Some comments to the lectures on taxation

March 2015

Intention

This note attempts to provide a higher perspective - a birds view - on the topics covered so far; tax incidence, excess burden of taxation, and the design of optimal taxes. I also hope to connect the different topics. The papers that accompany each topic cover more than the core insights that we want you to learn in this course. The intention of this note is to cover the most important insights in the lectures. You should read the papers to get more meat on the bone.

Tax incidence

Who pays the taxes? How is the tax burden distributed among different groups (consumers - producers - workers - capital owners) in the economy? That question is addressed in the analysis of “tax incidence”.

Partial equilibrium

A partial analysis considers only what happens to prices in the market where there is a tax change. A partial equilibrium is incomplete, but is approximately correct if the taxed good is (i) not closely linked to other products (substitutes/complements) and (ii) makes up only a small fraction of the budget of consumers (income effect is not important).

A partial analysis also assume that the money that is collected in taxes disappears from the economy. If not we would have to discuss how the use of the revenue the government collects may affect the supply and demand for the good that is taxed and therefore also the incidence of the tax.

Suppose consumers must pay a tax τ per unit of a good they purchase. This means that consumers pay a price $q = p + \tau$, when producers obtain the price p. In a market equilibrium the producer-price will – in general – be a function of the tax rate; $p(\tau)$. So even if consumers formally pays the price, the producers will bear part of the tax burden if p declines when a tax $\tau > 0$ is introduced.

In a partial equilibrium it is the supply and demand elasticities that determine how the "tax burden" is divided between consumers and producers. It is easy to show that the change in the consumer price is given by

$$\frac{dq}{dt} = \frac{\varepsilon_S}{\varepsilon_S - \varepsilon_D}$$

Hence the consumers pay the whole burden of the tax, if supply is totally elastic, or if demand is completeley inelastic. If consumers have no alternatives (inelastic demand) they will pay the the whole tax, if producers have a lot of alternative markets they can deliver to (elastic supply) they will not pay the tax.

Based on this, what do you think happens to the cigarette price in Maine if the local government levy a local tax on cigarettes?

Salience A fundamental insight from tax incidence with well informed rational producers and consumers is that it does not matter which side of the market that formally has to pay the tax; suppliers or consumers. This is no longer true if the adjustment to taxes depends on their salience.[1] If taxes are not salient for consumers, but are salient - or more salient - for producers, it matters where the tax is levied, on producers or on consumers.

Lack of salience is, in one way, a blessing in terms of minimizing the efficiency loss associated with taxation - if consumers are unaware that they have to pay a tax on a good, demand is essentially inelastic and the efficiency loss of the tax is low. But reality bites. Consumers may misjudge, or be ignorant, of the taxes *they have to pay*, but they still have a budget constraint. This then, implies that consumers must at the end of the day adjust other consumption. The paper by Chetty at al demonstrates that the fact that sales taxes are not included on the price tag of certain goods in the US induce many consumers to ignore the tax. They also discuss the implication for welfare.

General Equilibrium

A general equilibrium analysis takes account of the fact that markets are tied together. A tax that changes the price in one market will change demand and prices in other markets. In a general equilibrium model almost "everything" can happen; it is even possible that more than 100 % of a tax levied on capital income can be shifted to wages.

It is possible to illustrate this general equilibrium effect in a simple two factor, two sector model developed by Harberger. But before we write down the equations of that

[1] In non-competitive markets or if the government regulates prices it may also matter on which side - buyers or sellers - the statutory tax burden is levied.

model, we should note that workers may also loose from a capital tax in a small open economy with a high capital mobility.

Suppose capital is fully mobile between home and abroad and that the production function at home is given by $F_H(K_H, L_H)$, satisfying standard assumptions. Being small implies that the amount of capital that is moved from home to abroad does not affect the return to capital in the "world market"

$$r^* = \frac{\partial F_H}{\partial K_H}$$

If a tax (t) is levied on capital income earned at home, the equilibirum condition becomes

$$r^* = \frac{\partial F_H}{\partial K_H}(1-t)$$

With a capital income tax, less capital is invested at home and this leads to a decline in the value of the marginal product of labour and to lower wages at home. Workers loose if a capital tax is introduced, capital owners are not affected.

The Harberger model. Something similar may happen in a closed general equilibrium model if a tax is levied on capital used in one sector of the economy. To get a taste of what is going on here, consider a closed economy with two sectors, both using capital and labour to produce output. There is constant returns to scale in production, competitive markets and capital and labour are fully mobile between sectors. Finally, the model assumes that there is a fixed amount of labour and capital in the economy.

- $X_1 = F_1(K_1, L_1)$
- $X_2 = F_2(K_2, L_2)$
  - crs $\Longrightarrow X_i = L_i f_i(k_i)$, where $k = \frac{K}{L}$
- competitive markets: $w = p_i \frac{\partial F_i}{\partial L_i}$ & $r = p_i \frac{\partial F_i}{\partial K_i}$
- homothetic preferences: $X_1 = g(p_1/p_2)I$ & $X_2 = h(p_1/p_2)I$
- $L_1 + L_2 = \bar{L}$, $K_1 + K_2 = \bar{K}$

We have ten equations to determine ten endogenous variables.

Capital tax in sector 2

Suppose a small tax t is imposed on capital income earned from renting capital to sector 2. This tax alters the equilibrium condition on capital returns in sector 2: $r = (1-t)\, p_2 \frac{\partial F_2}{\partial K_2}$. All other equations are the same, but it is of course not the same prices and quantities that satisfies the equilibrium equations.

Since K and L are exogenously given, the incidence on capital and labour is fully specified by $\frac{dw}{dt}$ and $\frac{dr}{dt}$. It would be more complicated if leisure and future consumption were endogenous variables, in that case the elasticity of capital and labour supply would also have an impact on tax incidence. To characterize the wage and rental effect of a tax on capital used in sector 2, we have to do comparative statistics on the equilibrium equations; total differentiate the system of equations with respect to the tax change. It involves quite a bit of algebra.

If we focus on intuition, there are two main forces that determine how this tax will affect pre-tax rental price of capital and wages.

Substitution effect: Producers in sector 2 will substitute capital for labour and this will reduce the rental price of capital. To see this suppose those who rent capital to sector 2 require the same return as before $\rightarrow$ relative price of capital to labour has increased and producers want to reduce their use of capital and use more labour. Since capital is fixed and demand falls there is a pressure towards a lower rental price for capital.

Output effect Although the producers of good 2 substitute capital for labour, production costs increase. This means that good 2 becomes more expensive and demand shifts towards sector 1. How this output effect impacts on the demand for capital, and therefore in the price of capital (the tax incidence) depends critically on the relative capital intensiveness in the two sectors.

If sector 2 is most capital intensive ($\frac{K_2}{L_2} > \frac{K_1}{L_1}$), overall demand for capital goes down when output shifts towards product 1 and the rental price of capital falls further. It is possible (if demand elasticities are high (1 and 2 are close substitutes) that the overall negative effect on the rental price is higher than the tax rate.

