# Exam - Tax Policy - Summer 2019

## Read carefully before you start:

The exam consists of three parts each with a number of subquestions. You are supposed to answer ALL questions and subquestions. Good luck!

## Part 1: Tax salience

(1A) In the United States, price tags in retail stores generally indicate prices before sales tax. This implies that sales taxes have a low degree of salience to consumers when they make spending decisions. Chetty, Looney and Kroft (2009) make an experimental intervention that increases the salience of taxes by adding post-tax prices to the price tags. The intervention is conducted on a selected series of *treated categories* in a selected series of *treatment stores* in a specific *experiment period*. The researchers collected data on daily sales at the product-level. Sales data is also collected for products that are never treated, *control categories*, stores where no products are ever treated, *control stores*, and a time period where no products are treated, the *baseline period*.

The DiD analysis uses sales data for the treatment stores only. The DiD estimate of -2.14 is given by the *change* in mean sales of treated categories between the baseline period and the experiment period (-1.30) minus the *change* in mean sales of control categories between the baseline period and the experiment period (0.84). The second change is assumed to be the *counterfactual* change in sales of treated categories absent any experimental intervention. The DID estimate is the difference between the actual change and the counterfactual change in sales of treated categories. The identifying assumption is that sales of treated categories would have evolved in the same way as control categories without the experimental intervention. The DiD estimator is generally robust to store-specific shocks: an identical change in treatment and control categories due to factors external to the experiment does not affect the DiD estimator. The estimator is not robust to all product-specific shocks: shocks affecting treated categories differently than control categories invalidates the DiD estimate.

The DiDiD analysis uses sales data from both treatment and control stores. The DiDiD estimate of -2.20 is given by the difference between the DiD estimate from treatment stores (-2.14) and an equivalent placebo DID estimate for control stores where no intervention takes place (0.06). The second DID estimate is assumed to be the *counterfactual* differential change in sales of treated categories relative to control categories absent any experimental intervention. The DIDiD estimate is the difference between the actual and the counterfactual differential change in sales of treated categories relative to control categories. The identifying assumption is that differential sales of treated categories relative to control categories would have evolved in the same way in the treated stores as in the control stores without the experimental intervention. The DiDiD estimator is generally robust to store-specific shocks: an identical change in treatment and control categories within stores due to factors external to the experiment does not affect the two DiD estimates. It is also generally robust to product-specific shocks: a differential change in treatment categories relative to control categories that is identical in both types of stores affects both DiD estimates in the same way and therefore leaves the DiDiD estimate unaffected. The estimator is not robust to all product-store-specific shocks: a differential change in treated categories relative to control categories that varies systematically across treatment and control stores invalidates the DiDiD estimate.

(1B) The excess burden is the area of the triangle under the price-demand curve, which tracks the consumer's true valuation of the commodity. The baseline of the triangle is the distance between the quantity chosen without taxes $x_0$ and the quantity chosen with taxes $x_1$. This can be written in terms of the tax-demand curve as $-(dx/dt^S)t^S$. The height of the triangle is the baseline $-(dx/dt^S)t^S$ multiplied by the slope of the price-demand curve $-1/(dx/dp)$. The excess burden is therefore:

$$EB = -\frac{1}{2}\frac{(t^S)^2(dx/dt^S)^2}{dx/dp}$$

Using the definition $\theta \equiv (dx/dt^S)/(dx/dp)$, we can rewrite as:

$$EB = -\frac{1}{2}(\theta t^S)^2 dx/dp$$

This expression can again be written in terms of elasticities

$$EB = -\frac{1}{2}(\theta t^S)^2 \varepsilon_{D,q|p}\frac{X}{q}$$

The excess burden of a tax at the rate $t$ with a degree of misperception $\theta$ is identical to the excess burden of a fully perceived tax at the rate of $\theta t$.

When $\theta = 0$, the consumer by definition buys the same amount of the taxed good regardless of the tax rate. In the absence of income effects, this implies that the consumer chooses the exact same bundle as under a lumpsum tax. It follows that the excess burden is zero. In the presence of income effects, it implies that the consumer buys more of the taxed good and less of the untaxed good than under a lumpsum tax. It follows that the excess burden is positive.

## Part 2: Income taxation

(2A) Increasing tax $T_i$ to $T_i + dT_i$ affects welfare though these channels:

Mechanical revenue effect:

$$\Delta M_i = h_i dT_i$$

Social welfare cost:

$$\Delta W_i = g_i h_i dT_i$$

Behavioral revenue effect:

$$\Delta B_i = \frac{dh_i}{d(c_i - c_0)} dT_i (T_i - T_0) = \eta_i h_i dT_i \frac{(T_i - T_0)}{(c_i - c_0)}$$

The mechanical revenue effect captures the increase in government revenue holding behavior (i.e. labor supply decisions) constant. The social welfare cost captures the decrease in private disposable income holding behavior constant and expressed in units of government revenue. The behavioral revenue effect captures the decrease in government revenue deriving from behavioral responses to the tax change (i.e. changes in labor supply decisions).

