Answers to exam in Tax Policy, January 2013.

Part 1

(1A) **Q**: The parameter δ expresses standard time discounting: the tendency to attach more weight to utility gains that are closer in time. Under this form of discounting, the consumer is indifferent between one unit of consumption in period t and $1/\delta > 1$ units of consumption in period $t+1$ for any t. This form of discounting implies that consumers make *time consistent* consumption plans because the discount factor between two time periods is constant over time. The parameter β expresses hyperbolic discounting: an additional discounting of future consumption relative to current consumption or, in other words, a "present-bias" in the intertemporal preferences. This form of discounting implies that consumers make *time inconsistent* consumption plans because the discount factor between two time periods changes over time. For instance, when making consumption plans at $t = -1$, consumption at $t = 1$ is discounted at the rate δ relative to consumption at $t = 0$. When making consumption plans at $t = 0$, consumption at $t = 1$ is discounted at the rate $\beta\delta$ relative to consumption at $t = 0$, hence the plan made at $t = -1$ is no longer optimal. Hyperbolic discounting is frequently used to model self-control problems whereby consumers over-consume "sin goods" that are associated with current utility gains and future utility losses (cigarettes, sweets, fatty foods etc.).

Q: Combine the intertemporal and instantaneous utility functions to obtain:

$$\begin{aligned} U^0 &= \rho \ln(x_0) + \sigma \ln(y_0) + z_0 - \gamma \ln(x_{-1}) + \\ &\quad \beta\delta \{\rho \ln(x_1) + \sigma \ln(y_1) + z_1 - \gamma \ln(x_0)\} + \\ &\quad \beta\delta^2 \{\rho \ln(x_2) + \sigma \ln(y_2) + z_2 - \gamma \ln(x_1)\} \end{aligned} \tag{1}$$

Construct the Lagrangian $\mathcal{L}$ and differentiate with respect to x_0, y_0 and z_0 to obtain the following first-order conditions for optimal consumption in period 0:

$$\frac{\partial \mathcal{L}}{\partial x_0} = \frac{\rho}{x_0} - \frac{\beta\delta\gamma}{x_0} - \lambda p_x = 0 \tag{2}$$

$$\frac{\partial \mathcal{L}}{\partial y_0} = \frac{\sigma}{y_0} - \lambda p_y = 0 \tag{3}$$

$$\frac{\partial \mathcal{L}}{\partial z_0} = 1 - \lambda = 0 \tag{4}$$

where λ denotes the Lagrangian multiplier. It follows from (4) that $\lambda = 1$. Insert into (2) and (3) to obtain:

$$x^* = \frac{\rho - \beta\delta\gamma}{p_x}$$

$$y^* = \frac{\sigma}{p_y}$$

Insert x^* and y^* into the budget constraint and rearrange to obtain:

$$z^* = B - (\rho - \beta\delta\gamma + \sigma)$$

(1B) **Q:** The government maximizes the long-run preferences of the consumer, that is utility derived from consumption in periods $0, 1$ and 2 as evaluated in period -1. This is equivalent to maximizing U^0 while correcting for the present-bias, that is setting $\beta = 1$.

Q: The government thus maximizes U^0 evaluated at $\beta = 1$:

$$\begin{aligned} &\rho \ln(x_0) + \sigma \ln(y_0) + z_0 - \gamma \ln(x_{-1}) + \\ &\delta \left\{\rho \ln(x_1) + \sigma \ln(y_1) + z_1 - \gamma \ln(x_0)\right\} + \\ &\delta^2 \left\{\rho \ln(x_2) + \sigma \ln(y_2) + z_2 - \gamma \ln(x_1)\right\} \end{aligned} \tag{5}$$

Construct the Lagrangian $\mathcal{L}$ and differentiate with respect to the tax rate on potato chips in period 0 to obtain the following first-order condition for optimal taxes:

$$\frac{d\mathcal{L}}{dt_x} = \frac{(\rho - \delta\gamma)}{x_0}\frac{\partial x_0}{\partial t_x} + \frac{\sigma}{y_0}\frac{\partial y_0}{\partial t_x} + \frac{\partial z_0}{\partial t_x} + \mu\left\{t_x\frac{\partial x_0}{\partial t_x} + x_0 + t_y\frac{\partial y_0}{\partial t_x}\right\} = 0 \tag{6}$$

where μ denotes the Lagrangian multiplier.

