Optimal Taxation : (a) Ramsey's Rules for Optimal Commodity Taxation

The theory of optimal taxation is really an extension of the material in the previous chapter on efficiency : taxes have an excess burden, unless they are lump–sum taxes ; lump–sum taxation is impossible ; so how should taxes be designed so as to minimize the excess burden?

Much of the work to answer this question has already been done. In the previous section (in the fourth lecture, "Formulae for Excess Burden") , a formula for the excess burden of a tax was derived : that formula gave some hints as to which commodities should be taxed, and how much. What that analysis missed was the inter–relation between different taxes. But in that section there is a brief discussion of that issue : how an increase in the tax on tea (a substitute for coffee) would reduce the excess burden of any excise tax in place on coffee, and might even reduce the overall excess burden of the tax system. That result will be applied here as well.

Optimal commodity taxation can be started with the following, somewhat artificial, problem : the government needs to raise a fixed amount of tax revenue, R. The only way the government can get the revenue is through excise taxes, on goods X and Y. So the government has to choose excise tax rates t_X and t_Y on the two goods, so as to collect R dollars in total — and they want to do that at the least possible damage to the taxpayer. Or, to put it another way, they want to set taxes so as to raise the required amount R of revenue with the smallest possible excess burden.

Why is this problem "artificial"? First, I'm restricting the government to use excise taxes to raise its revenue. And second — and this is a bigger issue — the government is collecting the tax revenue from a single taxpayer.

That is, the concept of excess burden, or of compensating variation, refer to a single individual. And the basic problem, described above, uses those concepts. So, taken literally, this basic problem applies only in a country with a single taxpayer, or a country in which all taxpayers are identical, or in a country where we can ignore differences among taxpayers. This formulation is a limitation of the basic model. But it turns out that some of the results and formulae can be applied in a more realistic, complicated setting in which we care about differences among taxpayers.

I'm also going to assume that the "net of tax" prices p_x and p_y of the goods are given. The tax–included prices are

$$P_X = p_x + t_x \tag{1}$$

$$P_Y = p_y + t_y \tag{2}$$

The tax revenue collected will be t_xX from the tax on good X, and t_yY from the tax on good Y. Now the quantity consumed of each good will depend on the prices of both goods. I'll use the **compensated** demand functions here, because those are the demand functions I used in deriving the formulae in the section on efficiency. That means that the total revenue collected is

$$R = t_xX^H(P_X, P_Y, u) + t_yY^H(P_X, P_Y, u) \tag{3}$$

Equation (3) is the revenue constraint which the government must meet : they must choose tax rates t_X and t_Y which satisfy (3), for the fixed revenue target R which they have been given.

The government wants to minimize the cost to the taxpayer of these taxes. From the first lecture in the previous section on efficiency, that cost, in dollars, is

$$E(p_x + t_x, p_y + t_y, u) - E(p_x, p_y, u) \tag{4}$$

Therefore, the government's problem is to minimize the cost (4) to the taxpayer, subject to meeting the revenue constraint defined by equation (3).

To solve this problem, the method of Lagrange can be used. Minimization of (4) subject to the constraint (3) can be solved by setting up the Lagrange multiplier problem

$$E(P_X, P_Y, u) - E(p_x, p_y, u) + \lambda[R - t_x X^H(P_X, P_Y, u) - t_y Y^H(P_X, P_Y, u)] \tag{5}$$

where λ is the Lagrange multiplier attached to the revenue constraint (3). We then minimize expression (5) with respect to t_x, t_y and λ, by setting the derivatives of expression (5) with respect to t_x, t_y and λ equal to 0.

So, the government's optimal excise tax rates t_x and t_y solve

$$\frac{\partial E(P_X, P_Y, u)}{\partial P_X} - \lambda[X + t_x \frac{\partial X^H}{\partial P_X} + t_y \frac{\partial Y^H}{\partial P_X}] = 0 \tag{6}$$

$$\frac{\partial E(P_X, P_Y, u)}{\partial P_Y} - \lambda[Y + t_x \frac{\partial X^H}{\partial P_Y} + t_y \frac{\partial Y^H}{\partial P_Y}] = 0 \tag{7}$$

$$R - t_x X^H(P_X, P_Y, u) - t_y Y^H(P_X, P_Y, u) = 0 \tag{8}$$

where I have used the facts that $P_X = p_x + t_x$ and $P_Y = p_y + t_y$ and p_x and p_y are fixed, so that changing t_x by 1 dollar is the same thing as changing P_X by 1 dollar.

