UNIVERSITY OF ALBERTA
FACULTY OF ARTS
Department of Economics

Working Paper No. 2010-15

Teach a Man to Fish? Education vs. Optimal Taxation

Eric Stephens
University of Alberta

Revised February 2012

Teach a Man to Fish? Education vs. Optimal Taxation

Eric Stephens*

Department of Economics, University of Alberta †

February 27, 2012

Abstract

In models of redistribution, differences in human capital are often the relevant source of heterogeneity amongst individuals. Presumably, the distribution of human capital can be manipulated through education spending. This paper examines the use of education as a redistributive tool when there is a nonlinear tax system in place. The results show that taxation, whether under full or asymmetric information, substantially reduces the redistributive role of education spending in maximizing social welfare. This points to a conflict between the equalization of utility and human capital outcomes.

Keywords: Optimal nonlinear taxation, redistribution, education

JEL Classification Numbers: D63, H21

*I am grateful to Robin Boadway, Charles Beach, David Byrne, Mike Hoy, Marcelin Joanis, Adnan Khan, Pierre Pestieau, Sumon Majumdar, Afrasiab Mirza and James Thompson for helpful comments and discussions as well as and seminar participants at Queen's University, the University of Alberta, Australian National University, Singapore Management University, Wilfrid Laurier University, Les Journées Louis-André Gérard-Varet (2010) and the Canadian Public Economics Group meetings (2010). I am especially indebted to the referees who provided excellent feedback and helped to substantially improve the paper.

†Address: Department of Economics, University of Alberta, Edmonton AB, Canada.T6G 2H4. Telephone: 1-780-492-4405, E-mail: *eric.stephens@ualberta.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

Education has long been seen as integral to the health of a society. Schooling increases productivity and can also facilitate social mobility. In this way it seems a natural tool for redistribution 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 unclear.

Following Arrow (1971), denote policies which spend more (less) resources on those with lower individual endowments as progressive (regressive). A more progressive education policy generally represents an attempt to create more equality in individual productivity. There are numerous reasonable justifications for such policies, however these tend to extend beyond conventional models of redistribution. This paper is an attempt to characterize the role of education spending in a standard (welfarist) redistribution problem. To do this we analyze education spending decisions when individuals receive subsequent transfers through the tax system. In an initial period human capital is determined by a technology which maps individual characteristics and public spending on education into productivity. Thus, the distribution of skills is influenced by the policy maker through the education spending decision. Once the education spending choice is made the government employs cash transfers through the tax system.

If there were only an education spending choice and no subsequent transfers, all redistribution would come through education. In this case a progressive policy will be optimal for a planner with a sufficient aversion to inequality.[1] When individuals are subject to a redistributive tax system once educated, as they generally are, the analysis becomes more complicated. If welfare is ultimately determined by outcomes under the tax system, then we must characterize the impacts of changes to the distribution of skills in the tax problem. Generally considered exogenous, the distribution of skills clearly plays an important role in the optimal income tax problem. In his seminal paper, Mirrlees (1971) conjectures:

> "The results seem to say that, in an economy where there is more intrinsic inequality in economic skill, the income tax is a more important weapon of public control than it is in an economy where the dispersion of innate skills is less. The reason is, presumably, that the labour-discouraging effects of the tax are more important, relative to the redistributive benefits, in the latter case."

It is plausible that greater equality in productivity, which implies less redistribution through taxation, could mean less distortion to labour markets and an increase in social welfare. This

[1] For a general discussion of the case with a utilitarian social objective see Arrow (1971).

suggests that an education policy which results in a more equal distribution of human capital may be optimal. The analysis presented below suggests the contrary in that optimal education policy, which may be progressive in the absence of cash transfers, is regressive due to taxation. In other words, transfers through the tax system imply a reduced role for education in equalizing utility. This is true whether taxation is distortionary or not.

The intuition for this result is fairly straightforward when cash transfers are made under full information. In this case the role of education in equalizing utility is reduced because taxation is not distortionary and high levels of education spending on those with low endowments is costly and unnecessary. When cash transfers are made under asymmetric information (so that taxation is distortionary) the role of education spending is less clear. Contrary to the above conjecture, increasing the relative productivity of low types can actually increase the distortion caused by taxation. With only two types of individuals this is unambiguous and is explained by the rising cost of incentive compatibility in the tax problem between increasingly homogeneous individuals. In a richer framework with many types and multiple sources of individual heterogeneity this may not be the case. For example when tax treatment can be conditioned on education spending, as in the literature on tagging, the effect of asymmetric information is ambiguous so that distortionary taxation can imply a more progressive education policy than the full information case. Even so, optimal funding is still regressive under both informational assumptions.

The results described above suggest that more progressive education policies are not optimal because they reduce inequality in human capital outcomes which hinders the redistribution of cash. Given this, we also perform a related comparative statics exercise which directly considers the welfare effect of a mean-preserving reduction in the dispersion of the wage distribution. In accordance with the previous analysis we see that a reduction in wage inequality (which can be interpreted as resulting from a progressive education policy) is indeed welfare reducing in the presence of cash transfers under both full and asymmetric information.

It is important to stress that there are many reasonable arguments to support progressive education policies that go beyond the framework that is employed in this paper. For example, we may view human capital as intrinsically valuable for a variety of reasons such as self-esteem or health outcomes, or consider the impacts of the distribution of human capital on externalities such as voter savvy or crime.[2] Furthermore, progressive policy prescriptions may result from political considerations or from the use of different normative criteria as in the so-called "equal opportunity" literature.[3] Since we purposely abstract from these and other important issues to gain clear results we must be careful not to interpret the conclusions too literally. That said, consumption and leisure outcomes are presumably important to any social objective and the analysis points to a conflict between the equalization of educational attainments and the maximization of social welfare.

[2] For example see Usher (1997).

[3] For an introduction to many of the relevant issues see Roemer (1998) or Fleurbaey and Maniquet (2005).

Review of the literature

In the analysis of non-linear taxation pioneered by Mirrlees (1971) the relevant heterogeneity in the population is over individual productivity or skills. In the literature the distribution of skills is generally considered fixed, which is relaxed in this paper. In particular, the model is used to analyze redistributions through individual skills or wages as well as taxes. This has implications for the literature on optimal taxation but can also be viewed as a contribution to the discussion on education policy.

The distribution of skills has important implications for the optimal tax problem.[4] Using a simulation approach, Kanbur and Tuomala (1994) analyze the effect of inequality in skills on the shape of the optimal tax scheme. More recent theoretical contributions include Brett and Weymark (2008) and Simula (2010), who consider the impact of changes in productivities to the optimal tax scheme in a model with quasi-linear preferences (the former linear in leisure, the latter in consumption) and a weighted utilitarian objective function. In contrast, the focus of this paper is in characterizing the effects of changes to the distribution of skills on social welfare rather than the shape of the optimal tax scheme per se.

There is a large literature on education and redistribution that is relevant to the analysis presented here. In a seminal paper, Arrow (1971) analyzes the division of program spending when there is no further redistribution and describes the progressivity/regressivity of optimal policy (although his analysis is not specific to education). Bruno (1976), Ulph (1977) and Hare and Ulph (1979) extend this to focus on education and assume that schooling takes place in an initial period and taxation/redistribution in a second stage, which is assumed here as well. This framework generally implies that the optimal policy is that which maximizes the size of the output which is then redistributed through the tax system. As increasing total output often means spending on those with better endowments these models imply more regressive education spending. In deriving their results, the authors make strong informational assumptions and/or assume a fixed labour supply. These conclusions are similar in spirit to those presented in this paper, particularly the case where tax redistributions take place under perfect information and lump-sum taxes are employed. This paper extends their conclusions to the more standard income taxation setting in which individuals make labour supply decisions along an intensive margin and the tax scheme must be incentive compatible.

In a more recent paper, Krause (2006) shows that subsidies for "white collar" education, which can be viewed as a regressive education policy, may be consistent with an optimal redistributive program. He argues that subsidies may be optimal as they facilitate incentive compatibility in labour supply decisions (i.e. they increase differences in skills which in turn reduces the desire of high types to mimic low types in the income tax problem). Cremer et al. (2011) also consider a redistribution problem with endogenous productivity which is closely related to this paper. They

[4] A good illustration of this is found in Boadway et al. (2000), who derive qualitative properties of the tax function for various skill distributions.

show that with two identical individuals whose wages are endogenous, an equal wage policy will be inferior to one with extreme inequality in which all resources are spent on only one individual. The authors use this example to argue that inequality in productivities can be welfare improving due to non-convexities in the optimal tax problem. This is the force behind the results presented in this paper, as well as Krause (2006) and is generalized here in a variety of important ways. Relative to Cremer et al. (2011), we characterize education policy more generally rather than focusing on corner solutions. Furthermore, we allow for diminishing returns to education spending which can have non-trivial implications in the two-type case. More importantly the model incorporates many individual types (although the two-type case is discussed in-depth), which is interesting for a few reasons. With only two types, as in both Krause (2006) and Cremer et al. (2011), the optimal marginal income tax is regressive. The low-productivity type faces a positive marginal income tax and the high-productivity type does not, which is the standard "no distortion at the top" result. With many types this is not the case and we may have a progressive marginal income tax scheme (at least over some range of types), which is what we observe in all developed countries. With progressive marginal income taxation the labour supply of more productive types is more distorted, which is presumably when increasing the relative productivity of lower types will be most beneficial. Along with many types, more than one type of individual heterogeneity in the ability to benefit from education is also incorporated in the model. This allows for a more interesting set of informational assumptions; in particular we can allow the policy maker to condition taxes on educational policy as in the literature on tagging.

