Answers

Exam in Public Finance - Spring 2016

3-hour closed book exam

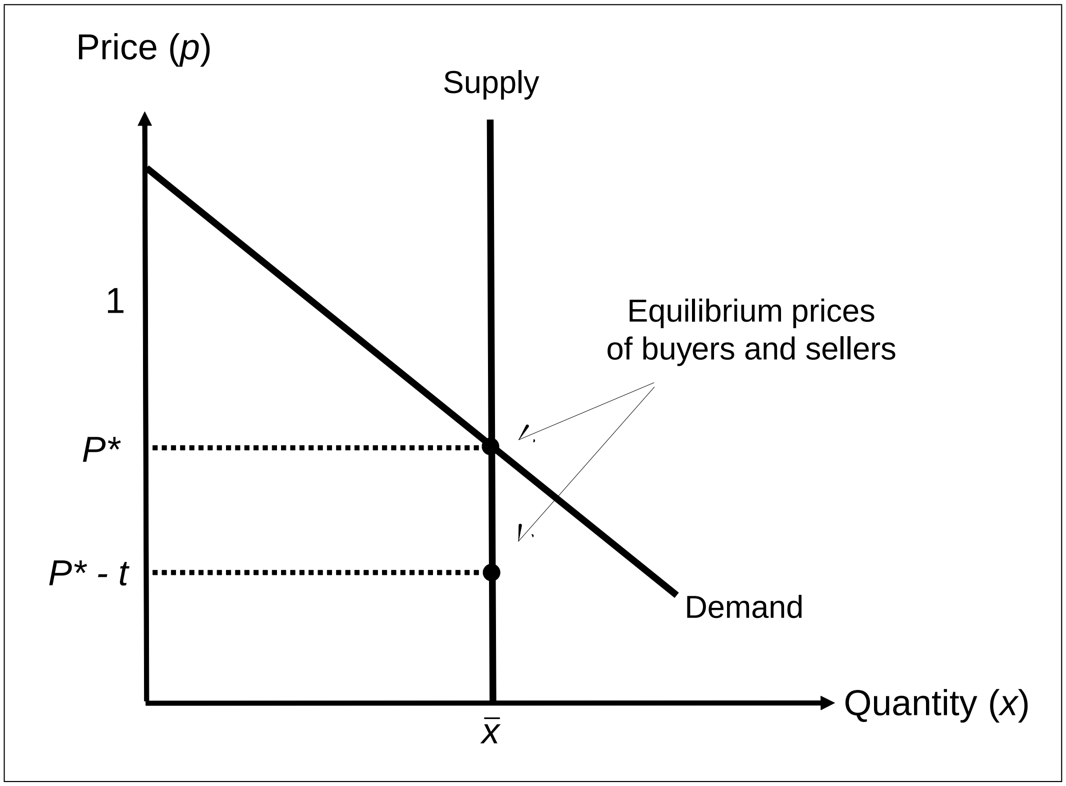
Claus Thustrup Kreiner

Part 1

(**1A**) No, the elasticity of taxable income and the elasticity of labor supply are not the same concepts. The elasticity of taxable income measures the percentage change in taxable income (the income reported to the tax authorities on the tax return) with respect to a percentage change in the net-of-tax rate. Mathematically, the elasticity may be defined as $\frac{dz}{d(1-m)}\frac{1-m}{z}$, where z is taxable income and m is the (marginal) tax rate. The elasticity of labor supply measures the percentage change in the input of labor supplied, for example measured in number of hours, with respect to a percentage change in the after-tax hourly wage rate. Mathematically, the elasticity may be defined as $\frac{dl}{d[(1-m)w]}\frac{(1-m)w}{l}$, where l is hours-of-work, m is the (marginal) tax rate, and w is the hourly wage rate before tax. The elasticity of taxable income includes the effect of taxes on labor supply through l, but includes also effects on labor supply behavior along other dimensions – work effort, job type and job location – that may influence productivity without affecting hours, and effects on reported income due to tax avoidance and tax evasion. All of these behavioral responses give rise to tax distortions, which may be underestimated if only looking at estimates of the labor supply elasticity. Since the elasticity of taxable income captures all behavioral responses to taxation, it may be argued that this elasticity is a sufficient statistic to compute the deadweight loss of taxation.

It may also be mentioned that there are different types of labor supply responses/elasticities (compensated vs uncompensated, intensive vs extensive margin responses).

(**1B**) This is correct. The economic incidence of the tax measures how the economic burden of the tax is shared among buyers and sellers in the market. This is different from the formal/statutory/legal tax incidence stating who has the legal obligation to pay the tax. Figure 1 illustrates the incidence of a tax in a supply-demand diagram when supply is fixed, i.e. the supply curve is vertical at $\bar{x}$. This implies that sellers are willing to sell $\bar{x}$ at any (positive) price. Buyers will demand exactly $\bar{x}$ if the before-tax price is p^*, which therefore becomes the equilibrium price before the tax. At this price the buyers pay p^*, while the sellers receive $p^* - t$, where t is the tax. Without the tax, the equilibrium price also becomes p^*. This implies that the sellers bear the full burden of the tax as described in the statement.



