Exam - Tax Policy - Fall 2016 - RESIT - Answers

Part 1: Commodity taxation

(1A) **Q:** The langrangian to the government problem writes

$$\mathcal{L}_G = V(q, Z) + \lambda[\sum_j t_j X_j(q, Z) - T]$$

The first-order condition for q_k equals

$$\frac{\partial \mathcal{L}_G}{\partial q_k} = \frac{\partial V}{\partial q_k} + \lambda[X_k + \sum_j t_j \partial X_j/\partial q_k] = 0$$

Use $\alpha = \partial V/\partial Z$ and Roy's identity to rewrite as:

$$(\lambda - \alpha)X_k + \lambda \sum_j t_j \partial X_j/\partial q_k = 0$$

Now insert the Slutsky equation to obtain:

$$(\lambda - \alpha)X_k + \lambda \sum_j t_j (S_{jk} - X_k \partial X_j/\partial Z) = 0$$

Insert $\mu \equiv \alpha + \lambda(\sum_j t_j \partial X_j/\partial Z)$ to obtain:

$$(\lambda - \mu)X_k + \lambda \sum_j t_j S_{jk} = 0$$

Rearrange as:

$$\frac{\lambda - \mu}{\lambda} = -\frac{\sum_j t_j S_{jk}}{X_k}$$

Q: The numerator on the right-hand side is the revenue effect of the compensated behavioral responses to a small increase in the tax on good k. This can be interpreted as the marginal excess burden of the tax increase. The denominator is the mechanical revenue effect of a small increase in the tax on good k. The right hand side thus expresses the share of the potential revenue gain from a small tax increase that is lost. The result shows that this share should be equalized across all instruments. This ensures that the total excess burden is minimized given the revenue constraint. The optimal commodity tax system thus maximizes economic efficiency.

(1B) **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.

(1C) **Q:** Doyle and Samphanthrak (2008) exploit that two U.S. states, Illinois and Indiana, first repealed and later reinstated, their gasoline taxes whereas the neighboring states left the gasoline taxes unchanged. The timing of these tax changes give rise to three natural experiments: (i) the simultaneous repeal of the gasoline tax by Illinois and Indiana; (ii) the reinstatement of the gasoline tax by Indiana; (iii) the reinstatement of the gasoline tax by Illinois. The causal impact on the consumer price is estimated by estimating the price change in the "treatment" state over and above the price change in the "control states" in a short time window around the tax change. The identifying assumption is that the average percentage change in gasoline prices across treatment and control states would have been identical in absence of the policy interventions in the treatment states (conditional on covariates).

Q: The study identifies the incidence over a time window of a few days. The short time window renders identification more convincing by making it more plausible that unobserved factors are constant over the time of the study. However, the incidence parameter that is relevant for policy is really the one that applies over the long term. It is plausible that the main finding of less-than-full shifting of gasoline taxes to the consumers only holds in the short term where supply is somewhat sticky whereas there is full shifting in the long term where the supply adjusts flexibly.

Part 2: Income taxation

(2A) **Q:** Disabled individuals cannot work and therefore make no choices over leisure and consumption. They simply consume G_A; hence, their utility can be expressed as $V(\omega, G_A) = U(G_A, 0) = G_A$.

Workers maximize $U(C,L) = C - g(L)$ subject to the constraint $C = G_B + (1-t)wL$. Inserting the constraint into the objective, workers choose the labor supply L in order to maximize:

$$U(C,L) = G_B + (1-t)wL - g(L)$$

The first-order condition reads: $\omega = g'(L)$ where $\omega = (1-t)w$ is the after-tax wage rate. This condition implicitly defines the labor supply as a function of the after-tax wage, $L = L(\omega)$. To derive the properties of the labor supply, differentiate the first-order condition with respect to ω to obtain:

$1 = g''(L)\frac{dL}{d\omega}$. It follows that $\frac{dL}{d\omega} = \frac{1}{g''(L)} > 0$ so that the labor supply is unambiguously increasing in the after-tax wage rate. Hence, the indirect utility of workers is given by:

$$V(\omega, G_B) = G_B + \omega L(\omega) - g(L(\omega))$$

Q: If inability were not immutable, individuals would additionally need to "choose" whether to be disabled or not while weighing the cost of changing their disability status against the utility benefit of the change.

