

In this Section we study the optimal design of top income taxes.¹² We have already covered optimal top income taxation in a simple Mirrlees framework in Section 2. Today, we will start from the “trickle down” model with endogenous wages introduced by Stiglitz (1982). We then analyze an example of optimal taxation in a general equilibrium model where workers choose between different occupations/sectors (Rothschild Scheuer, 2016). Finally, we present a model where top earners respond to taxes on three margins: labor supply, tax avoidance, and compensation bargaining (Piketty Saez Stantcheva, 2014).

6.1 Trickle Down: A Model With Endogenous Wages

Stiglitz (1982) studies a model with two unobservable types and endogenous wages. Suppose there are two types of workers: H (high skill) and L (low skill). For simplicity we assume they have equal mass and work l_i hours. The utility for a generic type i is $u(c_i, l_i)$. Work is the only input in the constant return to scale (CRS) production function $F(l_L, l_H)$. With competitive labor markets wages are equal to the marginal product of labor:

$$w_i = \frac{\partial F(l_L, l_H)}{\partial l_i} \quad (40)$$

The standard Mirrlees model implicitly assumes a linear production $F(l_L, l_H) = \theta_L l_L + \theta_H l_H$ where θ_i is the ability of agent i so that wages are $w_i = \theta_i$.

The resource constraint of this economy is:

$$\sum_i c_i \leq F(l_L, l_H)$$

The government assigns linear welfare weights ψ_L and ψ_H to the two types. If $\psi_H < \psi_L$ the government wants to redistribute to low types and we know that in equilibrium the incentive constraint for the high type is binding:

$$u(c_H, l_H) = u\left(c_L, \frac{w_L l_L}{w_H}\right) \quad (41)$$

We solve the problem with the following Lagrangian:

$$\begin{aligned} L = & \psi_L u(c_L, l_L) + \psi_H u(c_H, l_H) + \lambda \left[F(l_L, l_H) - \sum_i c_i \right] \\ & + \mu \left[u(c_H, l_H) - u\left(c_L, \frac{w_L l_L}{w_H}\right) \right] + \sum_i \eta_i (w_i - F_i(l_L, l_H)) \end{aligned} \quad (42)$$

λ is the marginal value of public funds, μ is the value of relaxing the incentive constraint for type H and η_i are the multipliers on the constraints in (40).

We derive the optimal marginal tax rate for the high type by optimally choosing c_H and l_H . The FOCs are respectively:

$$[\psi_H + \mu] u_c(c_H, l_H) = \lambda \quad (43)$$

$$[\psi_H + \mu] u_l(c_H, l_H) = -\lambda F_H(l_H, l_L) + \sum_i \eta_i F_{iH}(l_L, l_H) \quad (44)$$

¹²The first two paragraphs of this section are based on notes by Florian Scheuer.

The optimal labor supply choice implies the following labor wedge:

$$T'(z_H) = 1 + \frac{u_l(c_H, l_H)}{u_c(c_H, l_H) w_H}$$

Using (43) and (44) we can rewrite the labor wedge as follows:

$$T'(z_H) = 1 + \frac{-\lambda F_H(l_H, l_L) + \sum_i \eta_i F_{iH}(l_L, l_H)}{\lambda F_H(l_L, l_H)} = \frac{\sum_i \eta_i F_{iH}(l_L, l_H)}{\lambda F_H(l_L, l_H)} \quad (45)$$

The sign of (45) depends on η_L and η_H . In order to sign them, we exploit the government optimal choice of w_i characterized by the following FOCs:

$$-\mu u_l\left(c_L, \frac{w_L l_L}{w_H}\right) \frac{l_L}{w_H} + \eta_L = 0$$

$$\mu u_l\left(c_L, \frac{w_L l_L}{w_H}\right) \frac{w_L l_L}{w_H^2} + \eta_H = 0$$

Since $u_l < 0$, they imply $\eta_L < 0$ and $\eta_H > 0$. The CRS technology and concavity imply $F_{HL} > 0$ and $F_{HH} < 0$, which means complementarity between the two factors of production. Therefore, $\sum_i \eta_i F_{iH}(l_L, l_H) < 0$ and we conclude that $T'(z_H) < 0$. Top earners are subsidized at the margin because their labor raises the wages of lower earners. By closing the gap between the two wages the government can relax the incentive constraint for the high type and allow for additional redistribution. The result is entirely driven by the complementarity of the two factors in the production function, which generates a “positive externality” of the high type on the low type. In the classical Mirrlees model with linear technology there is no complementarity and top incomes are not subsidized.

6.2 Taxation in the Roy Model and Rent-Seeking

In this paragraph we study a more general model introduced by Rothschild and Scheuer (2013, 2016) where individuals can choose the sector where they work, how much they work and have a multidimensional vector of skills (one for each sector). We present a simple example with two activities and a two dimensional skill vector. Workers can choose between two activities: a traditional productive activity where the wage reflects the social marginal product of labor and rent-seeking where the marginal product of labor is zero and workers compete for a fixed rent μ such that wages are proportional to μ/E , with E being the total effort in the rent-seeking sector. Every individual has a skill vector (θ, φ) such that θ is the ability in the productive sector and φ is ability in the rent-seeking sector. Suppose there are only two types of workers in the economy: productive workers with $\theta = \varphi = 1$ and rent-seekers with $\theta = 0$ and $\varphi = \varphi_R$.

The total rent-seeking effort is:

$$E = \varphi_R e_R + \lambda_P e_P$$

where λ_P is the fraction of productive workers working in the rent-seeking sector. Productive workers are indifferent between the two sectors when the wage in the rent-seeking sector is equal to 1 (the marginal product of labor in productive sector) and we have $\mu/E = 1$ implying $\mu = E$. If instead $E > \mu$ they would all work in the traditional sector; while when $E < \mu$ they would all choose the rent-seeking activity.

Suppose that preferences are quasi-linear $u(c, e) = c - h(e)$. It can be shown that the optimal allocation involves an interior equilibrium where productive workers are indifferent between the two occupations. If λ_P is the share of productive workers working in the rent-seeking sector, we have:

$$E = \varphi_R e_R + \lambda_P e_P = \mu$$

which implies that λ_P is:

$$\lambda_P(e_R, e_P) = \frac{\mu - \varphi_R e_R}{e_P}$$

Given that the share of productive workers employed in the productive sector is $1 - \lambda_P$, total output produced in the economy is:

$$Y = \mu + (1 - \lambda_P(e_R, e_P)) e_P = e_P + \varphi_R e_R$$

Suppose the government is utilitarian, the welfare function is:

$$W = e_P + \varphi_R e_R - h(e_P) - h(e_R)$$

If the government can observe and tax income through a non-linear tax schedule but cannot tax occupational choices, we can solve the problem by choosing an optimal effort level for each type. The FOCs for effort are:

$$h'(e_P) = 1$$

$$h'(e_R) = \varphi_R$$

Notice that the two conditions imply zero wedge on labor income for both types. Although rent-seekers are not productive at all they are not taxed in equilibrium. The reason is that in this model rent-seekers are “indirectly productive” by crowding out productive workers from the rent-seeking sector. If rent-seekers were taxed, productive workers would be attracted into the rent-seeking sector (e_R would fall but $E = \mu$, and λ_P would have to increase to balance the change) and would decrease total production. Notice that the result is not dependent on the assumption of utilitarian social preferences, but would hold for any other combination of social welfare weights.

This example shows how general equilibrium considerations might be extremely important in shaping optimal marginal tax rates. Even under the extreme assumption that all top earners are rent-seekers, general equilibrium considerations would put downward pressure on marginal tax rates at the top to avoid attracting productive workers into rent-seeking.

The model with occupational choices can also be employed to study the “trickle down” effects in Stiglitz (1982) (see Rotschild Scheuer, 2013). Allowing for occupational choices still pushes towards lower top marginal tax rate than in a standard Mirrlees model with linear production, but less so than in a world without occupational choice (as in Stiglitz 1982). The reason is that, unlike in the Stiglitz’s model, if the government subsidizes high types, effort increases in the high skill sector decreasing wages in the high sector and increasing wages for the low sector. In a Roy model this would attract to the low-skilled sector some workers who were indifferent between the two sectors, reducing the increase in the low-skilled wage. This effect works against the standard general equilibrium effect presented in the previous paragraph, making trickle down less effective.

6.3 Wage Bargaining and Tax Avoidance

In this paragraph we study a standard Mirrlees model with a linear production function where individual income can depart from actual output. We present two potential departures: wage bargaining and tax avoidance.

Wage Bargaining Suppose top earners have measure 1 and after bargaining get a fraction η of their output z (where we allow for $\eta > 1$) such that $y = \eta z$. Bargained earnings are $b = (\eta - 1)y$ and average bargained earnings in the economy are $E(b)$. In the aggregate, it must be the case that total product is equal to total compensation. Hence, if $E(b) > 0$, so that there is overpay on average,

$E(b)$ must come at the expense of somebody. The opposite is true if $E(b) < 0$. For simplicity, we assume that any gain made through bargaining comes uniformly at the expense of everybody else in the economy. Hence, individual incomes are all reduced by a uniform amount $E(b)$ if $E(b) > 0$.¹³ We further assume that individuals can exert effort to increase η and their preferences are:

$$u_i(c, \eta, y) = c - h_i(y) - k_i(\eta)$$

When the cost of bargaining $E(b)$ is uniformly distributed across all agents, the government can offset it with the demogrant $-T(0)$. It follows that earnings can be written as $z = \eta y = y + b$. Each individual chooses η and z to maximize $u_i(c, \eta, y) = \eta y - T(\eta y) - h_i(y) - k_i(\eta)$ and FOCs are:

$$(1 - \tau)\eta = h'_i(y)$$

$$(1 - \tau)y = k'_i(\eta)$$

with $\tau = T'(z)$. Let us denote the average reported income, output and bargaining as Walrasian demands $z(1 - \tau)$, $y(1 - \tau)$ and $b(1 - \tau)$. The implied elasticities are:

$$e_1 = \frac{1 - \tau}{y} \frac{dy}{d(1 - \tau)}$$

$$e = \frac{1 - \tau}{z} \frac{dz}{d(1 - \tau)}$$

$$e_2 = \frac{1 - \tau}{z} \frac{db}{d(1 - \tau)} = s \cdot e$$

with

$$s = \frac{db/d(1 - \tau)}{dy/d(1 - \tau)}$$

The definitions imply that $(y/z)e_1 = (1 - s)e$ and that $e = (y/z)e_1 + e_3$.

