TI 2001-090/3
Tinbergen Institute Discussion Paper

Education and Efficient Redistribution

Robert A.J. Dur
Coen N. Teulings

Faculty of Economics, Erasmus University Rotterdam, and Tinbergen Institute.

Tinbergen Institute
The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Amsterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Please send questions and/or remarks of non-scientific nature to wdriessen@few.eur.nl.
Most TI discussion papers can be downloaded at http://www.tinbergen.nl.

Education and efficient redistribution

Robert A.J. Dur and Coen N. Teulings

Tinbergen Institute, Erasmus University Rotterdam

P.O. Box 1738, 3000 DR Rotterdam, The Netherlands.

E-mail: dur@few.eur.nl; teulings@few.eur.nl

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. 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, a subsidy to education is an ingredient of a second-best policy due to its general equilibrium effects on relative wages.

JEL codes: H21, H52, J24, J31.

First draft: September 2001. **This version: June 2003.**

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 therefore favor predominantly the high ability types, leading to a widening instead of a compression of the income distribution. If anything, education should therefore be taxed. In the spirit of Tinbergen (1975), this paper argues that general equilibrium effects may make education subsidies an optimal redistributive instrument. An increase in the mean level of human capital reduces the return to human capital, by a simple substitution effect. The supply of high-skilled workers goes up, reducing their relative wages, while the supply of low-skilled workers goes down, increasing their relative wages. Hence, the return to human capital and pre-tax wage inequality go down. When this indirect, general equilibrium effect of education subsidies 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.

The rationale for education subsidies in our model rests on an externality in individual schooling decisions. Individuals do not take into account the effect on the pre-tax distribution of income and, therefore, on distortions arising from progressive taxation. 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. A constrained Pareto efficient redistribution policy faces a trade off between the distortionary effect of progressive taxation and the distortions arising from education subsidies.

Hence, this 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. Mirrlees analyzes the trade off between the distortion caused by increasing marginal tax rates vis-a-vis the extra redistribution that can be achieved. 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 government only observes gross income and the years of education attained. Our set up differs in three important aspects from Mirrlees' analysis.

First, Mirrlees' (1971) 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. Since 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. 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. Our theory predicts a correlation between the progressivity of the income tax and the level of education subsidies. We present data which give some support to this hypothesis. The level of this correlation and the average level of education subsidies, 6 % of GDP, correspond surprisingly well with the predictions of the model for reasonable parameter values. Also, our model explains why cross country differences in the dispersion of disposable income are primarily due to differences in the dispersion of gross income, not to differences in the progressivity of the tax system.

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), Allen (1982), Stern (1982), and Stiglitz (1982) have analyzed this problem. 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 only two types of labor was applied.[1] Teulings (2000) shows that using only two instead of a continuum of types seriously understates the general equilibrium effects on relative wages of policies, in his example a change of the minimum wage. The intuition is that only the between-type relative wage effects are accounted for, while the within-type effects are ignored. Our claim is that the same problem applies for general equilibrium effects of an increase in the mean level of human capital, since large shifts in relative wages within each type are ignored by considering only two broad types. Instead, we use an assignment model in the spirit of Rosen

[1] See Johnson (1984) for a model with three types.

(1974), Sattinger (1975), and Teulings (1995), applying the Ricardian concept of comparative advantage.[2]

In the type of world described by these assignment models, the return to human capital is negatively related to its supply, see Teulings (2002). There is substantial evidence that an increase in the stock of human capital indeed reduces income dispersion. Tilak (1989) provides some early cross country evidence. Katz and Murphy (1992) provide evidence that high-skilled and low-skilled workers are imperfect substitutes in production. In addition, there are case studies for various countries, e.g. Goldin (1999), Hartog, Oosterbeek, and Teulings (1993), Edin and Holmlund (1995), and Kim and Topel (1995). 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 mean level of education of the workforce. Their estimate of the size of this effect is broadly consistent with elasticity of substitution reported by Katz and Murphy (1992).

Finally, our paper deviates from Mirrlees (1971) by restricting the attention to log-linear policy rules and a utility function that is additive in consumption and effort.[3] Mirrlees' great theoretical contribution was that his analysis derived

[2] There is an alternative way to evaluate the difference between two and a continuum of types. In Mirrlees' seminal paper, workers choose their level of effort from a continuum of alternatives, with the convenient characteristic of differentiability. Hence, the optimal choice satisfies a first-order condition. Analogously, in a world where workers also choose their level of human capital, it is derisable to let them choose from a continuum of alternative levels, so that we can apply marginal analysis. Though unusual in optimal taxation analyses with endogenous schooling, this necessarily implies that we have to use a production function with a continuum of types.

[3] A commentator suggested that this restriction to linear income policies would imply that these policies are no longer incentive compatible. This is incorrect. The log linearity of the income policy is only a restriction to the options available to the government. Each individual worker is free to choose whatever level of effort or human capital she wants to set. It just turns

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. On top of that, we adopt Diamond's (1998) idea of an additive utility function, thereby ruling out income effects in labor supply and leading to a log linear supply function. Similarly, we carefully specify the education production function such that it yields a log linear supply of human capital. These simplifications contribute to a key characteristic of our economy, namely, that the mean is a sufficient statistic for the distribution of human capital. As soon as we allow higher moments of the human capital distribution to be affected, no closed form solution of the assignment model is available. Obviously, log linearity is a serious limitation to the generality of the model. Nevertheless, compared to the use of two-type production functions, which has been the standard in optimal taxation analyses with endogenous schooling hitherto, we think that the log linearity is a reasonable simplification. Moreover, log linearity links the model directly to the empirical practice in labor economics. For example, the slope of our log linear wage function is equal to the Mincerian rate of return to human capital. This direct link to empirical studies enables us to come up with empirical estimates for the crucial parameters in our model so as to give an indication of the optimal size of education subsidies for redistributive reasons.

Several other arguments have been proposed in favor 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, 1993 and 1996). Similarly, education subsidies may help to remove distortions in human capital accumulation arising from progressive

out that her optimal response satisfies a log linear labor supply and human capital acquisition function.

income taxation (Bovenberg and Jacobs, 2001). Our model captures these arguments. In endogenous growth models, any investment has external effects to future generations (Lucas, 1988, Tamura, 1991, and Aghion and Howitt, 1998). The evidence for positive externalities is mixed, see for example Acemoglu and Angrist (1999), Bils and Klenow (2000), Krueger and Lindahl (1999), and Teulings and Van Rens (2002). Extending these arguments, Saint-Paul and Verdier (1993), Perotti (1993), and Benabou (1999) stress the role of capital market imperfections. When some groups have to borrow at rates above the market value, they will underinvest in human capital. There are two reasons why capital market imperfections can not fully account for the widespread prevalence of education subsidies. First, recent empirical studies cast 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 have a much broader character.