Equation (8) is just equation (3), so there's nothing new there.

Equations (6) and (7) can be simplified using *Shephard's Lemma*, from lecture 1 in the section on efficiency : the partial derivative of the expenditure function with respect to the price of a good must equal the compensated demand for the good : $\frac{\partial E(P_X, P_Y, u)}{\partial P_X} = X^H$ and $\frac{\partial E(P_X, P_Y, u)}{\partial P_Y} = Y^H$.

That makes (6) and (7) into :

$$(\lambda - 1)X = -\lambda[t_x \frac{\partial X^H}{\partial P_X} + t_y \frac{\partial Y^H}{\partial P_X}] \tag{9}$$

$$(\lambda - 1)Y = -\lambda[t_x \frac{\partial X^H}{\partial P_Y} + t_y \frac{\partial Y^H}{\partial P_Y}] \tag{10}$$

Now if I divide both sides of equation (9) by X, and then by λ, and divide both sides of equation (10) by Y and then by λ, then we have (the same) $(\lambda - 1)/\lambda$ on both sides, so that

$$\frac{[t_x \frac{\partial X^H}{\partial P_X} + t_y \frac{\partial Y^H}{\partial P_X}]}{X} = \frac{[t_x \frac{\partial X^H}{\partial P_Y} + t_y \frac{\partial Y^H}{\partial P_Y}]}{Y} \tag{11}$$

One more adjustment : one of the nice properties of the compensated demand functions is that their partial derivatives are symmetric : $\frac{\partial X^H}{\partial P_Y} = \frac{\partial Y^H}{\partial P_X}$. [That result was presented at the end of the first lecture in the section on efficiency.] So equation (11) becomes

$$\frac{[t_x \frac{\partial X^H}{\partial P_X} + t_y \frac{\partial X^H}{\partial P_Y}]}{X} = \frac{[t_x \frac{\partial Y^H}{\partial P_X} + t_y \frac{\partial Y^H}{\partial P_Y}]}{Y} \tag{12}$$

Equation (12) is Ramsey's basic rule for the optimal system of commodity taxes : if the excise taxes t_X and t_Y are set so as to raise the required revenue at the minimum possible damage to the taxpayer, then equation (12) must hold.

Equation (12) is the exact rule for the "best" commodity tax system : the one which raises the given revenue with the lowest possible excess burden. In other words, if the government agency wants to see how efficient its commodity taxes are, it should be checking two things : (*i*) whether enough revenue is raised ($t_X X + t_y Y = R$) and (*ii*) whether, when they estimate the derivatives of the compensated demand functions, equation (12) is satisfied. If (12) holds, then the commodity tax system is the best possible, given the revenue needs.

There are several interpretations, and special cases, of the Ramsey rule. But (12) is the basic equation : the other forms of Ramsey rule derived below follow from (12).

Suppose that we start with no taxes at all, so that the prices of the two goods are p_x and p_y. Now if we put taxes t_X and t_Y on the two goods, their prices will increase to $P_X = p_x + t_x$ and $P_Y = p_y + t_y$ respectively. So the changes in the prices of the two goods are

$$\Delta P_X = t_x \tag{13}$$

and

$$\Delta P_Y = t_y \tag{14}$$

What would be the change in quantity demanded of the two taxed goods X and Y, due to the introduction of the excise taxes?

The quantity demanded of good X should change, if we introduce taxes, because the price of good X will have changed from p_x to $p_x + t_X$. This increase in the price of good X (if $t_X > 0$) must cause quantity demanded of good X to decrease.