When the social welfare weight on top incomes is zero, the government chooses the top tax rate to maximize total revenues:

$$\max_{\tau} \tau [z(1 - \tau) - \bar{z}] - N \cdot E(b)$$

and the FOC is:

$$[z - \bar{z}] - \tau \frac{dz}{d(1 - \tau)} = z - \bar{z}$$

rearranging

$$[\tau - s] \frac{dz}{d(1 - \tau)} = z - \bar{z} \Rightarrow \frac{\tau - s}{1 - \tau} \cdot e = \frac{z - \bar{z}}{z} = \frac{1}{a}$$

using $e_2 = s \cdot e$ we get:

$$\tau e - e_2 = \frac{1 - \tau}{a} \Rightarrow \tau^* = \frac{1 + e_2}{1 + ae} \quad (46)$$

If top earners are overpaid relative to their productivity $s > 0$ and $e_2 > 0$ implying that the optimal tax rate is higher than the one maximizing revenues in the standard model ($\tau^* > 1/(1 + ae)$). This is because of a *trickle up* effect that arise when a higher tax on high incomes reduces the cost of bargaining for low incomes. On the other hand, if $z < y$ and $e_2 < 0$ we would have a *trickle down* situation where a lower tax on top incomes shifts resources to individuals at the bottom.

¹³This assumption can be relaxed allowing for income reductions affecting only workers in the same occupation using a framework similar to the one we presented in Section 6.

Tax Avoidance Responses to tax rates can also take the form of tax avoidance. Define tax avoidance as changes in reported income due to changes in the form of compensation, but not in the total level of compensation. We observe tax-avoidance if taxpayers can shift part of their taxable income into another form that is treated more favorably from a tax perspective. Denote with x total sheltered income such that ordinary taxable income is $z = y - x$. Sheltered income is taxed at a constant marginal tax rate t . Suppose that the individual faces a utility cost for sheltering taxes and utility is $u_i(c, y, x) = c - h_i(y) - d_i(x)$ where $c = y - \tau z - tx + R = (1 - \tau)y + (\tau - t)x + R$ and $R = \tau \bar{z} - T(\bar{z})$ is virtual income. We can write Walrasian demands $z(1 - \tau, t) = y(1 - \tau) - x(\tau - t)$. Let us define e_3 the elasticity of sheltered income:

$$e_3 = \frac{1 - \tau}{z} \frac{dx}{d(1 - \tau)} = s \cdot e$$

where

$$s = \frac{dx/d(\tau - t)}{dy/d(1 - \tau) + dx/d(\tau - t)} = \frac{dx/d(\tau - t)}{\partial z/\partial(1 - \tau)}$$

and $e = (y/z)e_1 + e_3$.

The government problem is:

$$\max_{\tau, t} \tau [z(1 - \tau, t) - \bar{z}] + tx(\tau - t)$$

Suppose the government could only optimally set τ given some t , the FOC would be:

$$\begin{aligned} 0 &= [z - \bar{z}] - \tau \frac{\partial z}{\partial(1 - \tau)} + t \frac{dx}{d(\tau - t)} \\ &= [z - \bar{z}] - \tau \frac{\partial z}{\partial(1 - \tau)} + ts \frac{\partial z}{\partial(1 - \tau)} \\ &= [z - \bar{z}] - ez \frac{\tau - ts}{1 - \tau} \end{aligned}$$

rearranging:

$$\tau^* = \frac{1 + t \cdot a \cdot e_2}{1 + a \cdot e} \quad (47)$$

Notice how the tax is proportional to $t \cdot a \cdot e_2$ that captures the fiscal externality of tax avoidance. If $t = 0$ and the government cannot do anything to prevent income shifting, it is irrelevant whether e is due to real response or tax avoidance response (see Feldstein, 1999).

If instead the government could also optimally set t , we would have an extra optimality condition:

$$\begin{aligned} 0 &= \tau \frac{\partial z}{\partial t} + x - t \frac{dx}{d(\tau - t)} \\ &= x + (\tau - t) \frac{dx}{d(\tau - t)} \end{aligned}$$

since $x \geq 0$ and $dx/d(\tau - t) \geq 0$ the first order condition can only hold if $\tau = t$. If this is the case $x(\tau - t) = x(0) = 0$ and $z = y$ so that $e - e_3 = e_1$. If we replace this in (47) we obtain:

$$\tau^* = t^* = \frac{1}{1 + a \cdot e_1} \quad (48)$$

Intuitively, the government finds optimal to close any tax avoidance opportunity at the optimum. When this is the case the elasticity of income is the only one that matters.

References

- Feldstein, Martin. 1999. "Tax Avoidance and the Deadweight Loss of the Income Tax." *Review of Economics and Statistics* 81 (4): 674–80.
- Piketty, Thomas, Emmanuel Saez, and Stefanie Stantcheva. 2014. "Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities." *American Economic Journal: Economic Policy* 6 (1): 230–71.
- Rothschild, Casey and Florian Scheuer, "Redistributive Taxation in the Roy Model," *Quarterly Journal of Economics*, 2013, 128, 623–668.
- Rothschild, Casey and Florian Scheuer, "Optimal Taxation with Rent-Seeking," *Review of Economic Studies*, 2016, forthcoming
- Scheuer, Florian (2014), Lecture Notes
- Stiglitz, Joseph, "Self-Selection and Pareto Efficient Taxation," *Journal of Public Economics*, 1982, 17, 213–240.

7 Section 8: Optimal Minimum Wage and Introduction to Capital Taxation

In this Section we develop a theoretical analysis of optimal minimum wage policy in a perfectly competitive labor market following Lee and Saez (2012).

7.1 Optimal Minimum Wage

We study a model with extensive and intensive margin labor supply responses where wages are endogenous. Suppose output is produced through a constant return to scale production function $F(h_1, h_2)$ where h_1 and h_2 are low and high-skilled workers respectively. Profits are given by $\Pi = F(h_1, h_2) - w_1 h_1 - w_2 h_2$ and wages are equal to the marginal product of labor:

$$w_i = \frac{\partial F(h_1, h_2)}{\partial h_i} \quad (49)$$

A mass 1 of individuals has three labor supply options: i) not work and earn zero income, ii) work in occupation 1 and get w_1 , iii) work in occupation 2 and earn w_2 . Individuals are heterogeneous in their tastes for work. Every individual faces a vector $\theta = (\theta_1, \theta_2)$ of work costs that is smoothly distributed across the entire population according to $H(\theta)$ with support Θ . The government perfectly observes the wage w_i , but does not observe the cost of working. There are no savings and after tax income equals consumption such that $c_i = w_i - T_i$. Suppose there are no income effects and utility is linear in consumption:

$$u_i = c_i - \theta_i$$

The subset of individuals choosing occupation i is $\Theta_i = \{\theta \in \Theta | u_i = \max_j u_j\}$. The fraction of the population working in occupation i is $h_i(c) = |\Theta_i|$ and is a function of $c = (c_0, c_1, c_2)$. The tax system defines a competitive equilibrium (h_1, h_2, w_1, w_2) .

Equation (49) implies that $w_2/w_1 = F_2(1, h_2/h_1)/F_1(1, h_2/h_1)$. Constant returns to scale along with decreasing marginal productivity along each skill implies that the right-hand-side is a decreasing function of h_2/h_1 . Therefore, the function is invertible and the ratio h_2/h_1 can be written as a function of the wage ratio w_2/w_1 : $h_2/h_1 = \rho(w_2/w_1)$ with $\rho(\cdot)$ a decreasing function. Constant returns to scale also imply that there are no profits in equilibrium. Hence $\Pi = F(h_1, h_2) - w_1 h_1 - w_2 h_2 = 0$ so that $w_1 + w_2 \rho(w_2/w_1) = F(1, \rho w_2/w_1)$, which defines a decreasing mapping between w_1 and w_2 so that we can express w_2 as a decreasing function of w_1 : $w_2(w_1)$.

Labor supply and demand for the low-skilled labor market are $D_1(w_1)$ and $S_1(w_1)$ with $D'_1(w_1) \leq 0$ and $S'_1(w_1) \geq 0$ and are defined assuming that the market clears in the high-skilled labor market. The low-skilled labor demand elasticity is:

$$\eta_1 = -\frac{w_1}{h_1} D_1(w_1)$$

The resource constraint of the economy is:

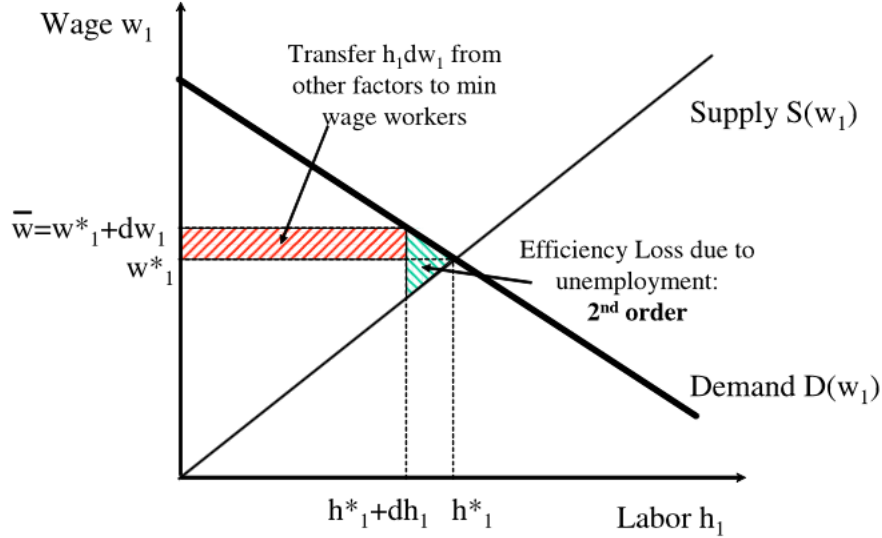
$$h_0 c_0 + h_1 c_1 + h_2 c_2 \leq h_1 w_1 + h_2 w_2 \quad (50)$$

The government weights individual utilities through a social welfare function $G(\cdot)$ and we can write the social welfare of the economy as:

$$SW = (1 - h_1 - h_2) G(c_0) + \int_{\Theta_1} G(c_1 - \theta_1) dH(\theta) + \int_{\Theta_2} G(c_2 - \theta_2) dH(\theta) \quad (51)$$

We define social marginal welfare weights as usual $g_0 = G'(c_0)/\lambda$ and $g_i = \int_{\Theta_i} G'(c_i - \theta_i) dH(\theta) / (\lambda h_i)$, where λ is the marginal value of public funds. The concavity of the SWF implies $g_0 > g_1$ and $g_1 > g_2$. Since there are no income effects the value of transferring \$1 to everyone in the economy is \$1 and we have $\lambda = g_0 h_0 + g_1 h_1 + g_2 h_2 = 1$.

