Answers to exam in Tax Policy, January 2015

Part 1: Income taxation

(1A) **Q**: The paper takes a heuristic approach to deriving the formula. Assume that the marginal tax rate is constant at τ above some income threshold z^* and consider a small increase in this top marginal tax rate to $\tau + \Delta\tau$. Such a policy change affects welfare through 3 channels. First, holding behavior constant, there is a transfer of funds from households to the government, which gives rise to a mechanical revenue gain ("ΔM") for the government and a social welfare cost ("ΔW") at the level of the tax payers. Behavioral responses to the tax give rise to a behavioral revenue loss ("ΔB") for the government whereas there is no direct effect on utility. Importantly, the latter result hinges on the argument that since individuals were initially optimizing, small behavioral changes only have second-order effects on utility (follows from the "envelope theorem").

At the optimal policy, it must hold that a small change does not increase nor decrease welfare, hence:

$$\Delta M = \Delta W + \Delta B$$

To derive ΔM, note that holding behavior constant, a small increase in the top marginal tax rate, increases the tax payments of a single individual with income z by $\Delta\tau(z - z^*)$ and hence the total tax payments by $N^*(z_m - z^*)\Delta\tau$ where N^* is the number of individuals with income above z^* and z_m is the average income of those individuals with income above z^*. This is the mechanical revenue gain

$$\Delta M = N^*(z_m - z^*)\Delta\tau$$

To derive ΔB, note that a single tax payer reduces his income in response to the tax change by $(dz/d\tau)\Delta\tau$. This can be rewritten as $-[ez/(1-\tau)]\Delta\tau$ using the definition of the elasticity of taxable income. Hence, the total behavioral revenue loss is:

$$\Delta B = N^*\tau \frac{ez_m}{(1-\tau)}\Delta\tau$$

Assuming that individuals with incomes above z^* have no weight in the social welfare function, the first-order condition for optimal policy becomes:

$$\Delta M = \Delta B$$

$$\Longleftrightarrow N^*(z_m - z^*)\Delta\tau = N^*\tau \frac{ez_m}{(1-\tau)}\Delta\tau$$

$$\Longleftrightarrow (z_m - z^*) = \tau \frac{ez_m}{(1-\tau)}$$

$$\Longleftrightarrow (z_m - z^*) = \tau \left[ez_m + z_m - z^*\right]$$

$$\iff \tau = \frac{1}{1 + e\frac{z_m}{z_m - z^*}}$$

which reduces to the expression in the question when using the definition of a.

(1B) **Q**: There are several important assumptions. First, the government does not care above high-income individuals, which may be far from reality. Second, the underlying income distribution is unaffected by tax policy, which may be justified in the case of a small tax change, but not when the formula is used to propose that a very large increase in the US top marginal tax rate is optimal. Third, the elasticity of taxable income is the same (on average) for all income levels above z^* - it is often invoked that the elasticity of taxable income is increasing in the income level, but the assumption may be justified by the fact that we are only considering top incomes.

Q: A higher e implies a lower optimal top marginal tax rate. Intuitively, a higher e means stronger behavioral responses to taxation and thus a higher cost of taxation in terms of lost tax base. A higher a implies a lower optimal top marginal tax rate. Intuitively, a measures the "thinness of the tail" of the income distribution above z^*. A higher value of a means a thinner tail above z^* and thus a smaller gain from taxation in terms of the mechanical revenue gain. The fact that the US income distribution has a long, thick tail of very high incomes implies that the optimal top marginal tax rate (everything else equal) is quite high.

Q: No. The objective of the government is effectively to maximize tax revenue from the top income earners. Hence, the government is not trading off equity against efficiency - it is maximizing equity. The only reason why the government does not wish to tax beyond τ^* is that τ^* is the revenue maximizing rate so that a further increase would lower the revenue (and thus reduce equity).

Part 2: Commodity taxation

(2A) **Q**: In the Ramsey model there are N commodities each taxed at a different tax rate. There is a single individual who chooses her consumption of each of the N commodities as well as her labor supply taking prices and the wage rate as given. There is a government that needs to raise a fixed amount of revenue with the N tax instruments at its disposal with the objective of maximizing the welfare of the individual. The key assumption is that leisure cannot be taxed. If the government could enforce a uniform tax on total consumption including the consumption of leisure, this would amount to a tax on the exogenously given potential income, that is a non-distortive lumpsum tax.

