Education vs. Optimal Taxation: The Cost of Equalizing Opportunities

Eric Stephens
Department of Economics, Queen’s University

December 31, 2008

Abstract

This paper analyzes the use of education spending as a redistributive tool. Individuals generally differ in their innate talents as well as the location they attend school, both of which affect the accumulation of human capital. If governments can target education funds to specific neighbourhoods/regions (i.e. low productivity ones), should they do so? The results suggest that educational transfers are inferior to money transfers as a tool to equalize utilities in a standard consumption/leisure framework. This implies that policies designed to equalize productivities must be justified on different grounds, for example the importance of self-esteem. Further, we show that even if “equalizing opportunities” is deemed optimal in the static problem, it may not be a reasonable policy goal when we extend the analysis to include dynamics. This is true because individuals are heterogeneous and because they receive benefits from living in rich areas. Once educated and earning, people segregate according to income and policy-makers face the same problem next period. Previous education spending decisions influence current human capital accumulation and an equal opportunity policy is seen to be too extreme, as it leads to lower outcomes in the next generation for all socio-economic groups.

Keywords: Optimal non-linear taxation, redistribution, equality of opportunity

JEL Classification Numbers: D63, D82, H21, I22, I28

* I am grateful to Robin Boadway, Sumon Majumdar, Charles Beach, Afrasiab Mirza, Adnan Khan, David Byrne and James Thompson for helpful comments and discussions as well as and seminar participants at Queen’s University and the CEA meetings 2008.

† Address: Dunning Hall, 94 University Ave., Kingston, ON K7L 3N6, Canada, Telephone: 1-613-533-6000 ext. 75955, E-mail: stephenss@econ.queensu.ca.
“Men are not born equal, they are not born free; they are born a most various multitude enmeshed in an ancient and complex social net.” [H.G Wells, *Outline of History*]

1 Introduction

What defines optimal policy depends crucially on the rule by which various outcomes are appraised. Recently, the concept of equality of opportunity as a normative criterion has gained some attention. This standard, it is argued, bridges the gap between egalitarianism and the ideals of responsibility and freedom. We consider the impact of policies designed to equalize productivities or “opportunities”, rather than utilities. Our focus, however, is more general than defining and arguing a specific policy or social objective. We wish to examine the merits of such an approach using the standard tools found in the optimal tax and redistribution literature. In particular, we investigate the solution to an education spending problem, which we can then use to define and analyze policies designed to bring about greater equality in human capital outcomes.

A quality education system has long been seen as integral to the health of a society. Schooling not only increases productivity, but can also facilitate social mobility. In this way, it seems a natural tool for redistribution and especially relevant when considering the importance of “equalizing opportunities” as it directly impacts the distribution of skills. Indeed, in most societies it is uncontroversial for the state to play an important role in the provision of education. The reasons for and the implications of this, however, are far from obvious. In particular, the distinction between economic and philosophical/ethical arguments is often blurred.

In the model, individuals vary in their innate talents and locations. We define a human capital technology which maps both characteristics, along with public spending on education, into individual productivity. Education policy shapes the human capital distribution, which in turn impacts welfare as defined by the solution to a social planners optimal tax problem. We consider optimal education spending when public funds are distributed between two different locations and when one location has more productive students, for example rich and poor neighbourhoods. It seems reasonable that greater equality in human capital between regions may be optimal, as welfare is defined in a second-best world. Greater equality in productivity, which implies less redistribution through taxation, could mean less distortion to labour markets and an increase in social welfare. The results suggest, however, that the equal opportunity policy, defined as that which compensates completely for local advantages, is never optimal. This is true regardless of the planners aversion to inequality. In fact, it is often the case that the optimal policy is that which spends nothing in the low-productivity region, regardless of social preferences. This implies that there is no “free lunch”, in that greater equality in human capital outcomes can be achieved only at the cost of lower social welfare.

---

1 Although often discussed, there is no consensus on the exact meaning of the term and there is little in the way of theoretical analysis. See, for example, Roemer (1998).

2 Many have condemned the often substantial inequality we observe in education budgets across various districts (largely funded by the local tax base) on grounds of efficiency, morality and legality. For a discussion of these and related issues, see Berne (1988) or Fernandez and Rogerson (1996).
Redistributing to individuals from low-productivity areas through education spending is limited for a variety of reasons. There will be a reduction in the amount available for money transfers as total output declines. Obviously, it is inefficient to spend too many resources on lower-productivity students. Less obvious is the role played by the tax system, and we see that for a given set of skills, increasing the productivity of low types tightens the incentive constraints in the tax problem which leads to greater distortions in the labour market\(^3\). Further, we note that to satisfy incentive compatibility in the tax problem, low types are discouraged from working. Thus any benefit from increasing their relative productivity, which is costly to begin with, is not fully realized.

Education policy obviously has important implications that extend outside of the simple framework used to derive the results discussed above. These include self-esteem, crime, voter savvy, general social cohesion, etc. Such considerations provide justification for equalizing opportunities even if these policies are not optimal in terms of a standard consumption/leisure view of utility. Whether or not we should equalize opportunities for such reasons is a moral/ethical question for which economic models provide little insight. Although as previously discussed, such policies may reduce welfare in the standard model, which is presumably of importance to the debate. We do not attempt to answer this question here but note that if such policies are indeed deemed optimal, it is not clear that they remain so when we extend the model to include dynamics. In particular, we see that equal opportunity type policies as defined in the static setting are not consistent with dynamic behaviour. This is because individuals are heterogeneous and free to live where they please. Once educated and working, they will segregate according to income and the problem repeats next period. Previous spending influences the composition of neighbourhoods/regions and we see that spending “too much” in the low-productivity area is sub-optimal as it leads to lower outcomes in the next generation for everyone. If the objective is to sever the link between socio-economic status and human capital outcomes, simply equalizing for local effects in each period is not reasonable. Optimal policy must account for the effects on the long run distribution of human capital, while being aware of the exodus of high productivity types from the low-productivity area.

1.1 Relation to the literature

Taxation is generally costly and redistribution schemes must weigh any distortionary effects against the benefits, which depend on the planner’s aversion to inequality. This is the often discussed efficiency/equity tradeoff. A legitimate criticism of this type of analysis is that there is rarely, if ever, much that can be said about the degree of inequality aversion. In particular, it depends crucially on the reasons for different outcomes as well as the way in which society aggregates individual preferences. Success or failure can be the result of luck, talent, effort, etc., each of which have different implications for any redistribution policy. We simply do not know why people have different outcomes, and informational asymmetries seriously limit the value of any proposed scheme designed to facilitate greater equality. Even in the extremely unrealistic case of perfect information, it is difficult to imagine a consensus on exactly what is deserving of transfers and what is not.

\(^3\)Although this effect may be ambiguous in some versions of the model
Theoretically, heterogenous outcomes can be divided into two categories. First, there are those such as preferences, for which we are deemed “responsible”, and do not call for compensation. Second, there are those for which we are held “not responsible”, such as talent or family wealth, which may call for compensation\(^4\). We assume here that there is an observable difference between the initial conditions of the two groups (e.g. rich and poor) for which they are not responsible. Specifically that each group transforms educational resources into human capital with varying effectiveness, depending on both the neighbourhood or region in which they attend school as well as their individual productivity (which is independent of location). There is a vast empirical literature on the impact of “local effects” on children’s academic outcomes. For instance see Goux and Maurin (2007), Ding and Lehrer (2007) or Hoxby (2000)\(^5\).

We take the view that the positive impact that parents’ socioeconomic status has on their children’s educational outcomes is “unfair”\(^6\). If a rich child is more productive than a poor one, then under an incentive-compatible tax scheme they are better off in terms of both standard utility and any non-pecuniary benefits of education (other things equal). The planner does not know why you are productive or not ex-post, and so even with redistributive transfers those born poor are always at a disadvantage\(^7\). In fact, even if the government could distinguish those born poor, it is not clear that they deserve compensation ex-post because their outcome is influenced by other factors that may be just as arbitrary as parents’ income, like talent or luck\(^8\). The question is whether it is reasonable to make up for a bad start ex-post. Consider talented and wealthy persons who came from modest beginnings. Should they be compensated at the expense of those were born privileged, but have low talent and are relatively poor? Justification for such a transfer ex-post is unclear\(^9\), but compensation may be warranted ex-ante.

There is a large literature on education policy that is relevant to the analysis here. In a seminal paper, Arrow (1971) analyzes the division of spending when there is no further redistribution and describes the progressivity/regressivity of optimal policy. In later papers (as in this one), education is assumed to take place in some initial period and taxation/redistribution in a second stage. This framework generally implies that the optimal policy is that which maximizes the size of the “pie”, which is redistributed through the tax system (for example, Ulph (1977) and Ulph and Hare (1979)).

Cremer et al. (2008) argue the merits of an unequal distribution of productivities. They show that inequality in abilities, which are control variables in the model, are valuable due to non-convexities in the optimal tax problem. In the same spirit, Krause (2006) shows that subsidies for skilled education can be consistent with an optimal redistributive program, as these may facilitate

---

\(^4\)For a comprehensive survey on these ideas, see Fleurbaey and Maniquet (2005).
\(^5\)Related theoretical contributions include C. de Bartolome (1990) and Benabou (1996a,b).
\(^6\)This seems to be the general consensus and can be seen in policies such as the “Head Start” program in U.S.
\(^7\)Throughout we will assume that the tax system is not conditioned on where you were educated (as in the literature on tagging). Relaxing this assumption is considered in Appendix A.
\(^8\)Education funds can be targeted towards the poor, poverty being something that the child is not responsible for. It should be noted that this type of redistribution may be more politically feasible than money transfers, that take place between adults whose income is determined by a variety of other factors. While politics are not modelled here, this is certainly of importance.
\(^9\)In particular, if such a transfer were possible, it would violate the so-called principal of horizontal equity.
incentive compatibility in labour supply decisions.

In another closely related paper, Fleurbaey et al. (2002) use a mechanism-design approach to consider the progressivity/regressivity of schooling expenditures where ability is private information and agents invest in education. They find that more aversion to inequality in consumption leads to more inequality in educational achievement. More generally, the impacts of taxation on educational investments are an important issue, but one not considered here as we focus on early education and do not allow for individual choice over the amount of human capital\textsuperscript{10}.

The concept of equal opportunity defined here draws on the work of Dworkin (1981), Sen (1985, 1995) and Roemer (1998) among others. Following Sugden (2004), we can decompose the basic ideas behind equality of opportunity in two ways. First, “starting-line equality” is the view that every person should have access to the same set of options from which one’s path can be chosen. This is stronger than just a lack of discrimination and generally implies a redistribution of resources. The second is that equal efforts should yield equal rewards (a principle Sugden argues is incompatible with a market economy). In what follows, by equal opportunity we shall refer to the relatively weak concept of “starting-line equality”, the exact meaning of which will be made precise below.

In a standard framework, we consider policies which allow us to bolster the relative productivity of those born less fortunate; the question is whether or not this is optimal. The results suggest that in fact this is not optimal when there is a tax system already performing the role of utility equalization. Further, the dynamic analysis suggests that if we employ an equal opportunity type policy (for reasons outside of the standard redistribution framework like the importance of self-esteem etc.), this may be inconsistent with dynamic goals and should be modified accordingly. The analysis makes clear the need to be precise about the objectives of education policy and to be aware of the effects it will have on the distribution of skills in a dynamic world. The rest of the paper is organized as follows. In section 2, the basic environment is described. Section 3 analyzes education policy in a static model. The framework is then extended to consider dynamic implications in section 4, and section 5 concludes. All proofs can be found in the appendix.

