DISCUSSION PAPER SERIES

No. 3357

AFFIRMATIVE ACTION AND EFFICIENCY IN EDUCATION

Gianni De Fraja

PUBLIC POLICY

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP3357.asp**

ISSN 0265-8003

AFFIRMATIVE ACTION AND EFFICIENCY IN EDUCATION

Gianni De Fraja, University of York and CEPR

Discussion Paper No. 3357
May 2002

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 3357

May 2002

ABSTRACT

Affirmative Action and Efficiency in Education

This Paper studies the optimal education policy in the presence of different groups of households, with groups differing in the distribution of the ability to benefit from education. The main result is that the high ability individuals from groups with relatively few high ability individuals should receive more education than equally able individuals from groups with a more favourable distribution of abilities. The interpretation of this conclusion is that affirmative action policies can find a rationale on efficiency grounds alone.

JEL Classification: D63 and I28
Keywords: affirmative action, education policy and minorities

Gianni De Fraja
Department of Economics &
Related Studies
University of York
Heslington
YORK
YO1 5DD
Tel: (44 1904) 433 767
Fax: (44 1904) 433 759
Email: gd4@york.ac.uk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=113071

Submitted 23 April 2002

NON-TECHNICAL SUMMARY

Affirmative action programmes use membership in those groups that have been the subject of discrimination (such as women, and racial, ethnic and national origin minorities) to expand the opportunity of the members of these groups. These programmes have often been criticised on the grounds of both equity and efficiency, and, indeed, some of them have been repealed, perhaps the most famous example being the referendum which approved Proposition 209 in California, which forbids 'discriminating against or giving preferential treatment to any individual or group in public employment, public education, or public contracting on the basis of race, sex, color, ethnicity or national origin'. The 'efficiency' criticism of these programmes is that by favouring someone on the basis of their appurtenance to a group it may happen that the best person is not chosen for the job, or for the university place, or the cheapest firm is not chosen for the award of a federal contract.

On the other hand, the economic literature has pointed out that there may well be 'efficiency' gains as a consequence of the implementation of affirmative action policies: society as a whole may gain if members of one group, who are held back, are allowed instead to express their full productive potential. Some instances of inefficient discrimination, such as an employer's dislike for hiring members of racial minorities, are likely to be temporary, as discriminating employers lose out to profit maximizing non-discriminating ones. This, however, need not happen: there are models displaying multiple equilibria, some equilibria are non-discriminatory and efficient, in other equilibria discrimination is inefficient, but because it is part of the equilibrium behaviour, is not eliminated without an explicit policy intervention. In these models, efficiency requires that otherwise identical individuals are treated in the same way even when they belong to different groups. This is not the case in this Paper. The optimal policy we derive here is an example of 'positive discrimination': it requires treating differently otherwise identical individuals from different groups.

The topic of the Paper is the optimal education policy in the presence of individuals who differ both in ability and in some observable characteristic, such as race, sex, religion and so on. Groups differ in the distribution of abilities among individuals: 'disadvantaged' groups have fewer high ability individuals. We do not take any position as to why this should be the case: as will become apparent, the analysis holds regardless of the cause of this difference in distribution. The benchmark for optimality is the maximization of the total utility in society; in the model this is equivalent to the maximization of total income, or the rate of return to the investment in education. A policy is constituted by the amount of education received by each individual and their contribution towards its cost. Our main result is that the optimal education policy is such that individuals from 'disadvantaged' groups pay a lower tuition fee for admission to a given education level than individuals with the same ability coming from households in more advantaged groups. While we do not consider admission policies as a separate policy instrument, individuals from disadvantaged groups are enrolled to higher education levels than individuals of the same ability from advantaged households: this is due simply to their paying a lower tuition fee. This clearly tallies with the practice of many US universities to alter admission standards according to the ethnicity of the applicant, and with financial programmes of the type mentioned in the opening paragraph.