Finally, Fleurbaey et al. (2002) consider a model with educational investments that is also relevant here. The authors use a mechanism-design approach to consider the progressivity/regressivity of schooling expenditures where ability is private information and agents have different talents. The planner is endowed with a fixed sum of money that is distributed either in cash or in kind (through educational "help" which reduces the cost of acquiring education). They show that when effort and help are substitutes the second-best optimum involves more education spending on low-types (at the expense of less effort from these types), along with money transfers to reduce consumption inequality. If effort and help are complementary, then the high-types receive more education help. Thus the regressivity/progressivity of education spending depends on the particular specification. In either case, however, education levels and individual efforts increase with talent, so that policy is always output regressive as defined by Arrow (1971). Interestingly, they find that the more averse to inequality is the planner the more inequality we see in educational attainment. So in this case as well equalization of individual educational attainments may conflict with that of utility levels, albeit for somewhat different reasons.

The paper proceeds as follows: Section 2 outlines the general environment. Section 3 considers a simplified model that portrays the intuition and motivation for sections 4 and 5, where a more general approach is discussed. Section 6 concludes.

2 Environment

Social preferences aggregate individual utility through a standard concave social welfare function denoted $\Psi(\cdot)$. Individual utility is defined by the concave function $u(c, l)$, where the bundle $(c, l) \in \mathbb{R}_+ \times \mathbb{R}_+$ denotes units of consumption and labour supplied. Consumption consists solely of a single good whose price is normalized to one. It is assumed that the labour market is competitive and individuals are paid a wage w. At a given wage each individual optimizes utility and supplies $l(w)$ units of labour, which implies a pre-tax income of $y(w) = wl(w)$. Pre-tax income could also be considered one's labour supply in efficiency units. As is common in the optimal tax literature we rewrite utility in terms of consumption and pre-tax income $u(c, l) = u\left(c, y/w\right)$.

Education spending is targeted to specific locations and takes place in an initial time period, after which students enter the working world where they pay taxes and receive transfers. Our focus is on the allocation of a fixed education budget across two districts, where e_j is defined as the amount of educational spending in district $j \in \{p, r\}$. The subscripts $\{p, r\}$ represent low-productivity "poor" and high-productivity "rich" respectively. The budget is normalized to one and the fraction of education funds spent in the low-productivity district is denoted δ, so that $e_p = \delta$ and $e_r = 1 - \delta$.[5]

Wages, denoted $w^j(\theta_i, e_j)$, are increasing and concave in government education spending e_j and individual characteristics θ_i, where $i = 1, ..., \frac{n}{2}$ and $\theta_1 < ... < \theta_{\frac{n}{2}}$.[6] Let individuals from the rich location have a higher earning capacity, so that $w^r(\theta, e) > w^p(\theta, e)$ for any (θ, e) pair. This notation is chosen to capture the positive empirical relationship between parent's socioeconomic status and their children's outcomes.[7] It is assumed that location is observable at the time of education spending. Individual characteristics θ_i are never known to the policy maker, either at the time of education or during working life in which redistributions can be made through the tax system.

[5]Allowing for an endogenous level of spending is relatively straightforward but yields little in the existing framework. Note also that education funds in the model are distributed by a central authority which is not the case in practice (at least not everywhere and not totally). A political model of local education funding is beyond the scope of this paper, so we simply note that spending can be interpreted as a set of transfers across districts imposed by a central government. Such transfers have been the source of heated debate in recent years since budgets are generally funded largely by the local tax base which can lead to sizable heterogeneity in resources across communities. For a discussion of these and related issues, see Berne (1988) or Fernandez and Rogerson (1996).

[6]It is assumed that wages are exogenous from the individual's perspective. This approach is in line with Roemer (1998) and is consistent with primary/secondary education if we take the view that children's aversion to schoolwork is irrelevant and may be more restrictive when considering higher education. Relevant papers analyzing the impact of taxation on educational investments include Maldonado (2008), Bovenberg and Jacobs (2005), Fleurbaey et. al (2002), Boadway et. al (1996).

[7]There is a vast empirical literature on the impact of "local effects" on children's academic outcomes (examples include Goux and Maurin (2007), Ding and Lehrer (2007) or Hoxby (2000)) as well as the impact on entrepreneurial success (see Gomez and Santor (2001)). Related theoretical contributions include C. de Bartolome (1990) and Benabou (1996a,b).

3 Two types

If there is no individual heterogeneity within each district, then wages are a function solely of the location in which one is educated. Thus, there are two types as there are two districts. Once working, students educated in the high-productivity district receive wage $w^r(1-\delta)$ while those from the low-productivity district receive $w^p(\delta)$. Further, define the elasticity of wages with respect to education funding as $\epsilon_w^j = e_j/w^j(e_j) \times dw^j(e_j)/de_j$. Throughout the analysis it is assumed that $\epsilon_w^j = \epsilon_w$ is constant and equal for both types, the implications of this are discussed below.

Along with the optimal education policy we are also interested in two other policies (which may or may not be optimal). The equal funding policy $\delta = 1/2$ and what we refer to as the egalitarian policy δ_E. When δ is greater (less) than 1/2 education spending is progressive (regressive). The egalitarian education policy is defined as that which generates the most equal distribution of wages possible across the two districts. When there are only two types this is simply the policy which equalizes wages across the two locations, so that $w^r(1-\delta_E) = w^p(\delta_E)$. Throughout, we restrict attention to $\delta \in [0, \delta_E]$ for obvious reasons. Note that $w^r(x) > w^p(x)\ \forall x \Rightarrow \delta_E > 1/2$.

In the optimal tax literature it is standard to assume there is asymmetric information. In particular, that the government can observe income but not the wage rate or labour supply of individuals. Indeed, tax schemes are generally a function of one's reported income and not earning capacity. Given this, it is instructive to study optimal education spending both when cash transfers are made under full and asymmetric information. We refer to the former as the first-best and the latter as second-best. Regardless of the informational assumptions in the tax problem it is assumed that the policy maker can observe location when undertaking education spending. This implies that educational transfers can always be made directly, unlike cash transfers which must be incentive compatible under asymmetric information.

3.1 No asymmetric information: first-best

Consider the problem recursively and note that the initial education choice impacts welfare which is defined as the solution to the optimal tax problem. Under perfect information this is characterized by

$$W^{fb} = \max_{c^p, c^r, y^p, y^r} \Psi(v^p) + \Psi(v^r) \quad \text{s.t} \quad y^p + y^r \geq c^p + c^r, \tag{1}$$

in which we denote individual utility $v^j = u(c^j, y^j/w^j)$. Using subscripts to denote partial derivatives, the first-order conditions can be manipulated to obtain.

$$u_1\left(c^j, \frac{y^j}{w^j}\right) + u_2\left(c^j, \frac{y^j}{w^j}\right)\frac{1}{w^j} = 0 \quad \text{for } j = p, r \tag{2}$$

This problem is analogous to that where the government uses lump-sum transfers to maximize welfare, as can be seen in expression (2) which is equivalent to individual utility maximization in the absence of distortionary taxes.[8] To analyze the education spending problem we characterize the effect of changes in δ on welfare, as defined by the solution to (1). The following lemma describes the optimal fraction of educational resources spent on each location when the policy-maker has full information.

Lemma 1 *The first-best education policy, denoted* δ^{fb}*, is regressive. Specifically,* $\delta^{fb} < \frac{1}{2}$.

Proof. *See Appendix.*

The optimal policy will depend on the social objective, the assumptions about the wage technology and the form of utility. Regardless of the specifics it will never be optimal to spend a larger fraction of educational resources on the low-productivity type.[9] This is ensured by the constant elasticity assumption made earlier and a less restrictive sufficient condition is described in the proof of Lemma 1. It is important to recall at this point that without cash transfers there is always a case for progressive spending anywhere up to δ_E (which would be optimal with a sufficient aversion to inequality). When cash transfers can be made under full information, optimal education spending is that which maximizes the size of the pie and consequently spends less on those who are less productive. In other words, progressive education spending is costly and unnecessary when cash transfers are not distortionary. With asymmetric information in the tax stage it is not clear that this will be the case as taxation is distortionary and it is plausible that there is a greater role for education in equalizing utility.

