

Dur, Robert A. J.; Teulings, Coenraad N.

Working Paper

Education and Efficient Redistribution

CESifo Working Paper, No. 592

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Dur, Robert A. J.; Teulings, Coenraad N. (2001) : Education and Efficient Redistribution, CESifo Working Paper, No. 592, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:

<http://hdl.handle.net/10419/75712>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



Working Papers

EDUCATION AND EFFICIENT REDISTRIBUTION

Robert A. J. Dur
Coen N. Teulings

CESifo Working Paper No. 592

October 2001

CESifo
Center for Economic Studies & Ifo Institute for Economic Research
Poschingerstr. 5, 81679 Munich, Germany
Phone: +49 (89) 9224-1410 - Fax: +49 (89) 9224-1409
e-mail: office@CESifo.de
ISSN 1617-9595



An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the CESifo website: www.CESifo.de

EDUCATION AND EFFICIENT REDISTRIBUTION

Abstract

Should education be subsidized for the purpose of redistribution? The usual argument against subsidies to education above the primary level is that the rich take up most education, so a subsidy would increase inequality. We show that there is a counteracting effect: an increase in the stock of human capital reduces the return to human capital and, therefore, pre-tax income inequality decreases. We consider a Walrasian world with perfect capital and insurance markets. Hence, in the absence of a strive for redistribution, the market generates the efficient level of investment in human capital. When there is a demand for redistribution, the general equilibrium effects on relative wages might make a subsidy to education an ingredient of a second-best optimal redistribution policy. Stimulating human capital formation results in a compression of the wage distribution, and hence reduces the need for distortionary redistributive taxation. We also study the political viability of education subsidies.

JEL Classification: H21, H52, J24, J31.

*Robert A. J. Dur
Tinbergen Institute
Erasmus University Rotterdam
P.O.Box 1738
3000 DR Rotterdam
The Netherlands*

*Coen N. Teulings
Tinbergen Institute
Erasmus University Rotterdam
P.O.Box 1738
3000 DR Rotterdam
The Netherlands
teulings@few.eur.nl*

1 Introduction

Should education be subsidized for the purpose of redistribution? Economists have usually argued against. In the absence of capital market imperfections and externalities, the market delivers an efficient level of investment in human capital. Subsidies to education would only create distortions. Moreover, the large literature on the ability bias in the return to education, see for example Angrist and Krueger (1991), shows that education and innate ability are complementary. Subsidies to education favor therefore predominantly the high ability types, leading to a widening instead of a compression of the income distribution. If anything, education should therefore be taxed.

This paper considers a new argument in this discussion. We argue that general equilibrium effects of education subsidies on pre-tax wage inequality may make education subsidies an optimal redistributive instrument. An increase in the mean level of human capital in the economy reduces the return to human capital, by a simple substitution effect. The supply of high-skilled workers goes up, reducing their relative wage, while the supply of low-skilled goes down, increasing their relative wages. Hence, the return to human capital and thus pre-tax wage inequality goes down. When this indirect, general equilibrium effect of education subsidies on relative wages is sufficiently large, it might offset the direct income effect due to the complementarity of education and ability. We derive the precise condition under which education subsidies contribute to redistribution. Given the fact that there is usually some demand for redistribution in a democracy, subsidies to education (next to a progressive income tax) might be the most efficient way to implement this redistribution. Optimal policy entails a trade-off between the distortionary effect of progressive taxation and the distortions arising from education subsidies. The rationale for education subsidies in our model rests on an externality in individual schooling decisions. Individuals do not take into account the effect of their schooling on the pre-tax distribution of income, and thus on tax distortions.

In our model, worker's skill level is determined by her innate ability and the years of education she attains. Workers decide on how long to go to school. There are perfect capital markets. Hence, individuals invest in human capital up to the point where the marginal cost equal the market rate of return to human capital. A subsidy to the education system therefore yields overinvestment in human capital. However, as investment in human capital reduces wage dispersion, subsidies to the education system reduce the need to impose high marginal tax rates for the purpose of redistribution. A constrained Pareto efficient redistribution policy sets equal the marginal cost of distortions in the acquisition

of human capital and in the choice of effort. We face the remarkable situation that the role of income and substitution effects in redistribution is reversed. Usually, redistribution is brought about by the income effects of a policy (e.g. progressive income taxation), while the substitution effects reduce their effectiveness. For education subsidies, it is the other way around. Substitution effects contribute to redistribution, while income effects work in the opposite direction.

From a theoretical point of view, our analysis stands in the tradition of Mirrlees' (1971) Noble prize winning paper on optimal income taxation. Similar to Mirrlees, we assume that the government can observe neither effort, nor ability, nor the skill level that is obtained by taking up education. The only variables that government can observe are the gross income and the years of education attained. Mirrlees then analyzes the trade off between the distortion caused by increasing marginal tax rates vis-a-vis the extra redistribution that can be achieved. Our set up differs in three important aspects from Mirrlees' analysis.

First, Mirrlees' analysis necessarily requires a welfare function. The gain in terms of redistribution can only be evaluated by means of a welfare function that enables intersubjective utility comparison. This requirement does not apply to our analysis. Where Mirrlees analyzes only a single policy instrument, income taxation, this paper adds a subsidy to education as a second instrument. When we take the distribution of utility as given, we can analyze the optimal mix of instruments for the implementation of this distribution. Where Mirrlees considers only a single instrument, this question would lead to a trivial answer in his case. We take it to be an advantage that the concept of a welfare function can be discarded from our analysis. There is no legitimation for this concept in positive theory. Hence, Mirrlees' (1971) analysis is a normative exercise. Following Becker's (1983) efficient redistribution hypothesis, our analysis contributes to the understanding of observed institutions. Insofar as the political system has an incentive to consume Pareto improving policy adjustments, observed institutions should be constrained Pareto efficient in equilibrium.

We deviate from Mirrlees (1971) in a second aspect. Mirrlees considers the case where worker types are perfect substitutes, so that relative wages for various ability types are independent of supply and demand. Imperfect substitution between worker types is crucial for our analysis. Previously, Feldstein (1973) has analyzed this problem, and a whole 1982 issue of the *Journal of Public Economics* is devoted to the issue (see, among others, Allen, 1982, Stern, 1982, and Stiglitz, 1982). The conclusion of these early contributions is that imperfect substitution between types of labor does not make a great deal of difference for realistic

values of the elasticity of substitution. Our claim is that this conclusion is largely due to an unresolved technical problem. Where Mirrlees applied a continuous type distribution for the perfect substitution case, a continuous type production function with imperfect substitution was not available. Hence, a production function with a discrete number of types (in practice: two types) was applied. In the context of spill-over effects of minimum wages, Teulings (2000) shows that using a production function with only two instead of a continuum of types seriously biases our conclusions. A two type production function greatly understates the importance of the shift in relative wages. Our claim is that the same problem applies for the topic considered in this paper. By using a production function with only two broad types, the shifts in relative wages within each type are ignored. We use recent results regarding the structure of substitution, applying the Ricardian concept of comparative advantage to a world where heterogenous workers have to be assigned to heterogenous jobs, see Teulings (1995, 2001). In this type of world, the return to human capital is negatively related to its supply. The market mechanism assigns highly skilled workers to the complex jobs where these skills get the highest return. When the supply of these workers is increased, all complex jobs are filled, so that highly skilled workers have to be assigned to jobs where their skills receive a lower return. This model yields an extremely simple, negative exponential relation between the return and the mean level of human capital, see Teulings (2001). This relationship is applied here.

Finally, our paper deviates from Mirrlees' (1971) exercise by restricting the attention to log-linear policy rules and a utility function that is additive in consumption and effort. Mirrlees' great theoretical contribution was that his analysis derived the optimal relation between net and gross income without any prior restriction on its functional form. Ideally, one would like to apply a similar framework in this paper. Ulph and Ulph (1982) have made an attempt in this direction for a similar production function, but without endogenous human capital formation. We decided to retreat one step by restricting the functional form of the income policy a priori. Similarly, we adopt Diamond's (1998) idea of an additive utility function, thereby ruling out income effects in labor supply. These simplifications contribute to a key characteristic of our economy, namely, that the mean is a sufficient statistic for the distribution of human capital. This characteristic enables us to apply the simple, negative exponential relation between the return and the mean level of human capital. As soon as we allow higher moments of the human capital distribution to be affected, no closed form solution of the equilibrium wage locus is available, see Teulings (2001). By restricting the attention

to the class of log-linear income policies, our conditions for constrained Pareto efficiency will only be necessary conditions. It might be the case that there is some non log linear policy that Pareto dominates the log linear policy.

The constrained Pareto efficiency of education subsidies does not guarantee their political viability. We study two issues in this respect. The first regards the level at which a decision on the education subsidy is made. Decision making has to be sufficiently centralized to internalize all external effects. Our model makes a case against decentralized decision making on the funding of education, as applies in the United States. Decentralization yields underinvestment. The general equilibrium effect of education subsidies on wage dispersion is not taken into account at the decentralized level. Similarly, to the extent that relative wages are equalized across countries, either by labor mobility or by trade, a country is too low a level for decision making. Empirically, there are rather persistent differentials in relative wages across countries, suggesting that mobility and trade are not that important. Our argument against decentralized decision making is complementary to Fernandez and Rogerson's (1996) argument that decentralized funding results in inefficiently low education subsidies in poor communities.

The second issue concerns the time consistency of subsidies to education. Investment in human capital are irreversible. As soon as workers have made their investment and the compression in its return has been achieved, the median voter has an incentive to renege on its income policy and to tax human capital. Since ability and education are complementary, and hence the rich have taken up most education, the direct effect of a tax on years of education contributes to redistribution. We show that when the median voter lacks an instrument to credibly commit to a future income policy, she will be able to capture only half of the potential gains from the indirect effect of human capital formation for redistribution (see Boadway, Marceau, and Marchand, 1996, and Konrad, 2001, for related arguments in a different set up).

There is substantial evidence that an increase in the stock of human capital does indeed reduce income dispersion. Tilak (1989) provides some early cross country evidence. Katz and Murphy (1992) estimate the elasticity of substitution between high and low skilled workers. This elasticity drives the negative relation between the return and the mean level of human capital. There are several case studies on this relation for a number of countries: Goldin and Margo (1992), Goldin and Katz (1999), and Goldin (1999) examine the United States between World War I and II, when the US took the lead in a massive expansion in secondary education; Hartog, Oosterbeek, and Teulings (1993) analyze

the Netherlands 1960-1985; Edin and Holmlund (1995) examine Sweden; Kim and Topel (1995) consider the case of South Korea in the sixties and seventies. Finally, Teulings and Van Rens (2001) analyze simultaneously the evolution of log GDP per capita and the Gini coefficient in a panel of 100 countries over the period 1960-1990. The evolution of both GDP and the Gini support the notion that the rate of return to education declines in the mean level of education of the workforce. Their estimate of the size of this effect is broadly consistent with the elasticity of substitution reported by Katz and Murphy (1992).