The result of the Paper is a typical second best result, and is due to asymmetric information: if the government had the same information as the private sector, individuals with the same ability would be treated in the same way regardless of their group. Moreover, the optimal policy is also a qualitative distortion from the market outcome: in the absence of public intervention, all individuals of the same ability would acquire the same education. The optimality of affirmative action policies in education is a consequence of the fact that higher ability individuals (in all groups) should receive higher education levels than they would acquire privately. To induce them to acquire these higher education levels, they are charged, in tuition fee, less than the (private) cost of education, and the difference is financed through general taxation. This subsidy has a social cost, which is increasing with the education level provided, and therefore with individuals' ability. But when there are relatively fewer higher ability individuals in a group, that is, when a group is 'disadvantaged', then the total cost to provide the high levels of education for individuals in that group is lower than it would be for a group with a larger number of high ability individuals, and therefore high ability individuals in disadvantaged groups should receive more education.

1 Introduction

In 1996, California approved Proposition 209, which forbids "discriminating against or giving preferential treatment to any individual or group in public employment, public education, or public contracting on the basis of race, sex, color, ethnicity or national origin". The year before, the US Supreme Court had ruled that the Banneker programme at the University of Maryland was in violation of the Constitution; this programme provided financial assistance to blacks, and had lower ability thresholds for eligibility than the colour-blind parallel programme.

These are but two examples of the repeal of "affirmative action" legislation, with the pendulum swinging back after the epochal changes brought about by the civil rights movements (see Holzer and Neumark 2000, for an extensive survey). Affirmative action programmes have typically an "equity" justification, to address past and present discrimination.[1] There may however well be "efficiency" reasons for these policies: society as a whole gains if members of one group, who are held back, are allowed instead to express their full productive potential. Some instances of inefficient discrimination, such as an employer's dislike for hiring members of racial minorities, are likely to be temporary, as discriminating employers lose out to profit maximising non-discriminating ones (eg Becker 1971, and Lundberg and Starz 1983). This however need not happen: there are models displaying multiple equilibria, some non-discriminatory and efficient, some where discrimination is inefficient, but because it is part of the equilibrium behaviour, is not eliminated without an explicit policy intervention.[2]

[1] An official definition of affirmative action can be obtained from the Clinton administration review of federal affirmative action programmes: "any effort taken to expand opportunity for women, and racial, ethnic and national origin minorities by using membership in those groups that have been subject to discrimination..." (Stephanopoulos and Edley 1996, cited in Holzer and Neumark 2000, p 488).

[2] A typical model is Coate and Loury's (1993) model of discrimination in the labour market: individuals acquire human capital, and their productivity in employment depends on ability and human capital, which are both unobservable: employers observe to which group an individual belongs and use statistics about the group to her characteristics; all groups have the same distribution of ability. Coate and Loury show that there may be multiple equilibria: in one equilibrium all groups are treated equally; in another, employers believe that individuals belonging to a given group acquire less human capital and therefore pay them less: but, because of this, the rewards to, and therefore the incentives towards, human capital acquisition are weakened for individuals belonging to this group, and so they will indeed acquire less human capital and the employers' belief turns out to be correct. Other examples, in a similar spirit, are in Loury 2002, pp 29-33. A different approach, leading to similar results is Milgrom

In these models, the socially preferred outcome is one where all groups are treated equally. Thus, if it were possible to "hide" the observable characteristics which distinguish groups, then statistical discrimination (or indeed discrimination based on tastes) would become impossible, and the efficient equilibrium, where all groups are treated equally, would clearly be unique. This is not the case in this paper. The optimal policy we derive here is an example of "positive discrimination": it requires treating differently otherwise identical individuals from different groups. The topic of the paper is the optimal education policy in the presence of individuals who differ both in ability and in some observable characteristic, such as race, sex, religion and so on. Groups differ in the distribution of abilities among individuals: "disadvantaged" groups have fewer high ability individuals. We do not take any position as to why this should be the case: as will become apparent, the analysis holds regardless of the *cause* of this difference in distribution. The benchmark for optimality is the maximisation of the total utility in society (and, because of the functional form we posit for the individual utility functions, this is equivalent to the maximisation of total income). A policy is constituted by the amount of education received by each individual and her contribution towards its cost. Our main result is that the optimal education policy is such that individuals from "disadvantaged" groups pay a lower tuition fee for admission to a given education level than individuals with the same ability coming from households in more advantaged groups. While we do not consider admission policies as a separate policy instrument, individuals from disadvantaged groups are enrolled, as a consequence of their paying a lower tuition fee, to higher education levels than individuals of the same ability from advantaged households: this clearly tallies with the practice of many US universities to alter admission standards and financial assistance according to the ethnicity of the applicant.