But the price of good Y has changed as well, and that change in P_Y may also affect quantity demanded for good X. So the overall change in quantity demanded of good X is

$$\Delta_X \approx \frac{\partial X}{\partial P_X}\Delta P_X + \frac{\partial X}{\partial P_Y}\Delta P_Y = \frac{\partial X}{\partial P_X}t_X + \frac{\partial X}{\partial P_Y}t_Y \tag{15}$$

Similarly,

$$\Delta_Y \approx \frac{\partial Y}{\partial P_X}\Delta P_X + \frac{\partial Y}{\partial P_Y}\Delta P_Y = \frac{\partial Y}{\partial P_X}t_X + \frac{\partial Y}{\partial P_Y}t_Y \tag{16}$$

Now if I plug equations (15) and (16) into the exact, general rule (12) for the optimal tax system, (12) becomes

$$\frac{\Delta X}{X} \approx \frac{\Delta Y}{Y} \tag{17}$$

Equation (17) is another version of the Ramsey rule. There will be 3 versions of the Ramsey rule in total in this section : the exact general formula (12), equation (17), which was just derived, and another special case which is yet to come.

Equation (17) is usually referred to as the **equi–proportional** version of the Ramsey rule.

EQUI–PROPORTIONAL RAMSEY RULE : If a commodity tax system is optimal, it should reduce the quantities demanded of each taxed good by approximately the same proportion.

Why the "approximately" in the statement of the equi–proportional version of the rule (and the approximation sign in equations (15), (16) and (17))?

There are two reasons why this rule is an approximation.

First, the (exact) equation (12) used the derivatives of the Hicksian demand functions. Like all derivatives, these are defined for very small — infinitesimally small – changes in the prices of the goods. But equations (15) and (16) use the changes in prices caused by the taxes : $\Delta P_x = t_X$ and $\Delta P_Y = t_Y$. These changes might not be so small. When taxes are relatively big, as a fraction of the prices of the goods, then the derivatives are just an approximation of the overall (not so small) effects of the price changes. [If the demand functions were linear in prices, so that the derivatives were constants, then these derivatives would provide an exact measure. But demand functions aren't always linear.]

Second, the demand functions used in equation (12) are the Hicksian, or compensated demand functions. These refer to the effects of price changes on quantities demanded when the consumer is compensated so as to remain on the same indifference curve. If the tax authorities do not actually compensate the consumers, then we are using these compensated demand derivatives to approximate the effect of price changes which will not be compensated. So a more precise version of the equi–proportional Ramsey rule would be : "If a commodity tax system is optimal, it should reduce the quantities demanded of each taxed good by approximately the same proportion, if the consumer were compensated so as to stay on the same indifference curve.".

The first two versions (equations (12) and (17)) of the Ramsey rule say nothing explicit about the actual tax **rates** themselves, only about the overall effects of those rates. But there is a fairly simple formula which does involve the tax rates themselves, also due to Ramsey.

Let $|\eta_X|$ and $|\eta_Y|$ be the **compensated own price elasticities** of demand for the two goods (defined so as to have a positive sign). That is

$$|\eta_X| = -\frac{\partial X^H}{\partial P_X}\frac{P_X}{X} \tag{18}$$

$$|\eta_Y| = -\frac{\partial Y^H}{\partial P_Y}\frac{P_Y}{Y} \tag{19}$$

That means that

$$-\frac{\partial X^H}{\partial P_X} = |\eta_X|\frac{X}{P_X} \tag{20}$$

$$-\frac{\partial y^H}{\partial P_X} = |\eta_Y| \frac{Y}{P_Y} \tag{21}$$

Now if I ignore the "cross–price" effects $\frac{\partial X^H}{\partial P_Y}$ and $\frac{\partial Y^H}{\partial P_X}$ in equations (9) and (10), they become

$$\frac{\lambda - 1}{\lambda} = -\frac{\partial X^H}{\partial P_X} \frac{t_X}{X} = -\frac{\partial X^H}{\partial P_X} \frac{P_X}{X} \frac{t_X}{P_X} = |\eta_X| \frac{t_X}{P_X} \tag{22}$$

and

$$\frac{\lambda - 1}{\lambda} = -\frac{\partial Y^H}{\partial P_Y} \frac{t_Y}{Y} = -\frac{\partial Y^H}{\partial P_Y} \frac{P_Y}{Y} \frac{t_Y}{P_Y} = |\eta_Y| \frac{t_Y}{P_Y} \tag{23}$$