If sector 1 is the most capital intensive ($\frac{K_1}{L_1} > \frac{K_2}{L_2}$), demand for capital increases as demand shifts towards the sector that uses more capital. In this case the output effect counters the substitution effect and may - at the end of the day - lead to an increase in the pre-tax rental price that is higher than the tax rate.

Literature

There are two pieces on the syllabus that covers tax incidence.

- L. Kotlikoff and L. Summers. "Tax Incidence," in A. Auerbach and M. Feldstein, Volume 2, 1043-1092. Required reading: Sections 0, 1, 2, 3.1, and 4.4.

- Chetty, A. Looney, and K. Kroft. "Salience and Taxation: Theory and Evidence." American Economic Review 99(4): 1145-1177, 2009. Section V.C.

The efficiency loss and excess burden of taxation

In a competitive economy without externalities (and with convex preferences and production technologies) we know from the 1. Welfare Theorem that there exists a decentralized equilibrium with prices that clears all markets and that is Pareto Efficient. The competitive price vector guarantees that all consumers have the same marginal rate of substitution between any pair of goods, which again is equal to the marginal rate of transformation between these goods on the production side.

A deadweight loss arise if prices are distorted from this utopian benchmark. The deadweight loss measures the economic decline (in terms of lower consumer and producer surplus) caused by the price distortion. When it is taxes that distorts prices, when taxes drive wedges between consumer and supplier prices, we talk about the excess burden of taxation (or the deadweight loss associated with taxation). The excess burden of a tax (or of a tax system) is the economic loss tax payers experience, over and above the tax revenue that is collected by the government. If consumers experience a loss, measured in NOK, of magnitude I when a tax is introduced and the revenue collected is R, then the deadweight loss is $I - R$.

Two immediate observations:

1. If the government collects its revenue through a lump sum tax there is no excess burden of taxation. There is a tax burden –also a gain of course, if the taxes are used to produce public or private goods, or to attain a more desirable distribution of economic resources – but there is no excess burden. A lump sum tax - for example the same tax on every household independent on their economic outcomes - is impractical and political impossible. In all economies taxes are levied on transactions, on tax bases that are endogenously determined by the behavior of economic agents.

2. If all goods that are consumed in the economy could be taxed at the same rate the tax system would not change relative prices, and there would not be an excess burden. But, since leisure is one of the goods that are consumed, this mean that such a tax system must be able to observe how much leisure a household consumes, and tax it. This is also practically (due to information constraints) impossible. Hence there will always be an efficiency loss associated with a tax system, and it is important to understand how we measure this loss and the factors that determine the magnitude of the loss.

3. If there are externalities in the production and consumption of some goods, a tax system that corrects market prices for these externalities may have a negative excess burden. A tax on externalities will raise funds that the government can use to produce public goods (or redistribute income), or the government can reduce other distorting taxes. In addition a tax on "pollution" may actually improve efficiency

by correcting the market price for the externality. This is sometimes called the double dividend of taxation. It is not covered in our course, but the intuition is straightforward.

Measuring the excess burden of taxation The excess burden is the monetary loss that consumers experience in addition to the tax revenue collected. There are two ways to measure the excess burden of taxation.

One is to ask what sum of money consumers would request in order to attain the utility they had before the tax was introduced. This is the compensating variation. If we subtract the tax revenue collected from this amount we get one measure of the excess burden.

The other thought experiment is to ask how much money consumers are willing to give up if the government abolish the tax. This is the equivalent variation. If we take this amount and subtract the revenue collected by the government we obtain another measure of the excess burden. Unless we impose structures on the preferences these two experiments will give different magnitudes of the excess burden. The mathematical expression of these measures are provided in the lecture notes.

We can illustrate these magnitudes in a price-quantity diagram, where we draw the compensated demand curves for the taxed good. We obtain the compensated demand curves by differentiating the expenditure function with respect to the price of the good. When a tax on one good is imposed we can trace out the compensated demand curve and integrate over the price change due to the tax and obtain the excess burden of a tax.

It is important to understand that it is the compensated demand curve we need to consider, the income effects that alter uncompensated demand (when income fixed) does not distort relative prices and will not create an excess burden of taxation. That is the reason why we in the lectures started with a simple case with quasi linear utility (in that model there are, by construction, no income effects for the taxed good). But in the more general case we must derive the compensated demand curves by differentiating the expenditure function.

Some important observations

- If a small tax $d\tau$ is introduced on one good, while the other goods are untaxed, there is, to a first order approximation, no excess burden associated with the tax. This is intuitive. Draw a figure and you will see that the the income collected by the government is approximatly equal to the value lost by the consumer if the tax is small. The efficiency loss is a triangle, and vanishes as the tax rate approaches zero.

- If we are considering a tax increase on a good that is already taxed, there will be a first order loss in efficiency, since the income collected by the government is lower than the value that is lost by the consumer (the price was distorted by the tax, at the outset). Deadweight loss of taxation is roughly (exact if compensated demand is linear) equal to the square of the tax. This is easy to see in a figure (see the lecture notes).

- It is also easy to see that the excess burden from taxing a good increases in the elasticity of the compensated demand. Excess burden increases roughly in proportion to the compensated elasticity of demand.
- Hence, if we consider different goods in isolation (no cross price effects) on would conclude that (i) in terms of minimizing excess burden of taxation it is wise to have a broad tax bases (many small taxes instead of one big). (ii) goods with a high own price compensated price elasticity should be taxed leniently. While there is something to these intuitions, they are altered if we also consider cross price effects. This is the problem analyzed in the Ramsey model of optimal linear tax structure.
- We got a taste of the Ramsey result in our analysis of the effects of introducing a tax on one good, when there are already other goods that are taxed. We showed that (i) Even if the new tax that is introduced is very very small it will have a first order effect on welfare since the tax typically will change demand of other goods that are taxed and, where there is a distortion. Put differently, the fact that the new tax changes demand for already taxed goods implies that it has a direct effect on the tax revenues collected by the government. (ii) If the tax that is introduced increases the compensated labour supply (if the taxed good is complementary to leisure (for example golf clubs) then there will actually be a first order reduction in the deadweight loss of this tax. This is the Corlett & Hague result that we will discuss in more detail in the Ramsey framework.

In the analysis of excess burden we used a Taylor approximation to measure the excess burden. This is a useful technique since it enables us to represent - albeit only approximately - the change in excess burden of changing a tax rate (or of introducing a tax) with a simple functional form. This simplification has also great empirical appeal since we need only local information to estimate the magnitudes (but of course it is only an approximation) of the excess burden. A first order approximation implies that we linearize the excess burden of taxation around the situation that prevailed before the tax is introduced. Higher order approximations give a better "fit", but require more information. A first order approximation only requires information about the slope of the compensated demand functions, a second order approximation requires information about the curvature etc.).

Literature There is only one paper on the syllabus that directly discuss the excess burden of taxation, but all papers on the Ramsey problem are relevant. In its simplest form the Ramsey problem is to choose a *tax system* that minimizes the excess burden of taxation.