Our analysis relates to two other strands in the literature. Persson and Tabellini (1994) study the relation between the pre-tax income distribution and political decision making on marginal tax rates. They build on studies (Romer, 1975, Roberts, 1977, and Meltzer and Richard, 1981) showing that pre-tax income inequality increases the political support for high marginal tax rates in a democracy (see Milanovic, 2000, for recent empirical evidence). Where high marginal rates reduces incentives and therefore growth, a dispersed pre-tax distribution reduces growth (see Alesina and Rodrik, 1994, for a related argument). These studies imply that for the purpose of growth, policies should be directed at achieving a flatter pre-tax income distribution. The model analyzed in this paper suggests that subsidies to education may serve this goal. In fact, the wide differences in inequality in disposable income between both sides of the Atlantic (or better, accounting for the UK: between both sides of the North Sea) are largely driven by differences in the dispersion of pre-tax income. This is consistent with the idea developed in this paper that differences in inequality might be driven by the level of human capital, rather than by marginal tax rates.

Several other argument have been proposed in favour of subsidies to education. If the direct cost of human capital acquisition are non-tax deductible, while the revenues are taxed, human capital decisions are distorted and a subsidy may correct for this (Trostel, 1996). Similarly, education subsidies may help to remove distortions in human capital accumulation arising from progressive income taxation (Bovenberg and Jacobs, 2001). Our model captures this argument. In endogenous growth models, any investment has external effects to subsequent generations, since this investment improves the starting position of these generations (Lucas, 1988, Tamura, 1991, and Aghion and Howitt, 1998). Investment in human capital is no exception to this general principle. The empirical evidence for positive externalities is weak, however (Acemoglu and Angrist, 1999, Bils and Klenow, 2000, and Krueger and Lindahl, 1999). Extending these arguments, Saint-Paul and Verdier (1993), Perotti (1993), and Benabou (1999) stress the role of capital market imper-

fections. When some groups have to borrow at rates above the market value, they will underinvest in human capital. Since the property rights of human capital cannot be transferred easily, capital market imperfections are particularly detrimental for this type of investment. There are, at least, two reasons why capital market imperfections can not fully account for the widespread prevalence of education subsidies. First, recent empirical studies cast serious doubt on the importance of borrowing constraints for educational choices (Cameron and Heckman, 1998 and 1999, Keane and Wolpin, 1999, Shea, 2000, and Cameron and Taber, 2000). Second, the argument is hard to reconcile with the comprehensiveness of government subsidies to education. If education subsidies only serve to attain equality of opportunity, subsidies targeted at the disadvantaged would be sufficient. In practice, government programs to encourage education have a much broader character.

Our analysis also relates to a recent paper by Heckman, Lochner, and Taber (1999) on the evaluation of training programs. The previous literature on this issue dealt with the implications of the selectivity in the participation in these programs for their measured return. Heckman et. al. ask attention for a second bias, the indirect, general equilibrium effects on relative wages. When the training program has only a marginal impact on the distribution of human capital, general equilibrium effects can be ignored. However, this no longer applies for programs that are available to the whole workforce. In fact, people will participate in these programs up to the point where their marginal private cost is equal to their marginal revenue. Our argument relies on the same type of general equilibrium effects, but now uses them to our advantage. By raising the human capital of the whole workforce, relative wages will be flattened. Education subsidies thus contribute to the political demand for redistribution. The least skilled benefit not so much from the direct effect of their own human capital accumulation, but from the indirect effect of raising the economy's human capital on their relative wage.

The analysis also helps to understand the persistence of unemployment / non-participation of low-skilled workers, both in Europe and the United States, and the lack of success of programs like the EITC in alleviating this problem. These programs intend to extract a larger supply of labor at the lowest skill levels by reducing the marginal tax rate at the relevant gross income levels. At the same time, these programs reduce the incentives for the investment in human capital, by increasing marginal tax rates at somewhat higher income levels. The latter increase is inevitable when one wants to avoid having to reduce the average tax rate for all income levels. The indirect, general equilibrium effects of both the increase in the supply of the least skilled and the reduction in

the incentives for investment in human capital run counter to the aim of the policy. They reduce the wages for the least skilled, and thereby their employment prospects. A more formal evaluation of this trade-off would require that we drop the restriction to a log linear policy. We leave that for a future paper.

The paper is organized as follow. Section 2 presents the model. In Section 3 we discuss some empirical studies to get an idea about the values of the key parameters of the model. Section 4 examines optimal redistribution policy. Section 5 is devoted to political economy issues. Section 6 concludes.

2 The structure of the economy

Consider an economy populated with individuals which are born with ability level a , $a \sim N(0, \sigma^2)$. The population grows at an exogenous rate ρ . Individuals die at a Poisson rate δ . At the beginning of their career, they take h years of education. Individuals choose h to maximize their expected lifetime utility. The optimal level of h may vary between ability types. After this investment in human capital, individuals start their working career endowed with a skill level s , that is a function of their innate ability a and their years of education h : $s = s(h, a)$, $s_h(h, a) > 0$, $s_a(h, a) > 0$. We ignore the cost of the education system itself, in line with the observation that this cost is of minor importance relative to cost of foregone labor income. For the sake of simplicity, we assume that workers do not accumulate human capital by work-experience.

Production is characterized by constant returns to scale and labor markets are perfectly competitive. Types of labor are the only factors of production. Hence, physical capital does not play a role in our economy. Total gross labor income is therefore equal to the value of production. We abstract from technological progress. The log gross labor income y of a particular skill type is determined by the log level of effort e and by the log wage per unit of effort w : $y \equiv w + e$. Like the years of education h , individuals choose e to maximize their expected lifetime utility. The optimal level of e may vary between ability types. For simplicity, the level of effort of students $e_{\text{education}}$ is supposed to be a fixed number. We conveniently choose: $e_{\text{education}} = -\infty$.

The wage rate of an individual depends on the individual's skill level s and the mean skill level weighted by units of effort among labour supply μ . μ is determined endogenously in the model. As will be explained in greater detail below, the model is structured as such that μ is a sufficient statistic for the characterization of the effect of economic policy on the skill distribution. Higher moments of the skill distribution are not affected by policy experiments. Due to the constant returns to scale pro-

duction technology, the size of the labor force is immaterial for the level of wages. However, due to imperfect substitution between skill types, relative wages depend on the composition of the skill supply distribution. As discussed before, the mean μ is a sufficient statistic to characterize this composition. As will also be explained below, a convenient implication of this set up is that log wages are a linear function of s :

$$w(s, \mu) = w_0(\mu) + w_s(\mu) s \quad (1)$$

with: $w_s(\mu) > 0$, $w'_s(\mu) < 0$. Without loss of generality, we impose the convenient normalization: $w_s(0) \equiv 1$. Equation (1) can be thought of as a simple Mincerian earnings equation, with $w_s(\mu)$ being the return to human capital which varies with the aggregate stock of human capital μ . The inequality $w'_s(\mu) < 0$ is due to the imperfect substitution between skill types: increasing the supply of highly skilled workers and reducing that of low-skilled workers (that is: increasing μ) reduces the wages of the former and raises the wages of the latter and therefore reduces the return to human capital.

In order to provide a clear cut separation between our model and models based on capital market imperfections, we assume perfect capital and insurance markets. Individuals can borrow sufficient funds to finance their consumption during their initial years of education h at the going interest rate. Also, they can insure perfectly the risk on their investment in human capital due to the uncertainty about their life expectancy related to the Poisson rate δ . The economy is supposed to be on a golden growth path: the interest rate is equal to the growth rate of the labor force ρ . Hence, individuals can borrow funds for a rate $\lambda = \rho + \delta$, ρ for the interest payments and δ as an insurance premium that covers the loan at the moment that individual dies. For convenience, the rate of time preference of individuals is assumed to be equal to ρ . We study the economy in its steady state equilibrium.

The government can observe neither a , nor s , nor e . Nor can it observe any borrowing or lending behavior of individuals. However, the government does observe log gross income y and the years of education h of an individual. It can therefore specify an income policy that provides working individuals a log disposable income d that is contingent on the two observable characteristics of the individual: $d = d(y, h)$. We make a simplifying assumption in that we consider only log-linear income policies:

$$d(y, h) = d_0 + d_y y + d_h h \quad (2)$$

Hence, tax levied on an individual with log gross income y and years of education h is equal to $e^y - e^{d(y, h)}$ (if positive, otherwise it is minus the

subsidy to the individual). In the redistribution free equilibrium without government intervention, this tax is zero. Hence, we have: $d_0 = d_h = 0$ and $d_y = 1$, and hence $d(y, h) = y$. More generally, $d_y = 1$ yields a proportional income tax; $d_y < 1$ ($d_y > 1$) yields a progressive (regressive) income tax; $d_h < 0$ is equivalent to a tax on education, while $d_h > 0$ represents an education subsidy. The government is assumed not to provide grants to students still at school. Their net income is zero and they must finance their consumption by borrowing. At first sight, this seems to be an important limitation to our analysis. However, it is not. Due to the perfect nature of capital markets, the introduction of a grant financed from a reduction of d_h would be offset by a reduction of the take up of credit by individuals during their years at school, leaving their lifetime consumption path, their years of education h and their level of effort e unaffected. Hence, the effect of grants for students is equivalent to $d_h > 0$. There are no other types of government spending than income policy. Since the government operates with a balanced budget in the steady state equilibrium, the sum of subsidies minus taxes for all working individuals at a particular point in time is equal to zero. We now proceed by discussing in greater detail the various building blocks of the model.

2.1 Utility, consumption, and effort

The individual's instantaneous utility at time x is assumed to be additively separable in consumption and effort, compare Diamond (1998):

$$u(c, e) = e^{\theta c} - e^{\pi e}$$

where $0 < \theta \leq 1$ and $\pi > 1$ and where c denotes log consumption. The additivity in this specification rules out income effects in the supply of effort. This is obviously an important restriction to the generality of our analysis, in particular when we would try to analyze log non-linear income policies. However, ruling out income effects in labor supply is a prerequisite for the tractability of our analysis. The special case $\theta = 1$ corresponds to the case of risk neutrality. For lower values of θ the individual is risk averse. However, risk aversion will play a minor role in our analysis. In line with Diamond (1998) we assume a constant elasticity of the supply of effort. With this specification of instantaneous utility, expected lifetime utility reads:

$$\begin{aligned} U &= \int_0^h \lambda e^{-\lambda x} e^{\theta g} dx + \int_h^\infty \lambda e^{-\lambda x} (e^{\theta(d+b)} - e^{\pi e}) dx \\ &= (1 - e^{-\lambda h}) e^{\theta g} + e^{-\lambda h} (e^{\theta(d+b)} - e^{\pi e}) \end{aligned} \quad (3)$$

where b is the log of one minus the fraction of disposable income that is used to repay the loans used for financing consumption when at school, denoted g . The first term of the first line reflects the utility during the years x that the individual spends at school, $0 < x < h$. The second term reflects the utility during working career of the individual, $x > h$. Future pay offs are discounted at a rate λ to account for the rate of time preference ρ and for rate of dying, δ . In the first term, we use: $e_{\text{education}} = -\infty$. Since the individual's rate of time preference is equal to the interest rate and since capital markets are perfect, consumption is constant both within the period spend at school and during the working career. This result is used in the specification of equation (3). The individual's budget constraint reads:

$$(1 - e^{-\lambda h}) e^g + e^{-\lambda h + d + b} = e^{-\lambda h + d} \quad (4)$$

where the left hand is net discounted expected lifetime consumption and where the right hand is net discounted expected lifetime income. Individuals maximize their utility with respect to b and g subject to their budget constraint, yielding:

$$\begin{aligned} \theta (1 - e^{-\lambda h}) e^{\theta g} - \Lambda (1 - e^{-\lambda h}) e^g &= 0 \\ \theta e^{-\lambda h + \theta(d+b)} - \Lambda e^{-\lambda h + (d+b)} &= 0 \end{aligned}$$

where Λ is the Lagrange multiplier of the budget constraint. Using the budget constraint, these relations can be simplified to:

$$\begin{aligned} g &= d - \lambda h \\ b &= -\lambda h \end{aligned} \quad (5)$$

