Answers
Exam in Public Finance - Spring 2013
3-hour closed book exam
Claus Thustrup Kreiner

Part 1: Tax pressure (in Danish: Skattetryk)

(**1A**) No. Tax pressure is defined as $TP = T/Y$, where T is total taxes and Y is total income. In order to show that tax pressure is not directly related to the efficiency loss from taxation consider two countries, A and B, with the same total taxes T and same total income Y, and therefore same tax pressure. Country A has a proportional tax system with $T = \tau Y$, where τ is the tax rate, while individuals in country B pay a poll tax (a constant tax paid per individual), giving the same revenue T. A person earning more income will have to pay more taxes in country A, which will distort behavior, but not in country B where the tax payment is independent of income (this may be illustrated graphically). Thus, two countries with the same tax pressure may have different efficiency losses from taxation. Tax distortions are related to the structure of the tax system and are not directly to tax pressure, although a higher tax pressure may be an indicator of a larger efficiency loss from taxation. It is possible to give other types of examples.

It may also be noted that two countries can have the exact same tax system but experience different efficiency losses because of differences in the elasticity of taxable income.

(**1B**) No. Consider the above example with country A and country B, and assume that tax revenue is used to finance, say, a public good that benefits everybody in society to the same extent. Country A will effectively have redistribution from high income persons to low income persons, while no redistribution is taking place in country B because everybody receives the same from the government and pays the same in taxes independent of income (this may be illustrated graphically).

Another obvious reason, that could be noted, is that two countries with the same tax system may have different transfer schemes, and therefore different degrees of redistribution. A related point is that two countries with the same degree of redistribution may have different measures of tax pressure if transfers are taxed in one country but are untaxed in the other country (e.g. tax pressure increases, ceteris paribus, when a country goes from having untaxed transfer to having taxable transfers with the same net-of-tax value, as Denmark did in 1995).

(**1C**) No, not if this is the only tax. By definition tax pressure equals $TP = T/Y$. In a proportional tax system with $T = \tau Y$, we have $TP = \tau Y/Y = \tau$, implying that tax pressure is constant, and therefore independent of business cycle variation. It may be noted that if the country also has other types of taxes then tax pressure may be procyclical or countercyclical. For example, with a constant consumption tax and a procyclical average propensity to consume, tax pressure will be procyclical.

Part 2: Social insurance

(**2A**) Moral hazard and adverse selection arise because of asymmetric information (between principal and agents) and will often lead to an inefficient outcome. Adverse selection refers to a situation where buyers of insurance know more about their risk level than the insurer and where risk differs across individuals. This implies that an insurance company may end up with customers with high risk levels who are costly for the company. The company will then have to raise the premium of the insurance contract implying that the selection of customers have even higher risk levels (i.e., adverse selection of customers). Moral hazard refers to adverse actions taken by agents because of the presence of insurance against adverse outcomes. For example, unemployment insurance may reduce the gain of going from unemployment to employment and thereby reduce search effort of unemployed. This is discussed more thoroughly in Chapter 12 of Hendricks and Myles (2006).

In the current context, the social planner faces a moral hazard problem. The social planner cannot observe search effort, which creates the moral hazard problem described above. There cannot be any adverse selection problems in the current context because individuals are (ex ante) identical.

(**2B**) It is stated that each worker chooses the effort level e that maximizes the expected utility

$$\Omega = e \cdot u(z-t) + (1-e)\,u(b) - v(e).$$

The derivative of this objective with respect to e gives

$$\frac{d\Omega}{de} = u(z-t) - u(b) - v'(e). \tag{*}$$

The first order condition for an optimum $d\Omega/de = 0$ gives the result

$$u(z-t) - u(b) = v'(e).$$

The right hand side (RHS) of this expression is the marginal disutility of search effort, which in the optimum equals the marginal benefit of search on the left hand side (LHS) of the expression. A marginal increase in the search effort level de increases the probability of employment by de

(by construction) and this gives a utility gain equal to $u(z-t) - u(b)$ where $z - t$ is the consumption/net income when employed, while b is the consumption level/benefit level of an unemployed.

(**2C**) The condition $z - t = b$ in equation (4) shows that net-income, and thereby consumption, is identical in the employment state and in the unemployment state. Thus, in the first best allocation, the tax t and the UI benefit b are set such that the consumption level of the individual when unemployed is same as the level when employed. This result of perfect insurance and perfect consumption smoothing maximizes individual utility because the utility function is strictly concave, i.e., workers are risk averse.