It may be noted that the incidence of a tax may be written approximately as

$$I_S \approx \frac{\varepsilon_B}{\varepsilon_S + \varepsilon_B},\ I_B \approx \frac{\varepsilon_S}{\varepsilon_S + \varepsilon_B},$$

where I_S and I_B are the incidence of the seller and the buyer, respectively, and where ε_B is the price-elasticity of the demand of the buyers, while ε_S is the price-elasticity of the supply of the sellers. A fixed supply corresponds to $\varepsilon_S = 0$, implying that $I_S = 1$ and $I_B = 0$, also showing that the full incidence is on the sellers.

(**1C**) No, Card et al. (2007) do not use a difference-in-difference method but a regression discontinuity method to estimate the effect of unemployment insurance on the duration of unemployment. The difference-in-difference method exploits that one group is exposed to a certain treatment, while others are not. This could for example be a reform that changes the unemployment insurance scheme for one group (treatment), but not for another group (control). The groups may be different and for example have different unemployment levels. The difference-in-difference method relies on an assumption of a "common trend", in this case a similar development over time in unemployment for the two group, implying that the effect of the treatment may be measured as the change in the treatment group from before the treatment to after the treatment compared to the change in the control group over the same time period.

Card et al. (2007) exploits that the length of the UI benefit period in Austria depends on the employment history of the individual with a jump in the length of the period (from 20 weeks to 30 weeks) when a person has been employed for more than a certain threshold number of months (36 months during the past 5 years). By comparing individuals with past employment just below and just above the threshold, assuming that this difference is due to randomness, it is possible to obtain a casual estimate of the effect of extending the UI benefit period. Card et al. (2007) shows, for example, that individuals just above the threshold are without a job 7 days longer than those just below the threshold.

It may be noted that a threat to identification is that the variation around the threshold is not fully random. For example, in the analysis of Card et al. firms may fire the least productive workers just before the 36 month threshold in order to avoid paying severance payment. However, the evidence in Card et al. does not indicate that this is a problem.

Part 2

(**2A**) Inequality measures the variation across individuals in economics outcomes, for example variation in income or wealth at a given point in time or differences in lifetime income across individuals. Intergenerational mobility measures how economic outcomes are related across generations. A high degree of intergenerational persistence (low degree of mobility) implies that a high degree of inequality is transmitted to the next generation, which implies that the concepts are related to each other. However, they are not the same. To see the difference between the two concepts, consider as an example two countries that have the same variation in income over time. One country has no intergenerational mobility, implying that a child get the same position in the distribution as the parents, while the other country has perfect intergenerational mobility, implying that the position of a child in the distribution is completely random. Thus, the two countries have the same distribution, but very different intergenerational mobility, with parents being crucial for outcomes of children in one country, but not in the other country.

(**2B**) Boserup et al. divide individuals into treatment and control groups depending on whether a parent dies at a given point in time, denoted by 0 in the graphs, or do not die. The aim is to measure the impact on the wealth distribution of the next generation of receiving bequest, which is not directly observable in the data. The left panel displays the percentage difference in the variance of the wealth distribution between the treatment group and the control group over time. The difference in the seven years before death of a parent is close to zero, and then increases by around 1/3 after death of the parent. Thus, bequests increase the variance of the wealth distribution of the next generation.

The right panel shows the share of wealth owned by the one percent most wealthy in the treatment group and the control group, respectively. In the seven years prior to death of the parent, the top 1% share of the control group lies somewhat higher than the treatment group, and the wealth share varies over time with the business cycle, but the wealth shares of the two groups co-vary reasonably well with a fixed difference over time. After death of the parent, the gap between the two curves increases (actually, it occurs somewhat already the year before, which may be due to wealth transfers in order to avoid inheritance taxation). This implies that the wealth share of the treatment group falls compared to the control group (by 6 percentage points), implying that bequests reduce the share of wealth owned by the top 1% wealthiest (and equivalently that the wealth share increases of those not in the top 1% group).

(**2C**) As described above the conclusion from the left panel is that bequests increase wealth inequality, while the conclusion from the right panel is that bequests reduce wealth inequality. The reason for this difference in conclusions is the use of different inequality concepts. The variance of the distribution is a measure of absolute inequality, focusing on absolute differences across people. Inheritances stretch the distribution because the rich inherits larger amounts, which increases absolute inequality. The top 1% wealth share measures one groups wealth out of total wealth, which is a measure of relative inequality. If the rich inherit a smaller percentage of their initial wealth than the less rich (which is the case here) then relative wealth inequality decreases. To conclude, whether bequests increase or decrease wealth inequality depends on how we measure wealth inequality.