Hence, consumption during the working career is equal to consumption during the period at school. This result is due to the separability of the utility function in effort and consumption. This relation allows the utility function to be simplified:

$$U = e^{\theta(d-\lambda h)} - e^{-\lambda h + \pi e}$$

Applying equation (2) and using $y \equiv w + e$, the first-order condition for optimal effort reads:

$$\frac{dU}{de} = \theta d_y e^{\theta[d(w+e, h) - \lambda h]} - \pi e^{-\lambda h + \pi e} = 0 \quad (6)$$

Hence:

$$e = \frac{\ln\left(\frac{\theta d_y}{\pi}\right) + \theta(d_0 + d_y w + d_h h) + (1 - \theta)\lambda h}{\pi - \theta d_y} \quad (7)$$

$$\equiv \varepsilon_0 + \varepsilon_w w + \varepsilon_h h$$

The second-order condition requires: $\pi - \theta d_y > 0$. Hence, effort is increasing in wages. This conclusion should not come as a surprise, since there is no income effect in the supply of effort. The elasticity of supply is equal to $\varepsilon_w = \frac{\theta d_y}{\pi - \theta d_y}$.

2.2 The education production function

The production function of education reads:

$$s(a, h) = \alpha + a + (\beta h - \xi a) - \frac{1}{2}\psi(\beta h - \xi a)^2 \quad (8)$$

where $0 < \psi$ and $0 \leq \xi < 1$. By the equations (1) and (3), the first-order condition for optimal years of education can be written as:

$$\frac{dU}{dh} = \theta [d_y w_s(\mu) s_h(a, h) + d_h - \lambda] e^{\theta[d(w+e, h) - \lambda h]} + \lambda e^{-\lambda h + \pi e} = 0$$

Applying (6), this relation can be solved for $s_h(a, h)$:

$$s_h(a, h) = \beta [1 - \psi(\beta h - \xi a)] = \frac{\pi(\lambda - d_h) - \lambda d_y}{\pi d_y w_s(\mu)} \quad (9)$$

The log linearity of income policy therefore implies that the marginal return to a year of schooling, $s_h(a, h)$, is independent of a and h in market equilibrium. We therefore drop its arguments. The parameter α shifts the distribution of s relative to that of a . Without loss of generality, we choose the parameters α and β such that $\mu = 0$ and $\beta h - \xi a = 0$ in the redistribution free equilibrium, $d_y = 1, d_h = 0$. Hence:

$$\beta = \frac{\pi - 1}{\pi} \lambda \quad (10)$$

Equation (9) provides a road map for the interpretation of the parameters. Since s_h is independent of a and h , so must be $\beta h - \xi a$ in market equilibrium. Hence, when $\xi = 0$, h is equal for all ability types, so that they all take up the same years of education. For $\xi > 0$, high ability types benefit more from schooling than low ability types and therefore take up more education. The parameter ξ measures the degree of complementarity of ability a and years of schooling h . This parameter plays therefore a crucial role in the analysis as it measures the direct

distributional effect of subsidies for education. There is an alternative interpretation of ξ , which is helpful for its empirical interpretation. Since $\beta h - \xi a$ is constant, $\text{Var}[\beta h] = \xi^2 \text{Var}[a] = \xi^2 \sigma^2$. Since $s_h = \beta$ in the redistribution free equilibrium, $\text{Var}[\beta h]$ is the variance of log wages due to education, and hence, ξ^2 is the share of the variance of log wages due to the "true" return to education. The parameter $\frac{1}{\psi}$ is the elasticity of demand for schooling with respect to its cost, $s_h(a, h)$.

Equations (9) yields an expression for the optimal years of education chosen by the individual:

$$h = \frac{1}{\psi\beta^2} (\beta - s_h) + \frac{\xi}{\beta} a \quad (11)$$

2.3 The skill distribution among labor supply

The equations (7), (8), (11), and (1) imply that a , e , and h , are linear functions of the skill level s :

$$\begin{aligned} a(s) &= -\frac{1}{2\psi\beta^2} (\beta^2 - s_h^2) - \alpha + s \equiv a_0 + s \\ e(s) &= \varepsilon_0 + \varepsilon_w w_0(\mu) + \varepsilon_h h_0 + \left[\varepsilon_w w_s(\mu) + \frac{\xi}{\beta} \varepsilon_h \right] s \equiv e_0 + e_s s \quad (12) \\ h(s) &= h_0 + \frac{\xi}{\beta} s \equiv h_0 + h_s s \\ h_0 &\equiv \frac{1}{2\psi\beta^2} (\beta - s_h) \left[2 - \frac{\xi}{\beta} (\beta + s_h) \right] - \frac{\xi}{\beta} \alpha \end{aligned}$$

The function $a(s)$ is the ability type that obtains skill level s in market equilibrium. It exhibits the important characteristic that $\frac{da(s)}{ds} = 1$, so that in equilibrium, schooling affects the mean of the skill distribution, μ , but leaves higher moments unaffected. This justifies our approach of treating μ as a sufficient statistic for the effect of income policy on the skill distribution. This feature is crucial condition for obtaining a closed form solution for the wage function $w(s, \mu)$ and for the linearity of this solution (Teulings, 2001). This characteristic depends crucially on the log-linearity of the income policy, see equation (2). A more general characterization of income policy would invalidate this convenient implication. The parameter a_0 is the ability level that leads to a zero skill level, $s = 0$, or equivalently, it is minus the skill level of the median ability type $a = 0$. The functions $e(s)$, and $h(s)$ are the values of e and h chosen by the skill type s , also in market equilibrium. Effort is an increasing function of the skill-level of the individual because

better skilled workers earn higher wages and effort is an increasing function of wages. Using the normality of the distribution of innate ability, $a \sim N(0, \sigma^2)$, we can derive the skill distribution measured in units of effort. Its density function $f(s)$ satisfies:

$$f(s) = \frac{1}{\sigma} \phi\left(\frac{a(s)}{\sigma}\right) \frac{da(s)}{ds} e^{e(s) - \lambda h(s)}$$

where $\phi(\cdot)$ is the standard normal density function. The factor $\frac{1}{\sigma} \phi\left(\frac{a(s)}{\sigma}\right)$ is the density of the underlying innate ability $a(s)$. The factor $\frac{da(s)}{ds}$ is the Jacobian for the transform of variable. By equation (12), it is unity and therefore cancels. The factor $e^{e(s) - \lambda h(s)}$ is the supply of skill s measured in units of effort per worker of ability type $a(s)$ in distribution of ability. This factor captures two effects. First, $e(s)$ comes in to adjust for the variation in the amount of effort provided by various skill types: the more effort type s supplies, the higher its share in the total skill distribution. Second, $\lambda h(s)$ captures the fact that the longer the ability type $a(s)$ has to stay at school to obtain the skill level s , the lower is its supply. The density function of the age x among the workforce is $\lambda e^{-\lambda x}$: the age group $x + dx$ is a fraction λdx smaller, ρdx due to population growth and δdx due to some workers dying between x and $x + dx$. Hence, the number of individuals that left school and started their labor market career (that is: $x > h(s)$) is equal to $\int_{h(s)}^{\infty} \lambda e^{-\lambda x} dx = e^{-\lambda h(s)}$. Substitution of equation (12) and rearranging terms yields:

$$\begin{aligned} f(s) &\equiv \frac{1}{\sigma} \phi\left(\frac{s - \mu}{\sigma}\right) \exp(L) & (13) \\ \mu &= -a_0 + \sigma^2 (e_s - \lambda h_s) \\ L &= e_0 - \lambda h_0 - a_0 (e_s - \lambda h_s) + \frac{1}{2} \sigma^2 (e_s - \lambda h_s)^2 \end{aligned}$$

where $\exp(L)$ measures the size of the labor force weighted by unit of effort and where $\frac{1}{\sigma} \phi\left(\frac{s - \mu}{\sigma}\right)$ measures its composition. Hence, the distribution of skill among labor supply is also normal, with the same variance as the underlying ability distribution and with mean μ . The first term of the relation for μ captures the overall shift of the skill distribution compared to the underlying ability distribution ($-a_0$ being skill level obtained by the median ability type). The second term captures the effect that high skill types supply more effort -therefore shifting μ upward-, but take up more years of schooling to obtain that skill level (if $\xi > 0$) - therefore shifting μ downward. The larger the variance of the ability distribution, the greater variation in e and h , and the greater therefore

their effect on μ . Aggregate log labor supply in units of effort is equal to L . It is increasing in the level of effort for the median worker, e_0 , and decreasing in the years of schooling taken by the median worker, h_0 . The final two terms in this relation are less easily interpreted.

Using the relations (12) for e and h , and the relations (1) and (2) for w and d , the utility function in equation (3) can be reduced to:

$$\begin{aligned} \ln U &\equiv \theta (u_0 + u_a a) - \ln \pi & (14) \\ u_0 &= d_0 + d_y [w_0 (\mu) + e_0] + (d_h - \lambda) h_0 \\ &\quad + \frac{1}{\theta} \ln (\pi - \theta d_y) - u_a a_0 \\ u_a &= d_y [w_s (\mu) + e_s] + h_s (d_h - \lambda) \end{aligned}$$

The utility of ability type a is log linear function of a . The intercept u_0 measures the level utility for the median ability type $a = 0$, the slope u_a measures the dispersion of utility across ability types. The level of utility depends on the general generosity of the income policy d_0 , the elasticity of net versus gross income d_y multiplied by the log wage and log effort of skill type $s = 0$, and on the cost of education $d_h - \lambda$ times the years of education taken up by skill type $s = 0$, h_0 . These terms specify the utility for *skill* type $s = 0$, while we are looking for the utility of the median *ability* type $a = 0$. The final term $-u_a a_0$ corrects for this difference. Apart from d_0 , the same factors affect the dispersion of utility. If $\xi = 0$, then $h_s = 0$. All ability types take up the same amount of schooling, so a subsidy to the education system, $d_h > 0$, has no direct effect on the dispersion of utility.