Figure 3:
Desirability of Small Minimum Wage



Minimum Wage with No Taxes Suppose there are no taxes and transfers, we have $c_0 = 0$, $c_1 = w_1$ and $c_2 = w_2$. Suppose the economy is at the equilibrium and the government introduces a small minimum wage above the equilibrium wage of the low-skilled market such that $\bar{w} = w_1^* + d\bar{w}$. The change will generate a drop in employment h_1 . The workers who drop out of the low-skilled sector will move either to unemployment or to the high-skilled sector depending on their preferences. We will assume *efficient rationing*: the workers who involuntarily lose their low-skilled jobs due to the minimum wage are those with the least surplus from working in the low-skilled sector.¹⁴ This is clearly the most favorable case to minimum wage policy. We establish the first result of the paper:

Proposition 1: *With no taxes/transfers, if (i) efficient rationing holds; (ii) the government values redistribution from high-skilled workers toward low-skilled workers ($g_1 > g_2$); (iii) the demand elasticity η_1 for low-skilled workers is finite; and (iv) the supply elasticity of low-skilled workers is positive, then introducing a minimum wage increases social welfare.*

Consider the changes dw_1 , dw_2 , dh_1 and dh_2 following the increase in the minimum wage, we have $d\Pi = \sum_i [(\partial F / \partial h_i) dh_i - w_i dh_i - h_i dw_i] = -h_1 dw_1 - h_2 dw_2$ and the no profit condition implies:

$$h_1 dw_1 + h_2 dw_2 = 0 \quad (52)$$

Therefore, the earnings gain for low-skilled people $h_1 dw_1 > 0$ is compensated by a loss in the earnings of high-skilled workers $h_2 dw_2 < 0$. The government values the transfer of resources $[g_1 - g_2] h_1 dw_1$. Under efficient rationing, positive supply elasticity and finite demand elasticity, the welfare loss due to low-skilled individuals moving to unemployment is second-order (see Figure 3).

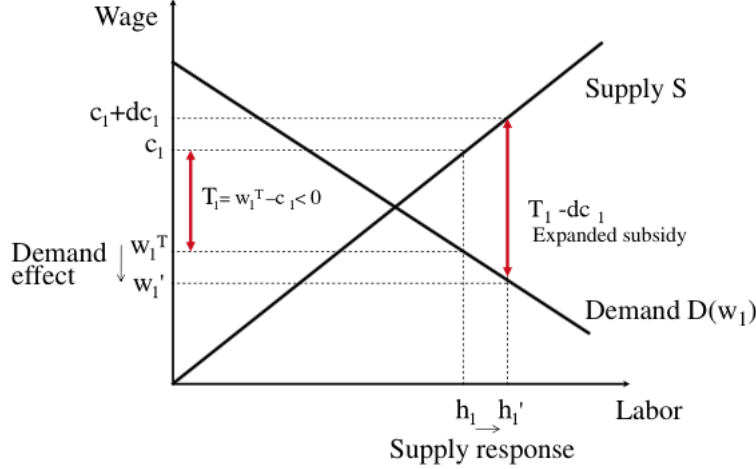
More formally, the first order condition wrt $d\bar{w}$ is:

$$\frac{dSW}{d\bar{w}} = \left[-\frac{dh_1}{d\bar{w}} - \frac{dh_2}{d\bar{w}} \right] G(0) + \frac{dh_1}{d\bar{w}} G(0) + \frac{dh_2}{d\bar{w}} G(0) +$$

¹⁴The assumption can be relaxed and a working paper version of the paper shows how the model can be derived under the assumption of uniform rationing.

Figure 4:

**a. Response to a low-skilled subsidy
with no minimum wage**



$$+ \int_{\Theta_1} G'(c_1 - \theta_1) dH(\theta) + \frac{dw_2}{d\bar{w}} \int_{\Theta_2} G'(c_2 - \theta_2) dH(\theta)$$

The second and third terms come from the assumption of perfect rationing: the workers moving to unemployment from the two occupations are those with zero surplus from working therefore the welfare loss associated to the change of occupation is zero. Also, those who drop out of occupation 1 and move to 2 are indifferent between the two and we can ignore the welfare effect associated to the change by envelope theorem. Using (52) we have $dw_2/d\bar{w} = -h_1/h_2$ and the FOC becomes:

$$\frac{dSW}{d\bar{w}} = h_1 \lambda [g_1 - g_2] > 0$$

which proves Proposition 1.

Minimum Wage with Taxes and Transfers We now assume that the government can use taxes and transfers jointly with the minimum wage policy.

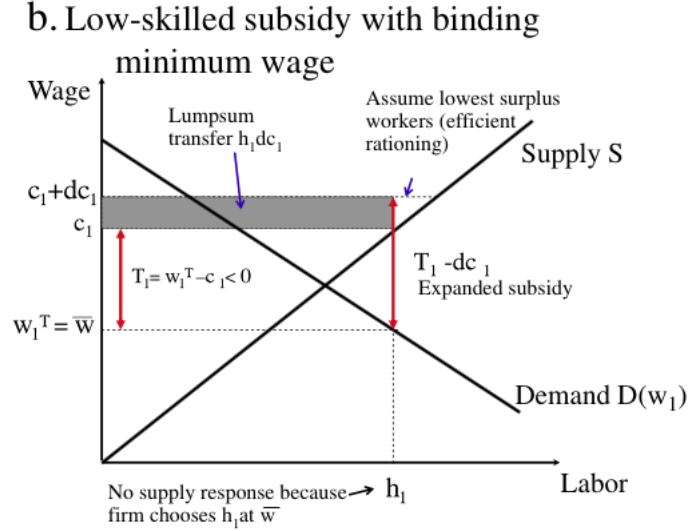
Proposition 2: *Under efficient rationing, assuming $\eta_1 < \infty$, if $g_1 > 1$ at the optimal tax allocation (with no minimum wage), then introducing a minimum wage is desirable. Furthermore, at the joint minimum wage and tax optimum, we have: (i) $g_1 = 1$ (Full redistribution to low-skilled workers); (ii) $h_0 g_0 + h_1 g_1 + h_2 g_2 = 1$ (Social welfare weights average to one).*

Suppose there was no minimum wage, an attempt to increase c_1 by dc_1 while keeping c_0 and c_2 constant through an increased work subsidy provides incentives for some of the non-workers to start working in occupation 1 (extensive labor supply response) and for some of workers in occupation 2 to switch to occupation 1 (intensive labor supply response). This leads to a reduction in w_1 through demand side effects (as long as $\eta_1 < \infty$). See Figure 4.

Consider the same increase in c_1 when the minimum wage was initially set at $\bar{w} = w_1^T$, where (w_i^T, c_i^T) is the optimal tax and transfer system which maximizes social welfare absent the minimum wage. Since w_1 cannot fall, labor supply responses are effectively blocked (Figure 5). Efficient rationing guarantees that individuals willing to leave occupation 1 are precisely those with the lowest surplus

from working in occupation 1 relative to their next best option. Therefore, the dc_1 change is like a lump-sum tax reform and its net welfare effect is simply $[g_1 - 1] h_1 dc_1$. If $g_1 > 1$, the introduction of the minimum wage improves upon the tax/transfer optimum allocation. This result shows that under the minimum wage policy, redistribution to low-skilled workers can be made lump-sum. Furthermore, raising the lump-sum transfer to occupation 1 improves welfare as long as $g_1 > 1$ and therefore the government will find optimal to do it until $g_1 = 1$. With no behavioral responses an increase of \$1 has a welfare effect of $h_0 g_0 + h_1 g_1 + h_2 g_2$ and at the optimal the two are equal.

Figure 5:



To prove it formally, rewrite consumption in occupation i as $\Delta c_i = c_i - c_0$ and the resource constraint as $h_1 \cdot (w_1 - \Delta c_1) + h_2 \cdot (w_2 - \Delta c_2) \geq c_0$. The Lagrangian of the problem is:

$$L = SW + \lambda [h_1 \cdot (w_1 - \Delta c_1) + h_2 \cdot (w_2 - \Delta c_2) - c_0]$$

Suppose there is a minimum wage and the government introduces a change dc_1 , the wage of occupation 1 does not change because of the minimum wage and so does w_2 given that $w_2(w_1)$ (as we showed above). As a consequence, there is no change in $h_1/h_2 = \rho(w_2/w_1)$ and no change in the levels of h_1 and h_2 since they cannot increase simultaneously. Therefore:

$$\frac{dL}{dc_1} = \int_{\Theta_1} G'(c_0 + \Delta c_1 - \theta_1) dH(\theta) - \lambda h_1 = \lambda [g_1 - 1] h_1$$

At the optimum it must be $g_1 = 1$. Taking the FOC wrt c_0 we have:

$$\begin{aligned} \frac{dL}{dc_0} &= (1 - h_1 - h_2) G'(c_0) + \int_{\Theta_1} G'(c_0 + \Delta c_1 - \theta_1) dH(\theta) + \int_{\Theta_2} G'(c_0 + \Delta c_2 - \theta_2) dH(\theta) - \lambda \\ &= \lambda [h_0 g_0 + h_1 g_1 + h_2 g_2 - 1] \end{aligned}$$

which proves Proposition 2.

Pareto Improving Reform In this section we review the last result in Lee and Saez (2012) that shows how minimum wage and low-skilled labor subsidies can be complementary. Suppose

there are extensive margin responses only, the participation tax rate of low-skilled workers τ_1 is $1 - \tau_1 = (c_1 - c_0)/w_1$, such that $c_1 = c_0 + (1 - \tau_1)w_1$.

Proposition 3: *In a model with extensive labor supply responses only, a binding minimum wage associated with a positive tax rate on minimum wage earnings ($\tau_1 > 0$) is second-best Pareto inefficient. This result remains a-fortiori true when rationing is not efficient.*

Suppose the government reduces the minimum wage by $d\bar{w} < 0$ while keeping c_0 , c_1 and c_2 constant. The change incentivizes unemployed individuals to enter occupation 1 generating a change $dh_1 > 0$ and increasing revenues since $\tau_1 > 0$. The change $dh_1 > 0$ induces a change $dw_2 > 0$. However, since $h_1d\bar{w} + h_2dw_2 = 0$ the mechanical effect of changes in wages is zero. Since c_0 , c_1 and c_2 are constant the total effect of the government policy is only given by the increase in revenues, which is positive. Proposition 3 implies that, when labor supply responses are concentrated along the extensive margin, a minimum wage should always be associated with low-skilled work subsidies such as the EITC.

To prove it formally notice that since consumption does not change at any occupation, the utility of those who do not switch jobs is not affected. From the demand side, we have $w_2(w_1)$ with $dw_2/dw_1 = -h_1/h_2 < 0$ and hence $dw_2 > 0$. This implies that relative demand for high-skilled work $h_2/h_1 = \rho(w_2/w_1)$ decreases as $\rho(\cdot)$ is decreasing. Because $c_2 - c_0$ remains constant, and labor supply is only along the extensive margin, the supply of high-skilled workers is unchanged so that $dh_2 = 0$, which then implies that $dh_1 > 0$. The dh_1 individuals shifting from no work to low-skilled work are weakly better-off because they were by definition rationed by the minimum wage (strictly better off in case of inefficient rationing). The government budget is $h_1(w_1 - \Delta c_1) + h_2(w_2 - \Delta c_2) - c_0 \geq 0$. Therefore, the net effect of the reform on the budget is: $dh_1 \cdot (w_1 - \Delta c_1) + h_1dw_1 + h_2dw_2 = dh_1\tau_1w_1 > 0$. Thus, with $\tau_1 > 0$, the reform creates a budget surplus which can be used to increase c_0 and improve everybody's welfare (with no behavioral response effects), a Pareto improvement.

