Answers
Final exam in Public Finance - Fall 2019
3-hour closed book exam

By
Claus Thustrup Kreiner and Jakob Egholt Søgaard

January 2020

Part 1: Intergenerational mobility

The article "Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States" by Chetty et al. (2014) provides empirical evidence on the degree of intergenerational mobility.

***(1A)** Describe the concept of intergenerational mobility and describe why it may be relevant for a society to care about the degree of intergenerational mobility.*

Intergenerational mobility measures to what extent outcomes such as income 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. This is not the same as inequality, which measures the variation in outcomes across individuals. To see the difference between the two concepts, consider as an example of two countries that have the same income inequality throughout an extended period of time. One country has no intergenerational mobility, implying that a child gets the same position in the income distribution as the parents, while the other country has perfect intergenerational mobility, implying that the position of a child in the distribution is completely unrelated to that of its parents.

A society can care about intergenerational mobility for different reasons. First, it provides information on the sources of inequality and the extent of equal opportunities, which may guide the design of school systems etc. Second, it may affect the perceived fairness of inequality. E.g people may be willing to tolerate higher levels of inequality if they perceive that everybody has equal opportunities in reaching the top of the income distribution (The American Dream). Third, a society with a low intergenerational mobility may have an inefficient allocation of talent if high-ability individuals from a poor background are prevented

from exploiting their talent.

It is common in studies of intergenerational mobility to estimate the equation

$$logy_g = \beta_0 + \beta_1 logy_{g-1} + \varepsilon, \tag{1}$$

where y_g denotes the income of individuals in generation g, y_{g-1} denotes the income of the parents (generation g-1) and ε is an error term.

***(1B)** What does the coefficient β_1 measure? Describe the empirical results shown in Figure 1 on the next page and why this evidence suggests that it may be problematic to use the regression (1) in studies of intergenerational mobility.*

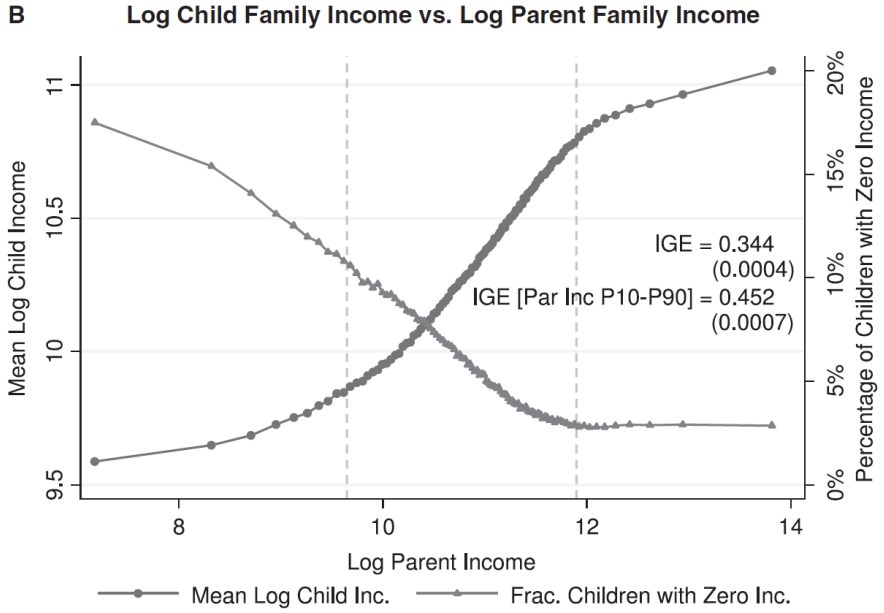
In equation (1), β_1 measures the effect of a one log-point (one percent) change in parental income on the log income of the child. As this is a log-log relationship, β_1 is an elasticity and typically labelled the Intergenerational Elasticity of Income (IGE).