2 Individuals

For each individual $k$, we denote units of labour supplied by $l_k$ and consumption by $c_k$, which consists solely of a single good whose price is normalized to one. Individual utilities are defined by the same quasi-linear in consumption function of the bundle $(l_k, c_k) \in \mathbb{R}_+ \times \mathbb{R}_+$.

$$\tilde{u}(c_k, l_k) = c_k - f(l_k)$$

Where the function $f(l)$ satisfies $f'(l) > 0$ and $f''(l) > 0$. The labour market is competitive and it is assumed that aggregate production is simply the sum of individual productivity. So that once

\textsuperscript{10}Boadway et al. (1996) show that when government cannot commit to tax policy once education is obtained, the result is too little investment in human capital. Similarly, Bovenberg and Jacobs (2005) argue for education subsidies as a way to alleviate tax distortions (although there is no issue of commitment in their model).
educated, human capital equals the wage $w_k$. At a given wage, each individual $k$ optimizes utility given by (1) and supplies $l_k$ units of labour, which implies a pre-tax income of

$$y_k = w_k l_k.$$  \hfill (2)

Pre-tax income could also be considered one’s labour supply in efficiency units. Using (2), utility can be written in terms of consumption and pre-tax income as is common in the optimal tax literature

$$u(c_k, y_k) = c_k - f\left(\frac{y_k}{w_k}\right).$$ \hfill (3)

3 Education policy

Wages and productivity are synonymous and a function of individual characteristics and government education spending. Importantly, we assume that human capital attainment is determined by purely exogenous factors from the individual’s perspective. Education spending is targeted to specific neighbourhoods/regions and takes place in an initial time period, after which students enter the working world. The planner not only makes an education spending choice, but also employs a tax and transfer scheme to achieve redistributive goals between individuals once they enter the workforce. Throughout, welfare is assumed to be aggregated by a standard concave welfare function. Education spending will impact welfare at the optimum as defined by the tax problem. Note that as a result of the timing, skills are considered fixed for a government solving an optimal tax problem affecting working adults.

Although integral to the results, our focus is not on the tax problem but over how to allocate a fixed education budget $B$ between the two neighbourhoods. Define $e_j$ as the amount of educational spending in location $j \in \{p, r\}$. The subscripts $\{p, r\}$ represent the low-productivity (“poor”) and high-productivity (“rich”) locations respectively. Let the fraction of $B$ spent in the low-productivity neighbourhood be denoted by $\delta$ and normalize $B$ to 1, so that $e_p = \delta$ and $e_r = 1 - \delta$.

Throughout the rest of section 3, we describe the education spending problem. The next two subsections consider simplified models in which there are no individual differences, so that there are two productivity types corresponding to the two locations. These frameworks portray the intuition and allow analytic results that are relatively easy to interpret. In section 3.3 we discuss the more general case where there is individual heterogeneity within each location.

---

11 The model considers primary/secondary education and takes the view that children’s aversion to schoolwork is irrelevant. This is in line with Roemer (1998) and is likely inappropriate if considering higher education.

12 To determine the size of this budget requires a general theory of social justice, which is ignored here as the focus is on the composition of spending. This was made endogenous in an earlier version and the optimal size of the education budget was seen to be positively related to the total value received from education. In particular, “sub-optimal” education policies like equal opportunity are generally accompanied by a reduction in the size of the budget.
3.1 Two endogenous wages

If there is no individual heterogeneity within each location, then human capital is a function solely of the location in which one is educated. Thus there are two types as there are two locations. The idea of location is only maintained here to be consistent with the later analysis. The interpretation can be more general and the basic assumption is that low and high productivity students can be observed and funded separately. Once working, students educated in the high productivity region receive wage \( w_r(1-\delta) \), while those from the low productivity region receive \( w_p(\delta) \). We assume that the functions \( w_j(e_j) \) are increasing and concave and that \( \forall x, w_r(x) > w_p(x), \frac{\partial w_r(x)}{\partial x} > \frac{\partial w_p(x)}{\partial x} \).

Before describing the optimal education spending choice, we define the equal opportunity policy \( \delta_E \) as that which equalizes the wages of students in the two locations. Thus \( w_r(1-\delta_E) = w_p(\delta_E) \).

3.1.1 First-best

In the first-best, the planner can observe the individual productivity of working adults and make lump sum transfers to achieve any redistributive goals. As individual utility is separable and linear in consumption, labour supply is solely a function of productivity and there is no income effect. Thus, regardless of the degree of inequality aversion, the optimal education policy in the first-best is simply that which maximizes “total surplus”\(^{13}\).

\[
S = w_pl_p(w_p) + w_rl_r(w_r) - f(l_p(w_p)) - f(l_r(w_r))
\]  

(4)

Differentiating with respect to \( \delta \), noting that the indirect impact of changes in \( \delta \) to the labour supply is zero at the optimum (by the envelope theorem), gives

\[
\frac{\partial S}{\partial \delta} = \frac{\partial w_p}{\partial \delta} l_p(w_p) + \frac{\partial w_r}{\partial \delta} l_r(w_r).
\]  

(5)

The conditions for a maximum are not necessarily satisfied here however, and in fact often will not hold. The second derivative is

\[
\frac{\partial^2 S}{\partial \delta^2} = \sum_{j=p,r} \left( \frac{\partial^2 w_j}{\partial \delta^2} l_j(w_j) + \left( \frac{\partial w_j}{\partial \delta} \right)^2 \frac{\partial l_j(w_j)}{\partial w_j} \right).
\]  

(6)

The first term in the sum is non-positive and the second positive. Thus the second-order condition implies a minimum if the curvature in \( w_j(e_j) \) is not too strong relative to the disutility of labour. For example, when productivity is linear, \( \frac{\partial^2 w_j}{\partial \delta^2} = 0 \), so we have a minimum and the optimal policy is \( \delta = 0 \).\(^{14}\)

\(^{13}\)There are no incentive problems and money can be transferred freely. As labour supply is not dependent on the income level (which includes the transfer), the objective is simply to maximize “total output” less “total labour disutility”, regardless of the degree of curvature in the welfare function.

\(^{14}\)Cremer et al. (2008) show with general individual preferences, that when the wage technology is linear, wage differentiation is preferred to wage equalization even when there are no local advantages.
3.1.2 Second-best

In the first-best we may redistribute incomes without imposing any efficiency costs. Now we consider the imperfect information case where only income (and not wages or labour supply) is observable and thus any redistribution scheme must be incentive compatible. This is the situation considered in most of the optimal tax literature. In our case, this is equivalent to the assumption that in the initial stage we can spend on education knowing which is the high and low productivity type, but once they enter the workforce we can no longer differentiate between the two. If we allow for the planner to observe each type in the second stage, then we are back to the first best where the planner imposes lump sum transfers. For the more general case in section 3.3, where there is individual heterogeneity, the results of relaxing this are less obvious and this is discussed in Appendix A\textsuperscript{15}.

We see in the first best it is optimal to maximize the size of the “cake” which may involve spending nothing (or certainly less) in the poor neighbourhood depending on the assumptions made regarding human capital acquisition. This strategy may no longer be optimal here as inefficiencies from redistribution will eat into the “bigger cake”. It is conceivable then that education spending be used to redistribute if inefficiencies from transferring utilities through the tax system are large relative to those brought about by transferring productivities. Before looking at the impacts of education on social welfare, we must first consider the second best tax problem which defines it. Social welfare can be represented by the solution to the following Pareto-optimizing problem\textsuperscript{16}.

\[
\max_{c_p, c_r, y_p, y_r} \quad c_p - f\left(\frac{y_p}{w_p}\right) \quad \text{subject to} \\
y_p + y_r \geq c_p + c_r \quad \text{(8)}
\]

\[
c_r - f\left(\frac{y_r}{w_r}\right) \geq c_p - f\left(\frac{y_p}{w_p}\right) \quad \text{(9)}
\]

\[
c_r - f\left(\frac{y_r}{w_r}\right) \geq m \quad \text{(10)}
\]

where (8) and (9) are the budget and incentive constraints respectively and \(m\) is the predetermined utility level of the high type. The incentive constraint requires that the optimal allocation for a high type provide at least as much utility as is attained by the high type mimicking the low (as types are not observable). Constraint (10) implements social preferences by forcing a minimum

\textsuperscript{15}This issue is considered in the literature on tagging; see for instance Boadway and Pestieau (2006)

\textsuperscript{16}The optimum is determined given the utility of the high type, characterized by \(m\). Changes in \(m\) amount to changes the planners aversion to inequality and this maps out the utility possibility frontier. For example, if \(m = 0\) we have a maxmin or Rawlsian objective.
utility requirement for the high type. The corresponding lagrangian expression is

$$\mathcal{L} = c_p - f\left(\frac{y_p}{w_p}\right) + \lambda (y_p + y_r - c_p - c_r) + \gamma (c_r - f\left(\frac{y_r}{w_r}\right) - c_p + f\left(\frac{y_p}{w_r}\right)) + \mu (c_r - f\left(\frac{y_r}{w_r}\right) - m).$$  (11)

The optimum is described by the constraints and the first-order conditions which reduce to

$$\gamma = 1 - \lambda, \quad \mu = 2\lambda - 1$$  (12)

$$\frac{f'(\frac{y_r}{w_p})}{w_p} = \lambda + (1 - \lambda)\frac{f'(\frac{y_r}{w_r})}{w_r}$$  (13)

$$f'(\frac{y_r}{w_r}) = w_r.$$  (14)

Note that (14) represents the familiar “no distortion at the top” condition and given the quasi-linear form for utility, completely determines \(y_r\). Using the optimal conditions, we can derive the following comparative static results:

$$\frac{\partial y_p}{\partial m} = \frac{2}{\left(1 - f'(\frac{y_p}{w_r})\right)^2} > 0$$  (15)

$$-\frac{\partial \gamma}{\partial m} = \frac{\partial \lambda}{\partial m} = \frac{2(f'(\frac{y_p}{w_p}) - (1 - \lambda)f'(\frac{y_r}{w_r}))}{\left(1 - f'(\frac{y_r}{w_r})\right)^2} > 0.$$  (16)

These inequalities hold as \(f(l)\) is convex, \(w_r \geq w_p\) and we know that \(\lambda \in \left[\frac{1}{2}, 1\right]\) from the first-order conditions. We see that as inequality aversion increases (\(m\) decreases), the before-tax income of the low type decreases as labour supply is more distorted. Also, we see that \(\lambda\) is increasing in \(m\) which implies that \(\gamma\) is decreasing in \(m\). As would be expected, the incentive constraint tightens as redistribution increases.