When e is unobservable, it is impossible to implement this system because the workers have no incentive to search for a job when unemployed, i.e., $e = 0$, implying that everybody would be unemployed and that it would be impossible to finance the UI system. To see that nobody would search for a job, note from the first order condition of the worker (*) that $z - t = b$ implies

$$\frac{d\Omega}{de} = -v'(e) < 0,$$

showing that it is optimal for the worker to reduce search effort until $e = 0$. The moral hazard problem arises because individuals do not bear the full income loss in case of unemployment. A part of the consequence is borne by other workers, which creates an externality through the UI program when e is unobservable.

(**2D**) Expression (5) is the so-called Baily-Chetty formula which characterized the second-best optimal UI benefit level when search effort is unobservable. The LHS represents the benefit side of the UI benefit system, i.e., the gain from consumption smoothing across states. A higher benefit level reduces the difference in consumption levels between the employment and unemployment states and reduces thereby the percentage consumption loss $\Delta c/c$ due to unemployment. The coefficient of relative risk aversion γ represents the degree of risk aversion of the worker. A high value of γ implies that workers are more risk averse which, ceteris paribus, will call for a smaller consumption difference between the employment state and the unemployment state, implemented by an increase in the benefit level b and a corresponding increase in the tax t. The RHS represents the efficiency loss from UI insurance due to moral hazard. The elasticity of unemployment with respect to the benefit level ε measures the size of the adverse effect on search effort behavior when the benefit level is increased. A high ε calls, ceteris paribus, for a low benefit level. On the other hand, with $\varepsilon = 0$ the moral hazard problem vanishes and it is possible to implement the first best insurance system.

With knowledge of the relevant parameters, it is possible to say whether UI benefits should be higher or lower. If, for example, the risk aversion parameter is $\gamma = 2$, the elasticity is $\varepsilon = 0.5$ and the employment rate is $e = 0.95$, which could be realistic parameters (an example from the lectures), then $\Delta c/c = 25\%$ is optimal. If consumption is equal to net-income, we would have $\frac{(y-t)-b}{y-t} = 1 - b/(y-t) = 0.25$ corresponding to the compensation rate $\frac{b}{y-t} = 75\%$. In practise, individual consumption does not have to equal net-income in each period because of the possibility of saving/dissaving. In this case, the optimal b may be significantly smaller than 75%.

(**2E**) Card et al. (2007) use a regression discontinuity approach to measure the impact of UI compensation (in their case measured as the maximum duration of unemployment insurance benefits) on unemployment duration. Consider the following model:

$$d_i = \beta_0 + \beta_1 b_i + \varepsilon_i,$$

where d_i is the duration of unemployment of a worker i, b_i is the maximum benefit duration of the individual, and ε_i is an error term. We are interested in whether a longer duration of UI benefits increases the duration of unemployment. In their data from Austria, individuals with a long work history (more than 36 months of employment during the last 5 years) prior to unemployment may receive UI benefits for a long period, while individuals with a short work history will receive UI benefit for only a short period of time. A simple estimation of the above relationship using the variation in maximum benefit duration across individuals to estimate β_1 will probably give a downward biased estimate of β_1 because unemployed individuals receiving benefits for a long period have had the strongest labor force attachment historically, implying that these individuals probably will be faster in obtaining a new job. In order to avoid this selection problem, the authors use a regression discontinuity approach where they estimate the relationship

$$d_i = \beta_0 + \beta_1 b_i + p(h_i) + \varepsilon_i,$$

where $p(h_i)$ is a continuous function of the length of the previous work history h_i (months of employment during the last 5 years), which captures that individuals with a strong labor attachment historically probably will be faster in finding a new job. The identifying variation comes from the fact that the maximum benefit duration is 30 weeks when h_i is larger than 36 months while the benefit duration is 20 weeks when h_i is below 36 months. This creates a discontinuity at 36 months, which is used for identification.

This is shown in Graph 1 where the mean unemployment duration in days is plotted against months of employment in the past (b_i is in this case just a dummy variable which is zero for a short benefit duration, i.e. when $h < 36$, and equal to one for a long benefit duration, i.e. when $h \geq 36$). The benefit duration is longer if the worker has had more than 36 months of previous employment, and as shown in the graph this increases unemployment duration by around 7 days on average for individuals close to the threshold (i.e. the curve jumps up from 153 to 160 days). The assumption behind identification is that the only discrete change taking place around $h = 36$ is the change in the benefit duration.