The result of the paper is a typical second best result, and is due to asymmetric information: if the government had the same information as the private sector, individuals with the same ability would be treated in the same way regardless of their group. Moreover, the optimal policy is also a qualitative distortion from the market outcome: in the absence of public intervention, all individuals of the same ability would again acquire the same education. The optimality of affirmative action policies in education is a consequence of the fact that higher ability individuals (in all groups) should receive higher education levels than they would acquire privately. To induce them to acquire these

and Oster (1987).

higher education levels, they are charged, in tuition fee, less than the (private) cost of education, and the difference is financed through general taxation. This subsidy has a social cost (as the shadow cost of public funds is greater than 1), which is increasing with the education level provided, and therefore with individuals' ability. Now note that, if there are fewer higher ability individuals in a group, that is, if a group is "disadvantaged", then the total cost to provide the high levels of education for individuals in that group is lower than it would be for a group with a larger number of high ability individuals, and therefore high ability individuals in disadvantaged groups should receive more education. The paper makes rigorous this loosely described intuition.

The paper is organised as follows: in Section 2 we describe the model, in Section 3 we solve the problem faced by a welfare maximising government, and Section 4 is a brief conclusion.

2 The Model

2.1 Household characteristics

There is a continuum of households, with measure normalised to 1. Each household comprises a mother and a daughter. Households differ according to two characteristics: the daughter's ability and an observable, exogenously given, feature, such as race, sex, ethnic origin, nationality, immigrant status, religion, age, sexual orientation, and so on. Formally, each household belongs to one of n groups, and each group is characterised by a particular value of the observable feature; groups are labelled by the subscript i, $i = 1, ..., n$. The number of households in group i is $h_i > 0$, with $\sum_{i=1}^{n} h_i = 1$.

Individuals' ability is given by a parameter $\theta \in \left[\underline{\theta}, \overline{\theta}\right]$, which affects that individual's expected income in the labour market. Specifically, a person's (possibly expected) labour market income is given by a function $y(\theta, e, E)$, whose argument are her ability, θ, her education, e, and the general education level in the economy, E.[3] The function $y(\theta, e, E)$ is increasing in all its three arguments: $y_\theta(\cdot) > 0$, $y_e(\cdot) > 0$ and $y_E(\cdot) > 0$. With regard to the second derivatives, we assume $y_{ee}(\cdot) < 0$, $y_{\theta\theta}(\cdot) < 0$, $y_{e\theta}(\cdot) > 0$. The last says that, given two individuals with the same education, the abler would benefit more from an increase in her investment in education. This is an important assumption, and a fuller

[3] This function can be derived from primitives: see De Fraja (2002). An example, special but nevertheless capturing the essence of our idea, is that income is $z_H(e, E)$ with probability θ, and $z_L(e, E) < z_H(e, E)$ with probability $(1 - \theta)$.

discussion is in De Fraja (2002). We also make two technical assumptions on the third derivatives to ensure that the appropriate second order conditions for an interior solution are satisfied: $y_{ee\theta}(\cdot) \geqslant 0$, $y_{e\theta\theta}(\cdot) \leqslant 0$.

Note that labour market income does not depend on an individual's group: two equally able, equally educated individuals from different groups will earn the same labour market income. Indeed, in our formalisation there is no independent definition of ability. Ability is simply the capacity to transform education into labour market income: a high ability individual is one, *tautologically*, whose future income, *ceteris paribus*, is higher.