- Stiglitz, J.E. (2000). Economics of the Public Sector, Norton, 3rd ed., Pp. 518-541 K

Optimal taxation

Given that any realistic tax system involves efficiency losses, it is important to design one that minimizes the deadweight loss, given a government revenue requirement. There are other costs than lost efficiency that matter for the design of an optimal tax system. One additional concern, a concern we will study, is the distributional outcomes of the tax system (and the welfare system more generally); does the tax system lead to a desirable distribution of income in the economy.

In addition there are administrative costs associated with designing and managing a tax system. In a full analysis of optimal tax systems, these costs should also be considered, but they are ignored here. There are also political costs associated with taxation. In the real world taxes are not chosen by a benevolent planner, but by politicians who are driven both by ideology, and by a desire to win the next election. The political costs of taxation will matter for which tax system that will be implemented. This problem is studied in the theory of Public Choice and in Political Economy.

In addition to the objectives of the government, how, for example, they trade off efficiency and equal distribution of income, there are constraints that shape the optimal tax policy. The Ramsey model considers - without any deeper justification - a restricted set of taxes, namely taxes that are proportional to the tax base.

The Mirrlees model allows for non-linear taxes. In one way this is what sets Mirrlees apart from Ramsey. A non-linear tax is especially relevant for labour and capital income (direct taxes). An important constraint in the Mirrlees framework is that it is impossible to tax individuals' income potential. The income potential is assumed to unobserved by the government; it is private information for the tax payers. The government can use non-linear taxes, but they can only tax the realized income of individuals, not their income potential. This information constraint imposes a "self selection" constraint on the governments tax problem. Individuals choose their income (labour supply) so as to maximize their own utility, and the government must implement a tax system that takes account of individuals response. This is a general principle. A sophisticated government must calculate how the public respond to a policy and then choose the optimal policy given this response; one must choose the optimal policy among those who are incentive compatible.

In the Mirrlees framework the incentive problem arises because individuals have private information about their productivity. An interesting question in the Mirrlees model is the interaction between an optimal non-linear income tax and a proportional consumption tax. Is it, for example, desirable to complement the income tax with a tax on the consumption of goods, and which goods should be taxed and why. Is it optimal to complement the non-linear income tax with a tax on capital income? And what implications does the income tax have for the provision of public goods?

In most countries the income tax system is indeed non-linear in income. But the tax liability of income earners is often a simple piecewise linear function of income (the marginal tax rate varies with income, but is constant within large intervals if income). A relevant question is how a change in one of the marginal tax rates affect efficiency and distribution when the income tax is piecewise linear. The simplest problem to analyze is a change in the the marginal tax rate of those in the top income bracket. If the marginal tax rate is changed for lower incomes there is an additional effect since a change of this rate changes the average tax rate for those who have income in a higher tax bracket. Another interesting question is how the tax and transfer system affect the extensive margin; whether or not individuals seek paid work. This decision is not addressed in the standard Mirrlees model.

It is these three models of optimal taxation that are covered in the course (Ramsey, Mirrlees and the piecewise linear income tax model). Let me (briefly) list some of the main insights from the models.

Ramsey

In the most basic version of this model there is only one household consuming $N + 1$ goods and N of these goods can be taxed. The good that cannot be taxed is often taken to be leisure. But it could be another good that is untaxed (in a model with exogenous income (no labour supply)). If all goods could be taxed the problem is trivial; tax all goods at the same rate, this will leave the relative prices unchanged and there will be no excess burden with taxation.

To see this assume that pre-tax prices are fixed (normalized to 1): $q_i = 1+\tau_i$ and wage is given by w, and the household has a time endowment of H. The budget constraint is $\sum_i (q_i x_i) = z + (H - l)w$, where z is exogenous income. We can rewrite the budget constraint and move all the consumed goods (leisure included) on the left hand side of the budget equation; $\sum_i (q_i x_i) + wl = z + Hw$. The right hand side is the full income of the household. It is easy to see that if all goods (leisure included) could be taxed at a rate τ, that would be equivalent with having a lump sum tax on the endowments $(z + Hw)$.

Let us now consider the problem with leisure as the untaxed good (this is realistic).

The household solves the following problem

$$Max\ L_H = u(x_1, ..x_N, l) + \alpha \left[z + (H - l)w - \sum_i (q_i x_i) \right]$$

The first order condition for good i is then

$$u'_{x_i} - \alpha q_i = 0$$

There are $N + 1$ first order equations. If we solve the problem we obtain demand choices for each good expressed as functions of after tax prices and the wage rate and exogenous income. If we insert optimal choices into the utility function we have the indirect utility function: $V(\mathbf{q}, w, z)$. A benevolent government with a revenue requirement R chooses taxes to solve

$$Max\ L_G = V(\mathbf{q}, w, z) + \lambda \left[\sum_i (\tau_i x_i) - R \right]$$

The first order condition for tax on good i is

$$\frac{\partial V}{\partial q_i} - \lambda \left(x_i - \sum_j \tau_j \frac{\partial x_j}{\partial q_i} \right) = 0$$

By applying the envelope theorem we obtain $\frac{\partial V}{\partial q_i} = -\alpha x_i$ and we can rewrite the first order condition as

$$(\lambda - \alpha) x_i + \lambda \sum \tau_j \frac{\partial x_j}{\partial q_i} = 0$$

Using the Slutsky decomposition $\frac{\partial x_j}{\partial q_i} = \frac{\partial h_j()}{\partial q_i} - x_i \frac{\partial x_j}{\partial z}$ and rearranging, using the fact that the substitution effect is symmetric $\frac{\partial h_j}{\partial q_i} = \frac{\partial h_i}{\partial q_j}$ allow us to write the formula for optimal tax policy as

$$\frac{1}{x_i} \sum_j \tau_j \frac{\partial h_i}{\partial q_j} = \frac{-\left(\lambda - \alpha - \lambda \frac{\partial}{\partial z}\left(\sum_j \tau_j x_j\right)\right)}{\lambda}$$

The term in the brackets in the nominator is the social value of collecting one unit of revenue through a lump sum tax: $\theta = \lambda - \alpha - \lambda \frac{\partial}{\partial z}\left(\sum_j \tau_j x_j\right)$. The social value (the increase in the social welfare function) of one NOK at the hands of the government is equal to λ; taking one NOK away from the consumer reduces her welfare with α utils. In addition, a consumer that becomes poorer will adjust consumption and this will have an impact on tax income for the government, the social value of this effect is $\lambda \frac{\partial}{\partial z}\left(\sum_j \tau_j x_j\right)$.

Using this notation a tax system that minimizes the excess burden can be written as:

$$\frac{1}{x_i} \sum_j \tau_j \frac{\partial h_i}{\partial q_j} = \frac{-\theta}{\lambda} \tag{1}$$

This must hold for any of the taxed goods i.

Interpretation Note that the rhs is independent of i (it is the same for all taxed goods). The lhs is approximately the percentage drop in compensated demand for good i caused by the tax system. The optimality condition is that this drop - often called the index of discouragement - should be the same for all goods. The tax system should be designed in such a way that the percentage drop in compensated demand should be the same across all taxed goods.