References

- [1] David Lee & Emmanuel Saez, 2012. "Optimal minimum wage policy in competitive labor markets," Journal of Public Economics, vol 96(9-10), pages 739-749.

8 Section 9: Linear Capital Taxation

In this section we introduce a framework to study optimal linear capital taxation. We first focus on a two-period model, define the concept of intertemporal wedge and derive optimal capital taxes using the Atkinson Stiglitz result. We then move to an infinite horizon model with aggregate uncertainty and derive optimal taxes. Finally, we study a model with capitalists and workers and show that only under some assumption about preferences we can recover a zero capital tax in steady state.¹⁵

8.1 A Two-Period Model

We introduce a two-period model with capital accumulation that will be useful to study the problem of capital taxation. The two time periods are $t = 0$ and $t = 1$. The preferences for individual i are $U^i(c_0, c_1, y_0)$. The individual can save period 0 income and earn gross interest rate R on savings. We start by constraining the government's instruments and focusing on a linear consumption tax, while keeping a non-linear tax on income. The budget constraints for the two periods read:

$$(1 + \tau_0) c_0 \leq y_0 - T(y_0) - k_1$$

$$(1 + \tau_1) c_1 \leq Rk_1$$

Combining the two we get:

$$(1 + \tau_1) c_1 \leq R(y_0 - T(y_0) - (1 + \tau_0) c_0) \quad (53)$$

Rearranging we have:

$$c_0 + \frac{1}{R} \left[\frac{1 + \tau_1}{1 + \tau_0} \right] c_1 \leq \frac{y_0 - T(y_0)}{1 + \tau_0}$$

Let us denote $1 + \tilde{\tau}_1 = (1 + \tau_1) / (1 + \tau_0)$ and $\tilde{T}(y) = (\tau_0 y + T(y)) / (1 + \tau_0)$. The budget constraint can be written:

$$c_0 + \frac{1}{R} [1 + \tilde{\tau}_1] c_1 \leq y_0 - \tilde{T}(y_0)$$

When the budget constraint holds with equality we can write consumption at time 1 as:

$$c_1 = \frac{R}{1 + \tilde{\tau}_1} s_0$$

where $s_0 = y_0 - \tilde{T}(y_0) - c_0$ is the total level of savings when there are no distortions in the economy. We can interpret $\tilde{\tau}$ as a capital income tax. It distorts inter-temporal consumption decisions by changing the relative price of c_0 and c_1 . It can be interpreted as a wedge on the optimal savings decision. Notice that whenever $\tilde{\tau}_1 = 0$ we have no distortion in the inter-temporal choice of the agent.

Suppose the agent has separable preferences in consumption and labor of the form:

$$U(c_0, c_1, y_0) = u(c_0) + u(c_1) - h(y_0/w)$$

We can rewrite the preferences as $U(g(c_0, c_1), y_0)$ so that the utility is weakly separable in $g(\cdot)$ and y_0 . It follows that Atkinson-Stiglitz applies: if non-linear income taxation is available the government finds optimal to set a flat zero tax on c_0 and c_1 .¹⁶

¹⁵The second part of these notes is based on lecture notes by Florian Scheuer.

¹⁶Here is a short proof of the Atkinson Stiglitz result provided in Kaplow (2006). Suppose all individuals have weakly

8.2 Infinite Horizon Model - Chamley (1986)

In this section we introduce a model where capital returns and wages are endogenous. We focus on linear capital and labor taxes in an infinite horizon economy. There is aggregate uncertainty in the economy and each period t a state s_t is realized so that the history of aggregate uncertainty is a sequence $s^t = (s_0, s_1, \dots, s_t)$. Output is produced according to a constant return to scale production function:

$$F(K(s^{t-1}), L(s^t), s^t, t) \quad (54)$$

the productive capital at time t is the stock that was chosen at time $t - 1$. The firm solves the following profit maximization problem:

$$\max_{K, L} F(K(s^{t-1}), L(s^t), s^t, t) - w(s^t) L(s^t) - r(s^t) K(s^t)$$

Competitive labor and capital markets imply that input prices are equal to their marginal product:

$$w(s^t) = F_L(K(s^{t-1}), L(s^t), s^t, t)$$

$$r(s^t) = F_K(K(s^{t-1}), L(s^t), s^t, t)$$

The economy is populated by a representative agent whose utility is:

$$\sum_{s^t} \beta^t \Pr(s^t) u(c(s^t), L(s^t))$$

where $\Pr(s^t)$ is the probability of history s^t . The aggregate resource constraint of the economy is:

$$c(s^t) + g(s^t) + K(s^t) - (1 - \delta) K(s^{t-1}) \leq F(K(s^{t-1}), L(s^t), s^t, t) \quad (55)$$

The output produced is employed to finance consumption, public spending and investments. The resource constraint implicitly assumes that aggregate uncertainty results from technology or government spending shocks. The government optimally chooses taxes on labor income $\tau^l(s^t)$ and taxes on capital income $\tau^k(s^t)$ and starts with initial debt B_0 .

We assume complete markets where the price of an Arrow-Debreu security is $p(s^t)$.¹⁷ The government budget constraint is:

$$\sum_{t, s^t} p(s^t) [g(s^t) - \tau^l(s^t) w(s^t) L(s^t) - \tau^k(s^t) (r(s^t) - \delta) K(s^{t-1})] \leq -B_0$$

Taxes on consumption and capital are employed to finance government layouts $g(s^t)$. Notice that the capital tax is levied on the capital gain net of the capital depreciation.

separable preferences $V(g(\mathbf{x}), y)$, where \mathbf{x} is a vector of commodities. Suppose we start from a situation where there are positive taxes on commodities and we implement a policy such that $\mathbf{t} \rightarrow \mathbf{0}$: zero flat tax on all commodities. Suppose the government offsets the utility change of the agent with non-linear income taxes such that labor supply is unchanged at the optimum and $V(g(\mathbf{x}'), y) = V(g(\mathbf{x}), y)$. By definition every agent has the same utility as before and no one is willing to imitate another individual (if they were not willing to imitate before the tax change). By revealed preference we know that $\sum_k p_k x_k > y - T^*(y)$: the agent cannot afford the old bundle under the current income taxation. Under the old policy scenario the agent could afford the bundle and we had $\sum_k (p_k + \tau_k) x_k^0 \leq y - T(y)$. Combining the two inequalities we find that $T^*(y) > \sum_k \tau_k x_k^0 + T(y)$ and the total revenue raised after the tax change is strictly higher than total revenues before the tax change. Since incentive compatibility holds and we have no welfare effect by construction, the new policy is welfare improving since it raises more revenues.

¹⁷An Arrow-Debreu security is a financial instrument that provides one unit of consumption in a state s_t and zero units in any other state. We talk about complete markets whenever we can price such an asset in every state of the world.

The household budget constraint reads:

$$\sum_{t,s^t} p(s^t) [c(s^t) + K(s^t) - w(s^t)(1 - \tau^l(s^t))L(s^t) - R(s^t)K(s^{t-1})] \leq B_0 \quad (56)$$

where $R(s^t) = 1 + (1 - \tau^k(s^t))(r(s^t) - \delta)$ is the gross interest rate net of taxes. We can set up the Lagrangian for the consumer problem:

$$\begin{aligned} L = & \sum_{s^t} \beta^t \Pr(s^t) u(c(s^t), L(s^t)) + \\ & + \lambda \left[B_0 - \sum_{t,s^t} p(s^t) [c(s^t) + K(s^t) - w(s^t)(1 - \tau^l(s^t))L(s^t) - R(s^t)K(s^{t-1})] \right] \end{aligned}$$

The first order conditions are:

$$\beta^t \Pr(s^t) u_c(c(s^t), L(s^t)) - \lambda p(s^t) = 0 \quad (57)$$

$$\beta^t \Pr(s^t) u_L(c(s^t), L(s^t)) + \lambda p(s^t)(1 - \tau^l(s^t))w(s^t) = 0 \quad (58)$$

$$-\lambda p(s^t) + \lambda \sum_{s^{t+1}} p(s^{t+1}) R(s^{t+1}) = 0 \quad (59)$$

On top of the FOCs, a non-arbitrage condition must hold between capital and Arrow-Debreu securities:

$$p(s^t) = \sum_{s^{t+1}} p(s^{t+1}) R(s^{t+1}) \quad (60)$$

We can define a competitive equilibrium as follows:

Definition: A competitive equilibrium is a policy $\{g(s^t), \tau^k(s^t), \tau^l(s^t)\}$, an allocation $\{c(s^t), K(s^t), L(s^t)\}$ and prices $\{w(s^t), r(s^t), p(s^t)\}$, such that households maximizes utility s.t. budget constraint, firms maximize profits, the government budget constraint holds and markets clear

Combining (57) and (58) we get the standard intratemporal condition for labor supply:

$$\begin{aligned} \frac{\beta^t \Pr(s^t) u_c(c(s^t), L(s^t))}{p(s^t)} &= - \frac{\beta^t \Pr(s^t) u_L(c(s^t), L(s^t))}{p(s^t)(1 - \tau^l(s^t))w(s^t)} \\ (1 - \tau^l(s^t))w(s^t) &= - \frac{u_L(c(s^t), L(s^t))}{u_c(c(s^t), L(s^t))} \end{aligned} \quad (61)$$

From (57) and (59) we derive the so called Euler equation that pins down the slope of the consumption path of the agent:

$$u_c(c_0, L_0) = \frac{\beta^t \Pr(s^t) u_L(c(s^t), L(s^t))}{p(s^t)} \quad (62)$$

Starting from (56), we can rewrite it using the optimality conditions and the no-arbitrage condition:

$$\begin{aligned}
\sum_{t,s^t} p(s^t) [c(s^t) - w(s^t)(1 - \tau^l(s^t))L(s^t)] &\leq B_0 + R_0 K_0 \\
\sum_{t,s^t} \frac{\beta^t \Pr(s^t) u_c(c(s^t), L(s^t))}{u_c(c_0, L_0)} \left[c(s^t) + \frac{u_L(c(s^t), L(s^t))}{u_c(c(s^t), L(s^t))} L(s^t) \right] &\leq B_0 + R_0 K_0 \\
\sum_{t,s^t} \beta^t \Pr(s^t) u_c(c(s^t), L(s^t)) \left[c(s^t) + \frac{u_L(c(s^t), L(s^t))}{u_c(c(s^t), L(s^t))} L(s^t) \right] &\leq u_c(c_0, L_0) [B_0 + R_0 K_0] \\
\sum_{t,s^t} \beta^t \Pr(s^t) [u_c(c(s^t), L(s^t)) c(s^t) + u_L(c(s^t), L(s^t)) L(s^t)] &\leq u_c(c_0, L_0) [B_0 + R_0 K_0] \quad (63)
\end{aligned}$$

We call the constraint in (63) implementability constraint since it captures the agent's optimal choices subject to their feasibility.