2.4 Production technology and wages

The production technology that we apply is originally derived in the context of an assignment model with heterogeneous jobs and heterogeneous workers. Jobs differ in their level of complexity and workers differ in their level of skill.¹ Skilled workers have a comparative advantage in complex jobs (Teulings 1995, 2001). When the mean of the skill distribution is shifted upwards (while preserving the spread of the distribution),

¹Note that the production technology uses the single index assumption: all worker characteristics can be meaningfully summarized in a single skill index, s . The assumption implies that innate ability and skill formation are perfect substitutes in production: the effects of lower innate ability can be offset by taking up more education (up to a certain limit, where taking up more years of education only reduces one's skill). Imperfect substitution diminishes the effectiveness of education subsidies in reducing inequality. The reason is that the return to innate ability depends less on the economy's stock of human capital, the less innate ability and schooling are substitutes in production.

a particular skill type s ends up in a less complex job, simply because there is not enough demand for highly complex commodities. It then follows from the first-order condition for optimal assignment of workers to jobs, that the return to skill, $w_s(\mu)$, goes down. Hence, different skill types are imperfect substitutes in this world. In the present paper, we skip the underlying assignment structure. Instead, we just apply the reduced form specification of wages that directly relates the return to skill $w_s(\mu)$ to the statistic μ , see Teulings (2001):

$$w_s(\mu) = e^{-\gamma\mu} \quad (15)$$

An increase in the mean skill level among labor supply by $d\mu$ will reduce the return to skill by $\gamma w_s(\mu) d\mu$. By the linearity of $w(s, \mu)$ in s , see equation (1), the normality of the distribution of s carries over to the distribution of w :

$$w \sim N [w_0(\mu) + w_s(\mu) \mu, w_s(\mu)^2 \sigma^2] \quad (16)$$

Hence, wages follow a log normal distribution, which fits the data well. Applying the formula for the expectation of the log normal distribution, the log average productivity per unit of effort as a function of μ , $H(\mu)$, can be calculated as:

$$H(\mu) \equiv \ln E[e^w] = w_0(\mu) + w_s(\mu) \mu + \frac{1}{2} w_s(\mu)^2 \sigma^2$$

The intercept $w_0(\mu)$ remains to be determined. Consider a policy that attributes the whole workforce with an equal amount of additional skill $d\mu$, that is, every skill type s is pushed to skill type $s + d\mu$. Hence, the mean of the skill distribution shifts from μ to $\mu + d\mu$. Suppose that we raise the skill of one worker of type s to $s + d\mu$. By marginal productivity theory, wages are equal to marginal productivity, so that the increase in output is equal to $e^{w(s+d\mu, \mu)} - e^{w(s, \mu)} = w_s(\mu) e^{w(s, \mu)} d\mu$. Aggregation over all skill types, taking logs, and the substitution of equation (15) for $w_s(\mu)$ yields an expression for the change in log average productivity H :

$$dH(\mu) = e^{-\gamma\mu} d\mu$$

Without loss of generality, $H(0)$ is normalized to zero. Then, the solution to the differential equation reads:

$$H(\mu) = -\frac{1}{\gamma} (e^{-\gamma\mu} - 1)$$

Combining these results yields:

$$w_0(\mu) = \frac{1}{\gamma} (1 - e^{-\gamma\mu}) - \mu e^{-\gamma\mu} - \frac{1}{2} e^{-2\gamma\mu} \sigma^2 \quad (17)$$

The compression elasticity is the percentage decrease in the return to human capital $w_s(\mu)$ per percent increase in its stock $e^{H(\mu)}$, see Teulings (2001):

$$-\frac{dw_s(\mu)}{w_s(\mu)} \frac{1}{dH(\mu)} = \gamma e^{\gamma\mu} \quad (18)$$

For $\gamma = 0$, the compression elasticity is equal to zero and skill types are perfect substitutes. In that case, relative wages do not depend on supply and demand:

$$w(s, \mu) = -\frac{1}{2}\sigma^2 + s$$

Aggregate output e^Y is equal to total labor supply in units of effort times average productivity per unit of effort, in logs:

$$\begin{aligned} Y &= L + H(\mu) \\ &= e_0 - \lambda h_0 - a_0(e_s - \lambda h_s) + \frac{1}{2}\sigma^2(e_s - \lambda h_s)^2 + H(\mu) \end{aligned} \quad (19)$$

Finally, we derive α . Since $s_h = \beta$ in the redistribution free equilibrium, equation (12) implies: $\alpha = -a_0$. Furthermore, since $\mu = 0$ in the redistribution free equilibrium, equation (13) implies: $a_0 = \sigma^2(e_s - \lambda h_s)$. Then, by equations (7), (10), (11):

$$\alpha = -\sigma^2 \frac{\theta - \xi\pi}{\pi - \theta} \quad (20)$$

2.5 The government's budget constraint

The model is closed with the budget constraint of the government:

$$\begin{aligned} e^Y &= \int_{-\infty}^{\infty} \frac{1}{\sigma} \phi\left(\frac{s+a_0}{\sigma}\right) \exp[d_0 + d_y \{w(s) + e(s)\} + (d_h - \lambda)h(s)] ds \\ &\equiv \int_{-\infty}^{\infty} \frac{1}{\sigma} \phi\left(\frac{s+a_0}{\sigma}\right) \exp(y_0 + y_1 s) ds \\ &= \exp\left(y_0 + \frac{1}{2}\sigma^2 y_1^2\right) \\ y_0 &= d_0 + d_y [w_0(\mu) - a_0 w_s(\mu) + e_0 - e_s a_0] + (d_h - \lambda)(h_0 - h_s a_0) \\ y_1 &= d_y [w_s(\mu) + e_s] + \xi(d_h - \lambda) \end{aligned}$$

The left hand side is total output, the right hand side is the sum of all income transfers. The final term $-\lambda h(s)$ in the exponent accounts for the fact individuals do not get any government support as long as they are still at school. Taking logs and substitution of equation (19) yields:

$$\begin{aligned} & e_0 - a_0 e_s + \frac{1}{2} \sigma^2 (e_s - \lambda h_s)^2 + \frac{1}{\gamma} (1 - e^{-\gamma \mu}) \\ &= d_0 + d_y [w_0(\mu) - a_0 w_s(\mu) + e_0 - e_s a_0] \\ & \quad + d_h (h_0 - h_s a_0) + \frac{1}{2} \sigma^2 [d_y \{w_s(\mu) + e_s\} + \xi (d_h - \lambda)]^2 \end{aligned} \quad (21)$$

Note that the term $-\lambda h(s)$ drops out again: the government saves money for every additional year of schooling since students get not paid, but at the same time this additional year also reduces the size of the labor force and therefore Y . Both effects cancel.

2.6 Combining the parts

We can substitute the budget constraint (21) and the monetary return to skill (15) into the utility function (14), yielding:

$$\begin{aligned} u_0 &= e_0 - a_0 (e_s - \lambda h_s) + \frac{1}{2} \sigma^2 (e_s - \lambda h_s)^2 \\ & \quad - \frac{1}{2} \sigma^2 [d_y (e^{-\gamma \mu} + e_s) + \xi (d_h - \lambda)]^2 \\ & \quad - \lambda h_0 + \frac{1}{\theta} \ln(\pi - \theta d_y) + \frac{1}{\gamma} (1 - e^{-\gamma \mu}) \\ u_a &= d_y (e^{-\gamma \mu} + e_s) + h_s (d_h - \lambda) \end{aligned} \quad (22)$$

Substitution of the equations (12) for a , e , and h , (7) for $\varepsilon_0, \varepsilon_w$, and ε_h , (13) for μ , (10) for β , and (20) for α , in the equations (9) and (22), yields after some rearrangement:

$$\begin{aligned} u_0 &= \frac{1}{\pi - \theta} \ln \frac{\theta d_y}{\pi} + \frac{1}{\theta} \ln(\pi - \theta d_y) - \frac{\pi}{\pi - \theta} \frac{1 - s_{h/\beta}}{\psi} \\ & \quad + \frac{\pi}{\pi - \theta} \frac{1}{\gamma} \left(1 - \frac{u_a}{1 - \xi s_{h/\beta}} \frac{\pi - \theta d_y}{\pi d_y} \right) - \frac{1}{2} \sigma^2 \left(\frac{\pi + \theta}{\pi} u_a^2 - \xi \frac{2\theta u_a + \xi \pi}{\pi - \theta} \right) \\ \ln u_a &= \ln(1 - \xi s_{h/\beta}) - \ln \left(\frac{\pi - \theta d_y}{\pi d_y} \right) \\ & \quad - \frac{\gamma}{2\psi} (1 - s_{h/\beta}^2) - \gamma \sigma^2 \theta \frac{(\pi - \theta) u_a - \pi(1 - \xi)}{(\pi - \theta) \pi} \\ s_{h/\beta} &= \frac{\pi(\lambda - d_h) - \lambda d_y}{\beta(\pi - \theta d_y) u_a + \xi[\pi(\lambda - d_h) - \lambda d_y]} \end{aligned} \quad (23)$$

where: $s_{h/\beta} \equiv \frac{s_h}{\beta}$.

This system of three equations provides a complete characterization of the equilibrium in terms of u_0 , u_a , and $s_{h/\beta}$, conditional on parameters of the income policy of the government, d_0 , d_y and d_h . This system is quite complicated and highly non-linear. Though it provides an explicit solution for $s_{h/\beta}$ and though it can be solved recursively (first $s_{h/\beta}$, then u_a , and finally u_0), the second equation does not provide an explicit solution for u_a . Nevertheless, the system has two convenient features from the point of view of policy evaluation. First, the parameter d_0 drops out. This result is due to the budget constraint of government: having set the elasticity of disposable income $d(\cdot)$ with respect to gross income d_y and years of education d_h , the level effect d_0 follows from the budget restriction. Hence, we only have to consider the policy parameters d_y and d_h . Second, due to our normalizations, the model has an analytical solution in the redistribution free equilibrium:

$$\begin{aligned} u_0 &= \frac{1}{\pi - \theta} \ln \frac{\theta}{\pi} + \frac{1}{\theta} \ln (\pi - \theta) \\ &\quad - \frac{1}{2} \sigma^2 \pi \frac{(1 - 2\xi) [\pi + \theta (1 - 2\xi)]}{(\pi - \theta)^2} \\ u_a &= \frac{\pi}{\pi - \theta} (1 - \xi) \\ s_{h/\beta} &= 1 \end{aligned} \tag{24}$$

This feature enables us to derive analytical results for the optimal policy mix for income redistribution, starting from the Walrasian equilibrium.

3 Empirical values for the key parameters

For the interpretation of the model, it is useful to have some idea about the values of its key parameters for the benchmark case of the redistribution free equilibrium, $d_y = 1$, $d_h = 0$. The parameter θ plays only a limited role in the subsequent analysis, and will therefore be set to unity in the most important case (the case where both $\gamma > 0$ and $\xi > 0$). The other five parameters of the model, π , β , σ , γ , ξ and ψ are discussed below; α and λ are determined residually from the normalizations; $\mu = 0$ in the redistribution free equilibrium.

The parameter π determines the wage elasticity of effort supply, which reads $\frac{d_y}{\pi - d_y}$, see equation (7), or for the redistribution free equilibrium $\frac{1}{\pi - 1}$. We deliberately use the term effort, not hours, since the supply of hours of work seems to be a bad proxy for what is economically relevant, the supply of effort. There has been an extensive debate on the elasticity of labour supply measured in hours, with little communis

opinion apart from the fact that this elasticity is much higher for females than for males and that for the latter it is difficult to provide evidence for any effect of wages on hours at all, see Pencavel (1986). However, effort is a different concept that might be more sensitive to monetary incentives. Similar to Diamond (1998), we assume the supply elasticity of effort in the redistribution free equilibrium to be equal to a half, yielding a value for $\pi \cong 3$.