The optimal tax scheme characterized above is the subject of a vast literature which we will draw on, but is not the focus here. Rather our interest is on the welfare impacts of manipulating the distribution of wages/productivities through the education system\(^\text{17}\). The impact of a change

\(^{17}\text{Recent related papers by Brett and Weymark (2007) and Laurent Simula (2007) consider the impact of changes in productivities to the optimal tax scheme in a model with quasi-linear preferences and a weighted utilitarian objective function (the former linear in leisure, the latter in consumption).}
in $\delta$ at the optimum is

$$\frac{\partial L}{\partial \delta} = f'(y_p) \frac{y_p}{w_p^2} \frac{\partial w_p}{\partial \delta} + \mu f'(y_r) \frac{y_r}{w_r^2} \frac{\partial w_r}{\partial \delta} + \gamma \left( f'(y_r) \frac{y_r}{w_r^2} - f'(y_p) \frac{y_p}{w_p^2} \right) \frac{\partial w_r}{\partial \delta}. \quad (17)$$

Direct impact $\geq 0$

Impact on the incentive constraint $< 0$

Education policy affects welfare directly through the first term which may be positive or negative as the impact on the poor and rich depends on the functional forms and the planner’s aversion to inequality. It is this term that represents the possible value in spending on the poor. If spending in the poor region is not too inefficient and the planner has a high degree of aversion to inequality, this term will be positive (with a maxmin this is unambiguously positive as $\mu = 0$). The effect on the incentive constraint is always negative however, so the question is whether this possible benefit will outweigh the costs?

If we are to say anything more concrete about the impacts of education spending, we must first characterize consumption and income under the optimal tax scheme. To this end, we impose some further structure on the problem which is assumed throughout the rest of this subsection\(^{18}\). Let $f(l) = \frac{1}{2}l^2$ and let the human capital function satisfy $\frac{\partial [wj]}{\partial e_j} \geq 0$. Finally, we assume the following (which is often implied by the latter condition on $w_j(e_j)$)

$$\frac{w_r(1)}{w_r(1 - \delta_E)} \geq \sqrt{2}. \quad (18)$$

We see that (18) holds when the efficiency loss from an equal opportunity policy is “large enough”. In particular, when (18) is satisfied, welfare under a Rawls (maxmin) objective is higher for $\delta = 0$ than for $\delta = \delta_E$ (see appendix for details). It would seem that (18) will cease to hold if diminishing returns are high enough. However, $\delta_E$ is also increasing with the curvature in $w_j(e_j)$, so it’s not immediately obvious when this will be violated. In fact, inequality (18) is satisfied by a large class of functions as discussed below in section 3.1.3. We now make the following, somewhat surprising, conclusion regarding the optimal second-best policy.

**Proposition 1** Given the assumptions above, for both extremes of inequality aversion (utilitarian and maxmin), the optimal education policy spends nothing in the low-productivity region ($\delta = 0$).

**Proof**: See appendix

We see that all educational resources are better spent in the high-productivity region which provides a larger cake to divide in the second period. This is true for both the utilitarian who has no preference for redistribution and the maxmin who cares only for the utility of the low type. Even with extreme aversion to inequality, the value of education spending on low types is outweighed by the costs. For intermediate levels of inequality aversion it is difficult to derive the globally optimal

\(^{18}\)The importance of these assumptions and the consequences of relaxing them is discussed below in section 3.1.3.
policy, although the results of proposition 1 appear to hold for all planners’ preferences between
the extremes (the details are discussed in the appendix).

Proposition 1 implies that education spending is inferior to money transfers as a tool for equal-
izing utility. In fact for a large class of functions, any spending in the low-productivity location
is inefficient, even when the planner only values the welfare of the these individuals. Spending on
the low-types reduces total output (and consequently transfers), as that of high-types is decreased,
while that of the low-types increases relatively less. For a fixed set of skills, low-types are discou-
aged from working by the tax scheme so that the benefit from increasing their human capital is not
fully realized. Finally, increasing the productivity of the low-types tightens the incentive constraint
in the tax problem which creates an even greater distortion in the labour market.

The proposition considers optimal spending for the two extremes of social preferences. It is in-
teresting however to consider how the impacts of policy changes differ for various levels of inequality
aversion. Using again the conditions for optimality in the tax problem, we have

\[
\frac{\partial^2 L}{\partial \delta \partial m} = \frac{\epsilon_w}{\delta(1-\delta)} \left[ \frac{\partial y_p}{\partial m} \left( \frac{2y_p}{w_p^2} - \delta \lambda \right) + \frac{\partial \gamma}{\partial m} \delta (y_p + y_r) \right]
\]

(19)

where \( \epsilon_w = \frac{e}{w(e)} \frac{\partial w}{\partial e} \) is the elasticity of productivity with respect to education funding (which is
assumed to be constant). We see there are two conflicting forces influencing the size of the impact of
education policy changes. Neither of these effects dominates for all \( \delta \), so it is impossible to describe
the importance inequality aversion has on a change in \( \delta \) without making further assumptions. We
can, however, gain some intuition by examining these effects in turn. The first is that the higher is
\( m \), the less redistribution and thus the less distortion on the low-wage type. Therefore, changes in \( \delta \)
have a larger impact on the poor because they are producing more. Noting that the “no distortion
at the top” condition (14) implies there is no impact on the output of the high type. Secondly, as
increases in \( \delta \) tighten the incentive constraint, we see that this effect is smaller the larger is \( m \) as
again we have less redistribution.

Before moving on, we consider the impact of relaxing the assumptions made earlier in the nu-
merical example below. Generally, if there is extreme diminishing returns in \( w(e) \), then a utilitarian
optimum (which is equivalent to the first best described above) will not be at the corner and it is
optimal to spend some funds in the low-productivity community. This is hardly surprising, and we
note especially that it is never optimal for this amount to exceed 1/2. More interesting is the case
of Rawlsian (maxmin) social preferences. We will see in section 3.3, that the equal opportunity
policy is never optimal in the more general model with individual heterogeneity, but that is not
true in this case. In fact, if inequality (18) does not hold, then the optimal policy in this case is
indeed to equalize productivity completely (which is the equal opportunity policy \( \delta_E \)).
3.1.3 Numerical example

To illustrate, continue to assume that utility be given by

\[ u(c_j, y_j) = c_j - \frac{1}{2} \left( \frac{y_j}{w_j} \right)^2 \]

Further, define the wage (human capital) functions as

\[ w_p = \delta^{\alpha}, \quad w_r = r(1 - \delta)^{\alpha} \]

where the constant \( r \geq 1 \) represents the advantage of the high productivity region.

First consider the problem facing a utilitarian planner, which is equivalent to the first best problem. For this specification, total surplus as described by (4) is

\[ S = \frac{1}{2} \sum_j w_j^2. \]

The second-order conditions imply that surplus \( S \) is convex whenever \( \alpha \geq \frac{1}{2} \) and concave otherwise. As mentioned above, if the wage function is linear \((\alpha = 1)\) this is always true, but generally the problem is convex when diminishing returns are not too large relative to the disutility of labour\(^{19}\). If the problem is convex then the optimal policy is \( \delta = 0 \). If it is concave, then there is some lower range over which spending on the low type is beneficial. Regardless of the parameter values, it is easy to show that the optimal policy always spends less than one half in the poor location (equal funding), which is of course less than the equal opportunity policy. Proposition 1 states that optimal policy in the second-best spends nothing in the low-productivity region when (18) holds, even with extreme aversion to inequality. For a Maxmin planner, given the functional forms in this subsection, we can describe the region over which this is true. Figure 1 describes the function \( \frac{w_r(1)}{w_r(1 - \delta_E)} \) over various combinations of \( \alpha \) and \( r \). When this function lies above the plane at \( \sqrt{2} \), the optimal policy with a Rawls objective is \( \delta = 0 \). However, the function lies below the plane at \( \sqrt{2} \) for some combinations of \( r \) and \( \alpha \). For such values, (18) does not hold and it is optimal to completely equalize productivity between the two types \((\delta_E)\). This range of parameters is easier to see in two dimensions. Figure 2 graphs the level curve of the function \( \frac{w_r(1)}{w_r(1 - \delta_E)} = \sqrt{2} \). With values of \( r \) and \( \alpha \) below the curve, there is little difference between locations and there is extreme diminishing returns. In this case, with a Rawls objective, equal opportunity (equal productivity) is optimal. Finally, we note that regardless of parameter values, when the first-best problem is convex (so that optimal policy is \( \delta = 0 \)), then assumption (18) always holds and the optimum in the second best is also \( \delta = 0 \).

\[^{19}\text{For instance, we could set } f(t) = \frac{1}{\epsilon} t^{\epsilon}, \text{ where } \epsilon \geq 2. \text{ Under this specification, surplus is convex when } \alpha \geq 1 - \frac{1}{4}. \]
Figure 1: $\frac{w_r(1)}{w_r(1-\delta_E)}$ for various $r$ and $\alpha$

Figure 2: $\frac{w_r(1)}{w_r(1-\delta_E)} = \sqrt{2}$
3.2 Two fixed wages

In this sub-section, we wish to highlight an important feature of policies designed to equalize productivities. This being that the beneficiaries of such policies are generally those in the low-productivity community who are able to make the best use of educational resources (i.e. the “smart ones”). This was impossible in the previous section as individuals were homogenous within locations. We will consider the more general problem where there is individual heterogeneity within locations in section 3.3, but we can highlight this particular point very starkly using a simpler model.

Let there be two types of jobs in the economy, high and low skill, which are characterized by two fixed wages $w_2$ and $w_1$ respectively. The fraction of high and low skill workers in the economy is denoted by $\pi$ and $1 - \pi$, and is a function of education spending in the initial period. In particular, define the total fraction of skilled workers $\pi$ as the sum of those originating from the low and high-productivity regions which are denoted by $p(e)$ and $r(e)$ respectively, so that

$$\pi(\delta) = p(\delta) + r(1 - \delta) \quad (20)$$

We assume that both $p(e)$ and $r(e)$ are increasing and concave in education spending $e$ so that the total fraction of skilled workers $\pi$ is concave in $\delta$. Although this is a reduced form approach and this is not explicit, the interpretation is that those on the “higher end of the distribution” within each location will attain a skilled job. Further, it is assumed that $r'(e) > p'(e) \forall e$ so that with the same resources, students from the high-productivity (rich) area are more likely to attain a skilled job.

3.2.1 First best

The first-best outcome is again that which maximizes “total surplus”, which in this case is given by

$$S = (1 - \pi)(w_1l(w_1) - f(l_1(w_1))) + \pi(w_2l(w_2) - f(l_2(w_2))). \quad (21)$$

In the interior, the first-order condition describes the first best $\delta^{fb}$.

$$\frac{\partial S}{\partial \delta} = \frac{\partial \pi}{\partial \delta} (w_1l(w_1) + w_2l(w_2) - f(l_1(w_1)) - f(l_2(w_2))) = 0. \quad (22)$$

Unlike the previous case, the second-order condition is satisfied and this indeed describes the optimal policy$^{20}$. Denote the first-best policy as that which satisfies

$$p'(\delta^{fb}) = r'(1 - \delta^{fb}). \quad (23)$$

Note that by the definition of $p(e)$ and $r(e)$, we have $0 \leq \delta^{fb} < \frac{1}{2}$.

$^{20}$Although if $p(e)$ and $r(e)$ are linear we are at a corner and $\delta^{fb} = 0$. 