3.2 Asymmetric information: second-best

The optimal taxation literature generally assumes that policy makers do not have information regarding individual wages. When this is the case the cash redistribution scheme must be incentive compatible, which results in the use of distortionary taxation. In the context of the model this implies that the government can spend on education knowing which is the high and low type (presumably this takes place in an earlier period), but once they enter the workforce they can not (or will not) differentiate between the two.

The limited record-keeping assumption is reasonable because tracking educational expenditures requires a very strong degree of coordination between various levels of government over a large period of time. Furthermore, conditioning taxation on childhood location in a more realistic model may violate the principle of horizontal equity. For example, imagine a situation with two individuals whose wages are a function of education obtained early in life and unobservable random "luck" later in life. Suppose the individual from the poor location was "lucky" and the individual from the rich location was "unlucky", so that both end up with equal earning capacity. It is not clear that the individual who attended school in the poor district deserves preferential tax treatment if good luck

[8] Lump-sum transfers will generally influence the supply of labour through income effects. Here we are referring to the distortions caused by positive marginal tax rates.

[9] Note that we cannot rule out corner solutions, and in fact for a large class of utility and human capital functions $\delta^{fb} = 0$.

is as arbitrary as place of birth. However, it is important to note that educational transfers may still be justified at the time of schooling if the poor individual is worse off ex-ante. In short, it is very difficult to implement or justify tax policy that is based on anything but current information. This is presumably why income tax systems are designed this way and limited record-keeping is commonly assumed in the literature. In the model presented here; if the planner were able to keep track of spending in the initial stages then with two types we are simply back to the first-best problem where lump-sum transfers are employed. For the more general case in section 4 where there is unobserved individual heterogeneity the results of relaxing this are less obvious and are modeled below.

As described in Lemma 1, education spending is regressive because spending on low types is inefficient and cash transfers are not. Is there is a greater redistributive role for educational transfers when cash transfers are distortionary? To analyze the impacts of education in the second-best we first define W^{sb} to be the solution to the optimal tax problem in the second stage. This is the same problem outlined in (1) with the following additional incentive constraint.

$$u\left(c^r, \frac{y^r}{w^r}\right) \geq u\left(c^p, \frac{y^p}{w^r}\right) \tag{3}$$

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.[10] Denoting the Lagrange multipliers for the budget and incentive constraints λ and γ respectively, the first-order conditions can be manipulated to obtain

$$u_1\left(c^p, \frac{y^p}{w^p}\right) + u_2\left(c^p, \frac{y^p}{w^p}\right)\frac{1}{w_p} = \gamma\frac{u_1\left(c^p, \frac{y^p}{w^r}\right) + u_2\left(c^p, \frac{y^p}{w^r}\right)}{\Psi'(v_p)} \geq 0 \tag{4}$$

$$u_1\left(c^r, \frac{y^r}{w^r}\right) + u_2\left(c^r, \frac{y^r}{w^r}\right)\frac{1}{w_r} = 0. \tag{5}$$

These expressions are the analogue to (2) in the first-best. With asymmetric information the left side of each expression is no longer set equal to zero for both individuals. The term on the right hand side of equation (4) is strictly positive for the low type whenever the incentive constraint binds ($\gamma > 0$), which implies a positive marginal tax rate. On the other hand, the expression for the high type is that found in the first-best, which is the familiar "no distortion at the top" condition.

Solving directly for the optimal second-best education policy, which we denote δ^{sb}, is not feasible without restrictive assumptions. However, we can characterize the salient features of the optimum in the following proposition. The proof requires some restrictions on the individual utility function which are not required for the other results in the paper and are discussed in the appendix. The difficulty arises because the effects on individual utility resulting from a change in δ depend on the

[10] We ignore the incentive constraint that ensures low types do not wish to mimic high types. It is straight-forward to show that this is redundant.

consumption/income bundle they receive under the optimal tax. These bundles can be characterized but not derived explicitly at this level of generality. A sufficient, but not necessary condition for the proposition to hold is $u_{12} \geq 0$. This captures a large class of functions commonly used in the tax literature, such as Cobb Douglas or additive separable utility.

Proposition 1 *When the first-best is not implementable, so that the incentive constraint in the tax problem is binding,* $\delta^{sb} < \delta^{fb} < \frac{1}{2}$. *Asymmetric information in the tax problem leads to a more regressive education policy.*

Proof. *See Appendix.*

In deriving this result, we decompose the effects of a change in δ into "direct" and "indirect" effects. The former being the effect from a change in wage on each individual utility directly. The latter representing the effect on the incentive constraint from the tax problem (see the proof of Proposition 1 for details). The direct effect is positive when the aversion to inequality is large enough but the effect on incentives is always negative. This can be explained by considering the solution to the optimal tax problem which implies a positive marginal tax on the low type (which discourages working) to maintain incentive compatibility. Thus, the value in increasing the relative wage of the low type cannot be fully realized in the second-best because of the downward distortion caused by the tax. In fact, increases in δ will tighten the incentive constraint in the tax problem as mimicking will be more attractive. Since more equal wages increase the deadweight cost of taxation, cash is even more attractive relative to education for redistribution in the second-best.[11]

The proposition holds when the first-best is not implementable and thus an important exception and a case where progressive funding can be optimal is at the corner δ_E. The egalitarian policy is a local optimum because there is no need for redistribution once productivities are completely equalized. Cremer et al. (2011) contrast the two extreme policies (corner solutions) that roughly correspond to $\delta = 0$ and $\delta = \delta_E$ in the model presented here. The authors argue that wage differentiation ($\delta = 0$) is preferred to the equal wage solution ($\delta = \delta_E$). This result follows from restrictions on the wage technology, without which we cannot rule out δ_E as an optimum.[12] At first glance this seems to provide some support for an egalitarian policy. However, this arises because there are only two types in the model and the planner can completely eliminate tax transfers. Section 4 considers a more realistic environment where there is unobservable heterogeneity across types so that there are always ex-post differences amongst individuals. When this is the case the policy maker can never completely eliminate cash redistributions, implying that the egalitarian policy is no longer a local optimum.

[11] Maldonado (2008) considers an environment similar to that outlined in Section 3 in which the elasticity of wages with respect to education is not fixed, but the result of individual choice. This elasticity is shown to have implications for both income taxes and education subsidies. Fixed elasticity is not vital to Proposition 1 because of the limited record keeping assumption. The analysis in Maldonado (2008) assumes that taxation decisions are made with knowledge of education choices in the first stage, which is not the case here. Furthermore, unlike Maldonado (2008), education of high types may be over-provided at the optimum.

[12] δ_E can be ruled out as an optimum when diminishing returns in the wage technology and the disutility of labour are not too large.

4 Individual heterogeneity

With two types the optimal marginal income tax is regressive. Specifically, we see a positive marginal tax on the low type and a zero marginal tax on the high type. This was shown to result in a more regressive education policy. Does this extend to the case with many types in which the optimal marginal tax schedule can take a variety of forms? In particular, the optimal marginal tax schedule may be increasing at high levels of income, see for example Diamond (1998) or Boadway et. al. (2000). It seems reasonable that the return to raising the relative wages of those with lower productivity is higher if they are less distorted by the tax system. When there are many types, progressive marginal income taxes may be optimal (at least over some range) and it is not clear that the conclusions drawn in Section 3 will remain valid.

This section introduces a second source of heterogeneity by means of the individual-specific endowment θ_i, where $i = 1, ..., \frac{n}{2}$ and $\theta_1 < ... < \theta_{\frac{n}{2}}$. This could represent innate talents, good family environment, charisma, or any such combination of the like. It is assumed for simplicity that the distribution of θ is the same in each district. The exact interpretation of θ is immaterial, what is important is that wages differ across individuals for reasons that are not observable. Individuals continue to differ with respect to location as described in section 3, which is observable to the policy maker when undertaking education spending. Wages will now depend on both individual endowments and location and are defined by

$$w^{ij} = \theta_i w^j(e_j), \tag{6}$$

where the function $w^j(e_j)$ is unchanged from section 3.[13] We continue to assume there are two locations so there are a total of n individual types. The egalitarian policy δ_E again represents the upper bound on education spending and is defined as in section 3 by $w^p(\delta_E) = w^r(1-\delta_E)$. Thus, δ_E completely equalizes the local component of wages so that the distribution is equal across districts (those with the same value of θ receive the same wage).