Our analysis calls for subsidies to all levels of education. This redistribution policy contrasts sharply with the usual idea of compressing the wage distribution via compression of the distribution of human capital, that is by putting special policy effort in raising the education of the least skilled. This latter policy, that relies on direct, partial equilibrium effects, might run into trouble due to adverse general equilibrium effects which are concentrated just above the bottom of the skill distribution. The empirical evidence supports these ideas. There is a strong negative relation between the first moment of the human capital distribution and the second moment of the wage distribution, but there seems to be no relation between the second moments of both distributions (Teulings and Van Rens, 2002). This points to promotion of education at all levels rather than at the low levels only.

The paper is organized as follow. Section 2 presents the model. Section 3 examines optimal redistribution policy. Section 4 provides some conclusions for policy making and some political economy extensions to the model.

2 The structure of the economy

Consider an economy populated with individuals who are born with ability level a, $a\tilde{}N(0,\sigma^2)$. 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: $s = s(h,a)$, $s_h(h,a) > 0$, $s_a(h,a) > 0$, $s_{ha}(h,a) \geq 0$. We ignore the cost of the education system itself, in line with the observation that this cost is of minor importance relative to the 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 log 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 labor supply μ. μ is determined endogenously in the model. Due to the constant returns to scale production 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. The model is structured as such that μ is a sufficient statistic for the skill distribution. Higher moments of the skill distribution are not affected by policy experiments. A convenient implication of this set up is that log wages are a linear function of s (see section 2.4):

$$w(s,\mu) = w_0(\mu) + w_s(\mu)\, s \tag{1}$$

where: $w_s(\mu) > 0, w_s'(\mu) < 0$. 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: as skills become more abundant in the economy, the return to skill falls.

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. The interest rate is exogenously determined at a global capital market. Hence, individuals can borrow funds at a rate $\lambda = \rho + \delta$, ρ for the interest payments and δ as an insurance premium that covers the loan at the moment that the 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.[4]

[4]The combination of a golden growth rule and a perfect international capital market implies

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:

$$d(y, h) = d_0 + d_y y + d_h h \tag{2}$$

Hence, the 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, we have: $d_0 = d_h = 0$ and $d_y = 1$, so that $d(y, h) = y$. The coefficient d_y equals one minus the marginal tax rate divided by one minus the average tax rate. It is Musgrave and Musgrave's (1973) coefficient of residual income progression; $d_y = 1$ 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

---

that we can ignore transition dynamics. E.g. suppose that the optimal policy requires an increase in the level of human capital. Hence, the loans for new investment in human capital exceed the interest payments on the outstanding debt in the transition phase. The economy can borrow additional funds on the capital market. This debt will never be serviced. However, since debt plus accumulated interest payments will remain a fixed fraction of GDP, this is not a problem.

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 and to have a constant marginal utility of consumption, compare Diamond (1998):

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

where $\pi > 1$ 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. Also 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:

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

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 the 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 the rate of dying, δ. In the first term, we use: $e_{\text{education}} = -\infty$. The individual's budget constraint reads:

$$e^{-\lambda h+d} = \left(1 - e^{-\lambda h}\right) e^{g} + e^{-\lambda h+d+b} \tag{4}$$
$$\text{implying } U = e^{d-\lambda h} - e^{-\lambda h+\pi e}$$

where the left-hand side of the first equation is discounted lifetime income and where the right-hand side is discounted lifetime consumption. The second equation uses the individual budget constraint to simplify the utility function. Since the individual's rate of time preference is equal to the interest rate, capital markets are perfect, and consumption and effort are separable in the utility function, only total discounted life time consumption matters; its distribution over life time is irrelevant.[5] Applying equation (2) and using $y \equiv w + e$, the first-order condition for optimal effort reads:

$$\frac{dU}{de} = d_y e^{d_0 + d_y(w+e) + (d_h - \lambda)h} - \pi e^{-\lambda h + \pi e} = 0 \tag{5}$$
$$\text{implying } U = \frac{\pi - d_y}{\pi} e^{d_0 + d_y(w+e) + (d_h - \lambda)h}$$

where we use the first-order condition to further simplify the utility function in the second equation. We can solve the first-order condition to obtain an expression for the optimal amount of effort for the individual, conditional on her wage rate, her choice of years of schooling, and the government's income policy parameters, d_0, d_y and d_h:

[5] In a previous version, instantaneous utility was characterized by a declining marginal utility of consumption:

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

where $0 < \theta < 1$. Then, constancy of consumption over the lifetime comes out as the optimum. However, θ did not play an important role in the analysis. Therefore, we simplified by setting it equal to unity.

$$
\begin{aligned}
e = & \frac{\ln\left(\frac{d_y}{\pi}\right) + d_0 + d_y w + d_h h}{\pi - d_y} \\
\equiv & \varepsilon_0 + \varepsilon_w w + \varepsilon_h h
\end{aligned}
\tag{6}
$$

The second-order condition requires: $\pi - d_y > 0$. 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 $\varepsilon_w = \frac{d_y}{\pi - d_y}$.

2.2 The education production function

The production function of education reads:

$$
s(a,h) = \alpha + a + (\beta h - \xi a) - \frac{1}{2}\psi(\beta h - \xi a)^2 \tag{7}
$$

where $\psi > 0$ and $0 \leq \xi < 1$. Before discussing the interpretation of the parameters, we first derive the optimal years of education. Solving the partial derivative of $s(a,h)$ with respect to h yields the demand for education of an individual with ability a:

$$
h = \frac{1}{\psi\beta^2}\left[\beta - s_h(a,h)\right] + \frac{\xi}{\beta}a \tag{8}
$$

This equation provides a road map for the interpretation of the parameters ξ and ψ. Holding constant $s_h(a,h)$, h is equal for all ability types when $\xi = 0$. All ability types take up the same years of education in that case. For $\xi > 0$, high ability types benefit more from schooling than low ability types and, hence, take up more education. The parameter ξ thus measures the degree of complementarity of ability a and years of schooling h. This parameter determines the direct effect of subsidies to education on the income distribution.