Appurtenance to a group, however, does matter, because it affects the probability of having a certain ability:[4] in group i, $\theta \in [\underline{\theta}, \overline{\theta}]$ varies according to the distribution function $\Phi_i(\theta)$ with $\Phi_i(\underline{\theta}) = 0$ and $\Phi_i(\overline{\theta}) = 1$, and density $\phi_i(\theta) = \Phi_i'(\theta)$, which satisfies the standard assumption that the hazard rate is monotonic: $\frac{d}{d\theta}\left(\frac{1-\Phi_i(\theta)}{\phi_i(\theta)}\right) \leqslant 0$, for every $i = 1, ..., n$.

Clearly, our definition of ability is consistent irrespective of (and sheds no light on) the cause of any observed difference in the ability distribution across racial or ethnic groups. The extreme view, that differences in observed ability are genetically determined, is obviously captured by the assumption that the probability that a given individual's ability is at least θ varies from group to group. A different viewpoint is that measurable outcomes (be they IQ, SAT scores, or labour market earnings) are affected both by "innate" ability and by environmental influences occurring after birth (or conception), and that the probability of an individual having a certain "innate" ability is independent of the group (whilst not necessarily independent of the "innate" ability of her parents). Individuals from disadvantaged groups are more likely to belong to households without the means (economic or cultural) to let "innate" ability express itself. Clearly included among the possible causes of difference in outcome among groups is the possibility of discrimination, either explicitly enshrined in the legal system ("in contract" Loury 2002, p 95), or due to "unequal treatment of persons on the basis of race in the associations and relationships that are formed among individuals in social life" ("discrimination in contact", Loury 2002, pp 95-6).

We make the assumption that the ability distributions of two groups do not cross.

[4] In Loury's (2002) terminology, there is no "reward bias" (that is, racial discrimination), but given that it is less likely for individuals from a disadvantaged group to have high ability, there is "development bias" (Loury 2002, p 160).

Assumption 1 *Given i and j, $i, j \in \{1, ..., n\}$, either $\Phi_i(\theta)$ first order stochastically dominates $\Phi_j(\theta)$, or $\Phi_j(\theta)$ first order stochastically dominates $\Phi_i(\theta)$.*[5]

Relaxing this assumption would simply make our conclusions less clear-cut and more verbose in their description: we would need to qualify statements by saying "for ability level up to...". Assumption 1 determines a natural ordering for the various groups: simply re-label the groups in such a way that Assumption 1 can be written, without further loss of generality, as follows.

Assumption 1 (a) *For every $i \in \{2, ..., n\}$, $\Phi_i(\theta)$ first order stochastically dominates $\Phi_{i-1}(\theta)$.*

That is, group 1 has the highest proportion of low ability individuals, and group n the lowest. We label group 1 as the most "disadvantaged"; this term, though perhaps charged in general, describes accurately the specific set-up of our model: the average ability, and therefore the average labour market income, of individuals born to households in a group with a low index are lower than for individuals born to households in groups with higher indices.

The next assumption is crucial to the results of the paper.

Assumption 2 *For every $i \in \{2, ..., n\}$, $\frac{1-\Phi_{i-1}(\theta)}{\phi_{i-1}(\theta)} < \frac{1-\Phi_i(\theta)}{\phi_i(\theta)}$ for every $\theta \in \left(\underline{\theta}, \overline{\theta}\right)$.*

In words, the hazard rate is higher for more advantaged groups. Whether Assumption 2 holds in practice is a matter that, given our definition of ability, can be determined empirically using standard techniques. In the rest of this section we argue that this is a realistic assumption. We begin by showing that Assumption 2 holds for a very natural relationship between the distributions of two groups.