Unit free measure - elasticities Multiply and divide the rhs of (1) with $(1+\tau_j)$ and denote ε^c_{ij} as the compensated demand elasticity of good i with respect to the price of good j. We can write the optimality condition as

$$\sum_j \frac{\tau_j}{1+\tau_j}\varepsilon^c_{ij} = -\frac{\theta}{\lambda} \tag{2}$$

Consider a case with two taxed good, indexed 1 and 2 and leisure is good 0. We have

$$\frac{\tau_1}{1+\tau_1}\varepsilon^c_{11} + \frac{\tau_2}{1+\tau_2}\varepsilon^c_{12} = -\frac{\theta}{\lambda}$$
$$\frac{\tau_1}{1+\tau_1}\varepsilon^c_{21} + \frac{\tau_2}{1+\tau_2}\varepsilon^c_{22} = -\frac{\theta}{\lambda}$$

If we denote $T_i = \frac{\tau_i}{1+\tau_i}$ and if solve the equations we get

$$\frac{T_1}{T_2} = \frac{\varepsilon^c_{22} - \varepsilon^c_{12}}{\varepsilon^c_{11} - \varepsilon^c_{21}} = \frac{\varepsilon^c_{22} + \varepsilon^c_{10} + \varepsilon^c_{11}}{\varepsilon^c_{22} + \varepsilon^c_{20} + \varepsilon^c_{11}} \tag{3}$$

The second equality in (3) comes from the fact that compensated demand is homogenous of degree 0 which implies that $\varepsilon^c_{10} + \varepsilon^c_{11} + \varepsilon^c_{12} = 0$.

Inverse elasticity rule: If the cross price elasticities are 0 we obtain the result that the tax rates should be inversely related to the compensated own price elasticity. This is a very restrictive assumption.

Corlett Hague rule: We have $\frac{T_1}{T_2} = \frac{\varepsilon^c_{22}+\varepsilon^c_{10}+\varepsilon^c_{11}}{\varepsilon^c_{22}+\varepsilon^c_{20}+\varepsilon^c_{11}}$: since $\varepsilon^c_{11} + \varepsilon^c_{22} < 0$ we have $T_1 > T_2$ if good 1 is relatively more complementary with leisure than good 2: $\varepsilon_{20} > \varepsilon_{10}$. This is an important result. The intuition is that due to the taxes on consumption (and not on leisure) the marginal rate of substitution between labour and leisure is distorted. It is then optimal to differentiate taxes on consumption goods – create a distortion in the choice between different consumption goods – in order to induce consumers to reduce their consumption of leisure. Households enjoy too much leisure in this model, because of the consumption tax.

The optimality of a differentiated tax on commodities is an example of a more general second best logic; if there are distortions in one market, for example in the supply of labour, it may be optimal to introduce a distortion in another market (distort prices on consumption goods). The policy recommendation is that the government should tax golf clubs or fishing rods with a higher rate than other goods, because these goods are leisure goods that increase the costs of working (increase the value of leisure).

Note also that a uniform tax on consumer goods is desirable in this model if all goods have the same compensated cross price elasticity with leisure. This will be the case if the utility function is quasi-separable in leisure and other goods. If we took into account other costs and constraints, there may also be other reasons for having a uniform tax; the administrative costs of having a system with uniform tax rates is probably lower than with s system with differentiated rates. Differentiated rates may also give producers an incentive to misclassify their products.

Heterogeneous households

The model above simplifies a lot by having only one household (or implicitly assumes that all households are equal). If that really was the case it seems overly artificial to assume that the government cannot use a lump sum tax to collect revenue.

If we remove the "all households are equal" assumption we get a more interesting situation in which the efficiency rule for optimal taxes is adjusted by the fact that the government also have distributional objectives. Although the algebra gets a bit involved here it is pretty obvious what will happen. The efficiency rule will be adjusted by distributional concerns: if a high tax on good i is recommended on efficiency grounds, the tax will be adjusted downwards if this good is consumed relatively intensively by groups that have a high marginal utility of income (high α).

A reinterpretation of Ramsey result Before we consider the heterogenous household model, it is instructive to look at the problem we solved above (Ramsey model with only efficiency concerns) from a slightly different angle. We can rewrite the formula for optimal taxation with a representative household. Recall that the first order condition for optimal taxes is (here we have moved the quantity of the good over to the rhs)

$$-\sum_j \tau_j \frac{\partial h_j}{\partial q_i} = \frac{\theta}{\lambda} x_i \,\forall i \tag{4}$$

The left hand side in (1) is equal to the increase in excess burden of taxation of introducing a tax on good i (it is equal to $\frac{dEB}{d\tau_i}$ when there is no initial tax on good i). The marginal increase in revenues by introducing a tax on good i is given by $\frac{dR}{d\tau_i} = x_i + \sum_j \tau_j \frac{\partial h_j}{\partial q_i} \Longrightarrow x_i = \frac{dR}{d\tau_i} - \frac{dEB}{d\tau_i}$, we can therefore write the right hand side of equation (1) as $\frac{\theta}{\lambda}\left(\frac{dR}{d\tau_i} - \frac{dEB}{d\tau_i}\right)$ this then means that a first order condition for the optimal tax structure imply that the excess burden per NOK in revenue is equal across all taxed goods (there was a mistake in the seminar solution)

$$\frac{\frac{dEB}{d\tau_i}}{\frac{dR}{d\tau_i}} = \frac{\theta}{\lambda - \theta}.$$

This is quite intuitive: if the excess burden of collecting one NOK in revenue was lower for i than for j the government should increase the tax rate on i and lower it on j. The condition is parallel to the optimality condition in consumer theory where the marginal utility of per NOK spent on a good should be equal across all goods ($\frac{\frac{\partial U}{\partial x_i}}{p_i} = \alpha \ \forall i$)

Optimal consumption good taxes with many households

This rule will be modified when distributional concerns are introduced, that is, when households differ and a change in their consumption is assigned different social value. When the government evaluates a tax policy according to the welfare functions $W\left(V^1(\mathbf{q}), V^2(\mathbf{q})...V^H(\mathbf{q})\right)$ we should expect that the optimal tax structure will be modified by the fact that different households have different marginal utility of money ($\alpha^j \neq \alpha^k$) and different welfare weights ($\frac{\partial W}{\partial V^j} \neq \frac{\partial W}{\partial V^k}$) .

If a good i is consumed disproportionally much by individuals who have a high marginal utility of money (because they are poor) we expect the tax rate to be adjusted downwards compared to the optimal policy when only efficiency matters. This is exactly what will happen.