4.1 No asymmetric information: first-best

Welfare in the first-best is now characterized by the solution to the following:

$$W^{fb} = \max_{c^{ij}, y^{ij}} \sum_i \sum_j \Psi(v^{ij}) \quad \text{s.t} \quad \sum_i \sum_j y^{ij} \geq \sum_i \sum_j c^{ij}, \tag{7}$$

[13]The multiplicative form implies that θ and educational spending are complementary inputs to individual wages. Hence, within each location higher θ students will have a higher marginal return to educational expenditures. There has been some empirical analysis regarding this issue but little in the way of conclusive results. References include Ashenfelter and Rouse (1998), Arias, Hallock, and Sosa (1999), Tobias (2003) and Martins and Pereira (2004). A similar assumption has been made in a number of previous relevant theoretical contributions, such as Hare and Ulph (1979) and Bovenberg and Jacobs (2005).

where individual utility is $v^{ij} = u(c^{ij}, y^{ij}/w^{ij})$. Manipulating the first order conditions yields

$$u_1\left(c^{ij}, \frac{y^{ij}}{w^{ij}}\right) + u_2\left(c^{ij}, \frac{y^{ij}}{w^{ij}}\right)\frac{1}{w^{ij}} = 0 \quad \text{for } \forall i, j, \tag{8}$$

which is akin to the simpler model in that expression (8) characterizes individual utility maximization in the absence of marginal income taxes. Given this we obtain the analogue of Lemma 1 which characterizes the optimal education policy in the first-best with many types and two sources of heterogeneity.

Lemma 2 *The first-best education policy, denoted δ^{fb}, is regressive. Specifically, $\delta^{fb} < \frac{1}{2}$.*

Proof. *See Appendix.*

As with the simpler model the optimal policy will depend on the social objective, the assumptions about the wage technology and the form of utility. Similarly, it is never optimal to spend a larger fraction of educational resources on the low-productivity type when the planner has full information.

4.2 Asymmetric information: second-best

The introduction of asymmetric information in the simpler model led to an even *more* regressive education policy. With many types, the marginal tax schedule may take a variety of shapes and the effects of education spending on incentives is much less obvious. Denote welfare in the second-best by W^{sb}, which is characterized by (7) with the following additional set of incentive constraints that ensure each individual attains at least as much utility from their own bundle as from mimicking another type.

$$u\left(c^{ij}, \frac{y^{ij}}{w^{ij}}\right) \geq u\left(c^{hk}, \frac{y^{hk}}{w^{hk}}\right) \quad \forall h, k \neq i, j. \tag{9}$$

In the absence of unobservable individual heterogeneity, knowledge of previous education policy provides a tax-setter with perfect information. Thus, in the model with two types the second-best problem is equivalent to the first-best. With unobservable differences amongst individuals there is still asymmetric information in the second stage even when tax authorities can track educational expenditures. Below, we consider the optimal education policy when income taxes can be conditioned on location and when they can not.

Let us first consider the case for which the policy maker cannot condition the tax scheme on education history. As there is no way to determine the ordering of the wage distribution we simply denote each of the n individuals by s. Let γ_s be the multiplier on the incentive constraint that restricts individual s from mimicking individual $s-1$. Focusing only the downward incentive constraints[14], the following characterizes the effects of a change in δ at the optimum.

$$\frac{\partial W^{sb}}{\partial \delta} = \underbrace{-\sum_{s=1}^{n} \Psi'(v^s) u_2\left(c^s, \frac{y^s}{w^s}\right) \frac{y^s}{(w^s)^2} \frac{dw^s}{d\delta}}_{\text{“Direct effect”}} + \tag{10}$$

$$\underbrace{\sum_{s=2}^{n} \frac{\gamma_s}{(w^s)^2} \frac{dw_s}{d\delta} \left[-u_2\left(c^s, \frac{y^s}{w^s}\right) y^s + u_2\left(c^{s-1}, \frac{y^{s-1}}{w^s}\right) y^{s-1} \right]}_{\text{“Indirect (incentive) effect”}} \tag{11}$$

The sign of each term depends entirely on $dw_s/d\delta$, as gross income $y(w)$ is non-decreasing in the wage. For any individual s in the low-productivity (high-productivity) district, this will be positive (negative). As with the simpler case, we can decompose the impact of a change in δ into “direct” and “incentive effects” (see the proof of Proposition 1 for details). The magnitude of these effects depends on the various functional forms for utility, the wage technology and the social welfare objective. Note that the direct effects are always positive with a high social aversion to inequality, which is obvious when $\Psi(\cdot)$ is maximin. Importantly, the incentive impacts which are strictly negative when $n = 2$ are of ambiguous sign and we do not have an analogue to Proposition 1.[15] The ambiguous impact of spending on incentives makes it more difficult to characterize δ^{sb} under these informational assumptions (Section 5 employs a reduced-form model to capture the impact of a more equal distribution of skills in a similar problem). We can however, derive the following result regarding the egalitarian policy.

Proposition 2 *With many types and unobservable heterogeneity, the egalitarian policy is too progressive.*

Proof. *See Appendix.*

To interpret this result, first consider the situation in which there is no redistribution through the tax system. In this case any equalization by the planner must be done through education spending. With only the one instrument progressive policies up to and including δ_E are optimal under a social welfare objective with sufficient aversion to inequality. Proposition 2 states that in the more general case, the introduction of cash transfers implies that δ_E is *never* optimal regardless of social preferences (recall from the discussion following Proposition 1 that the corner solution δ_E can not be ruled out as the optimum in the two type case). In this sense the redistributive role for education is reduced when individuals are subject to a redistributive tax scheme, even when cash transfers are distortionary.

Proposition 2 states that with taxation a complete equalization across districts must be justified by considerations that go beyond the framework considered here. As mentioned earlier, there is a

[14]This is the “normal” case considered in the literature, in which redistributions go from the rich to the poor.

[15]The sign of $dw_s/d\delta$ completely determines the sign of the incentive effect only when the restrictions on utility discussed in Proposition 1 are met. Regardless, the effects on incentives from a change in δ are ambiguous here.

variety of perfectly reasonable justifications for δ_E (or progressive policy more generally) that come from outside the model even when cash transfers are available. For example, if we view individual utility as being derived directly from human capital and not from the income it is presumed to generate then δ_E could again be optimal. Alternatively, δ_E can be also be justified through the use of a different normative criterion as in the literature on "equal opportunity".[16] If these types of arguments are taken seriously (and there is no reason they should not be), then the results imply a trade-off between various objectives.

The preceding result assumes that taxes cannot be conditioned on individual location. The following proposition considers the case where the tax-setter can observe the location in which one was schooled, as in the literature on tagging.[17]

Proposition 3 *When individuals in each observable group receive separate tax treatment (tagging by location), the optimal second-best education policy is regressive, so that $\delta^{sb} < \frac{1}{2}$. However, unlike the two-type case, optimal policy may be more or less regressive than the first-best.*

Proof. *See Appendix.*

The results imply regressive education funding as in the simpler two-type model while capturing the intuition behind progressive arguments in a more interesting way. Specifically, we see it is indeed possible for education spending to be more progressive in the second-best. This is true when the relative output from the poor district is higher with income taxation, which is possible if marginal taxes are progressive enough (details in the proof of Proposition 3). However, even if this is the case the optimal policy will still be regressive in that less educational resources are always spent on the poor location. Furthermore, we see that regardless of whether or not taxes are conditioned on location the egalitarian policy is never optimal when there is income taxation.

5 Continuum of types

In the framework presented above, more progressive education policies represent an attempt to generate a more equal distribution of wages. In this sense we can interpret a higher δ (a more progressive policy) as one which results in a reduction in the dispersion of the wage distribution and possibly the mean depending on the assumptions regarding the wage functions.[18] When there

[16] Following Roemer (1998), it is relatively straightforward to characterize the equal opportunity policy here. Because we have assumed individuals have homogeneous preferences, the equal opportunity policy is simply that which equalizes wages. With two types this is the egalitarian policy δ_E. The introduction of a second source of heterogeneity complicates the issue and the link is more tenuous. Hypothetically, the equal opportunity policy should equalize wages across both dimensions of heterogeneity as these are beyond the individual's control. We have assumed that differences in θ simply cannot be observed at the time of education and can interpret δ_E as an equal opportunity policy in the sense that it produces the most equal distribution possible.

[17] The seminal contribution is Akerlof (1978).

[18] With respect to the framework presented above, if we have a continuum of both rich and poor individuals the overall distribution is simply the convolution of the two. In the spirit of the previous discussion an increase in δ represented a transfer of educational resources from the higher productivity to the lower productivity district. Appropriate restrictions on the human capital technology will ensure that such a transfer will indeed reduce the dispersion of the overall wage distribution.

are two types this is straightforward and involves a transfer of educational resources from those with relatively high to low wages. For example, the egalitarian policy δ_E represents the extreme case in which there is no dispersion in the resulting wage distribution. The results thus far suggest that while such policies may be optimal in the absence of income taxation, they are not generally optimal when cash transfers are available. This section extends the analysis to a continuum of types and considers the impact of a reduction in the dispersion of the wage distribution, which is presumed to result from a more progressive education policy.