Lemma 1 *Let $i > j$. Suppose there exists $\delta > 0$ and a bijective increasing function, $\eta : \mathbb{R} \longrightarrow \left[\underline{\theta}, \overline{\theta}\right]$,*[6] *such that $\Phi_j(\eta(x)) = \Phi_i(\eta(x-\delta))$ for every $x \in \mathbb{R}$. Then $\frac{1-\Phi_i(\theta)}{\phi_i(\theta)} > \frac{1-\Phi_j(\theta)}{\phi_j(\theta)}$.*

Proof. $\frac{1-\Phi_i(\theta)}{\phi_i(\theta)} > \frac{1-\Phi_j(\theta)}{\phi_j(\theta)}$ corresponds to

$$\frac{1-\Phi_i(\eta(x))}{\phi_i(\eta(x))} > \frac{1-\Phi_i(\eta(x-\delta))}{\phi_i(\eta(x-\delta))} \tag{1}$$

[5] Recall that, given two distribution functions $\Phi_i(\theta)$ and $\Phi_j(\theta)$ with common support $\left[\underline{\theta}, \overline{\theta}\right]$ we say that $\Phi_i(\theta)$ first order stochastically dominates $\Phi_j(\theta)$ if $\Phi_i(\theta) \leqslant \Phi_j(\theta)$ for every $\theta \in \left[\underline{\theta}, \overline{\theta}\right]$, with strict equality over a range (Hirschleifer and Riley 1992, pp 106).

[6] That is, η is strictly increasing and $\lim_{x\to-\infty} \eta(x) = \underline{\theta}$ and $\lim_{x\to\infty} \eta(x) = \overline{\theta}$.

Treating $\Phi_i(\eta(x-\delta))$ as a function of two variables, x and δ, use the mean value theorem to write $\frac{1-\Phi_i(\eta(x))}{\phi_i(\eta(x))} - \frac{1-\Phi_i(\eta(x-\delta))}{\phi_i(\eta(x-\delta))} = \delta \frac{\partial}{\partial \delta}\left(\frac{1-\Phi_i(\eta(x_0))}{\phi_i(\eta(x_0))}\right)$ for some $x_0 \in [x-\delta, x]$. Therefore (1) holds if $\frac{\partial}{\partial \delta}\left(\frac{1-\Phi_i(\eta(x_0))}{\phi_i(\eta(x_0))}\right) > 0$. Now simply note that $\frac{\partial}{\partial \delta}\left(\frac{1-\Phi_i(\eta(x_0))}{\phi_i(\eta(x_0))}\right) = -\frac{\partial}{\partial x}\left(\frac{1-\Phi_i(\eta(x_0))}{\phi_i(\eta(x_0))}\right) > 0$ by the assumption of monotonicity of the hazard rate. ■

That is, let the interval ability be stretched from $[\underline{\theta}, \overline{\theta}]$ to the real line $\mathbb{R}$ (this is an innocuous normalisation). Suppose that the distribution function (from the "stretched" ability interval) of an advantaged group is a horizontal right shift of the distribution function of a disadvantaged group. Then the hazard rate is higher for the advantaged group.

Clearly, it may not be the case that the distribution of one group is a horizontal shift of that of another group, and while we have not been able to prove that any two arbitrary distribution functions satisfy Assumption 2, we can show that Assumption 2 holds for all the commonly used distribution functions.[7]

Lemma 2 *Let $\Phi_i(\theta)$ and $\Phi_j(\theta)$ be two normal (lognormal, Beta, Gamma, exponential, Weibull, linear, power) distribution functions such that $\Phi_i(\theta)$ first order stochastically dominates $\Phi_j(\theta)$. Then $\frac{1-\Phi_i(\theta)}{\phi_i(\theta)} > \frac{1-\Phi_{i-1}(\theta)}{\phi_{i-1}(\theta)}$ for every $\theta \in (\underline{\theta}, \overline{\theta})$.*

Proof. See Appendix ■

In words, Lemma 2 says that, if, within each group, the ability distribution is normal, lognormal, Beta, Gamma, exponential, Weibull, linear, or power,[8] a parameter shifts which lowers the distribution also increases the hazard rate.

All households have income $Y > 0$. At face value, the assumption that income is constant is unrealistic; however, in the present set-up, it constitutes no restriction. Parental income matters for two kinds of reasons: firstly, when education has to be paid for, and the capital market is imperfect, less

[7] The motivation for making Assumption 2 is therefore the same as the justification of the assumption that the hazard rate itself is monotonic, which is routinely made in the papers which use the mathematical techniques we use here: "it is satisfied by most usual distributions" Laffont and Tirole (1993, p 66).