With heterogenous households the first order condition for optimal tax structure is given by

$$\sum_h \frac{\partial W}{\partial V^h}\frac{\partial V^h}{\partial q_i} + \lambda\left[\sum_h x_i^h + \sum_j \tau_j \sum_h \frac{\partial x_j}{\partial q_i}\right] = 0 \tag{5}$$

Using the envelope result $\frac{\partial V^h}{\partial q_i} = -\alpha^h x_i^h$, the Slutsky equation (decomposing the price effect on demand into a substitution and income effect (z is income) and letting $\sum_h \frac{\partial h_j^h}{\partial q_i} = H_{ji}$ and $X_i = \sum_h x_i^h$ we can write (2) as

$$-\sum_j \tau_j H_{ji} + X_i\left[\lambda - \left(\frac{\sum_h x_i^h\left(\frac{\partial W}{\partial V^h}\alpha^h + \lambda\sum_j \tau_j \frac{\partial x_j^h}{\partial z^h}\right)}{X_i}\right)\right] = 0 \tag{6}$$

The term $\frac{\sum_h x_i^h\left(\frac{\partial W}{\partial V^h}\alpha^h + \lambda\sum_j \tau_j \frac{\partial x_j^h}{\partial z^h}\right)}{X_i} = \beta_i$ is the social marginal welfare of income associated with good i. It is the social value of a marginal increase in income for household h :$\beta^h = \left(\frac{\partial W}{\partial V^h}\alpha^h + \lambda\sum_j \tau_j \frac{\partial x_j^h}{\partial z^h}\right)$ times the households share of the consumption of this good $\frac{x_i^h}{X_i}$. We can now write (3) as

$$-\sum_j \tau_j H_{ji} = \frac{\lambda - \beta_i}{\lambda} X_i \tag{7}$$

Comparing (4) with (7) we can see that the right hand side is no longer independent of i: It depends on the social marginal welfare of income associated with good i: β_i. Hence if we rewrite this equation as the ratio between the marginal excess burden of an increase in tax i and the marginal revenue of an increase in tax i we get

$$\frac{\frac{dEB}{d\tau_i}}{\frac{dR}{d\tau_i}} = \frac{\lambda - \beta_i}{\beta_i}$$

Which means it is no longer optimal to set the marginal excess burden equal for all sources of revenue; with heterogenous households we will adjust taxes according to which households that consume the good; goods with a high correlation between $\frac{x_i^h}{X_i}$ and β^h will have a high β_i and a lower tax is optimal.

The relevance of the Ramsey model

It is not easy to implement an optimal differentiated consumption tax in the spirit of Ramsey since it is difficult to estimate all the parameters that are needed to design an optimal system. The model should be considered as a framework that (i) specifies the forces that matter for designing a tax system that minimizes the efficiency loss, and (ii) illustrates how the optimality conditions are altered when distributional objectives are incorporated in the model

Another unrealistic feature is that this model disregards a very important feature of any advanced tax system, namely that a lot of the tax revenue is collected through a non-linear income tax. Hence a question of more practical relevance is if and how consumption good taxes should complement a non-linear income tax. In order to address that problem we first have to study the design of a nonlinear income tax (note that a system with proportional taxes on consumption is equivalent with a proportional tax on labour income). This is where the Mirrlees model comes in.

Non-linear income tax - the Mirrlees framework

Direct taxes are often non-linear with lump sum transfers to those who do not participate in the labour market and with a marginal tax rate that varies with income earned for those who participate. The optimal non-linear tax must balance distributional and efficiency concerns.

The efficiency loss associated with a non-linear (excess burden) depends on two behavioral responses. One response is whether or not an individual wants to work. This is the participation decision which depends on the extra money a person earns if he or she decides to participate in the labour market. The participation tax rate matters for this decision. Of course there is a bunch of other things that also matter if such as; does your spouse work, your friends work, work norms etc). *The participation tax rate* is defined by $\tau^p(y) = \frac{(T(y)-T(0))}{y}$. The participation tax rate matters for the incentives to participate in the labour market. We can express net earnings in terms of the participation tax rate

$$y - T(y) = -T(0) + y - (T(y) - T(0)) = -T(0) + y(1 - \tau^p(y))$$

Hence a person keeps a fraction $(1 - \tau^p(y))$ of his or her earnings (y).

The other response is the work effort a person supply if he or she who decide to participate. This is the intensive margin which depends on the marginal tax rate. *The Marginal tax rate* is given by the derivative of the tax-function. Let y be the pre-tax income and $T(y)$ the tax paid if income is y. Households after tax income is then given by $(y - T(y))$. $T'(y)$ is the marginal tax rate associated with income y: An individual keeps $1 - T'(y)$ of one extra NOK in income evaluated at y.

A tax scheme is progressive if the tax rate increases as the tax base (income) increases. That will be the case if the the marginal tax rate is (weakly) higher than the average tax rate. Some measures of progressiveness that are used in the literature. The elasticity of the tax bill with respect to pre-tax income: $\frac{T'(y)}{T(y)/y}$. The elasticity of the residual income with respect to pre-tax income: $\frac{1-T'(y)}{1-T(y)/y}$.

Optimal non-linear taxation with no behavioral response (exogenous income)
It is the behavioral (labour supply) response to the income tax that makes the problem of optimal income taxation difficult and interesting. To see this we start with a model where individuals have a fixed income that can be taxed (it is as if the government can tax the income potential of individuals).

Assume all individuals have the same strictly increasing and concave utility function $u(c)$. Income y is fixed (exogenous) and consumption is equal to income after tax: $c = y - T(z)$. Government maximizes Utilitarian objective: $\int_0^{\bar{y}} u(y - T(y)h(y)dy$, where $h(y)$ is the distribution (pdf) of income over the interval of income in the economy $[0, \bar{y}]$. Budget constraint $\int_0^{\bar{y}} T(y)h(y)dy = R$ (multiplier λ). Lagrangian is then

$$L = \int_0^{\bar{y}} (u(y - T(y) + \lambda [T(y) - R]) h(y)dy.$$

F.o.c:

$$\begin{aligned} 0 &= \frac{\partial L}{\partial T(y)} = -u'(y - T(y) + \lambda)h(y) \\ 0 &= u'(y - T(y) = \lambda \end{aligned}$$

$$\Longrightarrow y - T(y) = c = \lambda$$

$$c = \bar{y} - R$$

Equalization of after tax income, which means there is a 100% marginal tax rate of earnings above this level.

It is easy to understand that with equal social welfare weights (utilitarian) and diminishing marginal utility optimality it is equal consumption that maximizes social welfare. If the tax policy did not equate consumption if , say, j got more consumption than i, a transfer from j to i would, due to decreasing marginal utility of consumption, increase social welfare. The implication is that in a standard Mirrleesian framework, with endogenous income, the government would, if it could, equate everyone's income. It is information constraints that prevents this solution.

But is this reasonable? Suppose the government could observe the income potential (wage) of a person when he or she enters the labour market. Would we accept a lump sum tax that varied with the wage, but is independent of this persons actual income. Maybe, if we could control for individual effort costs, and if the income potential was independent of past choices (truly exogenous). A liberal egalitarian would for example argue that it is fair to redistribute income differences that are due to luck, but not differences that arise because of effort. Amartaya Sen and others argue that it is not ex post consumption we should equate, but the capabilities to live a full life.

Suppose income is not given, but earned by individuals who have the same utility function over leisure and consumption but who differs in productivity. In this case the utilitarian solution (if the government can observe individual productivity) is that those with a high earning capacity will work more, but consume the same amount. High productivity agents will then end up with a lower utility. We understand that this policy is not incentive compatible if individuals have private information about their productivity.