Optimal Taxes The government chooses taxes to maximize the welfare of the representative agent subject to the resource constraint and the implementability constraint. The problem reads:

$$\begin{aligned}
&\max_{c(s^t), L(s^t), K(s^t), \tau_0^k} \sum_{s^t} \beta^t \Pr(s^t) u(c(s^t), L(s^t)) \\
&\text{s.t.} \\
&c(s^t) + g(s^t) + K(s^t) - (1 - \delta) K(s^{t-1}) \leq F(K(s^{t-1}), L(s^t), s^t, t) \\
&\sum_{t,s^t} \beta^t \Pr(s^t) [u_c(c(s^t), L(s^t)) c(s^t) + u_L(c(s^t), L(s^t)) L(s^t)] \leq u_c(c_0, L_0) [B_0 + R_0 K_0]
\end{aligned}$$

We assume τ_0^k is fixed, we denote with μ the multiplier on the implementability constraint and define:

$$W(c(s^t), L(s^t)) = u(c(s^t), L(s^t)) + \mu [u_c(c(s^t), L(s^t)) c(s^t) + u_L(c(s^t), L(s^t)) L(s^t)]$$

The government problem becomes:

$$\begin{aligned}
&\max_{c(s^t), L(s^t), K(s^t)} \sum_{s^t} \beta^t \Pr(s^t) W(c(s^t), L(s^t)) - \mu u_c(c_0, L_0) [B_0 + R_0 K_0] \\
&\text{s.t.}
\end{aligned}$$

$$c(s^t) + g(s^t) + K(s^t) - (1 - \delta) K(s^{t-1}) \leq F(K(s^{t-1}), L(s^t), s^t, t)$$

For any period $t \neq 0$ the FOCs are:

$$\beta^t \Pr(s^t) W_c(c(s^t), L(s^t)) - \gamma(s^t) = 0$$

$$\beta^t \Pr(s^t) W_L(c(s^t), L(s^t)) + \gamma(s^t) F_L(K(s^{t-1}), L(s^t), s^t, t) = 0$$

$$-\gamma(s^t) + \sum_{s_{t+1}} \gamma(s^t, s_{t+1}) [F_k(K(s^{t-1}), L(s^t), s^t, t) + (1 - \delta)] = 0$$

Combining the FOCs we get an intratemporal condition:

$$-\frac{W_L(c(s^t), L(s^t))}{W_c(c(s^t), L(s^t))} = F_L(K(s^{t-1}), L(s^t), s^t, t)$$

Using the household FOC in (58), we can rewrite the optimality condition as a function of the tax:

$$\tau^{l*}(s^t) = 1 - \frac{u_L(c(s^t), L(s^t))}{W_L(c(s^t), L(s^t))} \frac{W_c(c(s^t), L(s^t))}{u_c(c(s^t), L(s^t))} \quad (64)$$

From the government FOCs we can also derive an intertemporal condition:

$$W_c(c(s^t), L(s^t)) = \beta \sum_{s_{t+1}} \Pr(s^{t+1}|s^t) W_L(c(s^{t+1}), L(s^{t+1})) R^*(s^{t+1})$$

where $R^*(s^t) = 1 + F_K(K(s^{t-1}), L(s^t), s^t, t) - \delta$ is the untaxed gross return on capital net of depreciation. Again, exploiting the household's Euler equation we can derive:

$$R(s^{t+1}) = R^*(s^{t+1}) \frac{W_c(c(s^{t+1}), L(s^{t+1}))}{u_c(c(s^{t+1}), L(s^{t+1}))} \frac{u_c(c(s^t), L(s^t))}{W_c(c(s^t), L(s^t))} \quad (65)$$

Proposition 1: Suppose that (i) there is no uncertainty (ii) there is a steady state. Then in the steady state $\tau^k = 0$ is optimal.

It is easy to see that in a steady state when there is no uncertainty $R(ss) = R^*(ss)$, which implies:

$$1 + (1 - \tau^k(ss)) (F_k(K(ss), L(ss)) - \delta) = 1 + F_K(K(ss), L(ss), ss) - \delta$$

and that $\tau^k(ss) = 0$.

Now consider a special case with separable preferences and constant intertemporal elasticity of substitution:

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

then we have:

$$\begin{aligned} W(c, L) &= \frac{c^{1-\sigma}}{1-\sigma} - v(L) + \mu [c^{-\sigma} c - v'(L) L] \\ &= \left(\frac{1}{1-\sigma} + \mu \right) c^{1-\sigma} - [v(L) + \mu v'(L) L] \end{aligned}$$

therefore:

$$\begin{aligned} W_c &= (1 + \mu(1 - \sigma)) c^{-\sigma} \\ &= (1 + \mu(1 - \sigma)) u_c \end{aligned}$$

$$\frac{W_c}{u_c} = 1 + \mu(1 - \sigma)$$

Equation (65) reduces to $R(s^{t+1}) = R^*(s^{t+1})$. Hence, we established that for separable preferences with constant intertemporal elasticity of substitution we have zero capital taxation even out of the steady state and in a model with uncertainty.

Tax Smoothing Take now the special case where $v(L) = \alpha L^\gamma / \gamma$ is isoelastic, we have:

$$W(c, L) = \left(\frac{1}{1-\sigma} + \mu \right) c^{1-\sigma} - \alpha \left(\frac{1}{\gamma} + \mu \right) L^\gamma$$

it follows that

$$\frac{W_L}{u_L} = 1 + \mu\gamma$$

the optimal linear labor tax becomes:

$$\tau^{l*}(s^t) = 1 - \frac{1 + \mu(1-\sigma)}{1 + \mu\gamma}$$

Therefore, labor taxes are constant across states and over time. The government finds optimal to smooth distortions to labor supply. This result depends on the possibility of setting state-contingent capital taxes. If the labor elasticity is constant and shocks can be offset using capital taxes, there is no residual reason to differentially tax labor.

8.3 Infinite Horizon - Judd (1985)

We now introduce the model by Judd (1985) where the famous “zero steady state capital tax” result arise. We then show that the result is not general and depends on the agent’s preferences as shown in Werning Straub (2015). Suppose there are two agents: capitalists and workers. The former save, get the return to capital and do not work; the latter supply one unit of labor inelastically and consume everything they earn. The government taxes return to capital and pays transfers to workers. Output is produced according to a constant return to scale technology with production function $F(k_t, n_t)$. Aggregate labor is $n_t = 1$ so that we can rewrite $f(k_t) = F(k_t, 1)$. Capitalists have utility $U(C_t)$ and workers’ utility is $u(c_t)$. The resource constraint of the economy is:

$$c_t + C_t + g + k_{t+1} \leq f(k_t) + (1 - \delta)k_t$$

Under the assumption of perfectly competitive labor markets, wages are:

$$w_t = F_L(k_t, n_t) = f(k_t) - k_t f'(k_t)$$

The after-tax return to capital is:

$$R_t = 1 + (1 - \tau_t)(R_t^* - 1)$$

where $R^* = f'(k_t) + 1 - \delta$.

Capitalists solve the following maximization problem:

$$\max_{C_t, a_{t+1}} \sum_{t=0}^{+\infty} \beta^t U(C_t)$$

s.t.

$$C_t + a_{t+1} = R_t a_t$$

The optimality condition delivers the standard Euler equation $U'(C_t) = \beta R_{t+1} U'(C_{t+1})$. Since total wealth must equal total capital stock in equilibrium, using the Euler equation:

$$C_t + k_{t+1} = R_t k_t = \frac{U'(C_{t-1})}{\beta U'(C_t)} k_t$$

rearranging:

$$\beta U'(C_t)(C_t + k_{t+1}) = U'(C_{t-1})k_t \quad (66)$$

Equation (66) is the implementability constraint. The government maximizes the following objective function, where γ is the Pareto weight on capitalists:

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

The Lagrangian of the problem is:

$$\begin{aligned} L = & \sum_{t=0}^{\infty} \beta^t (u(c_t) + \gamma U(C_t)) \\ & + \sum_{t=0}^{\infty} \beta^t \lambda_t (f(k_t) + (1 - \delta)k_t - c_t - C_t - g - k_{t+1}) \\ & + \sum_{t=0}^{\infty} \beta^t \mu_t (\beta U'(C_t)(C_t + k_{t+1}) - U'(C_{t-1})k_t) \end{aligned}$$

with $\mu_0 = 0$ since there is no implementability in the first period (τ_0 is taken as given). The first order conditions wrt to c_t , k_{t+1} and C_t are respectively:

$$u'(c_t) = \lambda_t \quad (67)$$

$$\frac{\lambda_{t+1}}{\lambda_t} (f'(k_{t+1}) + 1 - \delta) = \frac{1}{\beta} + \frac{U'(C_t)}{\lambda_t} (\mu_{t+1} - \mu_t) \quad (68)$$

$$\mu_{t+1} = \frac{\mu_t}{k_{t+1}} \left(C_t + k_{t+1} + \frac{U'(C_t)}{U''(C_t)} \right) + \frac{1}{\beta k_{t+1}} \left(\gamma \frac{U'(C_t)}{U''(C_t)} - \frac{\lambda_t}{U''(C_t)} \right) \quad (69)$$

It is straightforward from equation (68) that whenever a steady state exists it involves zero capital taxes and $R(ss) = f'(k_t) + 1 - \delta = R^*$. This result is extremely powerful since it is independent of the welfare weight attached to capitalists. However, the result does not hold for the case where $\sigma = 1$. Rewrite the FOCs (68) and (69) using the inverse intertemporal elasticity of substitution $\sigma_t = -U''(C_t)C_t/U'(C_t)$ and defining the ratio $v_t = U'(C_t)/u'(c_t)$:

$$(68) \Rightarrow \frac{u'(c_{t+1})}{u'(c_t)} (f'(k_{t+1}) + 1 - \delta) = \frac{1}{\beta} + v_t (\mu_{t+1} - \mu_t)$$

$$(69) \Rightarrow \mu_{t+1} = \mu_t \left(\frac{1 - 1/\sigma}{k_{t+1}} + 1 \right) + \frac{1}{\beta k_t \sigma_t v_t} (1 - \gamma v_t)$$

Take the case where $\sigma = 1$ (log preferences) and the allocation converges to a steady state, then:

$$\mu_{t+1} - \mu_t = \frac{R^* - \frac{1}{\beta}}{v}$$

$$\mu_{t+1} - \mu_t = \frac{1 - \gamma v}{\beta k v}$$

$$\Rightarrow R^* - \frac{1}{\beta} = \frac{1 - \gamma v}{\beta k}$$

As long as there is a low enough weight on capitalists, capital is taxed in steady state. For a long time we thought that this was simply an anomaly for the logarithmic case. However, Werning and Straub (2015) show that the result does not hold for any $\sigma > 1$ by noticing that the steady state does not necessarily exist.

Proposition 2: *If $\sigma > 1$ and $\gamma = 0$, then for any initial k_0 the solution to the planning problem does not converge to the zero-tax steady state, or any other interior steady state.*

Suppose capital taxes are raised in the future, capitalists will decrease savings today for the substitution effect. A capital tax increase will also reduce agent's wealth and lower capitalists' consumption through the income effect. When $\sigma > 1$ the income effect prevails and capitalists save more. The increase in the capital stock increases wages and is beneficial for workers. For this reason the government wants to set positive capital taxes in the long-term. The opposite is true when $\sigma < 1$: the substitution effect is larger than the income effect and zero taxes in the future increase savings in the short term increasing wages and workers' consumption.