Figure 1 shows the relationship between log parental income and log child income (left axis) and the share of children with zero income (right axis). It is constructed as a bin scatterplot, which depicts the average of the two outcome measures within bins of parental income. The figure illustrates two problems with estimating the IGE using equation (1). First, a significant share of children have zero income, both overall but in particular in the bottom of the parental income distribution. As you cannot take log of zero, these observations will not be included in the estimation, which will bias the IGE downwards (average income for children of poor parents will be higher when excluding children with zero income). Second, even looking at the children with positive income, we see that the relationship with parental income is non-linear, while equation (1) assumes a linear relationship. This implies that the IGE is sensitive to the range of parental income included in the estimation. The figure illustrates this point by providing two IGE estimations. The one for the whole sample is 0.344 (implying that an 1 percent increase in parental income increases child income by 0.344 percent on average), while the IGE for the sample with parental income between the 10th and 90th percentile is 0.453.

***(1C)** Describe the empirical results in Figure 2 on the next page and how it is informative about the degree of intergenerational mobility.*

Figure 2 illustrates the relationship between parental and child income using income ranks instead of log income, where income ranks are computed by lining individuals up according to their income and assigning them a number between 0 (lowest income) to 100 (highest

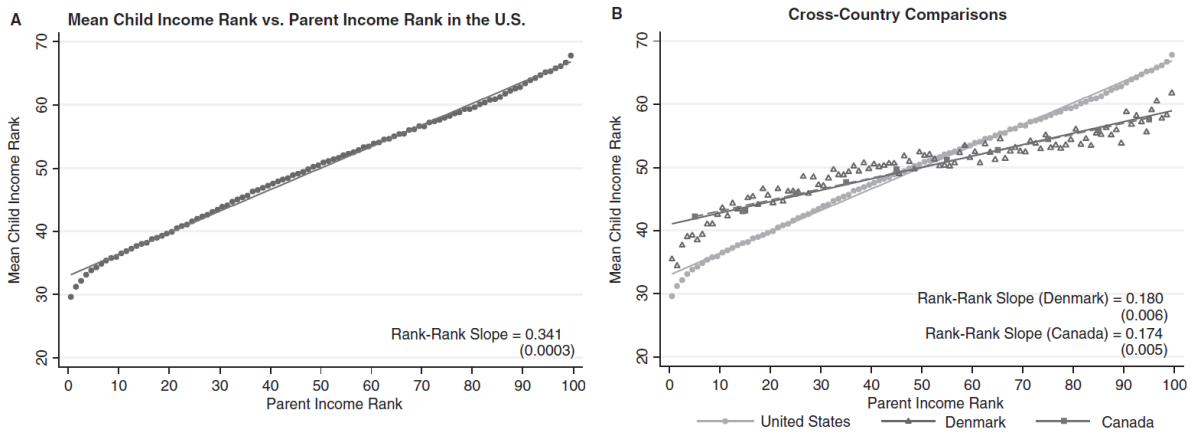
Figure 1: Relationship between child income and parental income in the US



Note: Reprinted from Chetty et al. (2014). "Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." Quarterly Journal of Economics 129(4).

income). Because we are using ranks, even children (and parents) with zero income can be included in the figure. Panel A shows the relationship for the US and we see an almost perfectly linear relationship between parent and child income with a slope of 0.341 (implying that an increase in parental income of one rank point increases child income by 0.344 ranks on average). Panel B compares intergenerational mobility in the US to that in Denmark and Canada and we see that the slope in these two countries are close to half the size of the slope in the US. From this figure, we would therefore conclude that intergenerational mobility is higher in Denmark and Canada than in the US.

Figure 2: Relationship between child income and parental income in the US, Denmark and Canada



Note: Reprinted from Chetty et al. (2014). "Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." Quarterly Journal of Economics 129(4).