13
3.2.2 Second best

As in the previous section, the second-best problem is characterized by the solution to the following pareto-optimizing problem.

\[
\max_{c_1, c_2, y_1, y_2} \quad c_1 - f\left(\frac{y_1}{w_1}\right) \quad \text{subject to} \\
(1 - \pi)(y_1 - c_1) + \pi(y_2 - c_2) \geq 0 \tag{25}
\]

\[
c_2 - f\left(\frac{y_2}{w_2}\right) \geq c_1 - f\left(\frac{y_1}{w_1}\right) \tag{26}
\]

\[
c_2 - f\left(\frac{y_2}{w_2}\right) \geq m. \tag{27}
\]

Where again (25) and (26) are the budget and incentive constraints respectively and \(m\) is the predetermined utility level of the high types. The incentive constraint requires that the optimal allocation for high types provide at least as much utility as is attained by mimicking the low type (as types are not observable). Constraint (27) implements social preferences by forcing a minimum utility requirement for the high type. The corresponding lagrangian expression is

\[
\mathcal{L} = c_1 - f\left(\frac{y_1}{w_1}\right) + \lambda((1 - \pi)(y_1 - c_1) + \pi(y_2 - c_2)) + \gamma(c_2 - f\left(\frac{y_2}{w_2}\right) - c_1 + f\left(\frac{y_1}{w_1}\right)) + \mu(c_2 - f\left(\frac{y_2}{w_2}\right) - m).
\]

We see that in the first-best, it is optimal to maximize the size of the cake. This may involve spending nothing, or certainly less, in the low-productivity area depending on the assumptions made regarding the function \(\pi\). Since the wages are assumed fixed, it would seem that the optimal fraction spent on the poor in the second-best problem is the same as that from the first-best as there is no redistributive motive for spending on the poor (i.e., there is no way to increase the wage of the low type and any deviation from \(\delta^{fb}\) shrinks the cake that can be redistributed).

**Proposition 2** The optimal second-best education policy is unique and equals that implied by the first best for any social preferences.

**Proof:** See appendix

The optimal education policy is independent of inequality aversion because the wages are fixed. Thus spending in the low-productivity region above the first-best level is never optimal, regardless of the planners’ objective. However, the impact on welfare from various education policies that are not optimal from this perspective is differing in the social objective. We saw above that the impact of changes in \(\delta\) were generally not the same for all social preferences. The exact relationship was ambiguous, however, as the impact, characterized in (19), was of indeterminate sign. In this case,
however, this is not true, and we can see how the degree of inequality aversion influences the impact of changes in $\delta$. The following proposition assumes that social preferences take an exponential form and in particular exhibit constant absolute inequality aversion.

**Proposition 3** *Education policies that spend a sub-optimally high proportion of funds in the low-productivity region have a larger negative impact on low types than high types.*

**Proof:** See appendix

An increase in the fraction of unskilled workers causes the utility possibility frontier to shift inward from the laissez faire (utilitarian) point. Low-skill types are hurt more by a deviation from the first best because it reduces the ability of the planner to redistribute. This is because there are fewer to tax and more recipients. There is no relative impact in the utilitarian case as there is no redistribution (of course total welfare is reduced simply because there are less skilled types). So we see in this simplified case that increasing education in the poor region shrinks the cake which has a negative impact on welfare directly as there are fewer skilled types and indirectly as the scope for redistribution is decreased. Not only is welfare decreased but the worst off are hit harder.

If the return to education funding is very low for those on the low end of a distribution of individual endowments, then equal opportunity policies will hurt these types the most. This is because they receive little benefit from education funding (in this case they receive none) and further they see a reduction in money transfers through the tax system. If our redistributive goals are to help the least fortunate (which is implied by most social welfare functions), policies that attempt to equalize opportunities will certainly not be optimal. It may be that the talented poor, who will likely achieve a reasonable outcome regardless, are the only beneficiaries of these policies.

### 3.3 Individual heterogeneity

We now extend the model of section 3.1 and let individuals differ within each region. Specifically, we allow for two types of heterogeneity amongst people. First, we define the individual specific endowment $\theta_i$. This can represent innate talents, good family environment, charisma, or any such combination of the like. Let $\theta_i \in \{\theta_l, \theta_h\}$, with $\theta_h > \theta_l$. It is assumed throughout that $\theta$ is not observable. Secondly, as with the simpler models presented above, individuals differ in their location. We will represent location by the constants $L_j$ where $L_j \in \{L_p, L_r\}$ and $L_r > L_p$. Wages (human capital) are defined by

$$w = \theta_i h(L_j, e_j)$$

(28)

where the function $h()$ is increasing and concave in both arguments and as above $e_j$ denotes education funding in region $j$. To reduce notation we define $h_j = h(L_j, e_j)^{21}$. In the simpler models we had only two types which were defined by their location. When we allow for individual specific

---

$^{21}$In the static model, as $L_r$ and $L_p$ are fixed, this is unnecessary but done to be consistent with the dynamic framework of section 4.

15
differences within locations, there are four types of persons with wages \(\{\theta_l h_p, \theta_h h_p, \theta_l h_r, \theta_h h_r\}\). We assume that students with no funding have zero productivity so that \(h(L, 0) = 0\). Finally let \(\frac{\partial h}{\partial L c} > 0\), so that for a given level of resources a rich student has a higher marginal return than a poor one

Our focus is on the optimal education policy, which again impacts welfare which as defined by the solution to an optimal tax problem. However, as we now have more than two types, we can no longer solve the tax problem using the pareto-optimizing characterization as above. We thus define social welfare explicitly as

\[
\sum_{k=1}^{4} \Psi \left( c_k - f \left( \frac{y_k}{w_k} \right) \right)
\]

where \(\Psi()\) is a standard concave social welfare function. The tax problem is to optimize social welfare through a tax and transfer scheme for a given set of individual productivities, where individuals are denoted by the subscript \(k\). In particular, we note that redistributions that occur through the tax system are accounting for differences across both \(L\) and \(\theta\). Spending on education however, accounts only for differences in \(L\) which can be important depending on the interpretation of \(\theta\). Below, we characterize the optimal \(\delta\), but we are also interested in the equal opportunity policy.

**Definition 1** Let the equal opportunity policy, denoted \(\delta_{E}\), be that which satisfies

\[
h(L^p, \delta_E) = h(L^r, 1 - \delta_E).
\]

The policy \(\delta_E\) equalizes the local component of productivity so that those with equal \(\theta\) have the same outcome. This policy accounts solely for differences in location and not for differences in individual characteristics \(\theta\). In this sense, \(\delta_E\) is a somewhat weak definition as it equalizes opportunities between locations, but not between individuals within locations. If we could target funds to the individual rather than just the location, we could define the equal opportunity policy as that which implies equal productivities for all types. In sections 3.1-3.2 above, we ignored individual heterogeneity so that definition 1 is consistent with this view. In what follows, the equal opportunity policy is taken as given, and is considered fundamentally different than any policy implied by optimizing a social welfare function.

### 3.3.1 First best

With perfect information about individual productivity, the planner again maximizes “total surplus”, which in this case is

\[
S = \sum_k \left( w_k l_k - f(l_k) \right).
\]

---

22The impacts of relaxing this are discussed below.
Differentiating with respect to \( \delta \) we have
\[
\frac{\partial S}{\partial \delta} = \sum_k \frac{\partial w_k}{\partial \delta} l_k(w_k).
\] (32)

As above, we consider the second order conditions.
\[
\frac{\partial^2 S}{\partial \delta^2} = \sum_k \left( \frac{\partial^2 w_k}{\partial \delta^2} l_k(w_k) + \left( \frac{\partial w_k}{\partial \delta} \right)^2 \frac{\partial l_k(w_k)}{\partial w_k} \right)
\] (33)

Analogous with the simpler case above, this is not generally a “nice” concave function; and depending on the assumptions made regarding \( h() \) and \( f() \), it may have a maximum at the corner where \( \delta = 0 \). Again, this is true when productivity is linear, so that \( \frac{\partial^2 w_k}{\partial \delta^2} = 0 \).

### 3.3.2 Second best

In the first best, redistribution imposes no efficiency costs. Consider now the imperfect information case, where only income and not wages or labour supply, is observable. The optimal policy in the first best may involve spending nothing on the education of the poor, and certainly less than on the rich. Will such a policy remain optimal once we introduce information asymmetry? In the general model with asymmetric information regarding types, the optimal tax problem is

\[
\max_{c_k, y_k} W = \sum_{k=1}^{4} \Psi \left( c_k - f \left( \frac{y_k}{w_k} \right) \right) \quad \text{subject to }
\]

\[
\sum_k y_k \geq \sum_k c_k \quad (35)
\]

\[
c_2 - f \left( \frac{y_2}{w_2} \right) \geq c_1 - f \left( \frac{y_1}{w_2} \right) \quad (36)
\]

\[
c_3 - f \left( \frac{y_3}{w_3} \right) \geq c_2 - f \left( \frac{y_2}{w_3} \right) \quad (37)
\]

\[
c_4 - f \left( \frac{y_4}{w_4} \right) \geq c_3 - f \left( \frac{y_3}{w_4} \right). \quad (38)
\]

The inequalities (35)-(38) represent the budget constraint and three incentive constraints which must be satisfied by the tax scheme. Again, the incentive constraints require that at the optimum, each individual attains at least as much utility from their own bundle as from mimicking the type
The corresponding lagrangian expression is

\[ L = \sum_k \Psi \left( c_k - f\left( \frac{y_k}{w_k} \right) \right) + \lambda_1 \left( \sum_k y_k - \sum_k c_k \right) + \lambda_2 \left( c_2 - f\left( \frac{y_2}{w_2} \right) - c_1 + f\left( \frac{y_1}{w_2} \right) \right) + \lambda_3 \left( c_3 - f\left( \frac{y_3}{w_3} \right) - c_2 + f\left( \frac{y_2}{w_3} \right) \right) + \lambda_4 \left( c_4 - f\left( \frac{y_4}{w_4} \right) - c_3 + f\left( \frac{y_3}{w_4} \right) \right). \]  

(39)

Again, we are interested in the welfare impacts of manipulating the distribution of wages/productivities through the education system. As in the simpler model, there will be a “direct effect” on utilities at the optimum which is of ambiguous sign (although again strictly positive for a maxmin planner). We note however, that the impact on the incentive constraints is no longer unambiguously negative in this case as it was above. Analyzing the conditions for an optimum we have the following.

**Proposition 4** *Regardless of the planner’s degree of inequality aversion, the equality opportunity policy spends more than is optimal on the low-productivity region.*

**Proof:** See appendix

Proposition 4 tells us that the equal opportunity policy involves spending more on the low-productivity area than is optimal for a social welfare function with *any* non-negative aversion to risk (including infinite). Thus in the general case there are no social preferences that give \( \delta_E \) as the optimum. Note that we do not require specific restrictions as above with Proposition 1. It is interesting, however, to note that when we relax the assumption that \( \frac{\partial^2 h}{\partial L \partial e} > 0 \) it is possible that the equal opportunity policy is optimal (at least locally). This is true because for some specifications of the human capital function, redistributive and efficiency goals coincide (at least locally). This is stated in the following corollary.

**Corollary 1** *When education funding and local effects are substitutes, such that \( h(L, e) = h(L + ke) \), where \( k \) is some positive constant, the equal opportunity policy is locally optimal for any social preferences.*

**Proof:** See appendix

We see that when neighbourhood quality and education funding are complementary, the equal opportunity policy is too extreme from a social welfare perspective. Given this, a reasonable question is whether education has any value as a redistributive tool when there is a redistributive tax system? In the simpler case we found that for a large class of functions it did not. Here we see that deviating from a first best at \( \delta = 0 \), is not welfare improving in the second-best as well.