There is an additional direct effect on the utility of the individuals who change behavior (i.e. stop working) in response to the tax change; however, this effect is second-order because these individuals are initially indifferent between working and not working. The effect can therefore be ignored given that the tax change is marginal.

(2B) In the optimum, a small change in $T_i$ should have no effect on welfare; hence, it must be satisfied that:

$$\Delta M_i = \Delta W_i + \Delta B_i$$

Inserting the expressions for $\Delta M_i, \Delta W_i$ and $\Delta B_i$ derived under (2A) and rearranging yields the following equation that characterizes optimal taxation at the earnings level associated with occupation $i$:

$$\frac{T_i - T_0}{c_i - c_0} = \frac{1}{\eta_i}(1 - g_i)$$

Since the average value of $g_i$ across all individuals equals one, $g_i$ must be below one at some income levels. It is possible that $g_0 > 1$ and $g_i < 1$ for all $i \geq 1$.In all other cases, it must hold that $g_1 > 1$ since the preference for redistribution implies a higher social marginal welfare weight on individuals with lower income. If $g_1 > 1$, the equation implies that $T_1 < T_0$: individuals at the lowest earnings level optimally receive more transfers than individuals with no earnings. In other words, the marginal tax rate is optimally negative at the bottom of the income distribution. Intuitively, redistributing to individuals with low earnings involves small efficiency losses because it only distorts the labor supply decision of individuals with low skills whereas redistributing to individuals with zero earnings is more costly because it distorts the labor supply decision of individuals at all skill levels. This intuition is partly an artifact of the feature that the model only has an extensive margin.

(2C) In the scenario with a large participation elasticity ($\eta = 1$), the slope of the simulated curve is larger than unity at low earnings levels, which implies a negative marginal tax rate. This resembles the actual tax schedule under the EITC where earnings at low levels are indeed subsidized. In the scenario with no participation elasticity ($\eta = 0$), the slope of the simulated curve is much smaller than unity even at low earnings levels, which implies a positive and rather high marginal tax rate. This resembles the policy prescriptions from the Mirless model and the actual tax schedule under a Negative Income Tax (NIT). The simulation results suggest that the EITC is closer to the optimal tax schedule than the NIT when there are large participation responses to taxation.

In the sample of wage earners, Emmanuel Saez (2010) finds essentially no bunching around the kinks in the EITC schedule where the marginal tax rate changes discretely. This is suggestive of a low tax elasticity of the labor supply on the *intensive margin,* but is not informative about the tax elasticity of the labor supply on the *extensive* margin. Hence, the bunching results suggest that the simulation should use a value of the hours elasticity, $\varepsilon$, close to zero, but does not inform the choice of participation elasticity, $\eta$.

## Part 3: Shorter questions

(3A) Labor demand depends on the wage rate including the cost of providing the benefit: $D(w+t)$. Labor supply depends on the cash wage rate plus the perceived value of the benefit: $S(w+\alpha t)$. Hence, equilibrium requires that:

$$D(w+t) = S(w+\alpha t)$$

Differentiate with respect to $w$ and $t$ to obtain:

$$D'(\cdot)(dw+dt) = S'(\cdot)(dw+\alpha dt)$$

Rewrite to obtain:

$$-dw(S'(\cdot)-D'(\cdot)) = dt(\alpha S'(\cdot)-D'(\cdot))$$
$$\frac{dw}{dt} = -\frac{\alpha S'(\cdot)-D'(\cdot)}{S'(\cdot)-D'(\cdot)} = -\frac{S'(\cdot)-(1-\alpha)S'(\cdot)-D'(\cdot)}{S'(\cdot)-D'(\cdot)}$$
$$\frac{dw}{dt} = -1+\frac{(1-\alpha)S'(\cdot)}{S'(\cdot)-D'(\cdot)}$$

Multiply each term in the fraction with $w/S(\cdot)$ and evaluate at $t=0$ to rewrite in terms of elasticities:

$$\frac{dw}{dt} = -1+(1-\alpha)\frac{\varepsilon^S}{\varepsilon^S-\varepsilon^D}$$

When $\alpha=1$, the benefit is valued at cost and the formula shows that $dw/dt=-1$ which implies that workers bear the full cost of the benefit through a reduction in the wage. Intuitively, when employers are fully compensated for the cost of the benefit with a reduction in the wage, their demand for labor does not change relative to a situation with no mandated benefit. Similarly, the supply of labor does not change because the gross wage (including the perceived value of the benefit) precisely offsets the reduction in the cash wage rate. Hence, there exists an equilibrium with mandated benefits where employment is the same as without mandated benefits, but the wage rate is reduced by the full cost of the benefit. To illustrate in a ($w$,$E$)-diagram, the mandate moves both the labor demand curve and the labor supply down by the cost of the benefit. The equilibrium employment remains the same but the wage rate drops by the cost of the benefit.