References

- Chamley, Christophe, "Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives," *Econometrica*, 1986, 54 (3), pp. 607–622.
- Judd, Kenneth L., "Redistributive taxation in a simple perfect foresight model," *Journal of Public Economics*, 1985, 28 (1), 59 – 83.
- Kaplow, Louis, 2006. "On the undesirability of commodity taxation even when income taxation is not optimal," *Journal of Public Economics*, Elsevier, vol. 90(6-7), pages 1235-1250, August.
- Scheuer, Florian (2014), Lecture Notes
- Straub, Ludwig and Ivan Werning, "Positive Long Run Capital Taxation: Chamley-Judd Revisited," Working Paper, MIT 2015.

9 Section 10: Education Policies and Simpler Theory of Capital Taxation

In this section we study education policies in a simplified version of framework analyzed by Stantcheva (2016). We then review a simpler theory of capital taxation proposed by Saez and Stantcheva (2016) in a continuous time model.

9.1 Education Policies

We study a static model with human capital investments based on Bovenberg and Jacobs (2005, 2011) and Stantcheva (2016). Suppose individuals are heterogeneous in ability θ distributed according to $f(\theta)$. Agents can invest in education paying a monetary cost $M(e)$ such that $M'(e) > 0$ and $M''(e) \geq 0$. Wages are a function of ability and human capital that we can write as $w(\theta, e)$. Denote with $\rho_{\theta e}$ the Hicksian coefficient of complementarity between ability and education defined as:

$$\rho_{\theta e} = \frac{w_{\theta e} w}{w_{\theta} w_e}$$

Consider two special cases for the wage function: (i) multiplicative $w(\theta, e) = \theta e$ implies $\rho_{\theta e} = 1$; (ii) constant elasticity of substitution $w(\theta, e) = [\alpha \theta^{1-\rho} + (1-\alpha) e^{1-\rho}]^{\frac{1}{1-\rho}}$ implies $\rho_{\theta e} = \rho$. The elasticity of the wage to ability is $\varepsilon_{w,\theta} = \partial \log(w) / \partial \log(\theta)$.

Suppose utility is quasi-linear in consumption such that:

$$u(c, l) = c - h(l)$$

The agent consumes everything that is left after taxes and education investments such that $c(\theta) = w(\theta, e(\theta)) l(\theta) - M(e(\theta)) - T(w(\theta, e(\theta)) l(\theta), e(\theta))$. Solving the individual maximization problem we can define income and education wedges as:

$$\tau_y(\theta) = 1 - \frac{h'(l(\theta))}{w(\theta, e(\theta))} \quad (70)$$

$$\tau_e(\theta) = w_e(\theta, e(\theta)) l(\theta) (1 - \tau_y(\theta)) - M'(e(\theta)) \quad (71)$$

The elasticity of labor to the net of tax wage is:

$$\varepsilon = \frac{h'(l)}{h''(l) l}$$

We can write the indirect utility as $u(\theta) = c(\theta) - h(l(\theta))$. Using the Envelope we can derive the local incentive constraint:

$$\frac{\partial u(\theta)}{\partial \theta} = l(\theta) \frac{w_{\theta}(\theta, e(\theta))}{w(\theta, e(\theta))} h'(l(\theta)) \quad (72)$$

It differs from the one derived in the standard problem by the term $w_{\theta}(\theta, e(\theta))$, that takes into account the effect of ability on the wage. In the standard problem we normalized it to 1.

The resource constraint is:

$$\int (w(\theta, e(\theta)) l(\theta) - u(\theta) - h(l(\theta)) - M(e(\theta))) f(\theta) d\theta \geq E \quad (73)$$

The government assigns welfare weight $\psi(\theta)$ to each θ and solves:

$$\max_{c(\theta), l(\theta), e(\theta), u(\theta)} \int \psi(\theta) u(\theta) f(\theta) d\theta$$

subject to (72) and (73).

The first order conditions wrt $u(\theta)$, $l(\theta)$ and $e(\theta)$ are respectively:

$$\psi(\theta) f(\theta) - \lambda f(\theta) = -\mu'(\theta) \quad (74)$$

$$\mu(\theta) \left[\frac{w_\theta}{w} h'(l) + l \frac{w_\theta}{w} h''(l) \right] + \lambda [w - h'(l)] f(\theta) = 0 \quad (75)$$

$$\mu(\theta) \left[-\frac{l}{w^2} w_e w_\theta h'(l) + \frac{l}{w} w_{\theta e} h'(l) \right] + \lambda [w_e l - M'(e)] f(\theta) \quad (76)$$

Taking the integral over (74) we derive $\mu(\theta) = \Psi(\theta) - F(\theta)$. Using (75), (70) and the definition of the elasticity of labor supply we get:

$$\begin{aligned} -\mu(\theta) \frac{w_\theta}{w} h'(l) \left[1 + l \frac{h''(l)}{h'(l)} \right] &= \lambda w \tau_y(\theta) f(\theta) \\ -\mu(\theta) w_\theta (1 - \tau_y(\theta)) \left[1 + l \frac{h''(l)}{h'(l)} \right] &= \lambda w \tau_y(\theta) f(\theta) \\ \frac{\tau_y(\theta)}{1 - \tau_y(\theta)} &= \frac{\Psi(\theta) - F(\theta)}{f(\theta)} \frac{w_\theta}{w} \left(\frac{1 + \varepsilon}{\varepsilon} \right) = \frac{\Psi(\theta) - F(\theta)}{f(\theta)} \frac{\varepsilon_{w,\theta}}{\theta} \left(\frac{1 + \varepsilon}{\varepsilon} \right) \end{aligned} \quad (77)$$

The optimal income wedge is similar to the one we studied in Section 3. The formula has an extra term proportional to the elasticity of the wage to ability. Labor distortions are higher at the optimum when income is highly elastic to ability: the government distorts labor more when income is mostly explained by ability and less by effort or investment in education.

Notice that:

$$w_e l - M'(e) = \frac{\tau_e + M'(e) \tau_y}{1 - \tau_y}$$

Rearranging (76) and using (70) and (71) we get:

$$\begin{aligned} \mu(\theta) l \frac{w_e w_\theta}{w^2} h'(l) \left[-1 + \frac{w_{\theta e} w}{w_\theta w_e} \right] + \lambda \frac{\tau_e + M'(e) \tau_y}{1 - \tau_y} f(\theta) &= 0 \\ \frac{\tau_e(\theta) + M'(e) \tau_y(\theta)}{(1 - \tau_y(\theta))^2} &= \frac{\Psi(\theta) - F(\theta)}{f(\theta)} l \frac{w_\theta w_e}{w} (1 - \rho_{\theta,e}) \end{aligned} \quad (78)$$

The optimal education wedge decreases in the Hicksian coefficient of complementarity. When education and ability are complements (i.e. $\rho_{\theta e} > 0$) the government wants to discourage human capital investments in order to redistribute income more. On the other hand, if the coefficient is negative or low, education benefits low ability individuals more and a government subsidy to education helps redistribution. Suppose the wage function is $w(\theta, e) = \theta e$ and the monetary cost is linear in e , the optimal wedge is $\tau_e(\theta) = \tau_y(\theta)$. This is the special case studied by Bovenberg and Jacobs (2005) whose result is that income and education taxes are ‘‘Siamese Twins’’ and both margins should be distorted the same way. They also prove that the optimal linear education subsidy is equal to the optimal linear income tax rate, which is equivalent to making human capital expenses fully tax deductible.

9.2 A Simpler Theory of Capital Taxation

We introduce in this paragraph a continuous time model with wealth in the utility function. We study the case where utility is quasi-linear in consumption that allows us to transform the problem in a static taxation problem.

Suppose individual i has utility $u_i(c, k, z) = c + a_i(k) - h_i(z)$ where $a_i(\cdot)$ is increasing and concave and $h_i(\cdot)$ is the standard disutility from labor. Agents have heterogeneous discount rates δ_i . The discounted utility is:

$$V_i(\{c_i(t), k_i(t), z_i(t)\}) = \delta_i \int_0^\infty [c_i(t) + a_i(k_i(t)) - h_i(z_i(t))] e^{-\delta_i t} dt \quad (79)$$

Capital accumulates according to:

$$\frac{dk_i(t)}{dt} = rk_i(t) + z_i(t) - T(z_i(t), rk_i(t)) - c_i(t) \quad (80)$$

where $T(z_i(t), rk_i(t))$ is the tax paid by individual i and is dependent on income and capital returns. Wealth accumulation depends on the heterogeneous individual preferences, as embodied in the taste for wealth $a_i(\cdot)$ and in the impatience δ_i . It also depends on the net-of-tax return $\bar{r} = r(1 - T_k)$: capital taxes discourage wealth accumulation through a substitution effect (there are no income effects).

The Hamiltonian for the individual maximization problem is:

$$H_i(c_i(t), k_i(t), z_i(t), \lambda(t)) = [c_i(t) + a_i(k_i(t)) - h_i(z_i(t))] e^{-\delta_i t} + \lambda_i(t) [rk_i(t) + z_i(t) - T(z_i(t), rk_i(t)) - c_i(t)]$$

Taking first order conditions we have:

$$\frac{\partial H_i}{\partial c} = e^{-\delta_i t} - \lambda_i(t) = 0$$

$$\frac{\partial H_i}{\partial z_i} = -h'_i(z_i(t)) e^{-\delta_i t} + \lambda_i(t) [1 - T_z(z_i(t), rk_i(t))] = 0$$

$$\frac{\partial H_i}{\partial k_i(t)} = a'_i(k_i(t)) e^{-\delta_i t} + \lambda_i(t) r(1 - T_k(z_i(t), rk_i(t))) = -\lambda'_i(t)$$

Rearranging:

$$\lambda_i(t) = e^{-\delta_i t}, \quad h'_i(z_i(t)) = 1 - T_z(z_i(t), rk_i(t)), \quad a'_i(k_i(t)) = \delta_i - r(1 - T_k(z_i(t), rk_i(t)))$$

Since utility is quasi-linear in consumption, the model converges immediately to a steady state. Denote (c_i, z_i, k_i) the steady state allocation, the problem collapses to a static optimization of the following objective function:

$$V_i(\{c_i(t), k_i(t), z_i(t)\}) = [c_i + a_i(k_i) - h_i(z_i)] + \delta_i (k_i^{init} - k_i)$$

where k_i^{init} is the inherited level of capital and $(k_i^{init} - k_i)$ is the utility cost of going from k_i^{init} to the steady-state level.

The government maximizes the following:

$$SWF = \int_i \omega_i U_i(c_i, k_i, z_i) di$$

with $\omega_i \geq 0$ is the Pareto weight on individual i . The social marginal welfare weight is $g_i = \omega_i U_{ic}$.