Let there now be a continuum of individuals who differ in their wages. Characterize the distribution of wages using the distribution function $F(w)$, for $w \in [\underline{w}, \overline{w}]$. Consider a subset of individuals $\mathcal{S} = [w_{\underline{s}}, w_{\overline{s}}]$, where $\underline{w} < w_{\underline{s}} < w_{\overline{s}} < \overline{w}$. Let the elements of $\mathcal{S}$ be centered around the mean w^m such that $w^m = \int_{\underline{w}}^{\overline{w}} w dF(w) = \int_{w_{\underline{s}}}^{w_{\overline{s}}} w dF(w)$. Normalize the mass of types belonging to the set $\mathcal{S}$ to one, so that $\int_{w_{\underline{s}}}^{w_{\overline{s}}} f(w)dw = 1$. To capture changes in the distribution, define the wage for individual $s \in \mathcal{S}$ as

$$w_s = (1 - \epsilon)w + \epsilon w^m. \tag{12}$$

Thus, $\epsilon = 0$ recovers the initial distribution of wages. For some $\epsilon > 0$, we will see a reduction in the dispersion of wages. The change is such that the average wage will remain constant, which can be interpreted as a costless redistribution of productivity from high to low types.[19] As with our earlier analysis we consider changes to the wage distribution when there are subsequent transfers through the tax system. However, this is considered directly rather than characterized through the first and second-best education policy. Specifically, we show that a mean-preserving reduction in the variance of wages reduces welfare whether taxation is distortionary or not. This is a valuable exercise because it allows us to capture the phenomena described in the previous two sections in a way that more closely relates to the large body of literature on taxation and the income distribution.

5.1 No asymmetric information: first-best

If individual type is known to the policy-maker in the tax stage then lump-sum transfers can be employed to achieve redistributive goals. Denote the lump-sum tax for type w (possibly negative) as T_w. Given T_w, individuals solve

$$\max_{l} \quad u(wl - T_w, l), \tag{13}$$

which yields labour supply $l(w)$ and indirect utility $V(w, T_w)$. For simplicity, we assume throughout this section that social preferences are utilitarian, thus transfers are chosen to maximize

$$W^{fb} = \max_{T_w} \int_{\underline{w}}^{\overline{w}} V(w, T_w) f(w) dw \quad \text{subject to} \quad \int_{\underline{w}}^{\overline{w}} T_w f(w) dw \geq 0. \tag{14}$$

[19] We ignore the efficiency loss that will likely arise when transferring educational resources amongst locations. Allowing for costly redistribution complicates the issue and would only serve to strengthen the results.

The following lemma considers the impact on welfare, described by the solution to (14), of a reduction in the variance of the wage distribution.

Lemma 3 *In the first-best, when individual wages are observable to the tax authority; an increase in ϵ, which results in a costless reduction in the variance of the wage distribution, is welfare-reducing.*

Proof. *See Appendix.*

The intuition for Lemma 3 is straightforward. An increase in ϵ results in a reduction in the wages of higher types while increasing that of lower types. Since higher types are producing more (with more labour supplied), this will reduce the size of aggregate output that can be redistributed, and have a negative impact on welfare.

5.2 Asymmetric information: second-best

That an increase in ϵ is welfare-reducing depends importantly on the first-best environment. In particular, redistribution takes place only over consumption outcomes and high wage types are providing relatively cheap labour to subsidize the consumption of lower types. As types are observable transfers ensure all individuals have the same marginal utility of consumption, but high types work more and thus have lower utility (see the proof of Lemma 3 for details). When the informational environment is changed so that individual type is unobservable to the tax-setter, this will not be the case. In particular, the incentive constraints ensure that higher types will be better off, and the implications of a lower variance on welfare are much less obvious.

The introduction of informational asymmetry in the continuous type case increases the complexity of the problem substantially. To simplify, in this sub-section we let utility be characterized by the quasi-linear in leisure function $u(c) - l$, where c is consumption, l is labour, and $u(\cdot)$ is increasing and concave. In solving the optimal tax problem the government chooses consumption-income bundles subject to the budget and incentive constraints. The budget constraint is represented by

$$\int_{\underline{w}}^{\overline{w}} [y(w) - c(w)]dF(w) \geq 0. \tag{15}$$

Before characterizing the incentive constraints we transform utility so that we can substitute $y(w)$ out of the problem. Define $V(w)$ as

$$V(w) = wu(c(w)) - y(w). \tag{16}$$

Using this transformation, write the incentive constraint for individual of wage type w

$$V(w) \geq V(w') + (w - w')u(c(w')) \quad \forall w'. \tag{17}$$

Incentive compatibility requires $V(w') + (w - w')u(c(w'))$ attain a maximum with respect to w' at $w' = w$. This yields the first-order incentive compatibility condition $\frac{dV(w)}{dw} = u(c(w))$. The second-order incentive compatibility condition requires that c be non-decreasing in w, or $\frac{dc(w)}{dw} \geq 0$.[20] The problem of a utilitarian planner is to maximize

$$W^{sb} = \max \int_{\underline{w}}^{\overline{w}} \frac{V(w)}{w} dF(w) \tag{18}$$

subject to

$$\int_{\underline{w}}^{\overline{w}} [wu(c(w)) - V(w) - c(w)] dF(w) \geq 0 \tag{19}$$

$$\frac{dV(w)}{dw} = u(c(w)) \tag{20}$$

$$\frac{dc(w)}{dw} \geq 0. \tag{21}$$

Define $z(w) \equiv \frac{dc(w)}{dw}$ and rewrite the second-order incentive compatibility constraint as $z(w) \geq 0$. In the control problem described in (18)-(21) we treat $z(w)$ as a control and $V(w)$ and $c(w)$ as state variables. The Hamiltonian function is

$$\mathcal{H}(w) = \frac{V(w)}{w} f(w) + \lambda[wu(c(w)) - V(w) - c(w)] f(w) + \pi(w) u(c(w)) + \mu(w) z(w) + \kappa(w) z(w),$$

where λ is interpreted as the shadow price of government funds (which is independent of w), $\pi(w)$ is the co-state variable associated with the equation of motion for $V(w)$, $\mu(w)$ is the co-state variable associated with $dc(w)/dw = z(w)$ and finally $\kappa(w)$ is the shadow value of the non-negativity constraint on $z(w)$. The following proposition considers the impact on welfare, described by the solution to the optimal tax problem outlined in (18)-(21), of a reduction in the variance of the wage distribution.

Proposition 4 *In the second-best, when individual wages are not observable to the tax authority; an increase in ϵ, which amounts to a costless reduction in the variance of the wage distribution, is welfare-reducing when individual utility is quasi-linear in leisure and social preferences are utilitarian.*

Proof. *See Appendix.*

In support of the previous results, we see that the direct increase in utility to lower types that arises from a transfer of productivity is more than offset by the loss in transfers through taxation. It is important to note that the result does not depend on the specific structure of the tax system.

[20] For an in-depth discussion of the optimal income tax problem with quasi-linear in leisure preferences, see Ebert (1992) or Boadway et al. (2000).

At this level of generality the marginal tax schedule can take any number of forms and in particular can be increasing at high levels of income.[21]

6 Conclusions

In the absence of transfers through the tax system there is always a case for a more equal distribution of productivity. When cash transfers are introduced the role for education spending in equalizing utility is diminished. This is most straight-forward when taxation takes place under full information. In this situation educational transfers to low types are costly and unnecessary as cash transfers are not distortionary. This leads to an optimal education spending policy that maximizes the size of the "pie" to be redistributed and consequently spends less on those with lower initial endowments (resulting in less equality in wages). In the more realistic case where cash transfers are made under asymmetric information the story is less clear. Intuition might suggest that greater equality in wages, which implies less redistribution through taxation, could mean less distortion to labour markets and result in an increase in social welfare. Contrary to this the results indicate that asymmetric information can lead to an even more regressive education policy. Even when asymmetric information leads to more progressive education spending, we still see a reduced redistributive role for education in the sense that policy is still regressive.

In conclusion it must be stressed that the merits of progressive education policies clearly extend beyond the framework employed in this paper. The value of education undoubtedly transcends its impact on consumption/leisure outcomes, both individually and in the aggregate. The point of this paper is not to argue for or against policies designed to bring about more equality in productivity, but rather to bring to light a potential conflict between various objectives.

7 References

Akerlof, G. (1978), "The economics of "tagging" as applied to the optimal income tax, welfare programs, and manpower training" *American Economic Review*, 68, 8-19.

Arias, O., Hallock, K. F., and Sosa, W. (1999), "Individual heterogeneity in the returns to schooling: Instrumental variables quantile regression using twins data", *Empirical Economics*, 26, 740.