Let the unit of time in our model be one year. The estimates of return to human capital corrected for ability bias float around 10 % per year. One might argue that this number is a bit high as an average. However, one should keep in mind that this is the return in the redistribution free equilibrium. Since most countries subsidize education, one can expect the actual return to be below its equilibrium value without government intervention. In the redistribution free equilibrium, β equals the return to education. Hence, we set $\beta \cong 0.10$. By equation (10), this implies $\lambda \cong 0.15$.

Since $w_s = 1$ in the redistribution free equilibrium, σ is equal to the standard deviation of wages in the redistribution free equilibrium. For the redistribution free equilibrium, $\sigma \cong 0.60$. This is the value that we observe in the Anglo-Saxon world and above the value for Continental Europe. However, in Continental Europe, education is subsidized much more, compressing $\text{Var}[w]$ below the value it would take otherwise. For $\mu = 0$, γ measure the compression elasticity, which determines the degree of substitutability between worker types. Katz and Murphy (1992) estimate the elasticity of substitution between high- and low-skilled workers to be 1.4. Teulings (2001) derives a simple transformation rule from this elasticity to the compression elasticity:

$$\gamma = \frac{1}{\text{Var}[w] \times \eta_{\text{low-high}}} \cong 2$$

Teulings and Van Rens (2001) estimate the compression elasticity from panel data on GDP per capita, income dispersion, and mean years of education for 100 countries over the period 1960-1990. They obtain similar values.

The parameter ψ is equal to the inverse of the price elasticity of the supply of human capital. By equation (11) we have:

$$\Delta h = -\frac{1}{\psi\beta^2}\Delta s_h = -\frac{1}{\psi\beta^2}\Delta w_h$$

where the latter equation follows from $w_s = 1$ in the redistribution free equilibrium; w_h is the usual measure for the cost of education (foregone earnings). Stanley (1999) analyzes the effects of the GI Bill education

subsidies for veterans from WW II and the Korean war on their educational attainment. This provides a natural experiment for testing the effect of financial incentives on educational attainment. The effects of these subsidies are necessarily limited to veterans with higher educational attainment, since they left the military at the age of 23. At that age, a substantial fraction has completed its investment in human capital. This fits the observation: only 40 % of the Korean veterans who were eligible took up any grants at all. Consistent with this observation, the veterans that took up grants descended predominantly from parents with a higher social economic status (SES). The 40 % that used the subsidies have increased their educational attainment on average by 1/3 year. Were the subsidy effective for all education levels, the effect would have been $\Delta h = \frac{1}{3 \times 40\%} \cong 0.85$ year. For WW II veterans, the data allow a more refined desaggregation by SES. There, the upper quintile of the SES distribution achieves a gain in educational attainment of even $\Delta h = 2$ years. The GI Bill reduced the cost of education by 50 % for Korean veterans, so from w_h to $\frac{1}{2}w_h$, and even by 60 % for WW II veterans. Since $s_h = w_h = \beta \cong 0.10$ in the redistribution free equilibrium, a reasonable estimate is: $\psi \cong 4$. One might suppose that for this group of veterans the cost of education has been higher than it would have been in case their educational career was not interrupted by the war. Hence, we view these estimates as an upper bound for ψ .

There is an alternative way to get an intuition for its magnitude, yielding a substantially lower value. Let h^* be the maximum number of years of education that still yields a positive effect on the skill level of the worker on top of the years of education taken in the redistribution free equilibrium. Further investment in education would only deteriorate the skill level of the worker. By equation (8) and since $h = 0$ in the redistribution free equilibrium, we have:

$$h^* = \frac{1}{\psi\beta}$$

Suppose that starting from the education attainment in the redistribution free equilibrium, four additional years of education might still be able to add to your level of skill, then $\psi = \frac{1}{4\beta} = 2.5$, suggesting a much larger price elasticity for the demand for education than the study by Stanley (1999). Given the fact that this is about the average increase in educational attainment since WW II, this does not seem to be an unreasonable number. Nevertheless, we use the higher value of $\psi \cong 4$ as a benchmark for the subsequent discussion.

Finally, the parameter ξ^2 measures the share of the variance in education measured at its ‘true return’ β as a share of the total variance of wages.

$$\xi = \frac{\text{std.dev.} [\beta h]}{\text{std.dev.} [w]}$$

Teulings and Hartog (1998) report values of about 0.30, at least for the United States. Angrist and Krueger (1991) find both the ‘true return’ and the ability bias to be 10 %. Our model implies that educational attainment and ability are perfectly correlated for $\xi > 0$, and that there are no other sources of variation in wages. Then, Angrist and Krueger’s results suggest $\xi \cong 0.50$. However, ability is not perfectly correlated with educational attainment. There are other sources of wage variation, apart from the ‘true return’ and ability bias. Hence, a value of a half is an upper bound for ξ . We use: $\xi \cong 0.33$. Table 1 gives an overview of the parameters.

Table 1 The benchmark parameters

π	β	σ	γ	ψ	ξ
3	0.10	0.60	2	4	0.33

4 Optimal income policy

The linear structure of utility as a function of the ability level of the worker is a convenient feature for the analysis of optimal redistribution policies. The amount of redistribution is captured by the slope of the utility function, u_a . A necessary condition for an efficient redistribution policy is therefore that it maximizes u_0 , taking u_a as given and using the parameters of the income policy, d_y and d_h as instruments. This condition is only a necessary condition, since we restricted attention a priori to log linear income policies. We have not checked whether there is Pareto improving non-linear policy. Since the ability level of the median worker is equal to zero, the outcome preferred by the median voter would simply maximize u_0 , using u_a , d_y and d_h as instruments. Since we restrict attention to log linear income policies, voters preferences are single peaked, so that we can apply the median voter theorem.

The optimality of subsidies to (or taxes on) education depends crucially on three parameters of our model: γ , which measures the degree of substitutability between skill types in production, ψ , which measures the price elasticity of educational attainment, and ξ , which measures the degree of complementarity of education and ability in skill formation. These parameters jointly determine the redistributive effects of education subsidies. The parameters γ and ξ have opposite effects on the role of education subsidies. For $\gamma > 0$, $\xi = 0$, the general equilibrium effects of an increase in educational attainment contribute to redistribution. An optimal redistribution policy will therefore include a subsidy to

education. For $\xi > 0, \gamma = 0$, the complementarity between ability and education makes that a subsidy to education raises inequality. Hence, an optimal redistribution policy will include a tax on education. In the empirically relevant case of $\gamma > 0, \xi > 0$, it depends on the relative magnitude of both effects.

We proceed as follows. First, we examine optimal income policy in the absence of complementarity effects of education subsidies ($\xi = 0$). Next, we study the empirically relevant case where both γ and ξ are positive. The calculations for this case are complicated. For the sake of convenience, we simplify the model for this case, by setting $\theta = 1$. We derive a simple condition that determines whether education should be taxed or subsidized for the purpose of redistribution.

4.1 The case $\gamma > 0, \xi = 0, \theta \leq 1$

Following (23), the reduced system with $\gamma > 0$ and $\xi = 0$ can be written as:

$$\begin{aligned}
u_0(d_y, s_{h/\beta}, u_a) &= \frac{1}{\pi - \theta} \ln \frac{\theta d_y}{\pi} + \frac{1}{\theta} \ln(\pi - \theta d_y) - \frac{\pi}{\pi - \theta} \frac{1 - s_{h/\beta}}{\psi} \\
&\quad + \frac{\pi}{\pi - \theta} \frac{1}{\gamma} \left(1 - \frac{u_a(\pi - \theta d_y)}{\pi d_y} \right) - \frac{1}{2} \sigma^2 \left(\frac{\pi + \theta}{\pi} u_a^2 \right) \\
F(d_y, s_{h/\beta}, u_a) &= -\ln u_a + \ln \left(\frac{\pi d_y}{\pi - \theta d_y} \right) - \frac{\gamma}{2\psi} (1 - s_{h/\beta}^2) - \gamma \sigma^2 \frac{\theta}{\pi} \left(u_a - \frac{\pi}{\pi - \theta} \right) = 0 \\
s_{h/\beta} &= \frac{\pi(\lambda - d_h) - \lambda d_y}{\beta(\pi - \theta d_y) u_a}
\end{aligned} \tag{25}$$

First, we show that, starting from the redistribution free equilibrium $d_h = 0, d_y = 1$, a subsidy to education contributes to redistribution. Substituting the expression for $s_{h/\beta}$ into the relation for $F(\cdot)$, setting $d_y = 1$, and totally differentiating the result with respect to u_a and d_h yields after some rearrangement:

$$\frac{du_a}{dd_h} \Big|_{d_h=0, d_y=1} = - \frac{\gamma \pi^4 (\pi - \theta) u_a^2}{(\pi - \theta)^2 u_a^2 \psi \pi + \gamma \pi^3 + \gamma \sigma^2 \theta (\pi - \theta)^2 u_a^3 \psi} < 0$$

Hence, starting from the redistribution free equilibrium, a subsidy to the education system contributes to redistribution. Education subsidies induce individuals to increase years of education. When skill types are imperfect substitutes in production, the resulting increase in the stock of human capital in society implies a decrease in the return to human capital and, hence, a reduction in the dispersion of wages. The dispersion of utility decreases.

Proposition I: Efficient redistribution without complementarity

When $\gamma > 0$ and $\xi = 0$, the parameters d_y and d_h of a constrained Pareto efficient log linear income policy are characterized by:

$$F(\cdot) = 0 \tag{26}$$

$$d_h = \lambda(1 - d_y) + \frac{\gamma\beta(1 - d_y)d_y^2}{u_a(\pi - \theta d_y) + \gamma d_y(1 - d_y)}$$

Proof:

A constrained Pareto efficient policy maximizes $u_0(\cdot)$ subject to the constraint $F(\cdot) = 0$. Since d_h shows up only in the final equation of (25), we can use $s_{h/\beta}$ instead of d_h as an instrument and solve the final equation ex post for d_h . The proof follows immediately from eliminating the Lagrange multiplier for the constraint from the first-order conditions for d_y and $s_{h/\beta}$:

$$u_{0,d_y}F_s = u_{0,s}F_{d_y}$$

where the subscripts refer to the relevant partial derivatives.

□

Before discussing the general case, two special cases deserve a separate discussion. First, when $d_y = 1$, then $d_h = 0$. The redistribution free equilibrium $d_y = 1, d_h = 0$ is therefore constrained Pareto efficient. This mirrors the first theorem of Welfare economics: investment in human capital in a market economy is Pareto efficient. If there is no demand for redistribution, the best a policy maker can do is not to intervene in the market mechanism. The second special case follows when types of labor are perfect substitutes: $\gamma = 0$. In that case: $d_h = \lambda(1 - d_y)$. This argument for education subsidies is discussed by Bovenberg and Jacobs (2001): education should be subsidized to offset the disincentive effects of the progressive income tax. This effect is a logically different reason for a subsidy to education than the general equilibrium effects on relative wages, which are the prime focus of the present paper. We shall therefore analyze taxes or subsidies on education net of this effect, by taking the variable X as a point of reference:

$$X \equiv \frac{d_h}{1 - d_y} - \lambda$$

X is the log subsidy per year of education divided by the marginal tax elasticity net of the subsidy necessary to correct for the distortionary effect of progressive taxation.