Optimal Linear Taxes Suppose the government sets linear income and capital taxes τ_L and τ_K . The individual chooses labor and capital according to $a'_i(k_i) = \delta_i - \bar{r}$ and $h'_i(z_i) = 1 - \tau_L$ with $\bar{r} = r(1 - \tau_K)$. The government balances the budget through lump-sum transfers for a total of $G = \tau_K r k^m(\bar{r}) + \tau_L \cdot z^m(1 - \tau_L)$, where $z^m(1 - \tau_L) = \int_i z_i di$ is the aggregate labor income and $k^m(\bar{r}) = \int_i k_i$ is aggregate capital. Total consumption is $c_i = (1 - \tau_K) r k_i + (1 - \tau_L) z_i + \tau_K r k^m(\bar{r}) + \tau_L \cdot z^m(1 - \tau_L)$ and the government maximizes:

$$SWF = \int_i \omega_i [(1 - \tau_K) r k_i + (1 - \tau_L) z_i + \tau_K r k^m + \tau_L \cdot z^m + a_i(k_i) - h_i(z_i) + \delta_i \cdot (k_i^{init} - k_i)] di$$

Using the Envelope theorem we get:

$$\begin{aligned} \frac{dSWF}{d\tau_K} &= \int_i \omega_i \left[-r k_i + r k^m + \tau_K r \frac{\partial k^m}{\partial \tau_K} \right] di \\ &= r k^m \left[\int_i \omega_i \left(1 - \frac{k_i}{k^m} \right) di - \frac{\tau_K}{1 - \tau_K} e_K \right] \end{aligned}$$

where e_K is the elasticity of aggregate capital with respect to the net of tax return \bar{r} . At the optimum $dSWF/d\tau_K = 0$ and the optimal linear tax is:

$$\tau_K = \frac{1 - \bar{g}_K}{1 - \bar{g}_K + e_K}$$

where $\bar{g}_K = \int_i g_i k_i / \int_i k_i$. This is the standard formula for optimal linear taxes that we studied in Section 2 applied to capital. Notice that whenever capital accumulation is uncorrelated with social marginal welfare weights (i.e. $\bar{g}_K = 1$) the optimal tax is zero. The reason is that if capital has no tag value the government does not find optimal to tax capital for redistributive purposes. We also know from previous sections that the revenue maximizing tax rate corresponds to the case of $\bar{g}_K = 0$ and it is $\tau_K = 1/1 + e_K$.

Optimal Non-Linear Separable Taxes Suppose the government optimally sets $T_K(rk)$ and $T_L(z)$. The individual budget constraint is:

$$c_i = r k_i - T_K(r k_i) + z_i - T_L(z_i)$$

Define with $\bar{G}_K(rk)$ the average relative welfare weight on individuals with capital income higher than rk . We have:

$$\bar{G}_K(rk) = \frac{\int_{\{i: rk_i \geq rk\}} g_i di}{P(rk_i \geq rk)}$$

Let $h_K(rk)$ be the distribution of capital income so that the Pareto parameter associated to the capital income distribution is:

$$\alpha_K(rk) = \frac{rk \cdot h_K(rk)}{1 - H_K(rk)}$$

Denote $e_K(rk)$ the elasticity of capital income with respect to the net of tax return $r(1 - T'_K(rk))$. Suppose the government introduces a small reform $\delta T_K(rk)$ where the marginal tax rate is increased by $\delta \tau_K$ in a small interval of capital income from rk to $rk + d(rk)$. The mechanical effect associated to the reform is:

$$d(rk) \delta \tau_K [1 - H_K(rk)]$$

The welfare effects just weights the mechanical effect by $\bar{G}(rk)$, the social marginal welfare weight associated to capital incomes above rk . Individuals who face the increase in the tax rate change their capital incomes by $\delta(rk) = -e_K \delta\tau_K / (1 - T'_K(rk))$. There are $h_K(rk) d(rk)$ individuals in the window affected by the tax change. Therefore, the total behavioral effect is:

$$-h_K(rk) d(rk) rk \frac{T'_K(rk)}{1 - T'_K(rk)} e_K(rk)$$

Summing up the three effects and rearranging we find:

$$\frac{T'_K(rk)}{1 - T'_K(rk)} = \frac{1}{e_K(rk)} \cdot \frac{1 - H_K(rk)}{rk \cdot h_K(rk)} \cdot (1 - \bar{G}_K(rk))$$

Using the definition of the Pareto parameter we derive:

$$T'_K(rk) = \frac{1 - \bar{G}_K(rk)}{1 - \bar{G}_K(rk) + \alpha_K(rk) \cdot e_K(rk)}$$

which looks like the standard optimal non-linear tax formula.

References

- Bovenberg, L. and B. Jacobs. 2005. “Redistribution and Education Subsidies are Siamese Twins,” *Journal of Public Economics*, 89(11-12), 2005-2035
- Bovenberg, L. and B. Jacobs. 2011. “Optimal Taxation of Human Capital and the Earnings Function,” *Journal of Public Economic Theory*, 13, (6), 957-971.
- Saez, Emmanuel, and Stefanie Stantcheva. Working Paper. “A Simpler Theory of Optimal Capital Taxation.” NBER Working Paper 22664.
- Stantcheva, Stefanie. “Optimal Taxation and Human Capital Policies over the Life Cycle,” *Journal of Political Economy*, forthcoming.

10 Section 11: Non-Linear Capital Taxation

In this section we first introduce a framework to study non-linear capital taxes and establish the inverse Euler equation relation at the optimum. We then discuss implications of the inverse Euler equation, the effect of skill shocks on consumption and the trend in consumption inequality required by a Pareto-optimal allocation.¹⁸

10.1 Non-Linear Capital Taxation: Two-Periods Model

We introduce in this paragraph a two-period model with idiosyncratic uncertainty that unfolds over time. The agent chooses savings and consumption today without knowing the realization of a stochastic variable s in period 1 that can be interpreted as the agent's skills. Preferences are represented by $U(c_0, c_1, y(s)/s)$. Notice that the two periods model we saw in section 9 had a time-0 shock, while in the current model uncertainty unfolds in period 1. We will show that in the current model when preferences are weakly separable, the Atkinson-Stiglitz result does not hold.

Individuals save through a linear storage technology with rate of return $R^* = 1/q$ and the aggregate resource constraint is:

$$c_0 + q \sum_s c_1(s) p(s) \leq q \sum_s y(s) p(s) \quad (81)$$

The individual maximizes utility:

$$\sum_s U(c_0, c_1(s), y(s)/s) p(s)$$

subject to (81).

First-Best The first-best allocation is characterized by the following two first-order conditions:

$$E[U_{c_0}(c_0, c_1(s), y(s)/s)] = \lambda$$

$$U_{c_1(s)}(c_0, c_1(s), y(s)/s) = \lambda q$$

which imply:

$$E[U_{c_0}(c_0, c_1(s), y(s)/s)] = R^* U_{c_1(s)}(c_0, c_1(s), y(s)/s) \quad \forall s \quad (82)$$

Since the condition holds for every realization of s and $E[U_{c_0}(c_0, c_1(s), y(s)/s)]$ is constant, it follows that in the first best the marginal utility of consumption across all states is equalized. This implies we have full insurance and consumption is the same in every state of the world. Assuming separability in consumption and labor and taking expectations on both sides of (82) we obtain the standard Euler equation:

$$u'(c_0) = \beta R^* E[u'(c_1(s))]$$

and $c_1(s) = c_1$ for every s .

¹⁸Part of these notes is based on the Lecture notes by Florian Scheuer.

Second-Best Suppose now that s is private information of the agent. We know that we can solve the problem by invoking the revelation principle. We assume the agent reports $r = \sigma(s)$, where σ is the reporting strategy and truthtelling implies $\sigma^*(s) = s$. We denote consumption and income under the reporting strategy σ with $c_1^\sigma(s) = c_1(\sigma(s))$ and $y^\sigma(s) = y(\sigma(s))$. Incentive compatibility requires:

$$E[U(c_0, c_1(s), y(s)/s)] \geq E[U(c_0, c_1^\sigma(s), y^\sigma(s)/s)] \quad \forall \sigma, s$$

which is equivalent to:

$$U(c_0, c_1(s), y(s)/s) \geq U(c_0, c_1(r), y(r)/s) \quad \forall r, s \quad (83)$$

An allocation is feasible whenever it satisfies (81) and (83).

Given that an allocation is feasible we can ask whether free savings is feasible, meaning that extra Δ savings in period 0 leave the incentive compatibility unchanged. Suppose that preferences are quasi-linear in consumption such that:

$$U(c_0, c_1(s), y(s)/s) = \hat{U}(c_0, c_1(s) - h(y(s)/s))$$

the incentive compatibility condition holds if:

$$\hat{U}(c_0 - \Delta, c_1(s) + R^*\Delta - h(y(s)/s)) \geq \hat{U}(c_0 - \Delta, c_1(r) + R^*\Delta - h(y(r)/s))$$

this is true if and only if:

$$c_1(s) + R^*\Delta - h(y(s)/s) \geq c_1(r) + R^*\Delta - h(y(r)/s)$$

therefore, only if $c_1(s) - h(y(s)/s) \geq c_1(r) - h(y(r)/s)$ that is implied by the feasibility of the original allocation. This is always true when there are no income effects. However, if there were income effects a change in savings in period 0 would have a negative effect on the labor supply in period 1.

We study the case with income effects assuming separable preferences in consumption and labor. Consider the following preferences:

$$U(c_0, c_1(s), y(s)/s) = u(c_0) + \beta u(c_1(s)) - h(y(s)/s)$$

Consider the variation:

$$(c_0 - \Delta, c_1(s) + \delta(\Delta, s), y(s))$$

such that

$$u(c_0 - \Delta) + \beta u(c_1(s) + \delta(\Delta, s)) = u(c_0) + \beta u(c_1(s)) + A(\Delta) \quad \forall s, \Delta \quad (84)$$

for some $A(\Delta)$ such that the variation is resource neutral:

$$-\Delta + q \sum_s \delta(\Delta, s) p(s) = 0 \quad \forall \Delta \quad (85)$$

Incentive compatibility after the variation holds if:

$$u(c_0 - \Delta) + \beta u(c_1(s) + \delta(\Delta, s)) - h(y(s)/s) \geq u(c_0 - \Delta) + \beta u(c_1(r) + \delta(\Delta, r)) - h(y(r)/s)$$

Using (84), we have:

$$\begin{aligned} u(c_0) + \beta u(c_1(s)) + A(\Delta) - h(y(s)/s) &\geq u(c_0) + \beta u(c_1(r)) + A(\Delta) - h(y(r)/s) \\ u(c_0) + \beta u(c_1(s)) - h(y(s)/s) &\geq u(c_0) + \beta u(c_1(r)) - h(y(r)/s) \end{aligned}$$

which holds since the original allocation was incentive compatible. It follows that the total utility from consumption is everything that matters for incentive compatibility when it is changed independently from s .