It follows from the demand functions that:

$$\frac{\partial x^*}{\partial t_x} = -\frac{\rho - \beta\delta\gamma}{(p_x)^2} \tag{7}$$

$$\frac{\partial y^*}{\partial t_x} = 0 \tag{8}$$

$$\frac{\partial z^*}{\partial t_x} = 0 \tag{9}$$

Insert (7)-(9) into (6) to obtain:

$$\frac{(\rho - \delta\gamma)}{\frac{\rho - \beta\delta\gamma}{p_x}}\left(-\frac{\rho - \beta\delta\gamma}{(p_x)^2}\right) + \mu\left\{t_x\left(-\frac{\rho - \beta\delta\gamma}{(p_x)^2}\right) + \left(\frac{\rho - \beta\delta\gamma}{p_x}\right)\right\} = 0$$

Reduce to obtain:

$$\frac{t_x}{1 + t_x} = \frac{\mu - \frac{\rho - \delta\gamma}{\rho - \beta\delta\gamma}}{\mu}$$

(1C) **Q:** The optimal tax formula ressembles the inverse elasticity rule where the elasticity of demand is one and the marginal social value of private income is one due to the quasi-linear form of the instantaneous utility function. The formula thus departs from the inverse elasticity rule only by having $-(\rho - \delta\gamma)/(\rho - \beta\delta\gamma)$ as the second term in the numerator rather than -1. In the special case where $\beta = 1$ and the consumer does not suffer from hyperbolic discounting, the term $-(\rho - \delta\gamma)/(\rho - \beta\delta\gamma)$ collapses to -1 and the standard inverse elasticity rule prevails. In cases where $\beta < 1$ and the consumer suffers from hyperbolic discounting, the tax on potato chips is higher, which reduces demand for this good. The optimal tax formula thus corrects for the optimization error made by the consumer. This is akin to a Pigouvian tax correcting for an externality. In this case, the tax corrects for an "internality", that is the harm inflicted by the consumer on himself due to the hyperbolic discounting.

Part 2

(2A) **Q:** The *mechanical revenue effect*: Holding behavior constant, the tax revenue increases by:

$$\Delta M = \Delta\tau\Delta z[1 - H(z)]$$

The mass of people with income above z, that is $1-H(z)$, each pay an increaed tax bill of $\Delta\tau\Delta z$ because the marginal tax is increased by $\Delta\tau$ over an income interval of length Δz.

The *behavioral revenue effect*: Behavioral responses to the tax reduce tax payments by:

$$\Delta B = -h(z)\frac{\partial z}{dT'(z)}\Delta\tau T'(z)$$

The mass of people who become subject to a higher marginal tax, that is $h(z)$, each reduce their taxable income by $-(\partial z/dT'(z))\cdot\Delta\tau$ and each dollar of reduction in taxable income generates a revenue loss of $T'(z)$. This equation can be rewritten in the following way using the elasticity of taxable income $e(z)$:

$$\Delta B = h(z)e(z)z\frac{T'(z)}{1-T'(z)}\Delta\tau\Delta z$$

The *social welfare cost*: The increased tax payment is associated with utility costs with a social value of:

$$\Delta W = \Delta\tau\Delta z[1 - H(z)]G(z)$$

This is simply the increased tax payment (absent behavioral responses) for persons with income exceeding z multiplied by the average social welfare weight for persons with income exceeding z, that is $G(z)$. As usual, behavioral responses to a small tax change have no first-order effect on individual utilities.

(2B) **Q:** In an optimum, it must hold that the social gain of a small tax increase equals the social cost of a small tax increase:

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

Insert the expressions derived under the previous question:

$$\Delta\tau\Delta z[1 - H(z)] = \Delta\tau\Delta z[1 - H(z)]G(z) + h(z)e(z)z\frac{T'(z)}{1-T'(z)}\Delta\tau\Delta z$$

Reduce and rearrange to obtain:

$$\frac{T'(z)}{1-T'(z)} = \frac{1-G(z)}{e(z)}\cdot\frac{1-H(z)}{zh(z)}$$

Q: A larger value of $e(z)$ implies that those who face increased marginal tax rates respond more strongly to taxation. This in turn implies a larger marginal efficiency loss and a lower optimal marginal tax rate.