Part 2: Breakdown of the second welfare theorem

Consider a society with two individuals: A high ability type (H) with a high hourly wage rate w_H *in the labor market and a low ability type (L) with a low hourly wage rate* w_L *in the labor market. They have the same utility function given by*

$$U(c,h) = u(c) - v(h), \tag{2}$$

where c *is consumption,* h *is hours of work,* $u(\cdot)$ *is a strictly concave function with standard properties, while* $v(\cdot)$ *is a strictly convex function with standard properties. The budget constraint for type* $i = H, L$ *is given by*

$$c_i = w_i h_i - T_i, \tag{3}$$

where T_i *is an individual lump sum tax. The optimal number of hours of the two individuals is characterized by*

$$w_i u'(c_i) = v'(h_i). \tag{4}$$

The social planner has a utilitarian objective function:

$$W = U(c_H, h_H) + U(c_L, h_L). \tag{5}$$

The social planner maximizes equation (5) with respect to T_H and T_L subject to the decision rules of the two individuals (4) and subject to

$$T_L + T_H = 0. \tag{6}$$

(2A) *Provide an economic interpretation of equation (6).*

Equation (6) represents the government's budget constraint and states that the total (net) transfers to the two individuals has to be zero. This implies that any transfer given to the low ability type has to collected as (lump sum) taxes from the high ability type.

(2B) *Show how to derive equation (4) and provide an economic interpretation of this equation.*

Equation (4) is obtained by inserting equation (3) into (2) and differentiate wrt. h_i

$$U(c_i, h_i) = u\left(w_i h_i - T_i\right) - v\left(h_i\right) \Rightarrow \frac{\partial U}{\partial h_i} = w_i u'\left(c_i\right) - v'\left(h_i\right) = 0.$$

This equation has a straight forward interpretation. Individuals are willing to supply extra hours worked until the marginal utility gain, given by the extra income/consumption per hour (w_i) times the marginal utility of consumption $u'(c_i)$, is equal to the extra disutility of working an additional hour $v'\left(h_i\right)$.

(2C) *Show that the allocation preferred by the social planner, as described above, is characterized by*

$$c_H = c_L \text{ and } h_H > h_L.$$

The social planner seeks to maximize the sum of utilities in society (equation (5)) subject to its budget constraint and individuals' behavioral responses. Plugging equations (2), (3) and (6) into equation (5), we get

$$W = u\left(w_H h_H - T_H\right) - v\left(h_H\right) + u\left(w_L h_L + T_H\right) - v\left(h_L\right),$$

and differentiating wrt. T_H yields

$$\frac{\partial W}{\partial T_H} = -u'(c_H) + \underbrace{u(c_H)\frac{\partial h_H}{\partial T_H} - v'(h_H)\frac{\partial h_H}{\partial T_H}}_{=0}$$
$$+ u'(c_L) + \underbrace{u(c_L)\frac{\partial h_H}{\partial T_H} - v'(h_H)\frac{\partial h_H}{\partial T_H}}_{=0} = 0.$$

In this equation, the two underlined components are the same as the individuals' first order conditions and hence zero, when individuals have optimized (the envelope theorem). Hence, the social planner will set taxes such that

$$u'(c_H) = u'(c_L) \Leftrightarrow c_H = c_L.$$

Taking this result to the first order condition in equation (4), we see that as $u'(c_H) = u'(c_L)$ and $w_H > w_L$ so must $v'(h_H) > v'(h_L) \Leftrightarrow h_H > h_L$ because the marginal disutility of work is increasing in the number of hours.