(2B) **Q:**The first-order conditions for G_A and G_B read:

$$\text{FOC } G_A: \qquad \Psi'(G_A) - \lambda = 0$$

$$\text{FOC } G_B: \qquad \int_{\underline{w}}^{\overline{w}} \Psi'(V(\omega, G_B)) f(w) dw - \lambda = 0$$

It follows directly that

$$\Psi'(G_A) = \int_{\underline{w}}^{\overline{w}} \Psi'(V(\omega, G_B)) f(w) dw$$

This states that the welfare gain from dividing a dollar between the disabled (LHS) should equal the welfare gain from dividing a dollar between the workers (RHS) in the optimum. Employed workers must be better off being employed than being unemployed and receiving utility G_B. Hence, average utility among the workers is strictly larger than G_B and the right-hand side of the equation is therefore strictly smaller than $\Psi'(G_B)$. It follows directly that $\Psi'(G_A) < \Psi'(G_B)$ and therefore, given the properties of Ψ, that $G_A > G_B$. Intuitively, starting from a situation where $G_A = G_B$, the welfare gain from splitting a dollar between the disabled is larger than splitting it between the workers because the former are poorer on average. Hence, in the optimum, $G_A > G_B$.

Q: If inability were not verifiable, the government would not be able to differentiate between disabled and workers; hence, the tax system would need to satisfy $G_A = G_B$.

(2C) **Q:** Tagging models show that if there exists some *immutable* and *verifiable* characteristic correlating with earnings ability, it is optimal to condition the tax schedule on this characteristic. Intuitively, a tax wedge between tagged and untagged causes no distortion of behavior because the characteristic is exogenous and conditioning the tax schedule on the tag improves equity because it correlates with earnings ability

Q: In principle, it is desirable to apply lower marginal taxes to women given that (i) gender is approximately immutable and costlessly verifiable and (ii) women earn less than men on average. Hence, a tax wedge between men and women would allow for more redistribution between single-men and single-women without distorting labor supply. Such a policy would, however, violate the principle of *horizontal equity* whereby two individuals with the same relavant characteristics should be treated identically.

Part 3: Shorter questions

(3A) **Q:** The *excess burden* of a tax is the dollar equivalent utility loss suffered by the consumers due to the tax *in excess of* the tax revenue. This is a measure of the efficiency loss of taxation. This can be illustrated in a partial-equilibrium diagram with a (down-ward sloping) compensated demand curve and a (flat) supply curve. The excess burden is the area between the compensated demand curve and the pre-tax supply curve taken between the pre-tax and the after-tax compensated demands.

Q: The excess burden of a tax t can thus be approximated with the area of the triangle:

$$EB = \frac{1}{2} t \Delta x$$

where Δx is the change in compensated demand induced by the tax. With a linear approximation for Δx this expression can be rewritten in the following way:

$$EB = -\frac{1}{2} \frac{x_0}{p+t} t^2 \varepsilon^C$$

where ε^C is the compensated elasticity of demand for x with respect to the tax inclusive price $p+t$. The excess burden of a tax is thus increasing linearly in the compensated elasticity of demand and increasing with the square of the tax rate.

(3B) **Q:** The cooperation is known as "information exchange on request" and implies that tax authorities can ask foreign tax authorities for tax relevant information, e.g. bank information, about specific tax payers when the "foreseeable relevance" of the requested information can be ascertained. This type of cooperation has a number of weaknesses. First, it is implemented through bilateral treaties; hence, tax evaders can avoid any increase in the detection risk by moving assets to a tax haven that does not have a treaty with their home country. Second, tax authorities can only obtain information in cases where they already have some evidence of tax evasion because of the requirement of foreseeable relevance.

Q: This first figure shows that havens signing more treaties with non-havens experienced a lower (sometimes negative) growth in the total deposits held by foreigners in their banks. The second figure shows that non-havens signing more treaties with havens experienced no less growth in the total deposits held by its residents in havens. The first correlation suggests that some tax evaders shifted deposits from the havens signing a treaty with their home country to the havens not signing such a treaty whereas the second correlation suggests that treaties did not induce tax evaders to repatriate deposits.

(3C) **Q:** Yagan (2015) exploits that corporations in the U.S. can choose between two fundamentally different tax treatments. They can elect that current profits are taxed at the corporate level at the rate t_c and that distributed profits are taxed at the shareholder level at the rate t_d ("C-corporations"). Alternatively, they can elect that current profits are taxed at the shareholder level at the personal income tax rate t_p and that no corporate or dividend taxes apply ("S-corporations"). Since C-corporations are subject to the dividend tax whereas S-corporations are not, Yagan (2015) can apply a difference-in-differences type of estimator where S-corporations work as a "control group" for the C-corporations that are "treated" with a large dividend tax reduction in the context of the 2003 tax reform. Empirically, C-corporations are overrepresented among large firms whereas S-corporations are overrepresented among small firms. This raises the concern that unobserved shocks that correlate with size will cause divergence between the "treatment group" and "control group". This concern is addressed in two ways. First, the estimation sample is restricted to firms of intermediate sizes. In this range, the size (and industry) distribution is shown to be reasonably balanced across C-corporations and S-corporations. Second, the regression weights are adjusted such that S-corporations have the precise same weight as C-corporations within each narrow size-industry bin. This ensures that size-industry specific shocks to investment decisions do not affect the difference-in-differences estimator.