Part 3

(**3A**) Tax evasion may be defined as a legal economic activity, not declared to the tax authorities although it is taxable. Thus, the reduction in tax liability is illegal. Examples are underreporting of income or overstating deductions on the tax return. Tax avoidance is also a reduction in tax liability, but it is legal and reflects "tax planning" (not intended by the policy makers). It may be mentioned that shadow/hidden economy activities include tax evasion, but also illegal economic activities where payments are made and not reported to the tax authorities. It may also be mentioned that the unmeasured economy covers the shadow economy plus do-it-yourself activities.

(**3B**) Eq. (1) states that in Model 1 the taxpayers maximize expected utility, which in this context is equivalent to expected income (i.e., an implicit assumption is that the agent is risk neutral corresponding to utility being linear in income). The first term in the eq. is the net-income if caught evading multiplied by the probability of being caught, while the second term is the income if not caught evading multiplied by the probability of not being caught. Eq. (2) states that in Model 2 the taxpayers maximize expected utility, which is the same as in Model

1 with the exception that the taxpayers face a loss of utility from evading due to moral, shame etc. Eq. (3) states that in Model 3 the taxpayers maximize expected utility, which is the same as in Model 1 with the exception that the probability of being caught is a function of the amount evaded. Eq. (4) defines the net-income if not caught evading. It is equal to the income after tax when reporting truthfully (the first term) and the taxes saved by evading the amount E (the second term). Eq. (5) defines the net-income if caught evading. It is equal to the income after tax when reporting truthfully (the first term) and the fine the taxpayer has to pay from the detected evasion, which equals the share F of the evaded tax payment.

(3C) The optimal behavior of a taxpayer in Model 1 is found by inserting eqs (4) and (5) in eq. (1) and differentiating with respect to E. After inserting eqs (4) and (5) in eq. (1), we have

$$U^e = (1-p)\left[(1-t)\,Y + tE\right] + p\left[(1-t)\,Y - FtE\right].$$

Differentiation with respect to E gives

$$\frac{dU^e}{dE} = (1-p)\,t - pFt \tag{1}$$

If this is positive then the taxpayer will evade taxes (because of the linear structure the taxpayer will evade on all income), and if it is negative then the taxpayer will not evade. The first term in the expression is the marginal benefit of evading one additional euro equal to the increase in net-income if not caught, while the second term is the marginal cost reflecting the increase in the fine paid if caught. A higher probability p of being caught will increase the marginal costs and reduce the marginal benefits and thereby reduce the incentive dU^e/dE. A higher fine F will increase the marginal costs of being caught and thereby reduce the incentive dU^e/dE. A higher tax rate t does not influence whether the incentive dU^e/dE is positive or negative, and does therefore not influence the decision to evade or not (if $(1-p) - pF$ is positive then a higher t will make dU^e/dE more positive and vice versa).

(3D) In Model 2, the incentive becomes

$$\frac{dU^e}{dE} = (1-p)\,t - pFt - \chi t.$$

This is the same as (1), with the only exception of the last term, which reflects a utility loss per euro of evasion due to moral, shame etc.

When χ varies across individuals then dU^e/dE becomes positive (negative) for individuals with $\chi < \hat{\chi}$ $(\chi \geq \hat{\chi})$ where $\hat{\chi}$ equals

$$\hat{\chi} = 1 - p\,(1+F)\,,$$

where $\hat{\chi}$ is the value of χ where $dU^e/dE = 0$ If p and F are small and χ is close to zero for all taxpayers then everybody will evade taxes. However, if many taxpayers have a "high" morale with χ larger than $\hat{\chi}$ then only few taxpayers will evade taxes. Thus, although the economic net-benefit of evasion is positive for all individuals then it may be the case that only few taxpayers evade because many people have a good tax morale (in a popular phrase "taxpayers are able but unwilling to cheat").