Optimal non-linear taxation with behavioral response (The Mirrlees problem) individuals maximize $u(c, L)$ s.t. $c = wL - T$, $w =$ wage rate, L is labour supply and T are taxes.individuals differ in wages (abilities) which is distributed with density $f(w)$.

Government maximizes a social welfare function $W\left(u(c,L)\right)$ (increasing and concave)

$$SWF = \int W\left(u(c,L)f(w)\right)dw$$

Subject to a budget constraint

$$\int T(wL)f(w)dw = R$$

and a behavior constraint (incentive compatibility (IC) constraint)

$$w(1-T')u_c + u_L = 0$$

Mathematically this is a more complex problem to solve than the Ramsey problem. The problem is not to find the optimal value of a variable (a tax rate), the problem is to find an optimal tax *function* $T(y)$. There are relatively few general insights we can draw from this model, unless we put more structure on parameters and functions.

It will typically be the case that with a concave social welfare function (W) we have $T < 0$ for individuals with low wages and $T > 0$ for individuals with higher income; the degree of redistribution depends on the concavity of W and the elasticity of labour supply. There is a trade off between efficiency and redistribution.

Another robust result is that it is never optimal to have a negative marginal tax rate. We should never have $T' < 0$. A negative average tax rate is, as we argued in the paragraph above, consistent with the model (those with low income may for example get a large lump sum transfer, but for every NOK they earn extra they should pay a positive tax). It is a negative marginal tax rate that is inconsistent with the model. This is a robust result in the framework laid out above. In another framework with a non-utilitarian government or with individuals who differ both in productivity and the value of leisure, the result may not hold.

Another robust insight is that $T' = 0$ for the individual with the highest ability (w) (again, it is the marginal tax rate that should be zero, not the average tax rate). This result is quite obvious and not very useful. Suppose the highest ability person earns an income y^{max}, suppose also that there is a positive marginal tax rate evaluated at y^{max}; that is $T'(y^{max}) > 0$. Now consider a reform that sets $T'(y^{max}) = 0$. There is no effect on tax income for the government since $T'(y^{max})$ is the tax rate of earning slightly above y^{max} and there is no one there. On the other hand the reform will (weakly) increase the welfare of the top person. Hence it is a Pareto improvement of set the $T'(y^{max}) = 0$.

The result is not very useful since it tells us nothing about the marginal tax rate slightly below the maximal income. Changing the marginal tax rate at a lower level will clearly reduce – given that we are on the right side of the laffer curve – the tax income for the government since it reduces the average tax of those with higher earnings.

The two-type model The basic principles of the non-linear income tax with information constraint (government cannot tax income potential only the earned income) can be illustrated in a simple model with two types of households.

The types of households are indexed $i = \{L, H\}$. In order to consume (C) a household needs to earn income (Y) and in order to earn income they must work. Suppose one unit of labour gives w_i with $w_H > w_L$ units of income. Normalizing the unit price of consumption to 1 we have $C_i = Y_i = w_i L_i$ in the absence of taxes and transfers. The households preferences over consumption and labour (leisure) is given by $U(C, L)$, with $U_C > 0$ (the marginal utility of consumption is positive) and $U_L > 0$.

Abstract from the the types of household for a moment (no subscript). A household solves $U(C, L)$ s.t. $wL = C$. The first order condition for an optimum is (the usual MRS = MRT conditions): $\frac{-U_L}{U_C} = w$. Later we will assume that the government only observes Y, not w and L separately. It is useful to rewrite utility in terms of what is observable for the government. With $L = \frac{Y}{w}$ we obtain $U(C, L) \equiv U(C, \frac{Y}{w}) = V(C, Y)$. Indifference curves increases in the Y, C space, and the indifference curve are flatter the higher the wage is: $\frac{-V_Y}{V_C} = \frac{-U_L}{wU_C}$. Hence without any taxes optimality requires that $\frac{-V_Y}{V_C} = \frac{-U_L}{wU_C} = 1$ (MRS=MRT). Draw figures!

Let us return to the two types and introduce a government that imposes a tax on the two types of households. The households budget constraint is then given by $C_i = Y_i - T(Y_i)$ and the governments budget constraint is $T(Y_L) + T(Y_H) \geq R$, where R is the revenue that government needs in addition to the amount it (may) redistribute to the L household ($T(Y_L)$ might very well be negative). We can write the budget constraint as $Y_H - C_H + Y_L - C_L \geq R$.

Direct implementation Let us wait before we characterize the optimal shape of the tax function (with only two types, we only need to consider two segments of this scheme (low and high)). Let us instead assume the following direct mechanism: The government asks which type the household is, and offers a bundle $\{C_L, Y_L\}$ to the household if it says it is a $L-$type and the bundle $\{C_H, Y_H\}$ if it is of the $H-$type. Note that if an $L-$type takes the bundle designed for her she has to work $\frac{Y_L}{w_L}$ hours while a $H-$type has to work only $\frac{Y_L}{w_H}$ hours. So unless the C_L is quite a bit lower than C_H a $H-$type may pretend to be L and consume a lot of leisure.

We will characterize Pareto efficient bundles using this direct method and then we will find a tax scheme that makes it possible to implement this solution. Pareto efficiency requires that it is impossible to improve the conditions for one person (group) without worsening it for another person (group). Hence in this framework, we characterize the Pareto optimal allocation by maximizing the utility of Hgiven a constraint that the utility of L should be at least at a certain level (what utility level one requires for L depend on the distributional concerns of the government).

In addition to this constraint, and the governments budget constraint, there are two "truth telling" constraints; that is, the government should make sure that the bundles offered induce the households to choose the bundle that is meant for them (they should not pretend to be of a different type) different type than they are (The incentive compatibility constraints):

Max $V^H(C_H, Y_H)$ subject to

1. $V^L(C_L, Y_L) \geq v$
2. $Y_H - C_H + Y_L - C_L \geq R.$
3. $V^H(C_H, Y_H) \geq V^H(C_L, Y_L) = V^{H(L)}$= utility of Hif she takes the bundle for L
4. $V^L(C_L, Y_L) \geq V^L(C_H, Y_H) = V^{L(H)}$

The two last equations are the incentive constraints, the self selection constraint in the information economics jargon, they simply say that a type (L or H) should not have incentives to pick the bundle meant for the other type. In this simple two type model the constraint will always bind for one of the types and not for the other, so there is a separating equilibrium - they choose different bundles. This is not always true in a more general model.