Q: The term $-\sum_j t_j S_{jk}$ is the marginal excess burden associated with the tax on good k. This is equivalent to the government loss caused by the consumer's responses to a small increase in the tax on good k. The term X_k is the mechanical revenue gain associated with a small increase in the tax on good k.

The right-hand-side of the equation thus captures the share of the mechanical revenue gain that is a deadweight loss. Because the left-hand-side is constant, the optimal tax system equalizes this fraction across all N tax instruments. This implies that the optimal commodity tax system minimizes the efficiency loss associated with taxation. Since there is only a single consumer in the economy there is no equity concern - only the concern to keep inefficiencies at a minimum.

On the left-hand side, λ is the multiplier on the government revenue constraint and, thus, the marginal cost to the individual of raising revenue with distortive commodity taxes whereas μ is the "net social benefit of private income" or, equivalently, the marginal cost of raising revenue with non-distortive lumpsum taxes. The difference between the two expresses the marginal excess burden expressed in utility terms. When divided by λ the difference expresses the marginal excess burden in revenue terms.

(2B) **Q**. Using the symmetry of the Slutsky matrix $S_{jk} = S_{kj}$, the Ramsey rule can be expressed as:

$$\frac{\lambda - \mu}{\lambda} = -\frac{\sum_j t_j S_{kj}}{X_k} \tag{1}$$

Defining the compensated elasticity of demand for good k with respect to the price of good j as $\varepsilon_{kj} = S_{kj}(1 + t_j)/X_k$, one obtains

$$\frac{\lambda - \mu}{\lambda} = -\sum_j \varepsilon_{kj} \frac{t_j}{1 + t_j}$$

Assuming that all cross-price elasticities are zero, $\varepsilon_{kj} = 0$ for $k \neq j$, one obtains:

$$\frac{t_k}{1 + t_k} = -\frac{\lambda - \mu}{\lambda} \frac{1}{\varepsilon_{kk}}$$

The equation states that the optimal tax rate on good k is inversely proportional to the elasticity of demand. Popularly, one should apply a higher tax rate to less elastic goods

Q: The assumption that all cross-price elasticities are zero has no empirical foundation and is wildly unrealistic. Hence, the inverse elasticity should not be used for practical policy purposes.

(2C) **Q**: In the special case with $k = 2$, (1) becomes:

$$\frac{\lambda - \mu}{\lambda} = -\frac{t_1 S_{11} + t_2 S_{12}}{X_1} \tag{2}$$

$$\frac{\lambda - \mu}{\lambda} = -\frac{t_1 S_{21} + t_2 S_{22}}{X_2} \tag{3}$$

Isolating t_1 in (2) yields:

$$-\frac{\lambda-\mu}{\lambda}\frac{X_1}{S_{11}} - t_2\frac{S_{12}}{S_{11}} = t_1$$

Inserting that expression into (3) yields:

$$t_2 = \frac{\lambda-\mu}{\lambda}\left[\frac{S_{21}X_1 - S_{11}X_2}{S_{22}S_{11} - S_{12}S_{21}}\right]$$

Inserting back into the expression for t1 yields:

$$t_1 = \frac{\lambda-\mu}{\lambda}\left[\frac{S_{12}X_2 - S_{22}X_1}{(S_{22}S_{11} - S_{12}S_{21})}\right]$$

Euler's theorem implies that:

$$S_{11}q_1 + S_{12}q_2 + S_{10}w = 0$$

$$S_{21}q_1 + S_{22}q_2 + S_{20}w = 0$$

Isolating S_{12} and S_{21} and inserting into the expressions for t_1 and t_2 yields:

$$t_1 = \frac{\lambda-\mu}{\lambda}\left[\frac{\left[-S_{11}\frac{q_1}{q_2} - S_{10}\frac{w}{q_2}\right]X_2 - S_{22}X_1}{(S_{22}S_{11} - S_{12}S_{21})}\right]$$

$$t_2 = \frac{\lambda-\mu}{\lambda}\left[\frac{\left[-S_{22}\frac{q_2}{q_1} - S_{20}\frac{w}{q_1}\right]X_1 - S_{11}X_2}{S_{22}S_{11} - S_{12}S_{21}}\right]$$