In the general case where $\gamma > 0$, the second term in equation (26) shows that for any $d_y < 1$, X must be positive for a constrained efficient

policy. Hence, in the empirically most relevant case of a political demand for redistribution from the rich to the poor with a progressive income tax ($d_y < 1$), a constrained Pareto efficient income policy requires a subsidy to education *above* the subsidy required to offset the distortions of the income tax. By encouraging schooling, pre-tax income inequality is reduced. Just like progressive income taxes, education subsidies entail distortions. The optimal subsidy to education induces individuals to *overinvest* in education. In the optimum, the marginal private cost of education exceeds the market rate of return to human capital ($s_{h/\beta} < 1$). The distortion in the schooling decision caused by the education subsidy is traded off against the distortion in the effort decision caused by marginal tax rates. The optimal redistribution policy mixes both distortions. The higher the compression elasticity γ , the more strongly relative wages are compressed by additional investment in human capital, and hence the higher is the optimal value of X . The higher the price elasticity of effort, $\frac{d_y}{\pi - \theta d_y}$, the higher the distortion caused by marginal tax rates, and hence the higher is the optimal value of X .

Proposition II: Median voter preferences without complementarity

When $\gamma > 0$ and $\xi = 0$, the parameters u_a , d_y and d_h that maximize the utility of the median voter satisfy the conditions in Proposition I and:

$$d_y (1 - d_y) (\pi + \gamma \sigma^2 \theta u_a) = \sigma^2 u_a^2 \pi (\pi d_y - \theta) \quad (27)$$

Proof:

The proof is obtained from eliminating the Lagrange multiplier for the constraint $F(\cdot) = 0$ from the first order conditions for u_a and $s_{h/\beta}$:

$$u_{0,ua} F_s = u_{0,s} F_{ua}$$

□

Due to the nonlinearity of the expression for u_a in (26), we do not obtain an explicit solution for d_y . However, equation (27) suggests that, starting from a value of σ of zero -where $d_y = 1$ - an increase in σ raises the tax rate. This is what one would expect intuitively. An increase in σ raises the difference between the mean and the median income, raising the median voter's financial interest in redistribution. Hence, she is prepared to accept a greater distortion in the effort and schooling decisions.

4.2 The case $\gamma > 0, \xi > 0, \theta = 1$

This section examines the empirically relevant case where both γ and ξ are positive. Allowing for both imperfect substitution between skill

types and complementarity between education and ability substantially increases the complexity of the model. Therefore, we simplify the model by setting the degree of risk aversion to zero ($\theta = 1$). For convenience, we rewrite the system given by (23). We substitute the equation for $s_{h/\beta}$ into the equations for u_0 and u_a and rearrange to obtain:

$$\begin{aligned}
U_0(d_y, d_h, U_a) &= \frac{1}{\pi} \ln \frac{d_y}{\pi} + \frac{\pi-1}{\pi} \ln(\pi - d_y) - \frac{1}{\psi} \left(1 - \frac{\pi}{\pi-1} \frac{\beta(\pi - d_y) - (\pi-1)d_h}{\beta(\pi - d_y)U_a - \xi\pi d_h} \right) \\
&\quad + \frac{1}{\gamma} \left(1 - \frac{\beta(\pi - d_y)U_a - \xi\pi d_h}{\beta\pi d_y} \right) \\
&\quad - \frac{1}{2} \frac{\sigma^2 (2\xi\pi + U_a - U_a\pi^2) (2\xi\pi + U_a - U_a\pi)}{\pi (\pi-1)^2} \\
F(d_y, d_h, U_a) &= -\ln \left(U_a - \xi \frac{\pi d_h}{\beta(\pi - d_y)} \right) + \ln \left(\frac{\pi - d_y}{\pi d_y} \right) - \frac{\gamma\sigma^2}{\pi} \left(U_a - \frac{\pi}{\pi-1} \right) \\
&\quad - \frac{\gamma}{2\psi} \left(1 - \left(\frac{\pi}{\pi-1} \frac{\beta(\pi - d_y) - (\pi-1)d_h}{\beta(\pi - d_y)U_a - \xi\pi d_h} \right)^2 \right) \\
&= 0
\end{aligned} \tag{28}$$

where

$$\begin{aligned}
U_0 &= \frac{\pi-1}{\pi} u_0 \\
U_a &\equiv u_a + \frac{\pi}{\pi-1} \xi
\end{aligned}$$

Since both U_0 and U_a are linear transformations of u_0 and u_a independent of the endogenous variables of the model, they play the same role as u_0 and u_a in the determination of efficient redistribution policies.

First, we analyze the redistributive effects of progressive income taxation and education subsidies separately. By differentiating totally the restriction $F(\cdot) = 0$ in the redistribution free equilibrium, we obtain after some rearrangement of terms:

$$\begin{aligned}
\frac{dU_a}{dd_y} \Big|_{d_y=1, d_h=0} &= \frac{\pi^2}{\pi-1} \frac{1}{\psi(\pi-1) + \gamma(\pi-1 + \psi\sigma^2)} > 0 \\
\frac{dU_a}{dd_h} \Big|_{d_y=1, d_h=0} &= \frac{\pi}{\beta} \frac{\xi\psi - (1-\xi)\gamma}{(\gamma + \psi)(\pi-1) + \gamma\sigma^2\psi} \geq 0
\end{aligned} \tag{29}$$

where we use the property that $U_a = \frac{\pi}{\pi-1}$ in the redistribution free equilibrium. Progressive income taxation, $d_y < 1$, reduces utility dispersion unequivocally. The effect of education subsidies depends on parameter values. The reason is clear. On the one hand, by stimulating human

capital formation, education subsidies reduce wage dispersion because skill types are imperfect substitutes in production. On the other hand, the complementarity between education and ability implies that individuals with high ability go to school longer. Since the amount of education subsidies is increasing in the years of education an individual takes up (see equation 2), education subsidies disproportionately favour the people with high ability. Hence, the complementarity of education and ability may cause education subsidies to increase the dispersion of utility. From (29) it follows that education subsidies are a redistributive instrument if and only if:

$$\xi\psi < (1 - \xi)\gamma$$

This equation gives the fundamental condition for the evaluation of the redistributive effects of education subsidies in the empirically relevant case where skill types are imperfect substitutes and education and ability are complementary. It has a simple economic interpretation. The parameter ξ is the share of wage dispersion that is attributable to the cost of human capital acquisition, while ψ is the inverse of the elasticity of educational attainment with respect to the cost of education. Hence, the left hand side is direct effect of the subsidy: the relative increase in inequality due to a subsidization of human capital. The right hand side measures the reduction in inequality: $1 - \xi$ is the share of wage dispersion that is directly attributable to ability differentials, while γ is the compression elasticity, measuring the relative decrease in the return to these ability differentials per value unit increase in the stock of human capital. Note that for the empirical values of the parameters discussed in Section 3, both sides of the inequality are just equal. Whether this condition applies is therefore extremely sensitive to the exact parameter values assumed.

Proposition III: Efficient redistribution with complementarity

When $\gamma > 0, \xi > 0$ and $\theta = 1$, the value for X for a constrained Pareto efficient log linear income policy in the neighbourhood of the redistribution free equilibrium is characterized by:

$$\frac{X}{\beta} = \frac{\gamma(1 - \xi) - \psi\xi}{(\pi - 1)(1 - \xi)} \times \frac{(\pi - 1)\psi - 2\pi(1 - \xi)}{\pi\psi + 2[\gamma(1 - \xi) - \psi\xi]}$$

Proof:

A constrained Pareto efficient policy maximizes $u_0(\cdot)$ subject to the constraint $F(\cdot) = 0$. Eliminating the Lagrange multiplier for the constraint from the first-order conditions for d_y and d_h results in:

$$U_{0,dy}F_{dh} = U_{0,dh}F_{dy}$$

Using the definition of X and taking the limit $d_y \rightarrow 1, U_a \rightarrow \frac{\pi}{\pi-1}$ yields after some rearrangement Proposition III.

□

As Proposition III shows, $\psi\xi < \gamma(1 - \xi)$ and $\frac{\pi-1}{\pi}\psi > 2(1 - \xi)$ are sufficient conditions for $X > 0$, implying that a progressive income tax goes along with subsidies to education in the case of efficient redistribution: $d_y < 1 \Leftrightarrow d_h > 0$. The first condition is the same condition as has been discussed above for the case where we considered the effect of education subsidies on the dispersion of utility. This condition is therefore easy to understand, see the discussion above. The second condition is less easy to appreciate. Suppose that the first condition is fulfilled, but not the second. This implies that an optimal progressive redistribution policy using either only taxation or only education subsidies would set $d_y < 1$ and $d_h > 0$. However, when both instruments are combined, one of the two inequalities is reversed. The following proposition is useful for the interpretation of this result:

Proposition IV: Efficient compressing redistribution requires progressive taxation

When $\gamma > 0, \xi > 0$ and $\theta = 1$ and when income policy satisfies the efficiency rule of Proposition III, then in the redistribution free equilibrium:

$$\frac{dU_a}{dd_y} \Big|_{d_y=1, d_h=0} > 0$$

Proof:

Differentiating the equations $F(\cdot) = 0$ and $d_h = (1 - d_y)(X + \lambda)$ totally yields:

$$\begin{aligned} 0 &= F_{U_a}dU_a + F_{d_y}dd_y + F_{d_h}dd_h \\ dd_h &= -dd_y(X + \lambda) + (1 - d_y)dX \end{aligned}$$

Eliminating dd_h from the first equation and taking the limit $d_y \rightarrow 1, d_h \rightarrow 0, U_a \rightarrow \frac{\pi}{\pi-1}$ yields proposition IV.

□

Proposition IV establishes that a progressive income tax is part of an efficient redistribution scheme from the rich to the poor. Hence, if condition $\psi\xi < \gamma(1 - \xi)$ is satisfied but condition $\frac{\pi-1}{\pi}\psi > 2(1 - \xi)$ is violated, then redistribution from the rich to the poor requires $d_y < 1$ and $d_h < 0$, despite the fact that by itself subsidies to education contribute to redistribution. The reason for this phenomenon is that the elasticity of labor supply $\frac{1}{\pi-1}$ is so low that the distortions caused by high

marginal tax rates are too small to be offset by subsidies to the education system, which cause by themselves also welfare losses. However, for the empirical values discussed in Section 3, the condition $\frac{\pi-1}{\pi}\psi > 2(1-\xi)$ is easily satisfied.

The median voter equilibrium can be derived in the same way as for the case $\gamma > 0, \xi = 0, \theta < 1$. However, the formulas do not yield further insights. Hence, we do not present them here.

5 Political economy aspects

5.1 Decentralized decision making

Our argument for subsidizing education rests on an externality in individual schooling decisions. Individuals do not take into account the effect of their schooling on pre-tax wage inequality and, thus, on the distortions arising from progressive income taxation. Clearly, to internalize the external effect, collective decision making must be sufficiently centralized. Labour mobility and trade make that local wage dispersion is to a large extent independent of the local stock of human capital. When decisions on funding of education are taken decentrally, the general equilibrium effect of education subsidies on wage dispersion will not be fully taken into account.