---

23Technically there are many more of these constraints but these can all be shown to be redundant here.
Proposition 5 If the solution to the first-best problem is $\delta = 0$, increasing education spending in the low-productivity region above the first-best is strictly welfare decreasing. This is true regardless of the degree of inequality aversion.

Proof: See appendix

Increasing education in the low-productivity region from $\delta = 0$ has a negative impact on welfare. Thus proposition (5) states that deviations from the first-best policy in a second-best world are welfare decreasing. So when diminishing returns to education funding are not too large (see equation 33), education has no value (locally) as a redistributive tool in the more general model.

We have thus far discussed the impact of education spending on welfare around certain important policies. Deriving the globally optimal $\delta$ in the second-best problem is more difficult in the general model than with the previous simplifications. We proceed by describing optimal education spending to the extent possible and then consider a numerical example with specific functional forms and parameter values in section 3.3.3.

When there is individual heterogeneity, we must allow for the fact that education policy will impact the ordering of wages between the types in each location. The following definition is useful.

Definition 2 Let $\tilde{\delta}$ be that policy which satisfies the following

$$\theta_h h_p(\tilde{\delta}) = \theta_l h_r(\tilde{\delta}).$$  \hfill (40)

The policy $\tilde{\delta}$ is that which equalizes the productivity of high-types in poor location with that of low-types in the rich location. Refer to values of $\delta < \tilde{\delta}$ as case 1. Over this range of expenditures, poor students will end up as the bottom two types in the income distribution. As $\delta \in (0, \tilde{\delta}]$, the wage ordering $0 \leq w_1 \leq w_2 < w_3 < w_4$ corresponds to $\theta_l h_p(\delta) < \theta_l h_p(1 - \delta) < \theta_l h_r(1 - \delta) < \theta_l h_r(1 - \delta)$. Now consider the effect of a change in $\delta$ on welfare under the optimal tax policy. After some manipulation we can show that welfare is increasing in $\delta$ whenever

$$\delta < \frac{(1 - T_2)(\frac{\lambda_1}{\lambda_1} + 1)y_2 + y_1}{\sum_k y_k}$$  \hfill (41)

and decreasing otherwise, where $y_k$, $\lambda_k$ and $T_2$ are determined by the optimal tax problem (denote the marginal tax on type $k$ as $T_k$). We see that the larger the share of output produced by those from the low-productivity neighbourhood, the more valuable is education spending in that region as we would expect. Also note that when $\delta = 0$, $y_1 = y_2 = 0$ which is consistent with the case where the second best problem has an optimum at $\delta = 0$. For the two extremes of social preferences (utilitarian and maxmin), which will be considered below, condition (41) is
Utilitarian:

\[ \delta < \frac{y_2 + y_1}{\sum_k y_k} < \frac{1}{2} \]  \quad (42)

Maxmin:

\[ \delta < \frac{3(1 - T_2)y_2 + y_1}{\sum_k y_k} \]  \quad (43)

The optimum for a utilitarian is equivalent to the first best as there is no redistribution and thus no concerns regarding information. Since spending on the low-productivity region is inefficient, we see that the impact of increasing \( \delta \) above equal funding is never optimal. At the other extreme, a maxmin planner places weight only on the welfare of type 1. Even so, it is not clear that education spending is beneficial as this has negative impacts on the incentive constraints and the amount which may be redistributed through the tax system.

Once spending in the low-productivity region exceeds \( \bar{\delta} \), this reverses the order of the untalented rich and talented poor in the income distribution, and in this range of expenditures, policy changes will have a different impact because of the effects on incentives in the tax problem. Refer to this as case 2. In case 2, \( \delta \in (\bar{\delta}, \delta_E) \), which implies the wage ordering is \( \theta_l h_p(\delta) < \theta_l h_r(1 - \delta) < \theta_h h_p(\delta) < \theta_h h_r(1 - \delta) \). For case 2, we have \( \frac{\partial C}{\partial \delta} > 0 \) (and negative otherwise) whenever

\[ \delta < \frac{\psi'(v_1)}{\lambda_1}(1 - T_1)y_1 + (1 - T_3)(\frac{\lambda_3}{\lambda_1} + 1)y_3 + y_2(1 - (1 - T_2)(\frac{\lambda_3}{\lambda_1} + 1))}{\sum_k y_k}. \]  \quad (44)

In case 2, the two extreme cases for social preferences are.

Utilitarian:

\[ \delta < \frac{y_2 + y_1}{\sum_k y_k} < \frac{1}{2} \]  \quad (45)

Maxmin:

\[ \delta < \frac{4(1 - T_1)y_1 + 2(1 - T_3)y_3 + (1 - 3(1 - T_2))y_2}{\sum_k y_k}. \]  \quad (46)

As in case 1, the larger the contribution of output of those from the poor neighbourhood, the more valuable spending is in that neighbourhood. Ideally, we could use these expressions and those above to characterize education policy for different social preferences. However, unless we have analytic solutions for the tax problem that are simple enough to work with, this yields little. To gain further insight, we solve the problem for a specific set of functions and parameters.
Figure 3: Case 1

Figure 4: Case 2
3.3.3 Numerical example

Let the functional forms be as in the example of section 3.1.3. Figures 3 and 4 describe welfare at the optimum in the second-best problem for different values of the education spending parameter $\delta$. The value functions represent the two extremes of social preferences, utilitarian (solid curve) and maxmin (dashed curve). These are generated with the function parameters $\alpha = 0.6$, $\epsilon = 2$. The heterogeneity between types is given by the parameters $L^p = \theta_l = 1$, $L^r = 3$ and $\theta_h = 2$. We see that, for this characterization, welfare in the second-best problem is strictly decreasing in $\delta$, even in the case with infinite aversion to inequality. The value function traced out in the graphs is one of a variety of combinations of parameters all of which have similar results.

The results in the more general case are in line with those found above in the simpler models. In particular, that arguments for the equalization of school funding, or more extremely equalization of opportunities, need to be based on reasons beyond the redistribution of utility in the consumption/leisure sense. Figures 3 and 4 imply that education is inferior to cash transfers even in the second-best problem and with extreme aversion to inequality.

4 Social mobility, equal opportunities and dynamics

Our results thus far suggest that education funding should not be used to redistribute utilities between individuals in the standard sense. However, there are many valid arguments supporting such policies that go beyond the analysis above, for example the importance of self-esteem. In this section, we take as given the validity of such a policy and assume that social welfare is concerned solely with the individuals human capital. Then we consider more deeply the importance of education policy in a dynamic setting. First, we modify the above framework slightly and let the population be defined by a continuum rather than four types and as above the distribution of $\theta$ is the same in the two communities. We continue to assume that the population is equally divided between the two locations. Thus the human capital of a child $i$ born in neighborhood $j$ is as described in section 3.3:

$$w = h_{ij} = \theta_i h(L^j, e_j)$$

with the exception that $\theta \in [0, \Theta]$ is continuous and distributed according to the measurable function $F(\theta)$, where the total population is normalized to one so that $F(\Theta) = \frac{1}{2}$.

4.1 Static education spending

We consider first the static spending problem in which the planner divides a fixed budget (again normalized to one) over the two neighborhoods. Define aggregate human capital as the sum of individual human capital in both neighborhoods$^{24}$.

$$H = \int_{\theta} [\theta h(L^p, \delta) + \theta h(L^r, 1 - \delta)] dF(\theta)$$

$^{24}$We can think of $H$ as equivalent to aggregate income given the assumptions made regarding wages in section 2.
Definition 3 Let $\delta^*$ denote the spending policy that maximizes aggregate human capital $H$, implicitly defined by

$$\frac{\partial h(L^p, \delta^*)}{\partial \delta} = -\frac{\partial h(L^r, 1 - \delta^*)}{\partial \delta}.$$  

We define the planners’ aversion to inequality over differences in human capital rather than utility as with a standard social welfare function. In this way the government has preferences solely over human capital and not over some subset of the results of human capital (i.e., consumption and leisure). Implicit in many arguments surrounding equal opportunity policy in education are issues rarely treated in economic analysis. For instance, non-pecuniary effects such as self-esteem, crime, voter savvy, general social cohesion, etc... Consumption is still important in that consumption is presumably rising in human capital, but this is not the only factor of importance. Thus we ignore redistributions over consumption in this section. Further, we put more structure on the planners’ preferences and assume they are given by a CES function with parameter $\rho = 1 - \sigma$, where $\sigma \in (-\infty, 1]$. Consequently, we define the static problem as

$$\max_\delta W = \left( \int \left( [\theta h(L^p, \delta)]^\sigma + [\theta h(L^r, 1 - \delta)]^\sigma \right) dF(\theta) \right)^{\frac{1}{\sigma}}.$$  

From the planners’ perspective, the unborn face a gamble with two outcomes: either born rich or born poor. Education spending influences the expected return of the gamble. Spending affects today’s children, who work and live tomorrow. This policy is undertaken ex-ante (before adult working life begins), unlike taxes and transfers (after working life begins). In this way, spending on the poor can be taken as insurance against poverty to the not-yet-born. This insurance is separate from that offered by a tax system which compensates for $\theta$ as well as socio-economic status. In short, larger aversion to differences in human capital across neighbourhoods results in more resources spent on the education of the poor. In particular, as $\rho \to \infty$ the optimal policy is the equal opportunity policy described in definition (1).

Lemma 1 Let $\hat{\delta}$ denote the optimal policy in the static problem, as defined by (49), then both $\delta_E \geq \hat{\delta} \geq \delta^*$ and $\frac{\partial \hat{\delta}}{\partial \rho} > 0$.

This bounds optimal policy choices and describes the “equity/efficiency” tradeoff in this problem. The larger is $\rho$, the larger is $\delta$ and the smaller is aggregate human capital relative to the possible maximum (at $\delta^*$). Furthermore, as $H$ is concave in $\delta$, this “inefficiency” is increasing and convex.

4.1.1 Example

Let human capital production take the following form

$$h(\theta, L, e) = \theta L^\alpha e^{1-\alpha}.$$  

23
Further let differences between neighbourhood spillovers be given by $L^p = 1$ and $L^r = \phi > 1$. The planners’ problem is

$$\max_\delta W = \max_\delta \left( \int \left( [\theta \delta^{1-\alpha}]^\sigma + [\theta \phi^\alpha (1-\delta)]^{1-\alpha} \right) dF(\theta) \right)^{\frac{1}{\sigma}}. \quad (52)$$

The optimal policy is given by

$$\hat{\delta} = \frac{a}{a + 1} \quad \text{where} \quad a = \phi^{\frac{\alpha}{1-\alpha}}. \quad (53)$$

From Lemma (1), the range of solutions for differing values of $\sigma \in (-\infty, 1]$ is given by $[\delta^*, \delta_E]$ which in this case is

$$\hat{\delta} \in \left[ \frac{1}{1 + \phi}, \frac{1}{1 + \frac{1}{\phi^{\frac{\alpha}{1-\alpha}}}} \right]. \quad (54)$$

We see that the larger are both $\phi$ and $\alpha$, the larger is the difference between efficiency and redistributive goals. We turn now to the dynamic problem.