Inverse Euler Equation Suppose a feasible allocation solves the second-best problem, the variation that we just studied should not improve the welfare of the agent. We have:

$$\begin{aligned}
0 &= \arg \max_{\Delta} \sum_s p(s) [u(c_0 - \Delta) + \beta u(c_1(s) + \delta(\Delta, s)) - h(y(s)/s)] \\
&= \arg \max_{\Delta} \sum_s p(s) [u(c_0) + \beta u(c_1(s)) + A(\Delta) - h(y(s)/s)] \\
&= \arg \max_{\Delta} A(\Delta)
\end{aligned}$$

where the FOC is $A'(\Delta) = 0$. Differentiating (84) at $\Delta = 0$ we have:

$$-u'(c_0) + \beta u'(c_1(s)) \frac{\partial \delta(\Delta, s)}{\partial \Delta} \Big|_{\Delta=0} = A'(\Delta)$$

That we can rearrange to get:

$$\frac{\partial \delta(\Delta, s)}{\partial \Delta} \Big|_{\Delta=0} = \frac{u'(c_0) + A'(\Delta)}{\beta u'(c_1(s))} = \frac{u'(c_0)}{\beta u'(c_1(s))} \quad (86)$$

using equation (85) that requires that the variation is neutral in terms of resources we find:

$$-1 + q \sum_s p(s) \frac{\partial \delta(\Delta, s)}{\partial \Delta} \Big|_{\Delta=0} = 0$$

Combining this equation with (86) we get:

$$\frac{1}{u'(c_0)} = \frac{1}{\beta R^*} \sum_s \frac{p(s)}{u'(c_1(s))} \quad (87)$$

This is the so called inverse Euler equation. Notice that we can rewrite the expression above as:

$$E\left(\frac{1}{u'(c_0)}\right) = \frac{1}{\beta R^*} E\left(\frac{1}{u'(c_1(s))}\right)$$

Using the Jensen's Inequality we obtain:

$$E(u'(c_0)) = E\left(\left[\frac{1}{u'(c_0)}\right]^{-1}\right) < E\left(\left[\frac{1}{\beta R^*} \frac{1}{u'(c_1(s))}\right]^{-1}\right) = \beta R^* E(u'(c_1(s)))$$

Since $u'(c_0) < \beta R^* E(u'(c_1(s)))$ we established that in the second-best the government will distort savings downwards such that there is a positive intertemporal wedge. The reason is that savings change the incentives to work in period 1 whenever there are income effects. By transferring money to period 1 the agent is less willing to work and more prone to imitate lower-skilled agents. By taxing savings the government can partially prevent this behavior. Therefore, it is impossible to achieve a Pareto optimal allocation when agents are allowed to undertake unlimited trading of bonds.

Dual Approach Consider the allocation in terms of utils:

$$\begin{aligned}
u_0 &= u(c_0) \\
u_1(s) &= u(c_1(s))
\end{aligned}$$

Suppose we change the allocation so that the agent is indifferent and we have:

$$u_0 + \beta u_1(s) = \tilde{u}_0 + \beta \tilde{u}_1(s) \quad \forall s \quad (88)$$

We can write:

$$\tilde{u}_0 = u_0 - \beta \Delta$$

and

$$\tilde{u}_1(s) = u_1(s) + \Delta \quad \forall s$$

Then the new allocation satisfies incentive compatibility only if the labor disutility under the new allocation is such that:

$$u_0 - \beta \Delta + \beta(u_1(s) + \Delta) - h(y(s)/s) \geq u_1(r) - \beta \Delta + \beta(u_1(r) + \Delta) - h(y(r)/s) \quad \forall r, s$$

this is true if and only if:

$$u_0 + \beta u_1(s) - h(y(s)/s) \geq u_1(r) + \beta u_1(r) - h(y(r)/s) \quad \forall r, s$$

which is true only if the initial allocation was feasible. The dual problem rewrites the objective function of the government so that at the optimum the total resource cost of the allocation is minimized:

$$\min_{\Delta} \left\{ C(u_0 - \beta \Delta) + q \sum_s p(s) C(u_1(s) + \Delta) \right\}$$

where $C(u)$ is the inverse function of $u(c)$. If an allocation is optimal, then $\Delta = 0$ must solve the problem. The first order condition evaluated at $\Delta = 0$ is:

$$-C'(u_0) \beta + q \sum_s p(s) C'(u_1(s)) = 0$$

Since $C(u)$ is the inverse function of $u(c)$, then $C'(u) = 1/u'(c)$ (we also know it from Section 3). It follows that:

$$\frac{1}{u'(c_0)} = \frac{q}{\beta} \sum_s \frac{p(s)}{u'(c_1(s))} \quad (89)$$

We can interpret $1/u'(c)$ as the resource cost of providing some incentives and therefore the inverse Euler equation is equalizing the expected resource cost of providing incentives across the two periods.

10.2 Infinite Horizon Model

The dual approach goes through very easily in an infinite horizon setting with separable preferences. The individual utility is:

$$\sum_{t, s^t} \beta^t [u(c(s^t)) - h(y(s^t)/s^t)] \Pr(s^t)$$

and s^t is the history of shocks up to time t . We implement a revelation mechanism where each agent must report her type through reporting strategies:

$$r_t = \sigma_t(s^t)$$

and truth-telling implies:

$$\sigma_t^*(s^t) = s^t \quad \forall s^t, t$$

The history of reports is:

$$\sigma^t(s^t) = (r_0, r_1, \dots, r_t) = (\sigma_0(s_0), \sigma_1(s^1), \dots, \sigma_t(s^t))$$

The dynamic incentive constraint is:

$$\begin{aligned} \sum_{t, s^t} \beta^t [u(c(s^t)) - h(y(s^t)/s^t)] \Pr(s^t) &\geq \\ \sum_{t, s^t} \beta^t [u(c(\sigma(s^t))) - h(y(\sigma(s^t))/s^t)] \Pr(s^t) \end{aligned}$$

for every σ . Start from a node s^t and set the following perturbation:

$$\tilde{u}(s^t) = u(s^t)$$

for any $s^\tau \neq s^t$ and $s^\tau \neq (s^t, s_{t+1})$ so that consumption utilities are unchanged at any node that is not s^t or any other of its direct successors. We therefore have:

$$\tilde{u}(s^t) = u(s^t) - \beta\Delta$$

and

$$\tilde{u}(s^t, s_{t+1}) = u(s^t, s_{t+1}) + \Delta \quad \forall s_{t+1}$$

The perturbation is incentive compatible if the starting allocation was incentive compatible. Moreover, the total expected utility after the perturbation is unchanged. At the optimal allocation we know that $\Delta = 0$ must solve the following problem:

$$\min_{\Delta} \left\{ C(u(s^t) - \beta\Delta) + q \sum_{s^{t+1}|s^t} \Pr(s^{t+1} | s^t) C(u_1(s^{t+1}) + \Delta) \right\}$$

The FOC is:

$$-\beta C'(u(s^t)) + q \sum_{s^{t+1}|s^t} \Pr(s^{t+1} | s^t) C'(u_1(s^{t+1})) = 0$$

which can be rearranged to get the inverse Euler equation:

$$\frac{1}{u'(c(s^t))} = \frac{1}{\beta R^*} E \left[\frac{1}{u'(c(s^{t+1}))} \mid s^t \right]$$

History Dependence When there is private information it is not Pareto-efficient to fully insure the agent in equilibrium. The government must provide the agent with the incentive to produce higher output by allowing her to receive a higher level of consumption in case of a positive skill shock. Suppose that $\beta R = 1$. Define as follows the innovation in agent's information about $1/u'(c_{t+1})$:

$$\Delta_{t+1} = \frac{1}{u'(c_{t+1})} - E_t \left[\frac{1}{u'(c_{t+1})} \right] \quad (90)$$

Define the change in future forecasts as:

$$E_{t+1} \left[\frac{1}{u'(c_{t+s+1})} \right] - E_t \left[\frac{1}{u'(c_{t+s})} \right] \quad (91)$$

The inverse Euler equation is a Martingale when $\beta R = 1$ and together with the Law of iterated expectations imply:

$$\frac{1}{u'(c_t)} = E_t \left[\frac{1}{u'(c_{t+1})} \right] = E_t \left[\frac{1}{u'(c_{t+s})} \right] \quad \forall s \quad (92)$$

Therefore the lhs of (91) becomes:

$$E_{t+1} \left[\frac{1}{u'(c_{t+s+1})} \right] = E_{t+1} \left[\frac{1}{u'(c_{t+2})} \right] = \frac{1}{u'(c_{t+1})}$$

while using (92) the rhs becomes:

$$E_t \left[\frac{1}{u'(c_{t+s})} \right] = \frac{1}{u'(c_t)} = E_t \left[\frac{1}{u'(c_{t+1})} \right]$$

Therefore:

$$E_{t+1} \left[\frac{1}{u'(c_{t+s+1})} \right] - E_t \left[\frac{1}{u'(c_{t+s})} \right] = \Delta_{t+1} \quad (93)$$

The change in future forecasts equals the the innovation in agent's information at $t + 1$. Any shock that generates a change Δ_{t+1} in $1/u'(c)$ leads to a change Δ_{t+1} in agent's forecasts of $1/u'(c)$ at any future date. This implies that skill shocks have a permanent effect on the reciprocal of marginal utilities. When utility is logarithmic permanent changes reflect directly in the level of consumption. The reason is that if an agent receives a positive skill shock, it is efficient for the government to require a higher level of output. However, since there is private information, the government must reward higher output with a lifetime increase in consumption to provide the right incentive to the agent. The inverse Euler equation governs how the increase is spread over time. In the special case of $\beta R = 1$ the increase is evenly spread over time.

Consumption Inequality Suppose that $\beta R = 1$, the inverse Euler equation implies:

$$\begin{aligned} \frac{1}{u'(c_t)} &= E_t \left[\frac{1}{u'(c_{t+1})} \right] \\ \frac{1}{u'(c_t)} + \varepsilon_{t+1} &= \frac{1}{u'(c_{t+1})} \end{aligned}$$

where ε_{t+1} has mean zero and is uncorrelated with $1/u'(c_t)$. Taking the variance:

$$Var \left(\frac{1}{u'(c_t)} \right) + Var(\varepsilon_{t+1}) = Var \left(\frac{1}{u'(c_{t+1})} \right)$$

We know from above that $1/u'(c_{t+1})$ depends on new information revealed at time $t + 1$ and therefore $Var(\varepsilon_{t+1}) > 0$. It follows that:

$$Var \left(\frac{1}{u'(c_{t+1})} \right) > Var \left(\frac{1}{u'(c_t)} \right)$$

Therefore, in a Pareto Optimal allocation the variance of the reciprocal of marginal utility grows overtime. If utility is logarithmic this is equivalent to an increase in the inequality of consumption over time. The reason for this result is that the government must provide through changes in consumption the incentives. Therefore, consumption will grow more for workers with high skill realizations.

References

Golosov, M., Tsyvinski, A. & Werning, I., 2006. New Dynamic Public Finance: A User's Guide. In NBER Macroeconomic Annual 2006. MIT Press.

Kocherlakota, Narayana R. (2010), The New Dynamic Public Finance, Princeton Press

Scheuer, Florian (2014), Lecture Notes