For the interpretation of the parameter ψ, it is useful to normalize the parameters α and β such that $s_h(a,h) = \beta$ and $\mu = 0$, and hence $w_s(\mu) = 1$ in the redistribution free equilibrium, $d_y = 1, d_h = 0$. As we show below, this normalization can be applied without loss generality. As $w_s(0) = 1$, the Mincerian rate of return to education, $\frac{dw}{dh}$, is equal to $s_h(a,h)$ in the redistribution free equilibrium. Hence:

$$\psi = -\frac{s_{hh}(a,h)}{s_h^2(a,h)} = -\frac{d\ln s_h(a,h)}{s_h(a,h)\,dh}_{|d_y=1,d_h=0}$$

The numerator is the relative change in the return to education. The denominator is the relative change in human capital, evaluated at market prices (i.e. the Mincerian rate of return $s_h(a,h)$). Hence, $1/\psi$ is the price elasticity of the demand for schooling.[6]

Individuals choose their years of education as to maximize their utility function as expressed in equation (5) subject to the optimal choice of effort (6) and the production function of education (7). The first-order condition can be simplified to yield:

$$s_h(a,h) = \frac{\pi(\lambda - d_h) - \lambda d_y}{\pi d_y w_s(\mu)} \tag{9}$$

The log linearity of income policy and linearity of the wage equation (1) in s therefore imply that the marginal return to a year of schooling, $s_h(a,h)$, is independent of a and h in equilibrium. We therefore drop its arguments. Since s_h is independent of a and h, so must be $\beta h - \xi a$. 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, $d_y = 1, d_h = 0, \mu = 0, w_s(0) = 1$. Hence, equation (9) implies:

$$s_h = \beta = \frac{\pi - 1}{\pi}\lambda \tag{10}$$

The constancy of $\beta h - \xi a$ yields an alternative interpretation of parameter ξ. 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

[6] This price elasticity is defined with respect to value of the human capital at market prices. This concept differs from the definition of the price elasticity of schooling usually applied in the literature, which reads:

$$\frac{d\ln h}{d\ln s_h(a,h)} = \frac{h^{-1}dh}{d\ln s_h(a,h)}$$

The difference is a factor $s_h(a,h)/h$. However, since the average years of education in developped economies is about 10 and since the Mincerian rate of return is about 10 %, both concepts happen to be about equal.

free equilibrium, 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. If $\xi = 0$, all individuals choose the same level of education and hence all variation in wages is due to differences in innate ability.

2.3 The skill distribution among labor supply

The equations (6), (7), (8), and (1) imply that a, e, and h, are linear functions of the skill level s:

$$\begin{aligned} a(s) &= -\frac{1}{2\psi\beta^2}\left(\beta^2 - s_h^2\right) - \alpha + s \equiv a_0 + s \\ e(s) &= \varepsilon_0 + \varepsilon_w w_0(\mu) + \varepsilon_h h_0 + \left[\varepsilon_w w_s(\mu) + \frac{\xi}{\beta}\varepsilon_h\right] s \equiv e_0 + e_s s \\ h(s) &= h_0 + \frac{\xi}{\beta} s \equiv h_0 + h_s s \\ h_0 &\equiv \frac{1}{2\psi\beta^2}(\beta - s_h)\left[2 - \frac{\xi}{\beta}(\beta + s_h)\right] - \frac{\xi}{\beta}\alpha \end{aligned} \tag{11}$$

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 characteristic depends crucially on the log-linearity of the income policy, see equation (2). Using the normality of the distribution of innate ability, $a\tilde{\ }N(0, \sigma^2)$, we can derive the density function $f(s)$ of the skill distribution measured in units of effort:

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

where $\phi(\cdot)$ is the standard normal density function. The factor $\frac{1}{\sigma}\phi\left(\frac{a(s)}{\sigma}\right)$ is the density function of innate ability $a(s)$. The factor $\frac{da(s)}{ds}$ is the Jacobian for the transform of variable. By equation (11), 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 skill type 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.[7] Substitution of equation (11) and rearranging terms yields:

$$\begin{aligned} f(s) &\equiv \frac{1}{\sigma}\phi\left(\frac{s-\mu}{\sigma}\right)\exp(L) \\ \mu &= -a_0 + \sigma^2 (e_s - \lambda h_s) \\ L &= e_0 - \lambda h_0 - a_0 (e_s - \lambda h_s) + \frac{1}{2}\sigma^2 (e_s - \lambda h_s)^2 \end{aligned} \tag{12}$$

where $\exp(L)$ measures the size of the labor force weighted by units of effort and where $\frac{1}{\sigma}\phi\left(\frac{s-\mu}{\sigma}\right)$ measures its composition. The distribution of skill among labor supply is also normal, with the same variance σ^2 as the underlying ability distribution and with mean μ. The first term of the relation for μ captures the overall shift of skill relative to ability ($-a_0$ being the 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. Aggregate log labor supply in units of effort is equal to L. It is increasing in the effort of the median worker, e_0, and decreasing in her years of schooling, h_0.

We are now in a position to derive an expression for α. Since $s_h = \beta$ in the redistribution free equilibrium, equation (11) implies: $\alpha = -a_0$. Furthermore, since $\mu = 0$ in the redistribution free equilibrium, equation (12) implies: $a_0 =$

[7] The density of age x among the workforce is $\lambda e^{-\lambda x}$: 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 share of individuals of type s that works (that is: $x > h(s)$) is equal to:

$$\int_{h(s)}^{\infty} \lambda e^{-\lambda x} dx = e^{-\lambda h(s)}$$

$\sigma^2(e_s - \lambda h_s)$. Then, by equations (6), (8), and (10):

$$\alpha = -\sigma^2 \frac{1-\xi\pi}{\pi-1} \tag{13}$$

2.4 Production technology and wages

The production technology that we apply is based on Rosen (1974), Sattinger (1975), and Teulings (1995). These papers consider the assignment of workers to tasks in an economy where both are heterogeneous and where all markets are perfectly competitive. The production of one unit of output requires the input of an infinite number of tasks, indexed by their level of complexity, c. The price of a unit of output is taken as the numeraire and hence normalized to unity. Like the skill level s, c varies continuously and can take any real number.[8] The transformation of tasks into output takes place by a Leontieff technology: tasks are required in fixed proportions.[9] The input requirements of c-type tasks per unit of output are described by a normal distribution, with the same variance as the skill distribution, $c\tilde{}N(0,\sigma^2)$. A c-type task can be produced by any s-type worker. However, the relative productivities of various worker types differ

[8] The production technology uses the single index assumption: all worker characteristics can be aggregated 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. Imperfect substitution between ability and education would diminish the effectiveness of education subsidies in reducing inequality, since the return to ability depends less on the economy's stock of human capital, the lower the substitutability of ability and schooling in production. Empirical evidence suggests the single index assumption to be a reasonable description of reality, see Teulings (2000).