### 4.2 Dynamic policy and mobility

The static model described in the previous section lays out a simple framework in which to analyze optimal education spending in this context. This provides a simple equity/efficiency tradeoff we can use to justify policies like those in definition (1). Now we consider the dynamic implications of such policies.

Education spending occurs in two periods: 0 and 1, which are denoted by subscripts. The problem is solved recursively and period 1, which is equivalent to the static problem above, is considered first. Up to this point, we’ve assumed that people are born into one of the two neighborhoods and this has impacted their human capital outcomes. Government spending on education was only concerned with a single generation of children. However, if local human capital externalities are truly important and education spending can generate “social mobility”, then we must consider any impacts on the distribution of skills of a “mobile” population.

Population size remains fixed and each person lives 2 periods, the first of which is childhood where education takes place and the second adulthood where one works and has one child. Once an adult, people choose a neighborhood in which to live. It is assumed that the rich neighborhood is more desirable (for reasons beyond but including better schools). Essentially, if one can afford it, they will live in the rich neighborhood and the poor are kept out. There are numerous papers both empirical and theoretical that describe this phenomenon, and it is simply assumed here\textsuperscript{25}.

\textsuperscript{25}Theoretical references include Benabou (1996a), Rogerson and Fernandez(1996). Recent empirical papers include Bayer, McMillan and Reuben (2005) and Bayer, Ferreira and McMillan (2007). In particular, it is assumed that moving between neighbourhoods is costless, which is appropriate within a city but perhaps not between countries (as in say the EU) or even states/provinces within a union. Adding moving costs will mitigate the negative impacts described below in Proposition 6 but will not eliminate them.
The idea of “social mobility” is commonly referred to as a policy goal which is akin to equal opportunity in that it is designed to sever the link between parents’ status and their children’s outcomes. However, when you increase “social mobility”, people must move in both directions, so it is unclear what the ethical rationale for such a policy would be unless there is some efficiency argument underlying it\textsuperscript{26}. In the dynamic case, however, education spending not only influences the outcomes of tomorrow’s adults’ incomes, but also neighborhood composition, which influences their children’s outcomes. Before considering optimal policies, we must characterize the period 1 local effects $L_1^p$, $L_1^r$ which are endogenous from a period 0 perspective.

As income is monotonically increasing in human capital and there are only two neighborhoods, we can divide the population into two groups: those below and those above the median in the human capital distribution. This implies cutoff levels of $\theta$ in each neighborhood labelled $\theta^m_p$ and $\theta^m_r$. If one’s $\theta$ is above the cutoff, they will earn enough (exceed the median) to live in the rich neighborhood, and vice-versa.

**Definition 4** The talent cutoff levels $\theta^m_p$ and $\theta^m_r$ are defined by the following conditions (where $L_0^p$, $L_0^r$ are exogenous initial differences between the neighborhoods):

$$\theta^m_p h(L_0^p, \delta_0) = \theta^m_r h(L_0^r, 1 - \delta_0)$$

and

$$F(\theta^m_p) + F(\theta^m_r) = \frac{1}{2}.$$  

Further, we assume that the local externality is given by the average human capital in the neighborhood,\textsuperscript{27} so that local effects in period 1 are defined by

$$L_1^p = \int_{\theta^m_p}^{\theta^*} [\theta h(L_0^p, \delta_0)] dF(\theta) + \int_{\theta^m_p}^{\theta^*} [\theta h(L_0^r, 1 - \delta_0)] dF(\theta)$$

$$L_1^r = \int_{\theta^m_r}^{\theta^*} [\theta h(L_0^p, \delta_0)] dF(\theta) + \int_{\theta^m_r}^{\theta^*} [\theta h(L_0^r, 1 - \delta_0)] dF(\theta).$$

Assuming $F$ is uniform over $[0, \overline{\theta}]$, the expressions for $L_1^p$ and $L_1$ ($L_1^r$ is given by default as $L_1 - L_1^p$) are:

$$L_1 = H_1 = \frac{\overline{\theta}}{2} (h(L_0^r, 1 - \delta_0) + h(L_0^p, \delta_0))$$

$$L_1^p = \frac{\overline{\theta}}{2} \frac{h(L_0^p, 1 - \delta_0) h(L_0^p, \delta_0)}{h(L_0^r, 1 - \delta_0) + h(L_0^p, \delta_0)}.$$  

Regardless of what education policy is in period 0, the rich will separate themselves and their children will benefit from being raised in the good neighborhood. This is true even if the planner is

\textsuperscript{26}For instance, models in which communities fund education and the poor community is equally productive but is credit constrained and so invests too little.

\textsuperscript{27}This is done for simplicity, but a more general specification could be a weighted sum that puts more emphasis on the lower tail (bad apples) or on the higher tail (role models), etc. It would be interesting to consider the impact of different specifications for the local effect.
able to completely eliminate neighborhood differences between rich and poor in period 0’s children. This is a fact of life in a free society and will always be true unless we force people to integrate, in which case it would cease to be a free society. A forward-looking policy will consider the impacts of spending on future neighborhood characteristics, which will impact the extent to which children will have the same chances in the future.

It can be shown that both $L_1$ and $L_r^1$ are strictly decreasing in $\delta_0$ (for $\delta \geq \delta^*$), which is not particularly surprising. It is generally accepted that equity may come at the price of efficiency, but at some point too much spending on today’s poor will hurt the next generation of poor.

**Lemma 2** Reducing the fraction spent on the poor in period 0 from $\delta_E$ will increase average human capital in both rich and poor neighborhoods in period 1.

**Proof**: See appendix

A static equal opportunity argument corresponds to extreme aversion to inequality in this model ($\sigma = -\infty$). In fact, the equal opportunity policy $\delta_E$ is so redistributive that the effect on tomorrow’s poor children is negative (let alone the impact on everyone else). Not only is it negative, but could be substantially so depending on the initial difference in neighborhood composition. Equal opportunity policy prescriptions based on static concepts of justice will be too extreme and inconsistent with optimal dynamic behaviour. The use of education as a redistributive tool should seek to optimize more than just this generation of poor. This reasoning extends beyond the extreme case of an equal opportunity policy as is shown below.

### 4.2.1 2 period problem

The period 0 problem is the choice of $\delta_0$ that provides spending for period 0 children but also considers the impact of spending on the distribution on human capital. Ignoring discounting, the problem is

$$W = \max_{\delta_0} \left( \int_\theta \left( \frac{1}{\sigma} \left[ \theta h(L_{00}, \delta_0) \right]^{\sigma} + \left[ \theta h(L_{01}, 1 - \delta_0) \right]^{\sigma} \right) dF(\theta) \right) \frac{1}{\sigma} + \left( \int_\theta \left( \frac{1}{\sigma} \left[ \theta h(L_{10}, \hat{\delta}_1) \right]^{\sigma} + \left[ \theta h(L_{11}, 1 - \hat{\delta}_1) \right]^{\sigma} \right) dF(\theta) \right) \frac{1}{\sigma}.$$  \hspace{1cm} (61)

$W_2(\delta_0)$ is welfare under the optimal static policy $\hat{\delta}_1$ (described above). Note that $L_{11}^1 = L_{11}^1(\delta_0)$ and $\hat{\delta}_1 = \hat{\delta}_1(\delta_0)$. While the general optimal policy is difficult to solve explicitly, we can gain some valuable insight from examining the optimal conditions, particularly the way in which our definition of equal opportunity may be inappropriate in the dynamic problem.

---

28Not to mention the negative impact on the tax base which is not modelled here.
**Assumption 1** Let \( h() \) satisfy the following condition

\[
\frac{\partial h(L^p_1, \delta_1)}{\partial L} \geq \frac{\partial h(L^*_1, 1 - \delta_1)}{\partial L}.
\]  
(62)

Assumption 1 requires that, at the period 1 optimum, an increase in community quality has at least as large an effect on the poor neighbourhood. This seems perfectly reasonable, and is in fact satisfied by a variety of functions \( h() \), including the Cobb Douglas production function considered above with any value \( \alpha \in (0, 1) \)\(^{29}\).

**Proposition 6** A planner maximizing a static objective (as defined by (49))

1. Will spend too little on the poor (with respect to the dynamic optimum) when \( \rho = 0 \).
2. Always redistributes too much (with respect to the dynamic optimum) For \( \rho \geq 2 \), the extent to which is increasing with \( \rho \).

**Proof:** See appendix

If there is no aversion to risk or ex-ante inequality, then optimal static policy in the first period \( (\delta^*_0) \) spends too little on the poor region. For \( \rho \in (0, 2) \) the difference between static and dynamic objectives is ambiguous and depends on the function \( h() \). For large enough aversion to inequality, \( \rho \geq 2 \), the static policy always spends too much on the poor.

**Corollary 2** If the human capital technology is defined such that \( h(L, e) = h(L + ke) \), where \( k \) is some positive constant, then the optimal static and dynamic policies are equivalent.

**Proof:** See appendix

We see that, if \( h() \) takes the form above, then the value of spending on the poor region is higher at the margin regardless of aversion to inequality. In essence, there is no conflict between redistributive and efficiency goals because investment in the poor neighbourhood is more efficient. The optimal policy for any degree of inequality aversion is in fact the equal opportunity policy\(^{30}\). To firm up the intuition, consider the following example.

**4.2.2 Numerical example**

We define a specific human capital production function as well as the other relevant parameters described above and solve numerically for the optimal spending mix \( \delta_0 \) for a variety of risk aversion parameters \( \rho \). Let the \( h() \) be given by

\[
h(L, e) = L^\alpha e^\beta.
\]  
(63)

\(^{29}\)Although not necessary, this simplifies the analysis substantially.

\(^{30}\)This case is more in line with models involving credit constraints in the investment of human capital.
The following table characterizes the solutions to the static problem (described in (49)) and dynamic problem (described in (61)), these are denoted by \( \delta_s^0 \) and \( \delta_d^0 \) respectively. The orders of magnitude are of no relevance of course but results are instructive nonetheless.

Table 1: Optimal policy under dynamic and static objectives

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>( \delta_s^0 )</th>
<th>( \delta_d^0 )</th>
<th>( \delta_s^0 - \delta_d^0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.355</td>
<td>.364</td>
<td>-.009</td>
</tr>
<tr>
<td>0.15</td>
<td>.384</td>
<td>.384</td>
<td>0</td>
</tr>
<tr>
<td>0.5</td>
<td>.439</td>
<td>.435</td>
<td>.004</td>
</tr>
<tr>
<td>1</td>
<td>.5</td>
<td>.495</td>
<td>.005</td>
</tr>
<tr>
<td>1.5</td>
<td>.545</td>
<td>.534</td>
<td>.011</td>
</tr>
<tr>
<td>2</td>
<td>.579</td>
<td>.559</td>
<td>.020</td>
</tr>
<tr>
<td>3</td>
<td>.627</td>
<td>.592</td>
<td>.035</td>
</tr>
<tr>
<td>5</td>
<td>.680</td>
<td>.630</td>
<td>.050</td>
</tr>
<tr>
<td>10</td>
<td>.733</td>
<td>.677</td>
<td>.056</td>
</tr>
<tr>
<td>100</td>
<td>.793</td>
<td>.735</td>
<td>.058</td>
</tr>
</tbody>
</table>

\((\alpha = 0.5, \beta = 0.3, \phi = 2, \theta = 50)\)

First, we see that Proposition 6 holds in that a planner with no aversion to risk spends too little on the poor and too much when \( \rho \geq 2 \). Also note that the difference between static and dynamic optima is increasing in \( \rho \). For this set of parameters, it is only for very low levels of aversion to risk that spending on the poor is too low in the static case. In particular, the planner with risk aversion parameter \( \rho = 0.15 \) has it “right”. Notably, for \( \rho = 1 \) (the log case which implies equal funding across neighbourhoods), we have overspending in the static problem. This case is often discussed in the literature and may represent a real constraint in this problem in that, for political reasons, a planner could never spend more on the rich.