Arrow, K. (1971), "A utilitarian approach to the concept of equality in public expenditures" *Quarterly Journal of Economics*, 85, 409-415.

Ashenfelter, O., and Rouse, C. (1998). "Income, schooling, and ability: Evidence from a new sample of identical twins", *Quarterly Journal of Economics*, 113, 253-284.

[21] This was shown by Diamond (1998) when preferences are quasi-linear in consumption. Boadway et. al (2000) show that this is also possible when utility is quasi-linear in leisure, as we have assumed here.

Benabou R. (1996a), "Equity and efficiency in human capital investment: The local connection", *Review of Economic Studies*, 63,237-264.

Benabou R. (1996b), "Heterogeneity, stratification, and growth: Macroeconomic implications of community structure and school finance", *American Economic Review*, 86, 584-609.

Berne R. (1988), "Equity issues in school finance", *Journal of Education Finance*, 14, 159-180.

Boadway R., Cuff, K., and M. Marchand (2000), "Optimal income taxation with quasi-linear preferences revisited", *Journal of Public Economic Theory*, 2 (4), 435-460.

Boadway R., Marceau, N. and M. Marchand (1996), "Investment in education and the time inconsistency of redistributive tax policy", *Economica*, 63, 171-189.

Bovenberg, L. and B. Jacobs (2005), "Redistribution and education subsidies are siamese twins", *Journal of Public Economics*, 89, 2005-2035.

Brett, C. and J. Weymark (2008), "The impact of changing skill levels on optimal nonlinear income taxes", *Journal of Public Economics*, 92 ,1765-1771.

Bruno (1976), "Equality, complementarity and the incidence of public expenditures", *Journal of Public Economics* 6, 395-407.

Caputo, M. (2005), *Foundations of Dynamic Economic Analysis*, Cambridge University Press.

Cremer, H., Pestieau, P. and M. Racionero (2011), "Unequal wages for equal utilities", *International Tax and Public Finance*, 18, 383-398.

de Bartolome, C. (1990), "Equilibrium and inefficiency in a community model with peer group effects", *Journal of Political Economy*, 98(1), 110-132.

Diamond, P. (1998), "Optimal income taxation: an example with a U-shaped pattern of optimal marginal tax rates", *American Economic Review*, 88(1), 83-95.

Ding, W. and S. Lehrer (2007), "Do peers affect student achievement in China's secondary school?", *Review of Economics and Statistics*, 89(2), 300-312.

Ebert, U. (1992), "A reexamination of the optimal nonlinear income tax", *Journal of Public Economics*, 49 (1), 47-73.

Fernandez, R. and R. Rogerson (1996), "Income distribution, communities and the quality of public education", *Quarterly Journal of Economics*, 111(1), 135-164.

Fleurbaey M. and F. Maniquet (2005), "Compensation and responsibility", in K. J. Arrow, A. K. Sen, K. Suzumura eds., *Handbook of Social Choice and Welfare*, vol. 2, North-Holland.

Fleurbaey M., Gary-Bobo R. and D. Maguain (2002), "Education, distributive justice, and adverse selection", *Journal of Public Economics*, 84, 113-150.

Gomez, R. and E. Santor (2001), "Membership has its privileges: the effect of social capital and neighbourhood characteristics on the earnings of microfinance borrowers", *Canadian Journal of Economics*, 34(4), 943-966.

Goux, D. and E. Maurin (2007), "Close neighbours matter: Neighbourhood effects on early performance at school", *Economic Journal*, 117, 1193-1215.

Hare P., and D. Ulph (1979), "Education and income distribution", *Journal of Political Economy*, 87, S193-S212.

Hoxby, C. (2000), "Peer effects in the classroom: learning from gender and race variation" NBER Working Paper no. 7867.

Kanbur, R., and M. Tuomala (1994), "Inherent Inequality and the Optimal Graduation of Marginal Taxes", *Scandinavian Journal of Economics*, 96(2), 275-282.

Krause, A. (2006), "Redistributive taxation and public education", *Journal of Public Economic Theory*, 8(5), 807-819.

Maldonado, D. (2008), "Education policies and optimal taxation", *International Tax and Public Finance*, 15(2), 131-143.

Martins, P., and P. Pereira (2004), "Does education reduce wage inequality? Quantile regression evidence from 16 countries", *Labour Economics*, 11, 355-371.

Mirrlees, J.A. (1971), "An exploration in the theory of optimal income taxation", *Review of Economic Studies*, 38, 175-208.

Roemer J. (1998), *Equality of Opportunity*, Harvard University Press.

Simula, L. (2010), "Optimal nonlinear income tax and nonlinear pricing: optimality conditions and comparative static properties", *Social Choice and Welfare*, 35(2), 199-220.

Tobias, J. L. (2003), "Are returns to schooling concentrated among the most able? A semiparametric analysis of the ability-earnings relationships", *Oxford Bulletin of Economics and Statistics*, 65, 1-29.

Ulph, D. (1977), "On the optimal distribution of income and educational expenditure", *Journal of Public Economics*, 8, 341-356.

Usher, D. (1997), "Education as a deterrent to crime", *Canadian Journal of Economics*, 30 (2), 367-384.

8 Appendix

Proof of Lemma 1. Welfare is denoted W^{fb}, defined by (1). The Lagrangian for this problem is given by

$$\mathcal{L} = \Psi(v^p) + \Psi(v^r) + \lambda(y^p + y^r - c^p - c^r). \tag{22}$$

The envelope theorem implies $\partial W^{fb}/\partial\delta = \partial\mathcal{L}/\partial\delta$. Using this, and substituting the expression for individual utility maximization (2), yields

$$\frac{\partial W^{fb}}{\partial\delta} = \lambda\left[\frac{y^p}{w^p}\frac{\partial w^p}{\partial\delta} + \frac{y^r}{w^r}\frac{\partial w^r}{\partial\delta}\right]. \tag{23}$$

We have assumed wages are given by a constant elasticity ϵ_w, so we can rewrite this

$$\frac{\partial W^{fb}}{\partial\delta} = \lambda\epsilon_w\left[\frac{y^p}{\delta} - \frac{y^r}{(1-\delta)}\right]. \tag{24}$$

Hence, $\partial W^{fb}/\partial\delta$ is negative whenever

$$\delta > \frac{y^p}{y^p + y^r} < \frac{1}{2}. \tag{25}$$

The second inequality holds as y is increasing in the wage (this is shown in the proof of lemma 3 below). Note that without the constant elasticity assumption, $\partial W^{fb}/\partial\delta$ is negative whenever

$$\delta > \frac{y^p\epsilon_w^p}{y^p\epsilon_w^p + y^r\epsilon_w^r}, \tag{26}$$

where ϵ_w^p and ϵ_w^r are the elasticities of the poor and rich respectively. Thus a less restrictive sufficient condition for regressivity in this case is $\epsilon_w^p(\delta) \leq \epsilon_w^r(1-\delta)$, for $\delta \in [1/2, \delta_E]$. In other words, the percentage change in wage from an increase in funding is lower for the poor type when they receive the majority of the funding. ■

Proof of Proposition 1. Employing the envelope theorem, we characterize the effects of a change in δ on welfare, as described in section 3.2.

$$\frac{\partial W^{sb}}{\partial \delta} = \frac{\partial \mathcal{L}}{\partial \delta} = \underbrace{-\Psi'(v^p)u_2\left(c^p, \frac{y^p}{w^p}\right)\frac{y^p}{(w^p)^2}\frac{\partial w^p}{\partial \delta} - \Psi'(v^r)u_2\left(c^r, \frac{y^r}{w^r}\right)\frac{y^r}{(w^r)^2}\frac{\partial w^r}{\partial \delta}}_{\text{Direct effect} \gtrless 0} +$$
$$\underbrace{\frac{\gamma}{(w^r)^2}\frac{\partial w^r}{\partial \delta}\left[-u_2\left(c^r, \frac{y^r}{w^r}\right)y^r + u_2\left(c^p, \frac{y^p}{w^r}\right)y^p\right]}_{\text{Indirect (incentive) effect} < 0}$$

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 and is captured in the first-best. If spending in the poor district is not too inefficient and the planner has a high degree of aversion to inequality this term will be positive (when the social objective is maximin this is unambiguously positive). The proposition holds when the second line is negative (the indirect effect). As w^r is decreasing in δ, this will be true if the bracketed term is positive. With two individuals, pooling (bunching) is never optimal so that $y^r > y^p$ and thus the incentive constraint implies $c^r > c^p$. We require the marginal disutility of labour to be higher for the high type when they choose their own bundle rather than mimicking. In the absence of income effects, this is certainly true as gross income is larger and utility is concave in leisure. As $c^r > c^p$, we see that a sufficient, but not necessary condition for the result to hold is $u_{12} \geq 0$. ■