The first order condition for the Lagrangian (G) of this problem with multipliers $(\mu, \gamma, \lambda_H, \lambda_L)$ are given by

$$\frac{\partial G}{\partial C_L} = \mu V^L_{C_L} - \lambda_H V^{H(L)}_{C_L} + \lambda_L V^L_{C_L} - \gamma = 0 \tag{8}$$

$$\frac{\partial G}{\partial Y_L} = \mu V^L_{Y_L} - \lambda_H V^{H(L)}_{Y_L} + \lambda_L V^L_{Y_L} + \gamma = 0 \tag{9}$$

$$\frac{\partial G}{\partial C_H} = V^H_{C_H} + \lambda_H V^H_{C_H} - \lambda_L V^{L(H)}_{C_H} - \gamma = 0 \tag{10}$$

$$\frac{\partial G}{\partial Y_H} = V^H_{Y_H} + \lambda_H V^H_{Y_H} - \lambda_L V^{L(H)}_{Y_H} + \gamma = 0 \tag{11}$$

With respect to the incentive constraint there are three possibilities. None of the constraints are binding, the constraint for Hbinds, but not the constraint for L, or it binds for $L$and not for H. The most natural and interesting case is when it binds for H bur not for L, that is when $\lambda_H > 0$ and $\lambda_L = 0$. In this case the $H-$type will be tempted to mimic the $L-$type and the government must take this into account when choosing its tax policy (which here means when the government offer the bundles $\{C_L, Y_L\}$ and $\{C_H, Y_H\}$). What we are particularly interested in is how this binding incentive constraint (the fact that

$\lambda_H > 0$ and $\lambda_L = 0$) affect the marginal tax rates. With some manipulation of the first order conditions we will show that the bundle offered to H assures that $\frac{-V_{Y_H}}{V_{C_H}} = \frac{-U_L}{w_H U_C} = 1$ and that $\frac{-V_{Y_L}}{V_{C_L}} = \frac{-U_L}{w_L U_C} < 1$. To implement this with a tax scheme there must be a zero marginal tax rate for the $H-$type and a positive marginal tax rate for the $L-$type.

It is easy to derive the condition for the $H-$type. Dividing equation 11 by 10 (remember that $\lambda_L = 0$ is zero), we obtain $\frac{-V_{Y_H}}{V_{C_H}} = \frac{-U_L}{w_H U_C} = 1$.

It is a bit more involved to characterize the bundle that is offered to the $L-$type. We have to work with equations 8 and 9. First we add these equations and get

$$\left(\mu V_{C_L}^L - \lambda_H V_{C_L}^{H(L)} - \gamma\right) + \left(\mu V_{Y_L}^L - \lambda_H V_{Y_L}^{H(L)} + \gamma\right) = 0.$$

Multiply the first term on the lsh with dC_L and the second term with dY_L. This is only allowed if both terms are of the same magnitude. We know that along L's indifference curve (and remember that we have the constraint that L should be on the indifference curve that gives utility v) $dC_L = dY_L MRS^L$. We can write this as

$$\left(\mu V_{C_L}^L - \lambda_H V_{C_L}^{H(L)} - \gamma\right) dY_L MRS^L + \left(\mu V_{Y_L}^L - \lambda_H V_{Y_L}^{H(L)} + \gamma\right) dY_L = 0.$$

Collecting terms, we get

$$\mu\left(V_{C_L}^L MRS^L + V_{Y_L}^L\right) - \lambda_H\left(V_{C_L}^{H(L)} MRS^L + V_{Y_L}^{H(L)}\right) - \gamma\left(MRS^L - 1\right) = 0$$

The first term disappears. The term $\left(V_{C_L}^{H(L)} MRS^L + V_{Y_L}^{H(L)}\right)$ can be written as $V_{C_L}^{H(L)}\left(MRS^L - MR\right.$ where $MRS^{H(L)}$ is the marginal rate of substitution between income and consumption for H at the bundle (C_L, Y_L). We now have

$$-\lambda_H V_{C_L}^{H(L)}\left(MRS^L - MRS^{H(L)}\right) = \gamma\left(MRS^L - 1\right)$$

Since we know that the slope of the indifference curve of the $L-$type is steeper in any point (C, Y) it follows that in the Pareto optimal solution to this problem $MRS^L < 1$, since we have

$$MRS^L = 1 - \frac{\lambda_H}{\gamma} V_{C_L}^{H(L)}\left(MRS^L - MRS^{H(L)}\right) \quad (12)$$

Although the algebra gets a bit involved the intuition is clear. In order to discourage the $H-$type from consuming the bundle intended for the $L-$type it is optimal to "give" the low type less income (require him to work less) and consumption than what is optimal if this person is considered in isolation. That is we tilt the leisure work decision for this person in such a way that he is kept at the indifference curve v. This will discourage the $H-$type to "take" this bundle since she is very efficient in producing income and hence the gain she gets by working less will not weigh up for the loss she gets from consuming less.

Make sure you understand this logic, it pops up many places where there are asymmetric information and "mimicking constraints". The best way to get a deep understanding of what is going on is by drawing figures. There are many nice illustrations in Stiglitz (1987) and Figure 1 in Broadway and Keen 1983 is very useful.

This "direct implementation" language is a bit abstract ("the government offers two bundles $\{C_L, Y_L\}$ and $\{C_H, Y_H\}$ .."). It is a nice way to characterize the Pareto optimal, information constrained, solution, but we are of course interested in how the government can use a tax policy to implement the solution.

Suppose the government levy taxes and $T(Y)$ is the tax a person with income Y must pay to the government. With two types in the economy, we only need to specify the level of taxes and the marginal tax rate at two income levels; Y_L and Y_H. From the analysis above we know that the marginal tax rate is positive for income Y_L , we know from (12) that the marginal tax rate is equal to $\frac{\lambda_H}{\gamma} V_{C_L}^{H(L)} \left(MRS^L - MRS^{H(L)} \right)$ - with this tax marginal rate the low wage type choose the bundle that we characterized above. This is now a decentralized decision; the household decide to earn this amount of income, given a tax scheme.

We also know that the marginal tax rate for the $H-$type is 0 (the no tax at the top result). For any income levels between the high and the low the average tax must be so high that it is not tempting for neither H nor L to choose any other income (labour supply). This is illustrated in Stiglitz (1987) figure 2.8.

Extensions This model has been extended in many different directions; more households, a tax on consumption (perhaps differentiated), a tax on savings, include public goods etc.

In order to understand the interactions between these extensions and the optimal nonlinear tax it is crucial to remember that the optimal non-linear income tax distorts the labour decision of the $L-$type in order to fulfill the incentive constraint for the $H-$type (make sure that Hwill not choose to earn Y_L and work (Y_L/w_H) hours and consume $Y_L - T(Y_L)$. Hence, everything that makes this incentive constraint less binding, that is, everything that makes it less tempting for a $H-$type to choose low income, will increase the objective that the government maximizes (improve the situation).

Public goods. Suppose that the government uses some of its revenue to produce a public good. If this good was financed with a lump sum tax, optimal provision implies that the marginal cost of providing one extra unit is equal to the marginal willingness to pay for this unit (as usual). The key with a public good is that the marginal willingness to pay is given by the sum of the individuals' willingness to pay (the sum of individuals' marginal substitution rate between the public good and the private good). This is the

Samuleson rule.