[9] Commentators suggested that the assumption of Leontieff technology is crucial for our results. That is not the case. Teulings (2000) shows that replacing the Leontieff technology by a Cobb Douglas technology is (almost) equivalent to halving the value of parameter γ that is introduced below. The advantage of using the Leontieff technology is that the differential equation (14) can be solved analytically.

according to the complexity of the task:

$$g(s,c) = -\frac{1}{\gamma}e^{\gamma(c-s)}$$

where $g(s,c)$ is the log productivity of skill type s in a task of complexity c. This specification implies comparative advantage of high skilled workers in complex jobs, since $g_{sc} > 0$: the productivity ratio of type s_1 and s_2, $s_2 < s_1$, is increasing in c. Teulings (1995) shows that this set up implies that every task type c is uniquely assigned to a single worker type $s(c)$, and vice versa. Furthermore, better skilled workers are assigned to more complex jobs, $s'(c) > 0$, due to comparative advantage. The equilibrium of supply and demand for each task type c requires, in logs:

$$L - \left(\frac{s(c)-\mu}{\sigma}\right)^2 + g[s(c), c] + \ln s'(c) = -\left(\frac{c}{\sigma}\right)^2 + Y \tag{14}$$

where Y is log output. The left hand side is the log supply of labor of type $s(c)$ (the log of the normal density function) plus its log productivity in task type c plus the log Jacobian $\frac{ds}{dc} = s'(c)$. The two terms on the right hand side measure the log demand for task type c: the Leontieff coefficient (again the log of the normal density) plus log output. Equation (14) is a differential equation in $s(c)$. The special case, where the variances of the skill distribution and the complexity demand distribution are equal, is the only for which this differential equation has an analytical solution:

$$\begin{aligned} s(c) &= c + \mu \qquad (15) \\ Y &= L - \frac{1}{\gamma}e^{-\gamma\mu} \end{aligned}$$

For the derivation of the (unique) wage equation that is consistent with this assignment, we define $x(s,c)$ to be the log production cost of c-type tasks by a s-type worker:

$$x(s,c) = w(s,\mu) - g(s,c) \tag{16}$$

that is, log production cost in log wages per worker minus log productivity per worker. In a market equilibrium, workers are assigned to tasks such that production cost for task c is minimized. The first-order condition reads:

$$\begin{aligned} x_s\left[s(c),c\right] &= w_s\left[s\left(c\right),\mu\right]-g_s\left[s\left(c\right),c\right]=0\Longrightarrow \\ w_s\left(\mu\right) &= g_s\left[c-\mu,c\right]=e^{-\gamma\mu} \end{aligned} \tag{17}$$

where we leave out the argument s of w_s since w_s is independent of s. The special case of equal variances for skill supply and complexity demand is the only case for which $w\left(s,\mu\right)$ is linear in s, as has been posited in equation (1). An increase in the mean skill level by $d\mu$ reduces the return to skill by $\gamma w_s\left(\mu\right)d\mu$. The intuition for this relation is that an increase in μ raises $s\left(c\right)$ for a given c, or equivalently, it reduces c for a given s: equally skilled workers end up in less complex jobs, simply because the demand for complex tasks is limited. Since the marginal productivity of skill $g_s\left(s,c\right)$ is lower in less complex jobs, $g_{sc}\left(s,c\right)>0$, this reduces the return to skill.

The parameter γ plays a crucial role in our analysis. In the redistribution free equilibrium, where $\mu=0$ and hence $w_s\left(\mu\right)=1$, it is equal to the compression elasticity, which is defined as the percentage decline in the return to skill $w_s\left(\mu\right)$ per percent increase in the supply skill of skill, evaluated at its market prices $w_s\left(\mu\right)$:

$$\gamma=\frac{d\ln w_s\left(\mu\right)}{w_s\left(\mu\right)d\mu}_{|d_y=1,d_h=0}$$

The numerator of the right hand side is the relative reduction in the return to skill, $w_s\left(\mu\right)$. The denominator is relative increase in the stock of skill, evaluated at its market price $w_s\left(\mu\right)$. Alternatively, where $1/\psi$ measures the price elasticity of the supply of human capital, $1/\gamma$ measures the elasticity of its demand.

The intercept $w_0\left(\mu\right)$ remains to be determined. This is derived from the numeraire. Production cost per task weighted by their Leontieff coefficient add up to unity:

$$1 = \int_{-\infty}^{\infty} \frac{1}{\sigma}\phi\left(\frac{c}{\sigma}\right) e^{x[s(c),c]} dc$$
$$= \int_{-\infty}^{\infty} \frac{1}{\sigma}\phi\left(\frac{c-\sigma^2 e^{-\gamma\mu}}{\sigma}\right) e^{w_0 + \left(\mu+\gamma^{-1}\right)e^{-\gamma\mu} + \frac{1}{2}\sigma^2 e^{-2\gamma\mu}} dc$$

where the second line follows from the substitution of equation (15), (16), and (17), and some rearrangement. Hence:

$$w_0(\mu) = -\left(\mu+\gamma^{-1}\right)e^{-\gamma\mu} - \mu e^{-\gamma\mu} - \frac{1}{2}e^{-2\gamma\mu}\sigma^2 \tag{18}$$

By the linearity of $w(s,\mu)$ in s, the normality of the distribution of s caries over to the distribution of log wages w, which fits the data well.

2.5 Government budget constraint

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

$$e^Y = \int_{-\infty}^{\infty} \frac{1}{\sigma}\phi\left(\frac{s+a_0}{\sigma}\right) \exp\left[d_0 + d_y\left\{w(s)+e(s)\right\} + (d_h-\lambda)h(s)\right] ds \tag{19}$$

The left hand side is total output, the right hand side is the sum of all income transfers. The final term $-\lambda h(s)$ accounts for the fact that individuals do not get any government support as long as they are still at school.