[8] This list is taken from Bagnoli and Bergstrom (1989). The other distributions they consider are not relevant here, either because there are no parameters which can vary with the group (this is the case for the logistic and the extreme value distribution), or because there are no parameters which imply a first order stochastic dominance (the Laplace, the Chi, the Chi-squared, the t distribution). Finally, it is worth noting that Bagnoli and Bergstrom show that the hazard rate of the "mirror image Pareto" distribution $\Phi_i(\theta) = (-\theta)^{-g(i)}$, $\theta \in (-\infty, -1]$, $g(i) > 0$, is increasing in θ.

well-off households are liquidity constrained, and hence acquire less education. However, given (2), from the point of view of the choice of the investment in education, a model where all households have the same income is equivalent to one where households have different incomes, but they can borrow at the market rate to finance the investment in education.[9] Secondly, parental income is important for the analysis of education choices because of the observed income correlation across generations: children from better-off household are more likely to be themselves better-off.[10] But, given that, as will be evident in the next subsection, an increase in income simply increases the monetary transfer from the mother to the daughter, and therefore does not alter the education choice, our set-up can incorporate households with different incomes, simply by introducing new groups, each characterised by a different income level.

Finally, the household income, Y, can be spent on current consumption, transferred to the daughter (at the market interest rate, normalised to 0), or invested in the daughter's education, which can be purchased privately at a unit cost $k > 0$, and, if publicly provided, requires the payment of the appropriate tuition fee determined by the government. The household utility function is given by:

$$u(c) + x \qquad u'(c) > 0, \quad u''(c) < 0, \quad \text{there exists } c^* \text{ such that } u'(c^*) = 1, \quad (2)$$

where c is the household's current consumption and x is the (possibly expected) amount of monetary resources enjoyed by the daughter, given by the sum of monetary transfer from the mother and her labour market income.

2.2 Decision in the absence of government intervention

In the absence of public intervention, the budget constraint is given by $Y = ke + c + t$, where t is the intergenerational transfer, so that the mother's optimisation problem is:

$$\max_{e,t} u(Y - ke - t) + y(\theta, e, E) + t \qquad t > 0. \quad (3)$$

Let $e^S(\theta; k)$ be the value of e satisfying $k = y_e(\theta, e, E)$.

[9] We are not, obviously, claiming that the assumption of perfect capital markets is realistic; the point of the paper is that affirmative action policies may be optimal even when other imperfections that affect different groups differently, such as capital market imperfections, have been eliminated by appropriate policy intervention.

[10] Some empirical studies suggest a correlation coefficient between father's income and son's income of up to 0.4 (Zimmerman 1992 and Solon 1992 for the US and Dearden et al 1995 for the UK).

Proposition 1 *Let* $Y > c^* + ke^S\left(\overline{\theta};k\right)$*. In the absence of public provision, a household where the daughter has ability* θ *chooses an education level given by* $e = e^S(\theta;k)$*, and an intergenerational transfer given by* $Y - c^* - ke^S(\theta;k)$*.*

Proof. Simply take the first order conditions for (3) and the definition of $e^S(\theta;k)$. Next note that if the households who spend the most on education, those where the daughter has ability $\overline{\theta}$ can afford its cost, then so can all other households.[11] ■

Unsurprisingly, the investment in education is carried out to the point where its marginal benefit (the increase in future income) equals the marginal cost, given by k. Given that, by construction, both the marginal benefit and the marginal cost are the same for all groups, so is the education level acquired by individuals of the same ability in different groups. Note that abler individuals acquire more education (this follows from $\frac{de^s(\cdot)}{d\theta} = -\frac{y_{e\theta}(\cdot)}{y_{ee}(\cdot)} > 0$). Therefore, on average, disadvantaged groups acquire less education: this is due to fact that they, on average, have fewer high ability individuals than more advantaged groups.

Finally, note that household utility, denoted by $P(\theta)$, is given by:

$$P(\theta) = u(c^*) + y\left(\theta, e^S(\theta;k), E\right) + Y - c^* - ke^S(\theta;k),$$

and is increasing in the daughter's ability: $P'(\theta) = y_\theta\left(\theta, e^S(\theta;k), E\right) > 0$.