(2D) *Argue why it is possible or not possible for the social planner to implement this solution depending on whether the social planner can observe (a) the hourly wage rates w_i or (b) only the income levels $z_i = w_i h_i$.*

When the social planner can perfectly observe w_i, she can assign taxes based on these exactly as in the solution to 2C above and as this solution take into account the individuals' first order conditions, the solution is incentive compatible and hence possible to implement. If the social planner cannot observe w_i, she can only try to assign taxes based on observed income $z_i = w_i h_i$. However, in this case the individuals can affect the taxes assigned to them by choosing a different h_i, and if the social planner tries to implement the solution from 2C, the high ability individuals will pretend to be a low ability type by choosing h_H so that $z_H = z_L$. To see this, note that because the high ability type in 2C has the same level of consumption as the low type, but works more, the high ability type will have lower utility if choosing the working hours intended by the social planner. Hence, the high ability type will instead mimic the low-ability type by working less so that $z_H = z_L$ (an example of adverse selection).

Part 3: Estimation of the elasticity of taxable income

In November 2007, the business man and politician Lars Kolind suggested a major reduction in the top tax rate for people living on Fyn (the large island in the middle of Denmark) for a five-year period as an experiment to study behavioral effects of income taxes. The experiment

was never carried out in practice, but Table 1 below shows some hypothetical results from the suggested experiment/reform. The reform reduced the effective top tax rate on Fyn from 70 percent to 61 percent corresponding to a change in the net-of-tax rate, $1-t$, by 30 percent, while tax rates for the rest of Denmark and non-top tax payers on Fyn remained unchanged.

Table 1 shows the average, logarithmic wage income for different income groups living on Fyn and living in the rest of Denmark before the reform (Pre) and after the reform (Post). The goal is to get an estimate of the income response to a lower top tax rate, which may be used to compute the elasticity of taxable income. The bottom panel in Table 1 provides different estimates of income response (the table does not provide standard errors, but we assume that E1-E9 are precisely estimated statistically).

Table 1: Wage income across groups and time

	2008	2010
Location/year	Pre	Post
Fyn, average log wage income		
A. Top tax payers	12,912	12,971
B. Other tax payers	12,612	12,626
Rest of Denmark, average log wage income		
C. Top tax payers	13,412	13,446
D. Other tax payers	13,122	13,141
Estimates of earnings responses		
E1 = A^{pre}-B^{pre} :		0,300
E2 = C^{pre}-D^{pre} :		0,290
E3 = A^{post}-A^{pre}:		0,059
E4 = B^{post}-B^{pre}:		0,014
E5 = (A^{post}-A^{pre})-(B^{post}-B^{pre}):		0,045
E6 = (C^{post}-C^{pre})-(D^{post}-D^{pre}):		0,015
E7 = (A^{post}-A^{pre})-(C^{post}-C^{pre}):		0,025
E8 = (B^{post}-B^{pre})-(D^{post}-D^{pre}):		-0,005
E9 = E5 - E6 :		0,030

(3A) *Provide a definition of the elasticity of taxable income.*

The elasticity of taxable income is defined as $\varepsilon = \frac{dz/z}{d(1-t)/(1-t)} = \frac{d\log z}{d\log(1-t)}$ and measures the percentage change in taxable income (z) following a 1 percentage point increase in the marginal net-of-tax rate. Using the elasticity of taxable income typically gives a better estimate of the deadweight loss from taxation because it captures a wider range of behavioral

responses than just hours worked. A higher tax might e.g. reduce the willingness to accept a higher paying job further away or give a higher incentive to transform earnings into fringe benefits. Behavioral responses across all of these dimensions cause distortions that should be included in a calculation of the marginal deadweight loss.

(3B) *Advisor #1 believes the best estimate of the income response to a lower top tax rate is E1. Do you agree with this advisor? Explain why or why not?*

No. E1 compares the level of wage income of top tax payers to the level of other tax payers on Fyn before the reform. Hence, this estimate does not measure the effect of the tax change caused by the reform and is very far from being a causal effect of a lower top tax (in fact, the implied elasticity would be negative because top tax payers have a lower net-of-tax rate than other tax payers). Instead, the estimate primarily reflects the differences in productivity/potential income between the two groups.