In Model 3, the expected utility of a taxpayer becomes

$$U^e = [1 - p(E)] [(1-t) Y + tE] + p(E) [(1-t) Y - FtE].$$

Differentiation with respect to E gives

$$[1 - p(E)] t - p(E) Ft - p'(E) (tE + FtE) = 0 \Leftrightarrow$$

$$1 - p(E) (1 + F) - p'(E) E (1 + F) = 0 \Leftrightarrow$$

$$p(E^*) (1 + \varepsilon) (1 + F) = 1,$$

where $\varepsilon = \frac{p'(\cdot)E}{p(\cdot)}$ denotes the percentage change in the probability of detection from a percentage change in evasion. The higher the evaded amount, the larger the probability of being detected. This equation determines the optimal amount evaded. The optimum is illustrated in Figure 2, showing that the probability of detection is small for low evasion levels–reflecting that the taxpayer underreports income that the tax agency has difficulty in uncovering (e.g. black-market work)–but that the detection probability is very high at higher evasion levels where the taxpayer reports less income than the tax agency already knows about from third-party information reports from employers and others. In the optimum, we may observe low evasion E^* and low detection probability $p(E)$ with small fines F as illustrated in Figure 2. However, additional tax evasion (moving to the right in the diagram) would make the detection probability very high. Thus, the low evasion rate basically reflects that it is impossible for the taxpayers to evade large amounts, because the tax agency obtains extensive third-party information about income (in a popular phrase "Taxpayers are willing but unable to cheat").

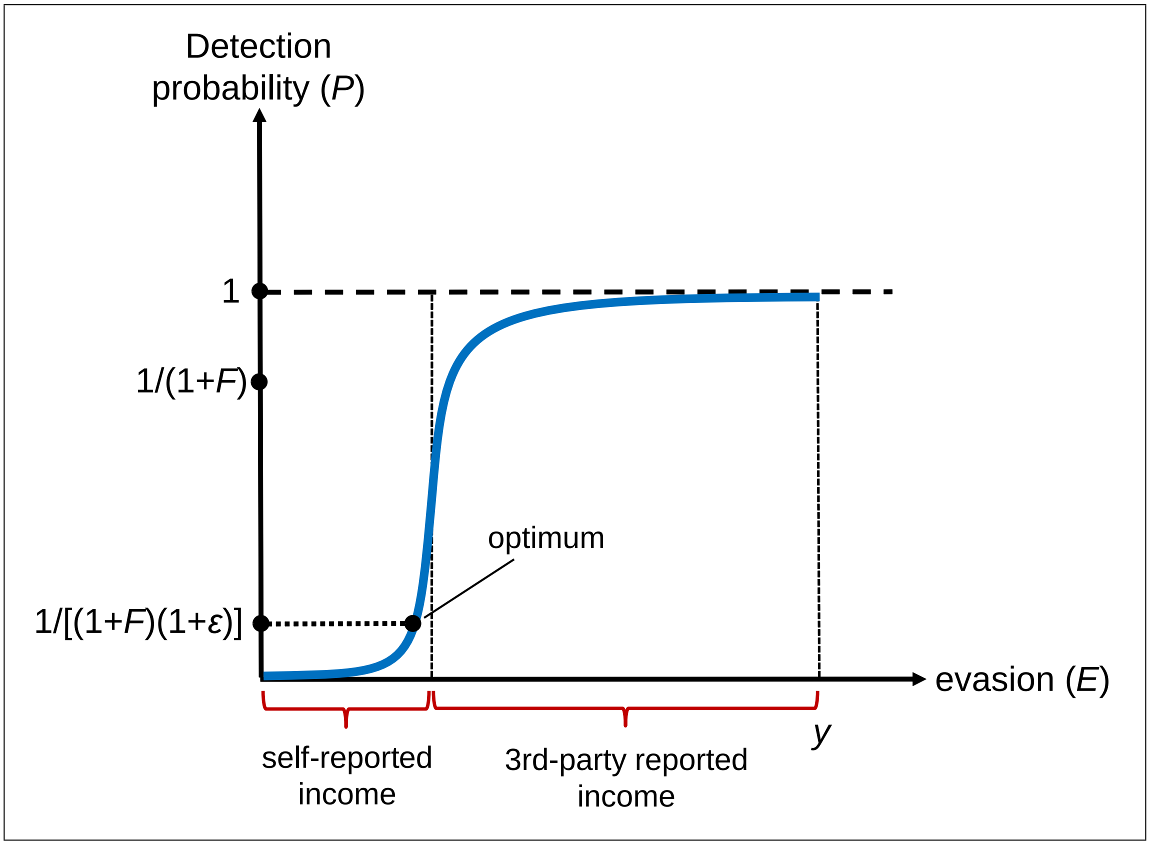


Figure 1: Endogenous detection probability

(**3E)** The empirical results in Kleven et al. are based on a large audit experiment conducted together with the Danish tax agency (SKAT), where a (stratified) random sample of taxpayers are audited thoroughly and homogeneously. The results most relevant for the question are: The overall tax gap is reasonably low (2-3%). On average, the tax agency knows a very high share of net-income of a taxpayer (95%) from third-party information. Underreporting of self-reported income is very high (40%), while underreporting of third-party reported income is very low (0.3%). These results indicate that third-party information is very effective in reducing tax evasion, and making it very difficult to evade large amounts in accordance with Model 3.

It may also be mentioned that an example of income with a high degree of third-party reporting is earnings of employees, while an example of income with no third-party reporting is self-employment income.