Now t_X/P_X is just the tax rate : the tax as a proportion of the price. I used the notation τ_X earlier (in section 2, "Tax Incidence") to denote that tax rate. With that notation, (22) and 23) become

$$|\eta_X|\tau_X = |\eta_Y|\tau_Y \tag{24}$$

or

$$\frac{\tau_X}{\tau_Y} = \frac{\eta_Y}{\eta_X} \tag{25}$$

Equation (25) is the third version (the first 2 versions were equations (12) and (17)) of the Ramsey rule. Equation (25) is often referred to as the **inverse elasticity** version of the Ramsey rule, because it says that tax rates on different goods should be inversely proportional to the goods' compensated own–price elasticities of demand.

That is, if good X had an own–price elasticity of demand of 0.5, and good Y had an own–price elasticity of 1.5, then equation (25) implies that the optimal tax rate should be higher on good X than on good Y, and that it should be 3 times higher on good X, because η_X is 1/3 of η_Y.

The advantage of the own–price elasticity rule is that it is defined directly in terms of the tax rates themselves. The equi–proportional version, equation (17), describes the effects of the tax rates on quantities demanded of the taxed goods, if the taxes are set optimally. It says nothing directly about the tax rates themselves. Equation (25) says that if $\tau_X = 2\tau_Y$, these taxes are optimal only if $\eta_X = (0.5)\eta_Y$.

There is also some intuition behind the prescription of the inverse elasticity rule. It says that we should levy the highest tax rates on goods which have inelastic (with respect to their own price) demands. The problem with excise taxes, as far as efficiency is concerned, is that they distort people's behaviour, inducing them to substitute away from the taxed good and towards untaxed goods, even when the actual cost of production of these goods have not changed. The bigger the substitution, the bigger the cost of the tax, and the bigger the excess burden of the tax relative to the revenue raised. In the extreme, if good X were a *perfect complement* to some untaxed goods, with L–shaped indifference curves, then there would be no excess burden from a tax on good X. So rule (25) says that if there is some good with a (nearly) inelastic demand, then the government should raise (nearly) all its revenue from taxing that good, because they can then raise their tax revenue with (nearly) no excess burden at all.

However there is a big problem with the inverse elasticity version of the Ramsey rule, a problem it does not share with the other versions of the rule. In deriving equations (22) and (23), which led to this inverse elasticity rule, I ignored the cross–price effects $\frac{\partial X^H}{\partial P_Y}$ and $\frac{\partial Y^H}{\partial P_X}$. Is that a legitimate simplification? **No**. In general, these cross–price effects are not 0, and ignoring them means that equation (25) is not actually equivalent to the general Ramsey rule.

In other words, the inverse elasticity rule is an approximation, and the approximation is valid only if the cross–price effects are close to zero. That is, rule (25) is only useful if we think that the taxed goods X and Y are not particularly good substitutes for each other, and not particularly good complements to each other. If we can only tax tea and coffee, then we should **not** use the inverse elasticity version of the Ramsey rule, because we think that tea and coffee may be pretty strong substitutes for each other. If we can only tax people's club memberships, and people's expenditure on sports equipment, then we should **not** use the inverse elasticity version of the Ramsey rule. It seems reasonable that membership in a sports club would be complementary with purchase of equipment to use at the club, so that in this case we'd expect $\frac{\partial X^H}{\partial P_Y} = \frac{\partial Y^H}{\partial P_X} < 0$.

So the inverse elasticity version of the Ramsey rule is in many ways more convenient and practical than the equi–proportional version. But it can be very misleading. The equi–proportional version is valid always, even when the goods are strong complements or substitutes. If the goods are not strong complements or substitutes, then the two versions of the rule will give the same prescription for which tax rates are optimal. If they are strong complements or substitutes, then the two versions of the rule may disagree. In this case, it's the equi–proportional version which is more accurate.