When $\alpha = 0$, the benefit is not valued at all and the formula implies that $dw/dt = \varepsilon^D/(\varepsilon^S - \varepsilon^D)$ which suggests that workers and employers share the cost of the benefit in proportions that depend on the relative size of the labor demand and labor demand elasticities. Intuitively, the mandate now works precisely like a tax because it represents a cost for the employers that does not represent a direct benefit for the workers. Employers therefore reduce labor demand, which reduces the wage rate and depresses labor supply. The new equilibrium exhibits a lower wage rate and a lower employment level than the equilibrium without the mandate. If labor supply is more elastic relative to labor demand, a smaller wage reduction is needed to restore equilibrium and a larger share of the cost is borne by the employers ($dw/dt$ approaches zero as $\varepsilon^S$ approaches infinity). To illustrate in a ($w$,$E$)-diagram, the mandate moves the labor demand curve down by the cost of the benefit whereas the labor supply curve does not move. The new equilibrium is where the new labor demand curve intersects the old labor supply curve.

(3B) The identification strategy exploits that U.S. corporations can choose between two fundamentally different tax treatments. If they elect to be C-corporations, current profits are taxed at the corporate level at the rate $t_c$ and distributed profits are taxed at the shareholder level at the rate $t_d$. If they elect to be S-corporations, current profits are taxed at the shareholder level at the personal income tax rate $t_p$ and no corporate or dividend taxes apply. There is a range of sizes where S-corporations and C-corporations co-exist in almost all industries.

In this institutional setting, the tax reform in 2003, which lowered the dividend tax rate from 35% to 15%, creates quasi-experimental variation by changing the tax environment for C-corporations but not for S-corporations.

Yagan thus uses S-corporations as a "control group" to establish how the investment rates of C-corporations would have evolved in a counterfactual world without the tax reform. The difference between the actual and counterfactual change in the investment rate of C-corporations amounts to a difference-and-difference estimate of the treatment effect of the reform.

The investment rates of C-corporations and S-corporations evolve similarly throughout both before and after the dividend tax reform. The implied elasticity of corporate investment with respect to the dividend tax rate is close to zero with relatively small standard errors suggesting that investment is essentially unaffected by the dividend tax rate. This is precisely the prediction delivered by the new view of firm taxation: if the marginal investment is financed by retained earnings (or debt), the dividend tax rate has no bearing on the cost of capital and should therefore not affect investment rates. Under the old view, marginal investment is financed by equity, in which case a decrease in the dividend tax rate lowers the cost of capital. Yagan's empirical findings are therefore only consistent with the old view if the elasticity of corporate investment with respect to the cost of capital is very low (or zero).

(3C) The first-order condition for utility maximization is:

$$(1-t)w = v'(L)$$

Differentiating with respect to the tax rate yields:

$$-wdt = v''(\cdot)dL$$

$$\frac{dL}{dt} = \frac{-w}{v''(\cdot)} < 0$$

A small tax increase has the following total effect on revenue where the first term is the *mechanical* effect (holding labor supply constant) and the second term is the *behavioral* effect (the revenue effect of the labor supply adjustment)

$$\frac{dR}{dt} = wL + tw\frac{dL}{dt}$$

A small increase has the following effect on utility:

$$dU/dt = -wL + [(1-t)w - v'(L)]\, dL/dt = -wL$$

where the second equality sign follows from the first-order condition for utility maximization, which implies that the expression in square brackets is zero.

The "marginal excess burden" equals the revenue effect of the labor supply adjustment $twdL/dt$. While the positive mechanical revenue effect is a pure transfer from the individual, the same term appears in the effect on utility with the opposite sign, the negative behavioral revenue effect has no positive counterpart at the level of the individual and is therefore a welfare loss. Intuitively, the behavioral response to the tax change only has a second-order effect on utility given that the individual is already at an interior optimum (an application of the envelope theorem), but has a first-order effect on government revenue since the marginal units of labor are taxed at the rate $t$. More generally, the marginal excess burden of a policy change equals the revenue effect of the behavioral responses to the change in models with optimizing agents.