A larger value of $1 - G(z)$ implies lower welfare weight on those facing increased tax bills. This in turn implies a higher optimal marginal tax rate. A larger value of $1 - H(z)$ implies larger mass of people who pay higher tax bills after reform but face no change in marginal tax rates and hence do not change behavior. This in turn implies a larger mechanical revenue gain relative to the behavioral revenue loss and therefore a higher optimal marginal tax rate. A larger value of $h(z)$ implies a larger mass of people who face an increase in the marginal tax rate. This in turn implies a larger marginal efficiency loss and a lower optimal marginal tax rate.

(2C) Assume that the marginal social welfare weight on those with the very highest incomes is zero and that $e(z) \simeq 0.25$. According to the figure in Annex A, $zh(z)/(1 - H(z)) \simeq 1.5$ for values of z larger than \$400.000. Inserting these values in the formula for the optimal marginal tax rate yields $T'(z)/1 - T'(z) = 2.67$ which implies that $T'(z) = 0.73$. These computations thus suggest that the optimal marginal tax rate on incomes above \$400.000 is around 73%. Since $zh(z)/(1-H(z))$ is decreasing in income over the range \$100.000 to \$400.000 and the marginal social welfare weight is decreasing in income, the optimal marginal tax rate implied by the formula must be strictly increasing in income over the range \$100.000 to \$400.000. In sum, the model suggests that optimal marginal tax rates on high incomes are high and increasing in income.

Part 3

(3A) **Q:** When there exists an *immutable* and *verifiable* characteristic, which *correlates with income*, it is optimal to condition the tax schedule on this characteristic. Examples include different types of disability and old age, but also to some extent single parenthood. **Q:** The intuition for the result is the following: Since the characteristic is correlated with income, some measure of redistribution can be achieved by letting persons with the "high-income" characteristic pay higher taxes than persons with the "low-income" characteristic. Moreover, the tax wedge between persons with "high-income" and "low-income"characteristics induces no behavioral responses and hence no efficiency loss, since the characteristic is exogenous. A controversial paper by Mankiw and Weinzierl (2010) shows that height satisfies the conditions laid out above and shows how the optimal system would optimally have tall people paying more taxes than short people for a given income.

(3B) **Q**: The paper conducts a controlled experiment to study the effect of tax salience on consumer demand. In many US states, sales taxes are not included in the posted price but are added at the register. The experiment increases the tax salience by adding price tags with the after-tax price to certain product categories ("treated products") in certain grocery stores ("treated stores"). In the first step, only treated stores are considered. The difference-in-difference estimate is the change in the demand for treated products between the baseline period and the experimental period relative to the change in the demand for non-treated products between the same periods. The estimate is the difference between the time difference in demand for the treated products (-1.30) and the time difference in demand for the non-treated products (0.84), hence the estimate -2.14. In a second step, the exact same computations are carried out for non-treated stores where all products are non-treated. This gives rise to a placebo estimate of 0.06. The difference-in-difference-in-difference estimate is the difference between the real difference-in-differences estimate and the placebo difference-in-differences estimate, hence the estimate is -2.20. **Q**: The difference-in-differences estimator controls for shop-specific shocks to demand and assumes that there are no product-specific shocks. The difference-in-difference-in-differences estimator controls for shop-specific shocks and product-specific shocks to demand and assumes that there are no product-shop-specific shocks.

(3C) **Q**: Under the new view of capital taxation, firms are cash-rich and the marginal source of finance is retained earnings. Retained earnings constitute "trapped cash", which will eventually be paid out as dividends and hit by the dividend tax. The dividend tax rate has no bearing on firms' optimal choice between immediate pay out on one hand and investment and later pay out on the other hand and therefore does not affect the optimal investment level and the optimal dividend level. Under the old view of capital taxation, firms are cash-poor and the marginal source of finance is new equity. The return to new equity will eventually be paid out as dividends and the investors have access to alternative investments, which are not hit by the dividend tax (such as bonds). By making new equity less attractive relative to alternative investments, the dividend tax depresses the investment level. Cash-constrained firms optimally pay no dividends regardless of the dividend tax rate, hence the dividend tax has no effect on dividend payments in the short run. In the long run, the dividend tax may, however, depress dividend pay outs by reducing the total capital stock of corporations.