Proof of Lemma 2. Denote welfare W^{fb} as the solution to (7). Using the envelope theorem and substituting the first order conditions on y^{ij} we have

$$\frac{\partial W^{fb}}{\partial \delta} = \frac{\partial \mathcal{L}}{\partial \delta} = \lambda \epsilon_w \left[\frac{\sum_i y^{ip}}{\delta} - \frac{\sum_i y^{ir}}{1-\delta}\right]. \tag{27}$$

Thus, $\partial W^{fb}/\partial \delta$ is negative whenever

$$\delta > \frac{\sum_i y^{ip}}{\sum_i y^{ip} + \sum_i y^{ir}} < \frac{1}{2}. \tag{28}$$

Analogous to Lemma 1, the second inequality holds when y is increasing in the wage. As in the proof to Lemma 1 we can relax the constant elasticity assumption and derive the weaker sufficient condition. ■

Proof of Proposition 2. Under an egalitarian policy there are no differences in wages across districts when individuals have the same θ, so that:

$$w^{ij} = \theta_i w^p(\delta_E) = \theta_i w^r(1-\delta_E) = w^i. \tag{29}$$

Incentive compatibility requires that each type i prefers their own bundle, which now implies fewer restrictions as there are effectively half as many types. The Lagrangian for this problem can be written

$$\mathcal{L} = 2\sum_i \Psi(v^i) + 2\lambda\left(\sum_i y_i - \sum_i c_i\right) + 2\sum_{i=2}^{\frac{n}{2}} \gamma_{i-1}\left[u\left(c^i, \frac{y^i}{w^i}\right) - u\left(c^{i-1}, \frac{y^{i-1}}{w^i})\right)\right], \tag{30}$$

where we leave factors of two to make explicit that there are indeed two individuals with each wage. Although rich and poor students with the same θ have the same wage they are affected differently by spending changes at the optimum. In particular, the impact on rich students of a change in δ is greater and of opposite sign than on poor students. After some manipulation we can write this effect as

$$\frac{\partial\mathcal{L}}{\partial\delta} = \left(-\sum \Psi'(v^i)u_2\left(c^i, \frac{y^i}{w^i}\right)\frac{y^i\theta_i}{(w^i)^2} + \sum_{i=2}^{\frac{n}{2}} \frac{\gamma_{i-1}\theta_i}{(w^i)^2}\left[-u_2\left(c^i, \frac{y^i}{w^i}\right)y^i + u_2\left(c^{i-1}, \frac{y^{i-1}}{w^i})\right)y^{i-1}\right]\right) \times \left(\frac{\partial w^p}{\partial\delta} + \frac{\partial w^r}{\partial\delta}\right).$$

This is negative whenever $(\partial w^p/\partial\delta + \partial w^r/\partial\delta) < 0$. Using the constant elasticity assumption and the fact that $w^p(\delta_E) = w^r(1-\delta_E) = w$, we have

$$\frac{\partial w^p}{\partial\delta} + \frac{\partial w^r}{\partial\delta} = w\epsilon_w\left[\frac{1}{\delta} - \frac{1}{1-\delta}\right]. \tag{31}$$

Which is negative at $\delta_E > \frac{1}{2}$. ■

Proof of Proposition 3. When there is tagging in the tax problem we can simplify the incentive constraints in that we only need be concerned with differences across θ within each district. Write the Lagrangian for the problem described in section 4.2 as follows:

$$\mathcal{L} = \sum_i\sum_j \Psi(v^{ij}) + \lambda\left(\sum_i\sum_j y^{ij} - \sum_i\sum_j c^{ij}\right) + \sum_{i=1}^{\frac{n}{2}-1}\gamma_i^p\left[u\left(c^{i+1,p}, \frac{y^{i+1,p}}{w^{i+1,p}}\right) - u\left(c^{ip}, \frac{y^{ip}}{w^{i+1,p}}\right)\right] + \sum_{i=1}^{\frac{n}{2}-1}\gamma_i^r\left[u\left(c^{i+1,r}, \frac{y^{i+1,r}}{w^{i+1,r}}\right) - u\left(c^{ir}, \frac{y^{ir}}{w^{i+1,r}}\right)\right].$$

The first order conditions from the tax problem are characterized by

$$c^{ij}: \Psi'(v^{ij})u_1\left(c^{ij}, \frac{y^{ij}}{w^{ij}}\right) - \lambda + \gamma_{i-1}^{j}u_1\left(c^{ij}, \frac{y^{ij}}{w^{ij}}\right) - \gamma_i^j u_1\left(c^{ij}, \frac{y^{ij}}{w^{i+1,j}}\right) = 0. \tag{32}$$

$$y^{ij}: \frac{\Psi'(v^{ij})}{w^{ij}}u_2\left(c^{ij}, \frac{y^{ij}}{w^{ij}}\right) + \lambda + \frac{\gamma_{i-1}^{j}}{w^{ij}}u_2\left(c^{ij}, \frac{y^{ij}}{w^{ij}}\right) - \frac{\gamma_i^j}{w^{i+1,j}}u_2\left(c^{ij}, \frac{y^{ij}}{w^{i+1,j}}\right) = 0. \tag{33}$$

for $1 \leq i \leq \frac{n}{2}$ and $j \in \{p, r\}$. The effect of a change in δ at the optimum is given by

$$\frac{\partial \mathcal{L}}{\partial \delta} = -\sum_i \sum_j \Psi'(v^{ij})u_2\left(c^{ij}, \frac{y^{ij}}{w^{ij}}\right)\frac{y^{ij}}{(w^{ij})^2}\frac{\partial w^{ij}}{\partial \delta} +$$

$$\sum_{i=1}^{\frac{n}{2}-1}\frac{\gamma_i^p}{w^{i+1,p}}\frac{\partial w^{i+1,p}}{\partial \delta}\left[\frac{-u_2\left(c^{i+1,p}, \frac{y^{i+1,p}}{w^{i+1,p}}\right)}{w^{i+1,p}}y^{i+1,p} + \frac{u_2\left(c^{ip}, \frac{y^{ip}}{w^{i+1,p}}\right)}{w^{i+1,p}}y^{i,p}\right]$$

$$+\sum_{i=1}^{\frac{n}{2}-1}\frac{\gamma_i^r}{w^{i+1,r}}\frac{\partial w^{i+1,r}}{\partial \delta}\left[\frac{-u_2\left(c^{i+1,r}, \frac{y^{i+1,r}}{w^{i+1,r}}\right)}{w^{i+1,r}}y^{i+1,r} + \frac{u_2\left(c^{ir}, \frac{y^{ir}}{w^{i+1,r}}\right)}{w^{i+1,p}}y^{i,r}\right].$$

Substitute $\frac{e}{w}\frac{dw}{de} = \epsilon_w$ and collect terms containing after-tax incomes y^{ij}. Then substitute the first order conditions on y and simplify to obtain

$$\frac{\partial W^{sb}}{\partial \delta} = \frac{\partial \mathcal{L}}{\partial \delta} = \lambda \epsilon_w \left[\frac{\sum_i y^{ip}}{\delta} - \frac{\sum_i y^{ir}}{1-\delta}\right]. \tag{34}$$

This is precisely the condition obtained in the first-best problem except that incomes are no longer those generated under a lump-sum tax. We immediately see that changes have a negative impact for any $\delta \geq 1/2$. Note that with many types partial pooling or bunching may occur so that y is not necessarily strictly increasing in the wage. However all that is required is that the inequality be strict somewhere. Whether or not this threshold value is larger or smaller than the first-best depends on the impact of marginal taxation. Unlike the two-type case, the second-best policy may be more progressive than the first-best if relative output from the poor district increases due to distortionary taxation. If marginal taxes are progressive enough this may indeed be the case. However, we cannot say more without fully describing the tax schedule (which would involve very restrictive assumptions). ■

Proof of Lemma 3. Characterize the effects of a change in ϵ on welfare defined by the solution to (14) as

$$\frac{\partial W^{fb}}{\partial \epsilon} = \int_{w_{\underline{s}}}^{w_{\overline{s}}} \frac{dw}{d\epsilon}\frac{\partial V}{\partial w}f(w)dw = \lambda \int_{w_{\underline{s}}}^{w_{\overline{s}}} (w^m - w)l(w)f(w)dw, \tag{35}$$

in which λ is the multiplier on the budget constraint. There are a few things to note. The first is that the transfer preserves the ordering of wages so that the mass of each type is unaffected by the change. Also we have used $dw/d\epsilon = w^m - w$. Ignoring λ, we now decompose this into effects above and below the mean

$$\frac{\partial W^{fb}}{\partial \epsilon} = \int_{w_{\underline{s}}}^{w^m} (w^m - w)\, l(w) f(w) dw + \int_{w^m}^{w_{\overline{s}}} (w^m - w)\, l(w) f(w) dw. \tag{36}$$