3 Optimal income policy

The model can be conveniently written in the following form:[10]

$$u(d_y, d_h, a) = u_0(d_y, d_h) + u_a(d_y, d_h)\, a \tag{20}$$

where $u(\cdot) \equiv \ln U(\cdot) + \ln \pi$. The expressions for $u_0(\cdot)$ and $u_a(\cdot)$ are given in the Appendix. They are quite complicated and non-linear. Moreover, there is only an implicit expression for $u_a(\cdot)$. However, the reduced system has three convenient

[10]Substitution of the equations (11) for a, e and h, (1) and (2) for w and d, (18) and (17) for $w_0(\mu)$ and $w_s(\mu)$, the output equation (15), and the government's budget constraint (19) in the utility function, equation (3), and applying the relations for α and β.

features. First, the equation is linear in a. This allows a simple interpretation of the parameters $u_0(\cdot)$ and $u_a(\cdot)$; $u_0(\cdot)$ measures the utility of the median voter ($a = 0$), $u_a(\cdot)$ measures inequality in welfare. Hence, we can think of both parameters as the goals of the policy maker. Second, the parameter d_0 drops out. This result is due to the budget constraint of the 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 of d_0 follows from the budget restriction. Hence, we only have to consider the policy parameters d_y and d_h. The parameters d_y and d_h are the instruments that the policy maker has available to attain its goal(s). Third, the model has an analytical solution in the redistribution free equilibrium, $d_y = 1, d_h = 0$, due to our convenient normalizations of α and β. It reads:

$$\begin{aligned} u_0(1,0) &= -\frac{1}{\pi-1}\ln\pi + \ln(\pi-1) \\ &\quad -\frac{1}{2}\sigma^2\frac{\pi}{(\pi-1)^2}(1-2\xi)(\pi+1-2\xi) \\ u_a(1,0) &= \frac{\pi}{\pi-1}(1-\xi) \end{aligned} \tag{21}$$

These features enables us to derive analytical results for the optimal policy mix for income redistribution, starting from the redistribution free equilibrium.

Given our restriction to log linear income policies, see equation (2), the condition for a constrained Pareto efficient redistribution policy is that it maximizes $u_0(\cdot)$, taking $u_a(\cdot)$ as given, and using the parameters of the income policy, d_y and d_h, as instruments.[11] That is, we characterize for any feasible level of inequality $u_a(\cdot)$, the mix of policy instruments d_y and d_h that minimizes efficiency losses, i.e. maximizes $u_0(\cdot)$. 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 in-

[11] It is well concievable that there is non-log linear income policy that Pareto dominates all log linear policies, but that issue is beyond the scope of this paper.

come policies, voters preferences' are single peaked, so the median voter theorem is applicable.

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 with complementarity ($\xi > 0$).

3.1 The case without complementarity, $\xi = 0$

In the redistribution free equilibrium $d_y = 1, d_h = 0$, a subsidy to education contributes to redistribution:

$$u_{a,dh}(1,0) = -\frac{\gamma\pi^3}{(\pi-1)(\psi+\gamma)+\gamma\psi\sigma^2} < 0 \tag{22}$$

where the second subscript refers to the relevant partial derivative. The derivation is in the Appendix. When skill types are imperfect substitutes in production ($\gamma > 0$), the resulting increase in the stock of human capital in the economy implies a decrease in the return to human capital and, hence, a reduction in the dispersion of wages. Hence, the dispersion of welfare decreases.

Proposition I: Efficient redistribution without complementarity

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

$$\begin{aligned} d_h &= \lambda(1-d_y) + \frac{\gamma\beta(1-d_y)d_y^2}{u_a(\pi-d_y)+\gamma d_y(1-d_y)} \\ u_a &= u_a(d_y, d_h) \end{aligned} \tag{23}$$

Proof:

See Appendix.

□

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, education should be subsidized if income taxes are progressive: $d_h = \lambda (1 - d_y)$. This argument for education subsidies is discussed by Bovenberg and Jacobs (2001): education should be subsidized to offset the disincentive effects of the increase in the marginal tax rate due to the higher income. Whether or not this effect applies is therefore contingent on the functional form of the tax scheme. For example, it does not apply for a linear instead of a log linear scheme, since then marginal rates are constant.

In the general case where $\gamma > 0$, the second term in equation (23) is positive if $d_y < 1$. 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. The distortion in the schooling decision due to the education subsidy is traded off against the distortion in the effort decision due to marginal tax rates. The optimal redistribution policy mixes both distortions, in line with the principles of tax smoothing. The higher the compression elasticity γ, the stronger the compression of relative wages by additional investment in human capital, and hence the higher is the optimal value of d_h. The higher the price elasticity of effort, $\frac{d_y}{\pi - d_y}$, the higher the distortion caused by marginal tax rates, and hence the higher is the optimal value of d_h. Lastly, note that the price elasticity of the demand for schooling, ψ^{-1}, does not show up in this equation. Since the schooling decision is distorted by both progressive taxation and subsidies to education, the elasticity (measuring the size of the welfare loss) does not affect the ratio between income taxes and subsidies to education.

The problem of equation (23) is that u_a depends on d_y and d_h, which makes it impossible to assess the overal effect of d_y on d_h. In the redistribution free equilibrium, we can use the explicit solution for $u_a(1,0)$ to simplify equation (23):

$$\frac{d_h}{\lambda} = \left[1 + \frac{\pi - 1}{\pi^2}\gamma\right](1 - d_y) \tag{24}$$

Since λ measures the cost of an additional year of education relative to consumption, d_h measures the subsidy for a year of education relative to consumption, so d_h/λ measures the subsidy rate for the cost of education. The two terms within square brackets reflect the two reasons for education subsidies: the distortionary effect of increasing marginal tax rates and the general equilibrium effect of education subsidies on wages. The latter term depends on two parameters, π and γ, just as in (23).