One may expect that a long benefit duration gives a better match between worker and firm because the worker on average uses longer time on finding a job. Hence, the worker may stay longer time in the next job, if he obtains a longer benefit duration while unemployed. Graph 2 does not provide evidence in favor of this hypothesis, since there is no discrete change at $h = 36$. (The evidence in the graph is based on a criteria for severance payment, which requires that the worker has worked in the same firm for at least 36 months but the basic principle is the same.)

Part 3: Redistribution policy

3A) The elasticity of taxable income (ETI) is defined as $\text{ETI} = \frac{dz/z}{d(1-m)/(1-m)}$, which is the percentage change in taxable income z with respect to a percentage change in the net-of-tax rate $1-m$. The elasticity measures the extent to which tax payers change their taxable income when the net-of-tax rate is changed. It may be noted that the ETI is often considered as a sufficient statistic for the evaluation of the efficiency effects of tax systems because it captures all relevant types of behavioral responses to tax changes (e.g. changes in hours-of-work, work effort, mobility, tax avoidance and tax evasion).

3B) No. Individuals with high ability and therefore high income will face a high marginal tax rate while individuals with low ability/income will face a low marginal tax rate. This would give a negative estimate of β_1 and it would not reflect the causal effect from the net-of-tax rate to income. The estimate is biased because marginal tax rates are not assigned randomly but determined by the size of income (reverse causality).

3C) The tax reform variation in (i) may be exploited to estimate β_1 with a difference-in-difference estimator. For example, by calculating

$$\beta_1 = \frac{\log\left(\frac{\overline{z}^{\text{High}}_{\text{Post}}}{\overline{z}^{\text{High}}_{\text{Pre}}}\right) - \log\left(\frac{\overline{z}^{\text{Low}}_{\text{Post}}}{\overline{z}^{\text{Low}}_{\text{Pre}}}\right)}{\log\left(\frac{1-m^{\text{High}}_{\text{Post}}}{1-m^{\text{High}}_{\text{Pre}}}\right) - \log\left(\frac{1-m^{\text{Low}}_{\text{Post}}}{1-m^{\text{Low}}_{\text{Pre}}}\right)},$$

where High (Low) denotes the group with high (low) income paying the high (low) tax rate, while Post (Pre) denotes after (before) the reform. In this case, only m_H is changed implying that $m_{\text{Post}}^{\text{Low}} = m_{\text{Pre}}^{\text{Low}}$. The above formula may then be written as

$$\beta_1 \approx \frac{\%\text{-change in } \overline{z}_{\text{High}} - \%\text{-change in } \overline{z}_{\text{Low}}}{\%\text{-change in } (1 - m_{\text{High}})},$$

where the numerator measures the percentage change in high incomes relative to low incomes. The identifying assumption is that %-change in $\overline{z}_{\text{High}}$ = %-change in $\overline{z}_{\text{Low}}$ without a reform, which is the so-called common trend assumption. With this assumption, the increase in $\overline{z}_{\text{High}}$ over and above $\overline{z}_{\text{Low}}$ is attributed to the the change in m_{High} generated by the reform.

The randomized experiment in (*ii*) would make it possible to assign the marginal tax rates m_L and m_H randomly to high-income individuals. Let T denote the treatment group with a high tax rate m_H and let C denote the control group with low tax rate m_L. It would then be possible to compute an unbiased estimate of β_1 by computing (this is done by comparing levels but it would also be possible to compute a difference-in-difference estimator)

$$\beta_1 = \frac{\log\left(\overline{z}^{\text{T}}\right) - \log\left(\overline{z}^{\text{C}}\right)}{\log(1 - m_H) - \log(1 - m_L)},$$

which measures how many percent z^{T} is lower than z^{C} relative to how many percent $1 - m_H$ is lower than $1 - m_L$. Because of the randomization, the expected income of the treatment group would be identical to the income of the control group if marginal tax rates would be identical. Hence, possibility (*ii*) provides a better estimate of β_1 because it does not rely on the common trend assumption for identification of β_1. (Carrying out such an experiment in practise may very well conflict with horizontal equality concerns and ethical concerns.)