Multiply both equations by $\frac{X_2X_1}{q_1q_2}$ and rewrite in terms of elasticities:

$$\frac{t_1}{q_1} = -\frac{\lambda-\mu}{\lambda}\frac{X_2X_1}{q_1q_2}\left[\frac{\varepsilon_{11}+\varepsilon_{10}+\varepsilon_{22}}{(S_{22}S_{11} - S_{12}S_{21})}\right]$$

$$\frac{t_2}{q_2} = -\frac{\lambda-\mu}{\lambda D}\frac{X_2X_1}{q_1q_2}\left[\frac{\varepsilon_{22}+\varepsilon_{20}+\varepsilon_{11}}{(S_{22}S_{11} - S_{12}S_{21})}\right]$$

Subtract $\frac{t_2}{q_2}$ from $\frac{t_1}{q_1}$ to obtain the "Corlett-Hague" rule

Q: The "Corlett-Hague rule" implies that goods that are more complementary to work, or equivalently, more substitutable with leisure, should be taxed at a lower rate than goods that are more substitutable with work, or equivalently, more complementary to leisure. This provides a rationale for the reduced effective taxation of public transport, repairs of private dwellings and restaurant meals observed in many countries. The information requirements to implement the "Corlett-Hague rule" (the cross-price elasticities with labor) is much lower than for the standard "Ramsey rule" (all own-price and cross-price elasticities). This makes the rule more operational in practice.

Part 3: Shorter questions

(3A) **Q**: The assumption is that changes in tobacco taxes do not correlate with unobserved factors that affect the well-being of people with different smoking propensitites in a systematically different manner.

Q: This assumption would fail, for instance, if the equation controls ineadequately for the business cycle; if the business cycle is correlated with tobacco taxes (for example because deficits lead states to increase taxes); and if the business cycle affects high-propensity individuals differently than low-propensity individuals (for example because the former have lower education and therefore are more exposed to job losses etc). It would also fail if the revenue raised with higher cigarette taxes lead to spending that raises the well-being of high-propensity individuals more (or less) than low-propensity individuals (for example ash trays or lung cancer treatments).

(3B) **Q**: The left-hand side is the marginal utility of public spending whereas the right-hand side is the perceived marginal cost of public funds, which is higher than one because the tax instrument - the capital tax - causes a capital outflow that reduces the tax base. This implies that the level of public expenditure, which is perceived as optimal by the individual country, is too low relative to the first-best level of public expenditure defined by $G'(r) = 1$. This can be interpreted as the result of an international positive externality of taxation. When the individual country sets its tax rate, it does not take into account the positive effects capital taxation has on other countries: what is perceived as a costly capital outflow from the perspective of the taxing country is also a beneficial capital inflow from the perspective of other countries. By not considering this positive external effect, capital taxes become lower than socially optimal.

Q: The individual country cannot by itself improve welfare - the tax rate defined by the first-order condition is by definition welfare-maximizing given the policy intrument at its disposal. But if all countries would cooperate on capital taxation, they would choose the first-best public expenditure level defined by $G'(r) = 1$. This can interpreted as the internalization of the positive externality of capital taxation. When countries make decisions about capital taxation jointly, they will take into account both the costly capital outflow and the beneficial capital inflow resulting from taxation and thus choose capital taxes at the socially optimal level.

(3C) **Q**: One type of taxes on alcohol - excise taxes - are included in the posted price so that tax changes should affect demand in exactly the same way as changes in the producer price whereas another type of taxes on alcohol - sales taxes - are not included in the posted price so that the effect of tax changes on demand depends on the importance of tax salience. By estimating the demand elasticities with respect to each of the two taxes, one can gauge the importance of tax salience for demand decisions. If demand responds much less strongly to changes in the less salient sales tax than to changes in the fully salient excise tax, it can be inferred that tax salience is an important issue.

Q: The salience parameter $\theta = \varepsilon_{x,1+\tau}/\varepsilon_{x,p}$ measures how responsive demand is to changes in the (unposted) sales taxes relative to changes in the posted price. A low (high) value therefore means that the tax requires large (small) cognitive and other costs for consumers to take into account when making demand decisions. It can be computed as the ratio of the coefficient on $\Delta \log(1 +$ sales tax rate) in the second row to the coefficient on $\Delta \log(1 +$ excise tax rate) in the first row. The value is between 0 and 0.25 depending on the exact specification. The coefficients are not very precisely estimated, however, which means that the confidence interval around θ is very large. In the last specification, for instance, the point estimate of θ is less than 0.1 but it is not significantly different from 1 at the 5% significance level.