The model allows a crude calculation of the optimal level of subsidies to education for redistributive purposes (we ignore the first reason for education subsidies in our calculations). 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 $\pi \cong 3$. Katz and Murphy's (1992) and Teulings and Van Rens's (2001) estimates imply a compression elasticity $\gamma \cong 2$. In order to get a feeling for what this value implies, suppose that the initial Mincerian rate of return to education is 10 %. Then, a one year increase in the average level of education increases the value of the stock of human capital by 10 %. A compression elasticity of 2 implies that this induces a decline of the return to human capital by $2 \times 10\% = 20\%$, that is from 10 % to 8 %. The mean value of income tax progression in OECD countries $1 - d_y \cong 0.15$ (OECD, 1997). Hence, imperfect substitution justifies a subsidy to education of approximately $\frac{2}{9} \times 2 \times 0.15 = 7\%$ of its total cost (in terms of foregone labor income). Subsidies as a share of GDP should be equal to this subsidy rate times the total cost of education, where the latter are equal to the labor share times the average years of education times the Mincerian rate of return. Suppose the average years of education to be 10 years, the Mincerian rate of return to be 10 % and the labor share to account for $\frac{2}{3}$ of GDP. Then, education subsidies should be equal to $\frac{2}{3} \times 10 \times 0.10 \times 7\% = 4.4\%$. Moreover, the optimal subsidy rate is increasing in the progressivity of the tax system. When we follow Becker (1983) and interpret our model as a positive theory of the policy mix used for redistribution, the model predicts that countries with a stronger progressivity of the tax system spend more on education. Figure 1, taken from Van Ewijk and Tang (2000), provides some evidence. There is a clear negative relation between d_y and d_h. Remarkably, the level of subsidies to education matches our crude rule of thumb closely: when $d_y = 1$, subsidies to education should be zero, when $d_y = 0.85$, subsidies to education should account for 4.4 % of GDP.

Proposition II: Median voter preferences without complementarity

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

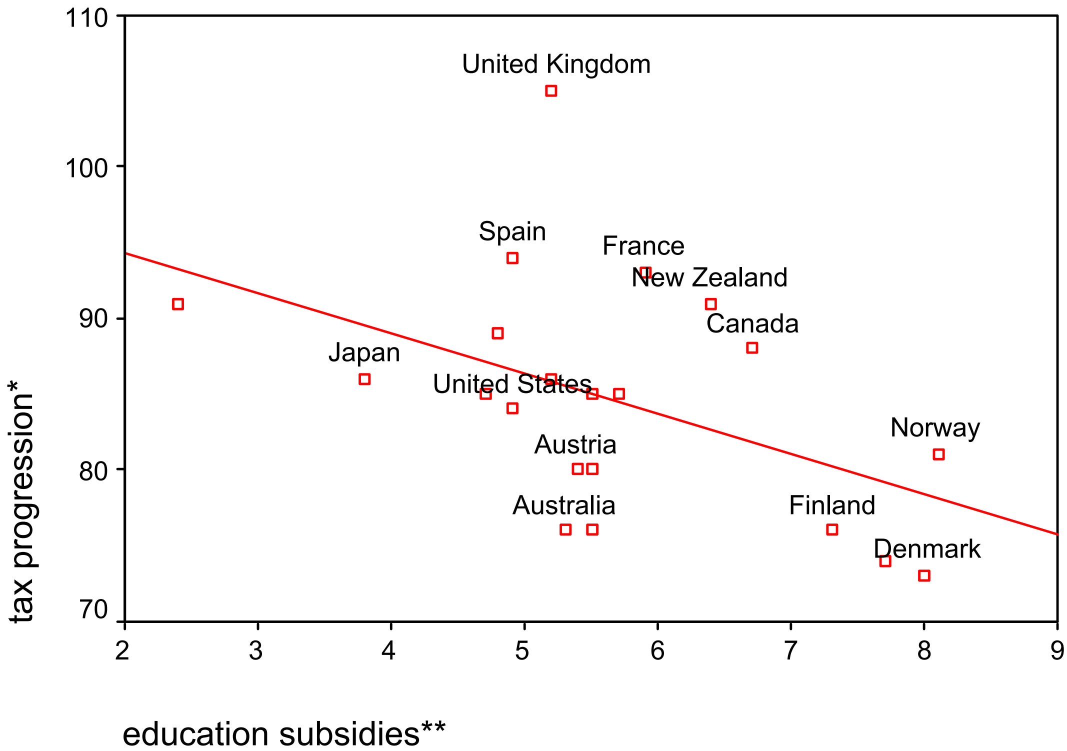
$$d_y (1 - d_y) \left(\pi + \gamma\sigma^2 u_a\right) = \sigma^2 u_a^2 \pi (\pi d_y - 1) \tag{25}$$

Proof:

See Appendix.

□

Due to the nonlinearity of the expression for u_a, we do not obtain an explicit solution for d_y. However, equation (25) 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.



* change in the after-tax wage, % of change in the before-tax wage.

** public expenditure, % of gross domestic product, in 1994.

Figure 1: Tax progression and education subsidies in OECD countries

3.2 The case with complementarity, $\xi > 0$

First, we analyze the redistributive effects of progressive income taxation and education subsidies in the redistribution free equilibrium separately:

$$u_{a,dy}(1,0) = \frac{\pi^2}{\pi - 1}\frac{1}{\psi(\pi - 1) + \gamma(\pi - 1 + \psi\sigma^2)} > 0 \tag{26}$$
$$u_{a,dh}(1,0) = \frac{\pi}{\beta}\frac{\xi\psi - (1-\xi)\gamma}{(\gamma + \psi)(\pi - 1) + \gamma\sigma^2\psi} \gtrless 0$$

The derivation is in the Appendix. Progressive income taxation, $d_y < 1$, reduces utility dispersion unequivocally. The effect of education subsidies, however, depends on parameter values. 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 disproportionally favor the people with high ability. Hence, the complementarity of education and ability may cause education subsidies to increase the dispersion of utility. From (26) it follows that education subsidies are a redistributive instrument if and only if:

$$\xi\psi < (1-\xi)\gamma \tag{27}$$

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.

Proposition III: Efficient redistribution with complementarity

When $\xi > 0$, the value for $\frac{d_h}{\lambda}$ for a constrained Pareto efficient log linear income policy in the neighborhood of the redistribution free equilibrium is characterized by:

$$\frac{d_h}{\lambda} = \left(1 + \frac{\pi - 1}{\pi^2} \frac{(1-\xi)\,\gamma - \xi\psi}{1-\xi}\right)(1 - d_y)$$

Proof:

See Appendix.

□

Proposition III shows that if (27) holds, and $d_y < 1$, the government should subsidize education for redistributive purposes.[12] If the condition does not hold, and $d_y < 1$, a tax on education is part of an optimal redistribution policy. Whether or not the condition holds depends on the values of γ, ξ, ψ. Clearly, the optimal subsidy increases in γ and decreases in ξ. A high elasticity of schooling, ψ^{-1}, makes it more likely that subsidies to education are optimal. The larger is the elasticity of schooling, the lower are the subsidies that are required for a given increase in the mean level of human capital, and hence the smaller the adverse effect of subsidizing education on the income distribution. Estimates of the elasticity of demand for years of schooling with respect to its cost vary from 0.2 to 0.8 (Lesley and Brinkman, 1987, Kane, 1995, Stanley, 1999), leading to values of ψ from 1.25 to 5. In Dur and Teulings (2003), we report empirical evidence for the UK, the US, and The Netherlands which suggests a value of ξ of about 0.3 to 0.6. Using these parameter values, and again a value of 2 for the compression

[12] The median voter equilibrium can be derived in the same way as for the case $\gamma > 0$, $\xi = 0$. However, the formulas do not yield further insights and, therefore, we do not present them here.

elasticity γ, it is clear that whether the condition holds is sensitive to the exact parameter values assumed.