(3C) *Advisor #2 believes the best estimate of the income response is E3. However, advisor #1 argues that estimate E4 is showing that the estimate E3 is likely to be upward biased. Which of the two advisors do you agree with? Explain why.*

Advisor #1 has a good point. E3 measures the change in wage income for top tax payers on Fyn from before to after the reform. The estimate includes the reform effect, but may be biased from other factors that change over time such as productivity growth and/or business cycles. One way to judge the effect of such other factors is to look at a group in the same macro economic environment, but who are not affected by the reform. This is exactly what E4 is doing, as E4 measures the change in wage income for other tax payers on Fyn. As E4 is positive, we would expect E3 to be upward biased compared to the true reform effect.

(3D) *The two differences-in-differences estimates E5 and E7 give two different results. Discuss the different assumptions underlying these two estimates and why the results may be different. Do you think these two estimates are equally good or do you prefer one estimate over the other estimate? Explain why.*

E5 is computed as E3 - E4. It hence builds on the discussion in 3C and uses the change in wage income for other tax payers on Fyn, who are not affected by the reform, to control for other factors such as productivity growth that change income over time. The identifying assumption in this estimate is the common trend assumption. I.e. absent the reform, top tax payers and other tax payers on Fyn should have experienced the same change in average (log) wage income. The common trend assumption cannot be proven, but if we had data available

for more pre-reform periods, we could validate it by checking if the trends are indeed the same in these periods (Placebo tests). One may note that estimate E6 in Table 1 offers another type of placebo test. It shows that the income of top tax payers increased by 1.5 percent more than the income of other tax payers in the rest of Denmark without any change in tax incentives (estimate equal to 0.015 in Table 1). This indicates that the common trend assumption is likely to be violated and that the estimate E5 may be upward biases, i.e. part of this estimate may be due to differential trends for the two groups.

E7 is computed as E3 minus the change in wage income for unaffected top tax payers in the rest of Denmark. Hence it use another control to control for the any underlying time trends. The identifying assumption is still the common trend assumption, but here wrt. top tax payers on Fyn relative to top tax payers in the rest of Denmark. Hence, E7 uses the same income class, but from another geographic area as the control group, while E5 uses another income class, but in the same geographic area, as the control group. Thus, which estimate is best depends on whether we expect same underlying income trends across income classes or across regions. As discusses above, estimate E6 could be seen as a placebo test of the parallel trend assumption underlying E5. Similarly, E8 may be seen as a placebo test underlying estimate E7. Estimate E8 shows that the income of other tax payers (not experiencing any change in tax incentives) grows nearly to the same extent in Fyn and in the Rest of Denmark (difference equal to -0.005 in Table 1). This indicates that the geographical variation underlying estimate E8 may provide a more reliable estimate than the variation across social classes underlying estimate E6.

(3E) *What would be your preferred estimate in Table 1? Explain why. Describe the potential threats to identification of this estimate.*

The best estimate of the reform effect is probably the triple difference (DiDiD) given by E9 = E5 - E6. This estimate uses both the fact that we have treated and untreated individuals within the treated area (Fyn), which makes is possible to control for other local growth and business cycle effects and the fact that we have treated and untreated areas, which makes it possible to control for differences in trends between treated (top tax payers) and untreated individuals (other tax payers). The identifying assumption is in this case common trend differentials. I.e. absent the reform the difference in wage income growth for top tax payers and other tax payers on Fyn should be the same as for the rest of the Denmark. Potential treats to this identification would be any factor, which specifically would affect the trend difference between top tax payers and other tax payers on Fyn or the rest of Denmark.

The estimate E9 measures the income response. It may be mentioned that the corresponding elasticity of taxable income equals $0.03/0.3 = 0.1$.

It may also be mentioned that a threat to the identification strategies is that some high-

income earners have moved to Fyn because of the reform without increasing their income. This behavioral effect of the reform (movement from one part of the country to another part) would not arise if the reform was carried out throughout the country. In this way, the result for Fyn would not be informative for tax reforms carried out at the national level.