5 Conclusions

The results suggest that redistribution through education is not optimal when policy objectives are defined by a standard consumption/leisure model. In particular, there are no social preferences under which an equal opportunity policy is called for. Further, under a variety of specifications, it is ideal to spend all educational resources on high-productivity students and make up for inequities through money transfers.

There are a variety of reasons why this is the case. Spending in the low-productivity location results in a reduction in total output (and consequently transfers), as that of high-types is decreased, while that of the low-types increases relatively less. It is obvious that spending on less productive students can reduce output, less obvious is the role played by the tax system. For a fixed set of skills,
low-types are discouraged from working by the tax scheme, so that any benefit from increasing their human capital is not fully realized. Further, increasing the productivity of the low-types tightens the incentive constraint in the tax problem, which creates an even greater distortion in the labour market.

A further important caveat is that the talented poor, who will possibly achieve a reasonable outcome regardless, may be the major beneficiaries of these policies. This will be true when the return to education funding is small for those on the low end of the distribution of individual endowments. We have have seen that these types are hurt the most by equal opportunity policies, as they will see little but a reduction in money transfers through the tax system. If our redistributive goals are to help the least fortunate (which is implied by the standard social welfare analysis), policies that attempt to equalize opportunities will certainly not be optimal in this case.

The point of this paper is not to argue against policies designed to bring about more equality in productivity, but rather to bring to light the conflict between various objectives. The merits of equal opportunity policies clearly extend beyond the standard limited concepts of utility maximization. The value of education undoubtedly transcends its impact on consumption outcomes both individually and in the aggregate. The impacts of educational spending on self esteem, crime, voter savvy, general social cohesion, etc., are central to the policy debate. Although sometimes difficult to model in an economic framework (at least in an interesting way), these concerns are of great importance none the less.

We have shown that if we accept the validity of such arguments and choose policies which create greater equality in productivities amongst socio-economic groups, then it is imperative that spending decisions take into account dynamic implications. In particular, when individuals are mobile and there are local externalities in the production of human capital, then a complete equalization of opportunities will be too extreme. If the objective is to sever the link between socio-economic status and human capital outcomes, simply compensating for local effects in each period is not reasonable. An alternative policy objective is to maximize human capital in the poor region over the long run. This would require a deeper understanding of both the importance of local externalities and of the human capital function in general.

An ideal education policy will depend on the relative importance of various objectives, which as we have seen, are generally in conflict. This paper highlights the need to be aware of both why we wish to equalize opportunities, and, if so, to be cognizant of the possible dynamic inconsistency inherent in such policies.
6 Appendix A: Tagging and Welfare

Throughout the analysis of section 3 we assumed that the planner could target education funds to neighbourhoods/regions but not implement separate tax policies due to limited information ex-post. Do the results of section 3 hold if we relax this?

Allowing for separate tax treatments between types is equivalent to assuming the planner can observe one’s socio-economic status ex-post. If there is no individual heterogeneity, this amounts to perfect information and the planner simply solves the first-best problem laid out in section 3.1.1.

In the more general case, the impact is less obvious. As with the tax problem laid out in section 3.3, it is difficult to attain an analytic solution to the education problem in this case. Technically, the only difference between the two problems is that there are only two incentive constraints as opposed to three. These incentive constraints apply to the high and low \( \theta \)'s within each neighbourhood/region (as \( \theta \) is still unobservable). It turns out that both Propositions 4 and 5 continue to hold when we allow tagging in the tax problem. Proposition 4 continues to hold because, at the equal opportunity policy, the only heterogeneity that exists is over \( \theta \), so the problems are equivalent. Further at \( \delta = 0 \), the impacts of a change in spending on the poor, although different in these two cases, can be shown to be unambiguously negative without much difficulty.

As with the general model above with no tagging, we can solve this for specific functional forms and parameters. Figures 5 and 6 map out welfare as the value function of the tax problem when we allow for differing tax treatments across socio-economic groups. The curves are formulated with the same parameters used in section 3.3.3. Again, the solid curve represents welfare under a utilitarian planner and the dashed curve welfare under a maxmin planner. We see that the results are similar and the optimal policy remains \( \delta = 0 \) for both.
7 Appendix B: Proofs

Proof of Proposition 1. After some manipulation we can write

\[
\frac{\partial L}{\partial \delta} = \frac{\epsilon_w}{\delta(1-\delta)} \left[ f'(\frac{w_p}{w_p}) y_p - \delta \lambda (y_p + y_r) \right]
\]

where \( \epsilon_w = \frac{e}{w(c)} \frac{\partial w}{\partial e} \) is the elasticity of productivity with respect to education funding (which equals \( \frac{1}{2} \) in this case). We can immediately see that this is negative when \( y_p = 0 \). We’ve assumed that \( f(l) = \frac{1}{2} l^2 \) so that we can solve for \( y_p \) and \( y_r \) explicitly using the first-order conditions on the tax problem.

\[
y_p = \frac{\lambda w_r^2 w_p^2}{w_r^2 - (1-\lambda)w_p^2}, \quad y_r = w_r^2.
\] (64)

Using these we can see that \( \frac{\partial L}{\partial \delta} \) is negative when

\[
w_r^2 + (2\lambda - 1)w_p^2 \geq \frac{\lambda w_p^2 w_r^2}{\delta(w_r^2 - (1-\lambda)w_p^2)}.
\] (65)

For a utilitarian objective, \( \lambda = 1 \) so that (65) reduces to

\[
\frac{[w_r(1-\delta)]^2}{1-\delta} \geq \frac{[w_p(\delta)]^2}{\delta}.
\] (66)

By assumption, the left (right)-hand side is decreasing (increasing) in \( \delta \). Since \( \delta \leq \delta_E \) showing this holds for \( \delta_E \) shows that the derivative is always negative and \( \delta = 0 \) is optimal. At \( \delta_E \), wages are equal, so this inequality can be written as

\[
\delta_E \geq 1 - \delta_E
\] (67)

which is true by definition. To see this, we know that \( w_r(1-\delta_E) = w_p(\delta_E) \) and that \( w_r(e) > w_p(e) \forall e \Rightarrow \delta_E > 1 - \delta_E \). Now turn to the maxmin problem.

For large enough \( \delta \), an increase in spending in the poor region can have a positive impact on welfare, depending on the planners’ aversion to inequality. For maxmin preferences, \( \lambda = \frac{1}{2} \), so that equation (65) implies that welfare is decreasing whenever

\[
\frac{2}{1+\delta} [w_r(1-\delta)]^2 \geq \frac{[w_p(\delta)]^2}{\delta}.
\]

As above, the left (right)-hand side is decreasing (increasing) in \( \delta \). However in this case, at \( \delta_E \) this is unambiguously positive. Since \( \frac{2}{1+\delta} [w_r(1-\delta)]^2 - \frac{[w_p(\delta)]^2}{\delta} \) is strictly increasing in \( \delta \), we know that if welfare is larger at \( \delta = 0 \) than \( \delta_E \) then this proves the result.

First show that welfare under the policy \( \delta = 0 \) is greater than welfare under the equal oppor-
tunity policy. With the assumptions given, we can derive analytic solutions for consumption and income by using the first-order conditions and the incentive and budget constraints (both of which bind for the maxmin problem). Welfare when \( \delta = 0 \) is simply \( w_r (1) \). Under the equal opportunity policy it is \( w_r (1 - \delta E) \), so that welfare is higher under \( \delta = 0 \) whenever \( \frac{w_r(1)}{w_r(1-\delta E)} \geq \sqrt{2} \) which is exactly the condition in (18).

At \( \delta_E \), there is no heterogeneity and thus there is no redistribution. It is reasonable to think that welfare is discontinuous at the point where there ceases to be a transfer between the two types. We now show that welfare is left continuous at \( \delta_E \). Note that \( y_p \to w_p^2 \) and \( w_p \to w_r \) as \( \delta \to \delta_E \).

Thus

\[
\lim_{\delta \to \delta_E} W = \lim_{\delta \to \delta_E} \frac{1}{2} \left( y_p + \frac{1}{2} \left( \frac{y_p}{w_r} \right)^2 - \left( \frac{y_p}{w_p} \right)^2 \right) + \frac{1}{4} w_r^2 = \frac{[w_r(1-\delta_E)]^2}{2}
\]

which is exactly the welfare under the equal opportunity policy (again where there is no redistribution). ■

More generally, we know that \( 2\lambda - 1 \geq 0 \) so that a sufficient condition for (65) to hold is

\[ w_r^2 \geq \frac{w_p^2}{\delta} \]

so that for any degree of inequality aversion, the second best problem is similar in that welfare is decreasing up to some threshold (which is increasing in the advantage of the rich) and possibly increasing for higher \( \delta \) (this is only a sufficient condition). Although the problem is similar in general, showing that welfare is monotonically increasing after this threshold is difficult and we can’t use the analysis as above with the maxmin. Generalizing this result is left to further work.

**Proof of Proposition 2.** A change in education policy is captured by

\[
\frac{\partial L}{\partial \delta} = \frac{\partial \pi}{\partial \delta} \lambda (y_2 - c_2 - y_1 + c_1) = \frac{\partial \pi}{\partial \delta} \lambda (t_2 - t_1) \Delta
\]

where \( t_2 \) and \( t_1 \) are the total taxes paid. As \( t_2 > t_1 \) we see that the term labelled \( \Delta \) is always positive. Thus an increase in \( \delta \) is positive (negative) when \( \frac{\partial \pi}{\partial \delta} > 0 \) \( (< 0) \). Thus the first-best policy is optimal. Further, we know that by definition the equal opportunity policy \( \delta_E > \frac{1}{2} > \delta^{fb} \) ■

**Proof of Proposition 3.** Any deviation from the optimum \( \delta^{fb} \) such as equal funding or the equal opportunity policy results in an increase in the fraction of unskilled workers. Boadway and Pestieau (2006) show that an increase in the fraction of unskilled workers has a negative impact on the welfare of both high and low types as would be expected. More importantly, they show that the
negative effect is larger for low types and that this is true for any positive aversion to inequality.