4 Concluding remarks

The general equilibrium effect of investment in human capital on relative wages provides a forceful argument for the subsidization of education for a government that wants to redistribute income. 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 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. An efficient redistribution policy should therefore combine progressive income taxation and subsidies to the formation of human capital. Crude calculations suggest that this model provides a rationale for subsidies to the education system of about the level that we observe empirically. Moreover, the model suggests positive cross country relation between the progressivity of income taxes and the rate of subsidization of the education system: the more redistributive a country's income policy, the higher will be both the progressivity of the tax system and the subsidy to education system. This relation is also born out by the data, with a slope that fits the theoretical predictions closely.

In the absence of complementarity between ability and education, a subsidy to education is always part of an optimal progressive redistribution policy, irrespective of the exact values of the elasticities of supply and demand for human capital. The reason is that the distortion of the schooling decision due to education subsidies is a second order effect, similar to the distortion of the effort decision due to progressive taxation. Hence, an optimal policy uses a bit of both instruments. However, the complementarity of ability and education is an empirically relevant phenomenon. Then, subsidies to education also have a degressive effect, since these subsidies favor the high ability / high income types. We have derived a precise and fairly intelligible condition for the progressive general equilibrium effect to dominate the degressive income effect. This condition involves only three parameters, namely the price elasticity of supply and of demand of human capital and the share of wage dispersion that is attributable to the complementarity of ability and education. In his discussion of our paper, Heckman (2003) argues that the supply of human capital is insufficiently price elastic to make subsidies to education a worthwhile policy. Our overview of empirical studies (see Dur and Teulings, 2003, for a more detailed discussion) suggests that both effects tend to cancel. How should we interprete this conclusion from a policy perspective? We offer two lines of reasoning.

A first line of reasoning argues that apart from the two arguments on the desirability of education subsidies discussed before, our model offers a third argument, which calls for subsidies to education. Increasing marginal tax rates discourage investment in human capital, since these investment reduce the return on effort. A subsidy to education can offset this distortion in the human capital acquisition, see Bovenberg and Jacobs (2001). Where the first two arguments more or less cancel, the third argument shifts the balance in favor of education subsidies. One could argue that this latter argument is not that strong, since it relies on increasing marginal tax rates. It is easy to design progressive income policies with constant marginal tax rates, like e.g. the negative income tax. Our analysis offers little guidance here, since its conclusions are contingent on the log linear structure of the income policy, which necessarily implies increasing marginal tax rates. We do not know whether the increasing marginal tax rate is part of a constrained Pareto efficient policy mix, and hence we do not know whether this argument for education subsidies would still apply when we allow for non-linearities in the log income policies. The increasing marginal tax implied by the log linearity offends the logic of the Sadka (1976) argument for low marginal rates at both ends of the income distribution. Interestingly, this argument can be extended towards 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 cause efficiency losses, here the substitution effects contribute to redistribution while the income effects work in 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. While high subsidies at the bottom fit the layman's intuition, its counterpart is more surprising. A subsidy for top education programs has little adverse income effects (since there are not many people taking up more years of education), while it raises the average level of education. The production function applied in this paper implies that all lower ability types will benefit from the general equilibrium effects of this policy, see Teulings (2002). We leave this issue for future research.

A second line of reasoning goes beyond the simple unconditional subsidy scheme that is considered in this paper. One can think of more sophisticated schemes, that entail larger substitution effects of education subsidies while at the same time smaller adverse income effects. We offer some suggestions. First, only the cost of education due to foregone earnings are analyzed in this paper, keeping the quality of the education system fixed. One could extend the analysis to the trade off between the quality and the direct cost of the education system. That would introduce an additional margin of substitution. 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 raise the average skill level in the economy. Another option is to include intergenerational information in the subsidization scheme. The social economic status of parents is a good indicator for the expected educational attainment of their children, partly by nature effects, partly by nurture, see Plug and Vijverberg (2002). Including intergenerational information might shift the balance between adverse income effects of education subsidies and the desired substitution effects in favor of the latter. In the practice of income policy, this boils down to subsidies that are conditional on parental income, an institution that is widely applied in practice. Again, this is an issue that deserves further research.

The analysis of the optimal functional form of taxes and education subsidies has strong policy implication for programs like the EITC and the New Deal, along the lines suggested by Heckman, Lochner, and Taber (1999). 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 will 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. This points to the need of a more formal analysis of the functional form of the optimal policy.

Lastly, the contrained Pareto efficiency of education subsidies does not guarantee their political viability. In a previous version of this paper, we have studied two issues (Dur and Teulings, 2001). First, we consider the level at which a decision on the education subsidy has to be made to achieve optimality. Decision making has to be sufficiently centralized. Decentralization yields too low subsidies as the general equilibrium effect of education subsidies on wage dispersion is not taken into account. 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 case 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 political economy 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. We show that when the median voter lacks an instrument to credibly commit to a future income policy, she is able to capture only half of the potential gains from the indirect effect of human capital formation for redistribution.

5 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.

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.

Bovenberg, A. Lans, and Bas Jacobs (2001), Redistribution and Education Subsidies Are Siamese Twins, CEPR Discussion Paper No. 3309.

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.

Dur, Robert A.J., and Coen N. Teulings (2001), Education and Efficient Redistribution, CESifo Working Paper No. 592, Munich.

Dur, Robert A.J., and Coen N. Teulings (2003), Are Education Subsidies an Efficient Redistributive Device?, Tinbergen Institute Discussion Paper No. 03-024/3, Amsterdam/Rotterdam.

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.), *Labor 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.

Heckman, J.J. (2003), The Supply Side of the Race between Demand and Supply: Policies to Foster Skill in the Modern Economy, *De Economist*, 151(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 No. 6384.

Johnson, George E. (1984), Subsidies for Higher Education, *Journal of Labor Economics*, 2(3), pp. 303-318.