From the definition of w^m, we have

$$\int_{w_{\underline{s}}}^{w^m} (w^m - w)\, f(w) dw = -\int_{w^m}^{w_{\overline{s}}} (w^m - w)\, f(w) dw. \tag{37}$$

Thus, if $l(w)$ is increasing in w we have the result. To show this note that first-best optimum is described by the first-order conditions from the problem defined in (14) as well as that for the individuals labour supply.

$$u_1(wl(w) - T_w, l) = \lambda \tag{38}$$

$$u_2(wl(w) - T_w, l) = -w\lambda \tag{39}$$

The optimum is characterized by the equalization of the marginal utility of consumption, so that $u_1(wl(w)-T_w, l) = \lambda$ is constant for all w. The individuals first-order conditions can be simplified to $u_2(wl(w) - T_w, l) = -w\lambda$. Totally differentiating with respect to w yields two equations describing $\frac{\partial C}{\partial w}$ and $\frac{\partial l}{\partial w}$. Eliminating the former yields

$$\frac{\partial l}{\partial w} = \frac{-u_1 u_{11}}{u_{11}u_{22} - u_{12}^2} > 0. \tag{40}$$

The inequality holds when $u(c, l)$ is concave. ■

Proof of Proposition 4. Employing a dynamic version of the envelope theorem[22], we characterize the effects of a change in ϵ on welfare, described in (18)-(21), as

$$\frac{\partial W^{sb}}{\partial \epsilon} = \int_{\underline{w}}^{\overline{w}} \left.\frac{\partial \mathcal{H}}{\partial \epsilon}\right|_{\text{Optimum}} dw. \tag{41}$$

To analyze this, we first differentiate the Hamiltonian function

$$\frac{\partial \mathcal{H}}{\partial \epsilon} = \frac{dw}{d\epsilon} f(w) \left[\lambda u(c(w)) - \frac{V(w)}{w^2}\right]. \tag{42}$$

[22] A relatively accessible statement and proof of this result can be found in Caputo (2005).

Noting that $dw/d\epsilon = 0$ for $s \notin \mathcal{S}$, $dw/d\epsilon = w^m - w$ for $s \in \mathcal{S}$ and that the mass of types is unaffected by the change, the welfare impact is[23]

$$\frac{\partial W^{sb}}{\partial \epsilon} = \int_{w_{\underline{s}}}^{w_{\overline{s}}} (w^m - w) f(w) \left[\lambda u(c(w)) - \frac{V(w)}{w^2} \right] dw, \tag{43}$$

where the state, costate and controls are evaluated at the optimum described in the tax problem above. Define

$$K(w) = \left[\lambda u(c(w)) - \frac{V(w)}{w^2} \right]. \tag{44}$$

Rewriting (43), we can analyze the effects on those above and below the mean separately

$$\frac{\partial W^{sb}}{\partial \epsilon} = \int_{w_{\underline{s}}}^{w^m} \underbrace{(w^m - w) K(w) f(w)}_{>,<,=0} dw + \int_{w^m}^{w_{\overline{s}}} \underbrace{(w^m - w) K(w) f(w) dw}_{<0}. \tag{45}$$

The transfer always results in a decrease in utility for those above the mean, as $K(w)$ is unambiguously positive in this range. To show this, first note that $\lambda = E(1/w)$. Rewrite $K(w)$, replacing $V(w) = wu(c(w)) - y(w)$, to obtain

$$K(w) = u(c(w)) \left[E\left(\frac{1}{w}\right) - \frac{1}{w} \right] + \frac{y(w)}{w^2}. \tag{46}$$

The first term is positive for $w > w^m$ and may be negative for some $w < w^m$. To see this, rewrite

$$\left[E\left(\frac{1}{w}\right) - \frac{1}{w} \right] = \int_{\underline{w}}^{\overline{w}} \left(\frac{1}{m} - \frac{1}{w\eta} \right) dF(m), \tag{47}$$

where the population mass is $\eta > 1$ (recall the redistribution is over a population of mass one that is a subset of the total population). If the distribution is symmetric around the mean then we see this is true. It is also true if the distribution is positively skewed as we might expect with a wage distribution. Thus, $K(w)$ is positive and the effect of an increase in ϵ is negative for those above the mean.

If the overall impact of an increase in ϵ for those below the mean is negative, the result is shown. This is true when $K(w)$ is negative for these types. Let $K(w) > 0$ for at least some below the mean, then we must compare the relative impacts of a change in ϵ. First, we show that $K(w)$ is everywhere increasing in w. Consider

$$K'(w) = \frac{d}{dw} \left[\lambda u(c(w)) - \frac{V(w)}{w^2} \right]. \tag{48}$$

As λ is constant (and positive) and $u(c(w))$ is at least weakly increasing in the wage, we focus on

[23] The reason we restrict the effect of changes to $\mathcal{S} \subset \mathcal{W}$ is for analytical convenience. In particular we may differentiate through the integral sign without concern for the limits of integration.

the second term

$$-\frac{d}{dw}\left[\frac{V(w)}{w^2}\right] = \frac{1}{w^3}[wu(c(w)) - y(w)] = \frac{V(w)}{w^3}, \tag{49}$$

which is non-negative so long as $u(0) \geq 0$. In deriving (49), we have used the definition of $V(w)$, as well as $V'(w) = u(c(w))$, from the first-order incentive compatibility condition. We saw earlier that the transfer of productivities is defined such that there is a zero-sum. In particular, that

$$\int_{w_{\underline{s}}}^{w^m} (w^m - w) f(w)dw = -\int_{w^m}^{w_{\overline{s}}} (w^m - w) f(w)dw. \tag{50}$$

Thus, $K'(w) \geq 0$ is a sufficient condition for the proposition to hold. ■

Department of Economics, University of Alberta
Working Paper Series

http://www.economics.ualberta.ca/WorkingPapers.aspx

2010-14: The Economic Choice of Participation and Time Spent in Physical Activity and Sport in Canada – **Humphreys, Ruseski**
2010-13: Do Gamblers Think That Teams Tank? Evidence from the NBA – Soebbing, **Humphreys**
2010-12: Would Hotelling Kill the Electric Car? – **Chakravorty,** Leach, Moreaux
2010-11: Residential Land Use Regulation and the US Housing Price Cycle Between 2000 and 2009 – **Huang,** Tang
2010-10: Government Revenue Volatility in Alberta – **Landon, Smith C.**
2010-09: Sports Participation and Happiness: Evidence from U.S. Micro Data – **Huang, Humphreys**
2010-08: Does Financial and Goods Market Integration Matter for the External Balance? A Comparison of OECD Countries and Canadian Provinces – **Smith, C**
2010-07: Consumption Benefits and Gambling: Evidence from the NCAA Basketball Betting Market – **Humphreys**, Paul, Weinbach
2010-06: On Properties of Royalty and Tax Regimes in Alberta's Oil Sands – **Plourde**
2010-05: Prices, Point Spreads and Profits: Evidence from the National Football League - **Humphreys**
2010-04: State-dependent congestion pricing with reference-dependent preferences - **Lindsey**
2010-03: Nonlinear Pricing on Private Roads with Congestion and Toll Collection Costs – Wang, **Lindsey**, Yang
2010-02: Think Globally, Act Locally? Stock vs Flow Regulation of a Fossil Fuel – Amigues, **Chakravorty,** Moreaux
2010-01: Oil and Gas in the Canadian Federation – **Plourde**
2009-27: Consumer Behaviour in Lotto Markets: The Double Hurdle Approach and Zeros in Gambling Survey Dada – **Humphreys, Lee, Soebbing**
2009-26: Constructing Consumer Sentiment Index for U.S. Using Google Searches – Della Penna, **Huang**
2009-25: The perceived framework of a classical statistic: Is the non-invariance of a Wald statistic much ado about null thing? - **Dastoor**
2009-24: Tit-for-Tat Strategies in Repeated Prisoner's Dilemma Games: Evidence from NCAA Football – **Humphreys, Ruseski**
2009-23: Modeling Internal Decision Making Process: An Explanation of Conflicting Empirical Results on Behavior of Nonprofit and For-Profit Hospitals – **Ruseski,** Carroll
2009-22: Monetary and Implicit Incentives of Patent Examiners – **Langinier,** Marcoul
2009-21: Search of Prior Art and Revelation of Information by Patent Applicants – **Langinier,** Marcoul
2009-20: Fuel versus Food – **Chakravorty,** Hubert, Nøstbakken
2009-19: Can Nuclear Power Supply Clean Energy in the Long Run? A Model with Endogenous Substitution of Resources – **Chakravorty**, Magné, Moreaux

Please see above working papers link for earlier papers

www.economics.ualberta.ca