Consider the case of a small district in a large country. Labor is immobile, but there is free trade of products between districts. Hence, by the Heckscher-Ohlin factor price equalization theorem, relative wages no longer depend on the distribution of skill in the district, but on the distribution of skill in the whole economy. Wages are thus given by:

$$w(s, \bar{\mu}) = w_0(\bar{\mu}) + w_s(\bar{\mu})s$$

where $\bar{\mu}$ is the economy-wide mean skill level. By marginal productivity theory, it holds that a change in wages as a result of an increase in the mean skill level is matched by an equal increase in log average productivity per unit of effort H . Hence:

$$dH(\mu) = w_s(\bar{\mu})d\mu$$

Normalizing $H(0)$ to zero, log average productivity per unit of effort is given by:

$$H(\mu) = \mu w_s(\bar{\mu})$$

The other parts of the model remain the same. The model in reduced form reads:

$$u_0 = \frac{1}{\pi - \theta} \ln \frac{\theta d_y}{\pi} + \frac{1}{\theta} \ln(\pi - \theta d_y) - \frac{\pi}{\pi - \theta} \frac{1 - s_{h/\beta}}{\psi} \quad (30)$$

$$\begin{aligned}
& + \frac{\pi}{\pi - \theta} w_s \frac{1}{2\psi} (1 - s_{h/\beta}^2) + w_s \sigma^2 \theta \frac{u_a (\pi - \theta) - \pi (1 - \xi)}{(\pi - \theta)^2} \\
& - \frac{1}{2} \sigma^2 \left(\frac{\pi + \theta}{\pi} u_a^2 + \xi \frac{2\theta u_a - \xi \pi}{(\pi - \theta)} \right) \\
u_a & = \pi d_y w_s \frac{1 - \xi s_{h/\beta}}{\pi - \theta d_y} \\
s_{h/\beta} & = \frac{\pi (\lambda - d_h) - \lambda d_y}{\beta \pi d_y w_s}
\end{aligned}$$

The exogeneity of the return to skill reduces the complexity of the model considerably. The expressions in (30) can be combined to solve explicitly for u_0 and u_a .

The redistributive effect of education subsidies is given by:

$$\frac{du_a}{dd_h} = \frac{\xi \pi}{\beta (\pi - \theta d_y)} \geq 0$$

Evaluated at the decentralized level, education subsidies increase the dispersion of utility unless $\xi = 0$. The reason is clear. Since the district is too small to have an effect on relative wages in the economy, the only distributive effect stems from the complementarity between ability and education in skill formation. Since people with high abilities go to school longer, they benefit more from education subsidies than people with low abilities.

Constrained Pareto optimal policy is found by maximizing u_0 to d_y and $s_{h/\beta}$ subject to the constraint on u_a given by the second equation in (30). The resulting first-order conditions imply:

$$d_h = \lambda (1 - d_y) \left(1 - \xi \psi w_s (\bar{\mu}) \frac{\pi - 1}{u_a \pi} \frac{d_y^2}{\pi - \theta d_y} \right)$$

Hence, if $\xi = 0$, education subsidies are only used to offset the distortionary effect of increasing marginal tax rates on schooling decisions. If $\xi > 0$, constrained Pareto efficiency implies that progressive taxation is combined with a subsidy to education which is *lower* than the subsidy needed to offset tax distortions. For high values of ξ , even a tax on education may become constrained Pareto efficient at the decentral level. Clearly, taxing education contributes to redistribution as high-ability types take up more education than low-ability types. The (local) distortionary effect on schooling decisions is traded off against the disincentive effect of the other redistributive instrument, progressive taxation. Since the general equilibrium effect of education subsidies on relative wages is not taken into account at the decentralized level, subsidies are inefficiently low. Hence, decentralization yields underinvestment in human capital.

The case discussed above matches closely the US institutional structure, where decisions on education are made at the level of school districts. The main difference is that the tax policy is decided predominantly at a federal level. This feature of the US system may strengthen our result that decentralized bodies provide too low subsidies to education. The reason is that central decision making on taxes introduces an additional externality in decentral decision making, discouraging investment in human capital. While in the analysis above the direct consequences of underinvestment in human capital for the government budget are fully taken into account, this is no longer the case if local income is subject to federal taxes. Studying these issue more fully would require the introduction of separate budget constraints for the school district and the local government.

5.2 Time consistency of subsidies to education

So far, we have studied optimal income policy from the perspective of an individual at the beginning of his life. Moreover, we have assumed that the optimal income policy is set once and for all. In this section, we relax both assumptions to gain insight into the political viability of education subsidies in a world where the decisive voter has already started his working career and can not commit to future policies.

We consider a dynamic economy where old generations die and new generations enter the labor force. Inhabitants differ along two dimensions. First, they are either at school or working. Second, they differ according to their ability level a . For simplicity, we assume that while at school, inhabitants vote as if they are working. In this way we ignore slight differences between the interest of those at school and those working. The main interest is within generations: the low income people have an interest in past accumulation of human capital (because of general equilibrium effects on relative wages) and today expropriation of the fruits of human capital (for redistribution). If income policy is bound to be uniform across generations, the incentive to expropriate the fruits of past human capital formation conflicts with the incentive to stimulate current human capital formation by young generations. As a result, the time-consistent income policy entails lower education subsidies than the education subsidies which maximize the median voter's life-time utility. Due to a credibility problem, she is less able to capture the gains from the indirect effect of human capital formation for redistribution.

Consider the economy at a particular point in time. In that economy, each cohort of workers make their own choice on the years of education to take up. Define s_t as the average skill level obtained by the cohort of workers born at time t . As discussed before, changes in income policy

affect only the mean skill level of a cohort of workers, so that s_t is a sufficient statistic for the skill distribution of a cohort. Furthermore, define S_t as the mean skill level across cohorts at time t .² Since the rate of inflow of new generations as a fraction of the labor force is equal to λ (workers die at a rate δ and the labor force grows at a rate ρ), the average skill level converges to the optimal skill level according to the following law of motion:

$$S'_t = \lambda [s_t - S_t]$$

Define d_t as the parameters $[d_0, d_y, d_h]$ of the income policy applying at time t and D_t as the set of all expected future income policies $d_\tau, \tau = t, \infty$. Furthermore, define $u(S_t, D_t)$ as the utility stream at time t of a working individual with ability $a = 0$ (the median voter) conditional on future income policy and the mean skill level at that point in time, and define $U(S_t, D_t)$ as the expected utility for the remaining lifetime of that individual from time t onwards. New cohorts choose their skill level depending on D_t . Let $s^*(D_t)$ be the optimal choice of the mean skill level by the cohort of workers born at time t conditional on future income policy D_t . Today's policy is set by the median voter as to maximize her utility, taking into account that future generations of median voters will set policy in the future so as to maximize their utility. Since S_t is the only relevant state variable in the economy, future policies D_t will be some function $D^*(S_t)$ of S_t . Hence, we can write $U[S_t, D^*(S_t)] \equiv U^*(S_t)$. Then, the Bellman equation for the utility of a working individual reads:

$$\begin{aligned} \lambda U^*(S_t) &= u(S_t, D_t) + U^{*'}(S_t) S'_t \\ &= u(S_t, D_t) + U^{*'}(S_t) \lambda [s^*(D_t) - S_t] \end{aligned}$$

The current pay off of the expected utility is equal to the sum of the instantaneous utility and the change in the expected utility. The optimal policy choice by the median voter at time t maximizes $U(S_t, D_t)$ with respect to d_t ,³ hence:

$$0 = u_d(S_t, D_t) + U^{*'}(S_t) \lambda s_d^*(D_t)$$

²Strictly, S_t is not a sufficient statistic for the aggregate skill distribution since mixing a number of normal distribution does not add up to normal distribution. Hence, the simple linear wage equation does not apply. The more the distributions for each cohort differ, the larger this problem. Hence, in the limit close to the steady state considered below, this problem vanishes since the distributions of various cohorts are almost identical. Note that S_t differs from the concept μ , in that the latter measures the average skill level weighted by units of effort.

³We ignore a slight complication here. Individuals start their career at school and work afterwards. The interest of workers and students diverge slightly, since a change

The Bellman equation applies identically for all S_t . Hence, its first difference must also apply. Differentiating the Bellman equation yields:

$$\lambda U^{*'}(S_t) = u_S(S_t, D_t) + U^{*''}(S_t) \lambda [s^*(D_t) - S_t] - \lambda U^{*'}(S_t)$$

In the steady state equilibrium, $S_t = s^*(D_t)$ and d_t and S_t converge to their steady state equilibria d and S respectively. Hence:

$$0 = u_d(S, d) + K u_s(S, d) s_d^*(d)$$

where $K = 1/2$ and where we replaced the set of all future income policies D_t by the single policy d for the sake of notational convenience. Other values for K yield the equilibria discussed before: $K = 1$ is the centralized equilibrium with commitment where all externalities of S_t are internalized; $K = 0$ is the decentralized equilibrium, where no externalities are internalized. Commitment problems bring us half way from perfect commitment to complete decentralization. The median voter does not expropriate past investments in human capital completely because she expects to live for another $1/\delta$ years. A complete expropriation would take away the incentives for investment in human capital too much, thereby reducing the future stock of human capital and hence raising its future rates of return, at the expense of the median voter.

6 Concluding

We have shown that, for a government that wants to redistribute income for whatever reason, the indirect, general equilibrium effect of investments in human capital provide a forceful argument for the subsidization of schooling. However, there is an effect working in the opposite direction. Since the take up of schooling is complementary to innate ability, the direct effect of a subsidy to education tends to favour high ability types. Our overview of some empirical studies fails to clarify which effect is most important: the direct effect or the indirect, general equilibrium effect. Much depends on what one believes about the price elasticity of the demand for education. The higher the elasticity, the more education should be subsidized. Previous studies on optimal taxation have always downplayed the importance of general equilibrium effects. The reason that these effects show up much more prominently in this study is that we use a more realistic production technology, based on comparative advantage of high skilled workers in complex job types. Contrary to for

in today's income policy does not affect students. We ignore this intergenerational complication and focus on intragenerational issue: the median ability type who cannot credibly commit vis-a-vis the high ability not to expropriate her investment in human capital at some future date.

example a two type CES technology, this production technology implies that the whole wage schedule becomes flatter as a result of an increase in the average stock of human capital.

In the end, after having spend all the effort, the reader might be a bit disappointed by the conclusion that the empirical evidence suggests that the direct and indirect effect cancel. Why trying to understand all the arguments, while we end up with a conclusion that might read as saying that a neutral policy is probably the best we can do.⁴ However, this conclusion is incorrect. The general notion that the general equilibrium effects of schooling offer an important contribution to redistribution is important. We can try to exploit this notion in a much smarter way than the simple log-linear income policy analyzed in this paper, merely for reasons of tractability.

Any policy that raises the average skill level without having the regressive income effects of a subsidy to years of education contributes to efficient redistribution. An interesting extension is to include inter-generational effects. Then, the social economic status of the previous generation can be used as an indicator for the expected educational attainment of the next generation. This might shift the balance between adverse income effects of education subsidies and the desired substitution effects to increase the take up of education in favor of the latter. In the practice of income policy, this boils down to subsidies that are conditional on parental income.