Kane, Thomas J. (1995), Rising Public College Tuition and College Entry: How Well Do Public Subsidies Promote Access to College?, NBER Working Paper No. w5164.

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.

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

Leslie, L.L., and P.T. Brinkman (1987). Student Price Response in Higher Education: The Student Demand Studies, *Journal of Higher Education*, 58(2), pp. 181-203.

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

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

Musgrave, Richard A., and Peggy B. Musgrave (1973), *Public Finance in Theory and Practice*, New York: McGraw-Hill.

OECD (1997), *Employment Outlook,* Paris.

Perotti, Roberto (1993), Political Equilibrium, Income Distribution, and Growth, *Review of Economic Studies*, 60(4), pp. 755-776.

Plug, Erik, and Wim Vijverberg (2002) Schooling, Family Background, and Adoption: Is it Nature or is it Nurture?, *Journal of Political Economy*, forthcoming.

Rosen, S. (1974) Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition, *Journal of Political Economy*, 82(1), pp. 34-55.

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.

Sattinger, M. (1975) Comparative Advantage and the Distribution of Earnings and Abilities, *Econometrica*, 43(3), pp. 455-468.

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, Coen N. (2002), Comparative Advantage, Relative Wages, and the Accumulation of Human Capital, Tinbergen Institute Discussion Paper No. 02-081/3, Amsterdam/Rotterdam.

Teulings, Coen N., and Joop Hartog (1998), *Corporatism or Competition? Labor Contracts, Institutions and Wage Structures in International Comparison*, Cambridge: Cambridge University Press.

Teulings, Coen N. and Thijs Van Rens (2002), Education, Growth and Income Inequality, Mimeo, Tinbergen Institute/Princeton University.

Tilak, Jandhyala B.G. (1989), *Education and its Relation to Economic Growth, Poverty and Income Distribution*, World Bank, Washington.

Tinbergen, Jan (1975). *Income Distribution: Analysis and Policies.* Amsterdam: North Holland.

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, Alistair, and David Ulph (1982), untitled paper presented at the Econometric Society meeting 1982.

Van Ewijk, C. and P.J.G. Tang (2000), Efficient Progressive Taxes and Education Subsidies, CPB Research Memorandum 170, The Hague.

6 Appendix

The budget constraint of the government (19) can be rewritten as follows:

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

Taking logs and substitution of equation (15) yields:

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

Substituting the government budget constraint and (1) and (2) for w and d, (18) and (17) for $w_0\left(\mu\right)$ and $w_s\left(\mu\right)$, and the output equation (15) in the utility function (3) yields:

$$
\begin{aligned}
u_0 &= e_0 - a_0\left(e_s - \lambda h_s\right) + \frac{1}{2}\sigma^2\left(e_s - \lambda h_s\right)^2 \\
&\quad - \frac{1}{2}\sigma^2\left[d_y\left(e^{-\gamma\mu} + e_s\right) + \xi\left(d_h - \lambda\right)\right]^2 \\
&\quad - \lambda h_0 + \ln\left(\pi - d_y\right) + \frac{1}{\gamma}\left(1 - e^{-\gamma\mu}\right) \\
u_a &= d_y\left(e^{-\gamma\mu} + e_s\right) + h_s\left(d_h - \lambda\right)
\end{aligned} \tag{28}
$$

where u_0 and u_a are defined by (20).

Substitution of the equations (11) for a, e, and h, (6) for $\varepsilon_0, \varepsilon_w$, and ε_h, (12) for μ, (10) for β, and (13) for α, in the equations (9) and (28) yields after some rearrangement:

$$\begin{aligned}
u_0 =& \frac{1}{\pi-1}\ln\frac{d_y}{\pi}+\ln\left(\pi-d_y\right)-\frac{\pi}{\pi-1}\frac{1-s_{h/\beta}}{\psi} \\
&+\frac{\pi}{\pi-1}\frac{1}{\gamma}\left(1-\frac{u_a}{1-\xi s_{h/\beta}}\frac{\pi-d_y}{\pi d_y}\right) \\
&-\frac{1}{2}\sigma^2\left(\frac{\pi+1}{\pi}u_a^2-\xi\frac{2u_a+\xi\pi}{\pi-1}\right) \\
F\left(d_y, s_{h/\beta}, u_a\right) =& -\ln u_a+\ln\left(1-\xi s_{h/\beta}\right)-\ln\left(\frac{\pi-d_y}{\pi d_y}\right) \\
&-\frac{\gamma}{2\psi}\left(1-s_{h/\beta}^2\right)-\gamma\sigma^2\frac{\left(\pi-1\right)u_a-\pi\left(1-\xi\right)}{\left(\pi-1\right)\pi} \\
=& 0 \\
s_{h/\beta}\equiv& \frac{s_h}{\beta}=\frac{\pi\left(\lambda-d_h\right)-\lambda d_y}{\beta\left(\pi-d_y\right)u_a+\xi\left[\pi\left(\lambda-d_h\right)-\lambda d_y\right]}
\end{aligned} \tag{29}$$

The equation $F\left(d_y, s_{h/\beta}, u_a\right)=0$ is an implicit equation in u_a. Substitution of the third equation for $s_{h/\beta}$ in the equations for u_0 and $F(\cdot)$ and solving $F(\cdot)$ for u_a yields a system of the form (20).

Derivation of equation (22)

Substituting the expression for $s_{h/\beta}$ into $F(\cdot)$, setting $d_y=1$, totally differentiating to u_a and d_h, and simplifying using the expression for u_a in the redistribution free equilibrium given by (21), results in equation (22) in the text.

Proof of Proposition I:

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 (29), 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 $s_{h/\beta}$:

$$u_{0,dy}F_{sh\beta}=u_{0,sh\beta}F_{dy}$$

where the subscripts refer to the relevant partial derivatives.

□

Proof of Proposition II:

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

$$u_{0,ua} F_s = u_{0,sh\beta} F_{ua}$$

□

Derivation of equation (26)

Substituting the expression for $s_{h/\beta}$ into $F(\cdot)$, totally differentiating to u_a and d_y (d_h, respectively), setting $d_h = 0$ and $d_y = 1$, substitution of the expression for u_a in (21), and some simplification yields equation (26) in the text.

Proof of Proposition III:

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 yields:

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

Taking the limit $d_y \to 1$, and using $u_a(1,0) = \frac{\pi}{\pi-1}(1-\xi)$ yields Proposition III.

□