**Proof of Proposition 4.** Under an equal opportunity policy there are only two wage/productivity outcomes:

\[ w_l = \theta_l h_p(\delta_E) = \theta_l h_r(1 - \delta_E) \]

\[ w_h = \theta_h h_p(\delta_E) = \theta_h h_r(1 - \delta_E) \]

Denote those with high and low productivity (those with high and low \(\theta\)) with the subscripts \(h, l\), respectively. The optimal tax problem now has only one incentive constraint, thus the lagrangian for this problem can be written

\[ L = \sum_{k \in \{l, h\}} 2\Psi \left( c_k - f \left( \frac{y_k}{w_k} \right) \right) + 2\lambda \left( \sum_k y_k - \sum_k c_k \right) + \gamma \left( c_h - f \left( \frac{y_h}{w_h} \right) - c_l + f \left( \frac{y_l}{w_h} \right) \right) \]

Note that although rich and poor students have the same productivity, they are affected differently by spending changes at the optimum. In particular, the impact on rich students of a change in \(\delta\) is greater and of opposite sign than on poor students. After some manipulation, we can write this effect as

\[ \frac{\partial L}{\partial \delta} = \left( 2\Psi' f'(\frac{y_l}{w_l}) \frac{y_l}{w_l^2} \theta_l + 2\Psi' f'(\frac{y_h}{w_h}) \frac{y_h}{w_h^2} \theta_h + \gamma \theta_h f'(\frac{y_h}{w_h}) \frac{y_h}{w_h^2} \right) \left[ \frac{\partial h_p}{\partial \delta} + \frac{\partial h_r}{\partial \delta} \right] \]

This is negative whenever \(\left[ \frac{\partial h_p}{\partial \delta} + \frac{\partial h_r}{\partial \delta} \right] < 0\). By assumption \(h(e)\) is concave and satisfies \(h_r(e) > h_p(e), \ h'_r(e) > h'_p(e) \ \forall e\) so that \(\delta_E > 1 - \delta_E\). Therefore \(\left[ \frac{\partial h_p}{\partial \delta} + \frac{\partial h_r}{\partial \delta} \right] = h'_p(\delta_E) - h'_r(1 - \delta_E)\) is negative.

**Proof Corollary 1.** We saw in the proof of Proposition (4) that at the equal opportunity policy, we have

\[ \frac{\partial L}{\partial \delta} = \left( 2\Psi' f'(\frac{y_l}{w_l}) \frac{y_l}{w_l^2} \theta_l + 2\Psi' f'(\frac{y_h}{w_h}) \frac{y_h}{w_h^2} \theta_h + \gamma \theta_h f'(\frac{y_h}{w_h}) \frac{y_h}{w_h^2} \right) \left[ \frac{\partial h_p}{\partial \delta} + \frac{\partial h_r}{\partial \delta} \right] \]

As \(h(L, e) = h(L + ke)\) and is concave, \(L^r > L^p\) implies that for all \(\delta \leq \delta_E\) we have \(\left[ \frac{\partial h_p}{\partial \delta} + \frac{\partial h_r}{\partial \delta} \right] \geq 0\)

**Proof of Proposition 5.** Differentiate the lagrangian (39) and set \(y_1 = y_2 = 0\) and note there is no longer an incentive constraint corresponding to \(\lambda_2\) and that the incentive constraint correspond-
Proof of Lemma 2. The result requires that human capital in both neighbourhoods be decreasing in \( \delta \) at the policy \( \delta_E \). We can solve for

\[
\frac{\partial L_p}{\partial \delta} = \frac{\bar{\theta}(h_p^2 \frac{\partial h_p}{\partial \delta} + h_r^2 \frac{\partial h_r}{\partial \delta})}{8(h_p + h_r)^2}
\]

At the policy \( \delta_E \) we have \( h_p = h_r \) by definition so that this term is negative whenever \( \frac{\partial h_p}{\partial \delta} + \frac{\partial h_r}{\partial \delta} < 0 \) which it is at \( \delta_E \). In fact as noted earlier, this is the impact on total human capital which is negative for any \( \delta \geq \delta^* \).

Similarly, we can derive the change in the externality in the rich neighbourhood evaluated at the equal opportunity policy

\[
\frac{\partial L_r}{\partial \delta}(\delta_E) = \frac{3\bar{\theta}}{8} \left( \frac{\partial h_p}{\partial \delta} + \frac{\partial h_r}{\partial \delta} \right)
\]

which is of course negative as shown above.

Proof of Proposition 6.

Total welfare consists of welfare in periods 1 and 2 where \( \hat{\delta}_1(\delta_0) \) is the optimal policy in period 1 that maximizes \( W_2^* \).

\[
W = \max_{\delta_0} \left( \int_{\delta_0}^{\delta_1} \left( [\theta h(L_p^0, \delta_0)]^\sigma + [\theta h(L_r^0, 1 - \delta_0)]^\sigma \right) dF(\theta) \right) + \int_{\theta}^{\hat{\delta}_1(\delta_0)} \left( [\theta h(L_p^1, \hat{\delta}_1)]^\sigma + [\theta h(L_r^1, 1 - \hat{\delta}_1)]^\sigma \right) dF(\theta)
\]

First we note that both \( W_1(\delta_0) \) and \( W_2^*(\delta_0) \) are concave in \( \delta_0 \). We omit this here, although this is not too difficult to show, the latter requires some manipulation. Define the optimal “static” policy in period 0 (choice of \( \delta_0 \)) as \( \hat{\delta}_0 \), which optimizes \( W_1 \) and ignores the impact on the future generation (through the impact on neighbourhood composition). Thus \( \hat{\delta}_0 \) is defined analogously to \( \hat{\delta}_1 \). Consider the impact of \( \delta_0 \) on period 2 welfare

\[
\frac{\partial W_2^*}{\partial \delta_0} = \frac{\partial L_p}{\partial \delta_0} \left( h(L_p^0, \hat{\delta}_1)^{\sigma - 1} \frac{\partial h(L_p^0, \hat{\delta}_1)}{\partial L} \right) + \frac{\partial L_r}{\partial \delta_0} \left( h(L_r^1, 1 - \hat{\delta}_1)^{\sigma - 1} \frac{\partial h(L_r^1, 1 - \hat{\delta}_1)}{\partial L} \right)
\]

The first (second) term is the impact of a change in \( \delta_0 \) period 2 welfare in the poor (rich) region. This effect is the result of education policy on the composition of each neighbourhood.

By definition, \( L_1 = L_p^0 + L_r^1 \) so that \( \frac{\partial h_r}{\partial \delta_0} = \frac{\partial L_r}{\partial \delta_0} \). Use this to substitute for \( L_r^1 \) and rewrite
\[
\frac{\partial W_2^*}{\partial \delta_0} = \frac{\partial L_1^p}{\partial \delta_0} \left[ h(L_1^p, \hat{\delta}_1)^{\sigma-1} \frac{\partial h(L_1^p, \hat{\delta}_1)}{\partial L} - h(L_1^r, 1 - \hat{\delta}_1)^{\sigma-1} \frac{\partial h(L_1^r, 1 - \hat{\delta}_1)}{\partial L} \right] + \frac{\partial L_1}{\partial \delta_0} \left[ h(L_1^r, 1 - \hat{\delta}_1)^{\sigma-1} \frac{\partial h(L_1^r, 1 - \hat{\delta}_1)}{\partial L} \right]
\]  
(70)

As \( \frac{\partial L_1}{\partial \delta_0} \leq 0 \) for any \( \delta \geq \delta^* \), the term on the right-hand side is less than or equal to zero. Further, the term in brackets on the left-hand side is positive by assumption 1. Thus a sufficient condition for part 2 of the proposition to hold is \( \frac{\partial L_1^p(\hat{\delta}_0)}{\partial \delta_0} \leq 0 \). Thus, whenever the impact on the poor neighbourhood is negative, the static optimal \( \hat{\delta} \) is too large. This is quite weak in that even when this term is positive it may still be outweighed by the negative impact on the rich.

Consider the following

\[
\frac{\partial L_1^p}{\partial \delta} = \frac{\bar{\theta}(h_2^p \frac{\partial h_2^p}{\partial \delta} + h_2^r \frac{\partial h_2^r}{\partial \delta})}{8(h_p + h_r)^2}
\]

so that \( \frac{\partial L_1^p}{\partial \delta} \) is negative when

\[
\left( \frac{h_r}{h_p} \right)^2 < \frac{\partial h_r}{\partial \delta} \frac{\partial h_r}{\partial \delta}
\]

The first-order condition from the static problem (which determines \( \hat{\delta}_0 \)) can be written

\[
\left( \frac{h^p}{h^r} \right)^{\sigma-1} = \frac{\partial h_r}{\partial \delta} \frac{\partial h^r}{\partial \delta}
\]

The condition is

\[
\left( \frac{h^p}{h^r} \right)^{\sigma-1} > \left( \frac{h^r}{h^p} \right)^2
\]

which is true whenever \( \sigma < -1 \) or \( \rho > 2 \). This proves the second part of the proposition. To show part 1, note that part 2 implies that the left-hand side of (70) is positive. Furthermore, \( \frac{\partial L_1}{\partial \delta_0} = 0 \) when \( \rho = 0 \) from the first-order conditions on the static problem. Thus \( \frac{\partial W_2^*}{\partial \delta_0} \) is positive when \( \rho = 0 \).
Proof of Corollary 2. As above, the impact on welfare of a change in $\delta_0$ is

$$\frac{\partial W_2^*}{\partial \delta_0} = \frac{\partial L_1^p}{\partial \delta_0} \left[ h(L_1^p, \hat{\delta}_1)^{\sigma-1} \frac{\partial h(L_1^p, \hat{\delta}_1)}{\partial L} - h(L_1^r, 1 - \hat{\delta}_1)^{\sigma-1} \frac{\partial h(L_1^r, 1 - \hat{\delta}_1)}{\partial L} \right]$$

$$+ \frac{\partial L_1}{\partial \delta_0} \left[ h(L_1^r, 1 - \hat{\delta}_1)^{\sigma-1} \frac{\partial h(L_1^r, 1 - \hat{\delta}_1)}{\partial L} \right]$$

Substituting $\frac{\partial h}{\partial L} = \frac{\partial h}{\partial e} K$, the first term on the left-hand side becomes

$$\frac{\partial L_1^p}{\partial \delta_0} \left[ h(L_1^p, \hat{\delta}_1)^{\sigma-1} \frac{\partial h(L_1^p, \hat{\delta}_1)}{\partial \delta_1} - h(L_1^r, 1 - \hat{\delta}_1)^{\sigma-1} \frac{\partial h(L_1^r, 1 - \hat{\delta}_1)}{\partial \delta_1} \right] K$$

which equals 0 by the f.o.c that defines $\hat{\delta}_1$. The sign of the second term depends solely on $\frac{\partial L_1}{\partial \delta_0}$. The optimal policy increases $\delta_0$ until the benefit from doing so equals the cost (to the poor neighbourhood). The f.o.c determining $\hat{\delta}_0$ can be written (using $\frac{\partial h}{\partial L} = \frac{\partial h}{\partial e} K$)

$$\left( \frac{h_p}{h_r} \right)^{\sigma-1} = \frac{\partial h_r}{\partial L}$$

Further, for any $\delta \in [0, \delta_E]$ we know that

$$1 \leq \left( \frac{h_p}{h_r} \right)^{\sigma-1} = \frac{\partial h_r}{\partial L} \leq 1$$

where the second inequality comes from the fact that $h(L, e)$ is concave in $L$ and $L^r \geq L^p$. Thus $h_p = h_r$ and the optimal policy is $\delta_E$. Also, we have $\frac{\partial h_r}{\partial L} = \frac{\partial h_r}{\partial e}$ which implies that $\frac{\partial L_1}{\partial \delta_0} = 0 \quad \square$
References