Next, only foregone earnings are analyzed in this paper. One could extend the analysis to the direct cost of education. Then, a typical policy parameter might be the quality of primary education: raising the quality affects everybody, but does not have the adverse income effects. This policy has no adverse income effects, but is likely to raises the average skill level in the economy.

The log linearity of the income policy imposes another strong restriction. It implies increasing marginal tax rates (for $d_y < 1$), offending the logic of the Sadka (1976) argument. Interestingly, this argument can be extended towards the education subsidies, but then reversed. Where in the case of income taxation, the income effects are desired for the purpose of redistribution while the substitution effects only cause efficiency losses, here the substitution effects contribute to the redistri-

⁴This is not to say that education should not be subsidized. The present paper considers only general equilibrium effects on inequality as a potential argument for education subsidies. To the extent that other arguments apply (capital market imperfections or production externalities), the weight of these arguments adds up to that of the general equilibrium effects. However, the empirical evidence on both issues is inconclusive.

bution while the income effects work in the opposite direction. Hence, the marginal rate of education subsidies should be high at the bottom and at the top, where they do not cause substantial income effects since there are no people earning less than the lowest or more than the highest income. The previous argument regarding the quality of primary education exploits this idea at the bottom. Where this idea fits the layman's intuition, its counterpart is more surprising. A subsidy for top education programs has little adverse income effects while at the same time raises the average education level. The DIDES (Distance Dependent Elasticity of Substitution) production function applied in this paper implies that all lower ability types will benefit from the general equilibrium effects of this policy, see Teulings (2001).

The proposed analysis of the optimal functional form of taxes and education subsidies will have strong policy implication for programs like the EITC and the New Deal. These programs aim at a reduction of marginal tax rates for the lowest ability types in order to combat low-skilled unemployment. The government budget constraint then dictates that marginal rates should be increased for higher ability types. The logic of the argument in this paper suggests that this policy will be victim of its own success. To the extent that the subsidies induce low ability types to go to work, the relative increase in low skilled labor supply will reduce their wages, thereby partially undoing the initial effect of the subsidy. Stated more crudely: there is limit to the demand for hamburger flippers. If we use tax policy to increase their supply, sooner or later their gross wages are going to fall. At the same time, the increase in marginal rates for somewhat higher skill types, which is necessary to satisfy the government budget constraint, reduces the incentive for investment in human capital, which further aggravates the problem. All this points to the need of a more formal analysis of this trade off.

7 References

Acemoglu, Daron, and Joshua Angrist (1999), How Large are the Social Returns to Education? Evidence from Compulsory Schooling Laws, NBER Working Paper No. W7444.

Aghion, Philippe, and Peter Howitt (1998), *Endogenous Growth Theory*, Cambridge, MA: MIT Press.

Alesina, Alberto, and Dani Rodrik (1994), Distributive Politics and Economic Growth, *Quarterly Journal of Economics*, 109(2), pp. 465-490.

Allen, Franklin (1982), Optimal Linear Income Taxation with General Equilibrium Effects on Wages, *Journal of Public Economics*, 17(2), pp. 135-143.

Angrist, Joshua D., and Alan B. Krueger (1991), Does Compulsory School Attendance Affect Schooling and Earnings?, *Quarterly Journal of Economics*, 106(4), pp. 979-1014.

Becker, Gary S. (1983), A Theory of Competition among Pressure Groups for Political Influence, *Quarterly Journal of Economics*, 98(3), pp. 371-400.

Benabou, Roland (1999), Tax and Education Policy in a Heterogeneous Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency?, *Econometrica*, forthcoming.

Bils, Mark, and Peter J. Klenow (2000), Does Schooling Cause Growth?, *American Economic Review*, 90(5), pp. 1160-1183.

Boadway, Robin, Nicolas Marceau, and Maurice Marchand (1996), Investment in Education and the Time Inconsistency of Redistributive Tax Policy, *Economica*, 63, pp. 171-189.

Bovenberg, A. Lans, and Bas Jacobs (2001), Redistribution and Education Subsidies, Mimeo, University of Amsterdam.

Cameron, Stephen V., and James J. Heckman (1998), Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males, *Journal of Political Economy*, 106(2), pp. 262-333.

Cameron, Stephen V., and James J. Heckman (1999), The Dynamics of Educational Attainment for Blacks, Hispanics, and Whites, NBER Working Paper No. W7249.

Cameron, Stephen V., and Christopher Taber (2000), Borrowing Constraints and the Returns to Schooling, NBER Working Paper No. W7761.

Diamond, Peter A. (1998), Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates, *American Economic Review*, 88(1), pp. 83-95.

Edin, Per-Anders, and Bertil Holmlund (1995), The Swedish Wage Structure: The Rise and Fall of Solidarity Wage Policy?, In: Freeman, Richard B., and Lawrence F. Katz (Eds.), *Differences and Changes in Wage Structures*, pp. 307-343, Chicago: University of Chicago Press.

Feldstein, Martin S. (1973), On the Optimal Progressivity of the Income Tax, *Journal of Public Economics*, 2(4), pp. 357-376.

Fernandez, Raquel, and Richard Rogerson (1996), Income Distribution, Communities, and the Quality of Public Education, *Quarterly Journal of Economics*, 111(1), pp. 135-164.

Hartog, Joop, Hessel Oosterbeek, and Coen Teulings (1993), Age, Wages and Education in the Netherlands, In: Johnson, Paul, Klaus F. Zimmermann (Eds.), *Labour Markets in an Ageing Europe*, pp.182-211, Cambridge: Cambridge University Press.

Goldin, Claudia (1999), Egalitarianism and the Returns to Education during the Great Transformation of American Education, *Journal of Political Economy*, 107(6), pp. S65-94.

Goldin, Claudia, and Lawrence F. Katz (1999), The Returns to Skill in the United States across the Twentieth Century, NBER Working Paper No. 7126.

Goldin, Claudia, and Robert A. Margo (1992), The Great Compression: The Wage Structure in the United States at Mid-century, *Quarterly Journal of Economics*, 107(1), pp. 1-34.

Heckman, J.J., L. Lochner, en C. Taber (1999), *Explaining rising wage inequality: explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents*, NBER working paper 6384.

Katz, Lawrence F., and Kevin M. Murphy (1992), Changes in Relative Wages, 1963-1987: Supply and Demand Factors, *Quarterly Journal of Economics*, 107(1), pp. 35-78.

Keane, Michael P., and Kenneth I. Wolpin (2001), The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment, Penn Institute for Economic Research Working Paper 01-018.

Kim, Dae-Il, and Robert H. Topel (1995), Labor Markets and Economic Growth: Lessons from Korea's Industrialization, 1970-1990, In: Freeman, Richard B., and Lawrence F. Katz (Eds.), *Differences and Changes in Wage Structures*, pp. 227-264, Chicago: University of Chicago Press.

Konrad, Kai A. (2001), Privacy and Time-Consistent Optimal Labor Income Taxation, *Journal of Public Economics*, 79(3), pp. 503-519.

Krueger, Alan B., and Mikael Lindahl (2000), Education for Growth: Why and For Whom?, NBER Working Paper No. W7591.

Lucas, Robert E., Jr. (1988), On the Mechanics of Economic Development, *Journal of Monetary Economics*, 22(1), pp. 3-42.

Meltzer, Allan H., and Scott F. Richard (1981), A Rational Theory of the Size of Government, *Journal of Political Economy*, 89(5), pp. 914-927.

Milanovic, Branko (2000), The Median-Voter Hypothesis, Income Inequality, and Income Redistribution: An Empirical Test with the Required Data, *European Journal of Political Economy*, 16(3), pp. 367-410.

Mirrlees, James A. (1971), An Exploration in the Theory of Optimum Income Taxation, *Review of Economic Studies*, 38(114), pp. 175-208.

O'Neill, Donal (1995), Education and Income Growth: Implications for Cross-country Inequality, 103(6), pp. 1289-1301.

Pencavel, John (1986), Labor Supply of Men: a Survey, in: O. Ashenfelter and R. Layard, *Handbook of Labor Economics*, Amsterdam, North Holland, 1986, pp. 3-102.

- Perotti, Roberto (1993), Political Equilibrium, Income Distribution, and Growth, *Review of Economic Studies*, 60(4), pp. 755-776.
- Persson, Torsten, and Guido Tabellini (1994), Is Inequality Harmful for Growth?, *American Economic Review*, 84(3), pp. 600-621.
- Romer, Thomas (1975), Individual Welfare, Majority Voting, and the Properties of a Linear Income Tax, *Journal of Public Economics*, 4(2), pp. 163-185.
- Roberts, Kevin W.S. (1977), Voting over Income Tax Schedules, *Journal of Public Economics*, 8(3), pp. 329-340.
- Sadka, Efraim (1976), On Income Distribution, Incentive Effects and Optimal Income Taxation, *Review of Economic Studies*, 43(2), pp. 261-267.
- Saint-Paul, Gilles, and Thierry Verdier (1993), Education, Democracy and Growth, *Journal of Development Economics*, 42(2), pp. 399-407.
- Shea, John (2000), Does Parents' Money Matter? *Journal of Public Economics*, 77(2), pp. 155-184.
- Stanley, Marcus (1999), College Education and the Mid-century GI Bills, Mimeo, Harvard University.
- Stern, Nicolas (1982), Optimum Income Taxation with Errors in Administration, *Journal of Public Economics*, 17(2), pp. 181-211.
- Stiglitz, Joseph E. (1982), Self-Selection and Pareto Efficient Taxation, *Journal of Public Economics*, 17(2), pp. 213-240.
- Tamura, Robert (1991), Income Convergence in an Endogenous Growth Model, *Journal of Political Economy*, 99(3), pp. 522-540.
- Teulings, Coen N. (1995), The Wage Distribution in a Model of the Assignment of Skills to Jobs, *Journal of Political Economy*, 103(2), pp. 280-315.
- Teulings, Coen N. (2000), Aggregation Bias in Elasticities of Substitution and the Minimum Wage Paradox, *International Economic Review*, 41(2), pp. 359-398.
- Teulings, C.N. (2001), *Comparative Advantage, Relative Wages, and the Accumulation of Human Capital*, discussion paper TI 99-049/3, Tinbergen Institute.
- Teulings, Coen N., and Joop Hartog (1998), *Corporatism or Competition? Labour Contracts, Institutions and Wage Structures in International Comparison*, Cambridge: Cambridge University Press.
- Teulings, C.N. and T. van Rens (2001), *Education, growth and income inequality*, Tinbergen Institute/Princeton University.
- Tilak, Jandhyala B.G. (1989), *Education and its Relation to Economic Growth, Poverty and Income Distribution*, World Bank, Washington.

Trostel, Philip A. (1993), The Effect of Taxation on Human Capital, *Journal of Political Economy* 101(2), 327-350.

Trostel, Philip A. (1996), Should Education Be Subsidized?, *Public Finance Quarterly*, 24(1), pp. 3-24.

Ulph, D. (1982) untitled paper presented at the Econometric Society meeting 1982.