Now let us reintroduce the two type model and assume that H appreciate the public good much less than L, and suppose government collects taxes through a non-linear income tax of the type shown above (incentive constraint binds for H). In this case it is optimal to over-provide (compared with the Samuelson rule) the public good. To understand this result, assume that the government produced public goods according to the Samuelson rule. Since the $L-$type value this good more, he is willing to pay more than the $H-$type to get one extra unit of the public good. If the government produced one extra unit public good and at the same time increased the tax liability just to keep both types on their indifference curves this would not – by definition – change the utility of the two types. But since the tax for L increases by more than it does for H (since L has a higher willingness to pay for public good) it becomes less tempting for H to mimic L. The incentive constraint for H becomes less binding and this increase the value of the Lagrangian.

Suppose all individuals have the same utility function in the public good example above. The case for providing public goods beyond the Samuelson rule is then that the valuation of the public good declines with the amount of leisure households consume.

Differentiated tax on commodities

This last observation makes it easy to understand how a differentiated consumer good taxes can make the incentive constraint in non-linear income taxation less binding. A $H-$type consumes a lot of leisure if she chooses to earn Y_L (she is so productive that she does not have to work very much to earn that income). This then means that any good that is complementary with leisure, goods that increases the marginal utility of consuming leisure, should be taxed since it makes the constraint on the income tax problem less binding.

A piecewise linear income tax system

Most tax systems have marginal tax rates that are piecewise linear. A more practical question then is how one should set the tax rate for different segments, different brackets, of the income distribution. Changing the tax rate within one income bracket has two effects, it changes the revenue for the government and it changes the welfare, the utility, of the tax payers who have an income within that bracket or above - those with an income below are not affected.

It is simplest to analyze a change in the top bracket, since a change here will only affect one group; the top earners. If the government were to change the tax rate of a bracket further down in the income distribution it would also affect the average tax rate of those who earn an income above this bracket.

Consider an economy with N individuals earning more than $\bar{y}$, which is the lower threshold for the top bracket of the income. Income above this level is at a marginal tax rate of τ. Assume that the average income of those who earn above $\bar{y}$ is equal to y^m, this average income will be a function of $(1-\tau)$, the net of tax income within this bracket. Consider no the effect of a small increase in the marginal tax rate $(d\tau)$ of top earners. For simplicity assume away all income effects so the elasticity of of labour supply is the compensated elasticity, denoted $\bar{\varepsilon}$. Note that while there will be adjustment of income within this bracket, no one will have an incentive to reduce their income blow $\bar{y}$ since the tax rate there is not changed.

Government Revenue effect (dR): We can decompose the Revenue effect into a mechanical effect (dM) (if there were no behavioral response) and a behavioral effect (dB): $dR = dM + dB$.

The mechanical effect is simplify $dM = d\tau\,(y^m - \bar{y})\,N$. To find the revenue change that arise because top earners adjust their work effort we need to note that y^m is a function of the marginal tax rate. We have $dB = N\tau dy^m = -N\tau \frac{dy^m}{d(1-\tau)} d\tau = -N\frac{\tau}{(1-\tau)}\bar{\varepsilon}\bar{y}d\tau$

The welfare effect (dW) is given by two terms; the change in utility of the individuals who are in this income bracket and the weight these individuals are given in the social welfare function. Since we assume that top earners have chosen their labour supply optimally given the marginal tax rate (τ), we know, from the envelope theorem, that the money metric value of their loss is equal to dM. Suppose that the government assign a weight $\bar{g}\,\epsilon\,(0,1)$ on this loss (if it is equal to 0 the income lost at the top does not have any effect on social welfare; in general $\bar{g}$ depends on the marginal utility of income for top earners and on the shape of the social welfare function: $dW = -\bar{g}\left(d\tau\,(y^m - \bar{y})\,N\right)$.

If we add together these terms we obtain the total social welfare effect of the small increase in the tax rate $(d\tau)$. The initial tax rate will be optimal if the total effect of increasing the rate is 0. Hence to characterize the optimum top income marginal tax rate $(\bar{\tau})$ we solve $dM + dB + dW = 0$. We get the following expression

$$\frac{\bar{\tau}}{1-\bar{\tau}} = \frac{(1-\bar{g})(\frac{y^m}{\bar{y}} - 1)}{\varepsilon\frac{y^m}{\bar{y}}}$$

The optimal tax rate of top earners

- decreases in $\bar{g}$; the more weight the loss of the top earners is given
- decreases in the compensated elasticity of labour supply
- increases in the thickness of the tale of the income distribution in the top bracket $(\frac{y^m}{\bar{y}})$

We can do the same exercise for lower brackets but, as noted above. An additional factor comes in, increasing the marginal tax rate for a bracket below the top reduces the tax income by lowering the average tax rate of those who belong to a higher income bracket.

Literature

Ramsey: The Ramsey model of optimal optimal consumer good taxes is discussed in many articles. Often these articles look at the problem from slightly different angles, and use different notation - this is challenging as it requires a deeper understanding of the problem in order to see past these differences. But hopefully the fact that different papers present different takes on the problem will also deepen your understanding of the analysis. The relevant literature here is

- Boadway, R (2012). From Optimal Tax Theory to Tax Policy, MIT Press. Chapter 3, pp. 47-58
- Keen, M. and S. Smith (1996). The Future of the Value Added Tax in the European Union. Economic Policy 23, pp. 378-379
- Christiansen, V.(2009) "The Choice between Uniform and Differentiated Commodity Taxation" in Blonder, J. (ed.): Yearbook for Nordic Tax Research 2009: The Non-fiscal Purposes of Taxation, DJØF Publishing Copenhagen, pp. 141-151
- Salany, B.(2003). The Economics of Taxation, MIT Press. Chapter 3.2, pp. 73-76
- Hindriks, J. and G. Myles (2006). Intermediate Public Economics, MIT Press. Chapter 14, pp. 443-475

Mirrlees (non-linear income tax (and its implications for...)): The papers covering this topic are

- Røed, K. and S. Strøm. Progressive taxes and the labour market – Is the trade-off between equality and efficiency inevitable? Journal of Economic Surveys 16
- Stiglitz, J. (1987). "Pareto Efficient Taxation and Optimal Taxation and the New New Welfare Economics in Auerbach" A. J. and M. Feldstein (ends): Handbook of public Economic, North Holland, pp. 991-1005
- Edwards, J., M. Keen and M. Tuomala (1994). Income tax, commodity taxes and public good provision: A brief guide. Finanzarchiv 51, pp. 472-487
- Kaplow, L. (2008). The Theory of Taxation and Public Economics, Princeton University Press. Chapter 4, pp. 53-72
- Boadway, R. and M. Keen (1993). Public goods, self-selection and income taxation. International Economic Review 34, sections 1-3, pp. 463-471

Piecewise linear income tax

This stuff is covered in

- E. Saez, “Optimal Income Transfer Programs: Intensive Versus Extensive Labor Supply Responses”, Quarterly Journal of Economics 117 (2002), 1039-1073
- Brewer, M., E. Saez and A. Shephard (2009). "Means-testing and tax rates on earnings" in Mirrlees, J., S. Adam, T. Besley et. al (ends). Reforming the Tax System for the 21st Century: The Mirrlees Review 2. Dimensions of Tax Design, London: Institute for Fiscal Studies, Ch. 2. Read Appendix 2A 1-3 of the main text. See also relevant parts of the main text being referred to.