3 The problem of a utilitarian government

The government maximises a utilitarian welfare function, given by the unweighted sum of the utility of all households in the economy: there is no weighting of utility according to the position in the utility distribution of households or individuals. This rules out any possible bias in favour of some of the groups in the government objective function and therefore ensures that in our set-up affirmative action is called for on efficiency grounds alone. Indeed, it is worth pointing out that the maximisation of a utilitarian welfare function is here

[11] The condition $Y > c^* + ke^S\left(\overline{\theta};k\right)$ in the proposition can be replaced by:

$$Y + y(\underline{\theta}, e, E) > c^* + ke^S(\underline{\theta};k) \qquad (*)$$

with t unconstrained: the mother can borrow in order to finance her daughter's education. In this case the poorest household is that where the daughter has ability $\underline{\theta}$, and if (*) holds, then this household can afford its combination of education and consumption and therefore all households can. The case where t is constrained to be non-negative and there are sufficiently poor households is analysed in De Fraja (2002).

equivalent to the maximisation of aggregate income,[12] and therefore the policy we derive maximises the monetary return of investment in education.

To achieve its goal, the government selects the education to be received by each individual and the associated tuition fee.[13] Both can be made conditional on the ability of the daughter, θ, and on the observable characteristics of the household. Let therefore $e_i(\theta)$ denote the education level offered to a household belonging to group i where the daughter has ability θ, and $f_i(\theta)$ the tuition fee charged to this household. Different amounts may of course be charged to households in different groups for the same education.

The government will, naturally, need to respect a number of constraints, derived below.

We begin to notice, that, given a policy $\{e_i(\theta), f_i(\theta)\}$, the household in group i where the daughter has ability θ will, as before, choose a transfer t to maximise $u(Y - f_i(\theta) - t) + y(\theta, e_i(\theta), E) + t$. This implies $u'(Y - f_i(\theta) - t) = 1$, that is

$$t = Y - f_i(\theta) - c^*. \tag{4}$$

Therefore, if $U_i(\theta) = u(Y - f_i(\theta) - t) + y(\theta, e_i(\theta), E) + t$ is the utility of a household in group i where the daughter has ability θ who accepts the government offer, $i = 1, \ldots, n$, $\theta \in [\underline{\theta}, \overline{\theta}]$, we can use (4) to write $U_i(\theta)$ as:

$$U_i(\theta) = u(c^*) + y(\theta, e_i(\theta), E) + Y - f_i(\theta) - c^* \qquad i = 1, \ldots, n \quad \theta \in [\underline{\theta}, \overline{\theta}]. \tag{5}$$

(5) holds provided that $f_i(\theta)$ does not exceed the household income minus the value of consumption c^*, $f_i(\theta) - c^* \leqslant Y + y(\theta, e_i(\theta), E)$: households can afford to pay the tuition fee charged by the government without having to reduce their current consumption. We do not impose this as an explicit constraint, but verify that it holds at the solution. We next note that the government cannot make a household accept a combination of education and tuition fee which makes the household worse-off than it would be if it opted out of public provision by choosing private education. Formally:

$$U_i(\theta) \geqslant P(\theta) \qquad i = 1, \ldots, n \quad \theta \in [\underline{\theta}, \overline{\theta}]. \tag{6}$$

This is the first constraint faced by the government. A second constraint follows from the fact that households have an information advantage vis-à-vis the

[12] This is a consequence of the fact that mothers are interested in the maximisation of the daughter's income and that all households consume the same amount, c^*.

[13] The paper applies to education levels beyond a certain minimum, the attainment of which is akin to a human right, and therefore not subject to policy decision. The existence of such a minimum is enshrined in the compulsory education legislation in force in most countries.

government: unlike the household group, the daughter's ability is private information. This implies, in the standard fashion, that the government must ensure that the chosen policy satisfies the incentive compatibility constraint, which in this case is given by:[14]

$$\frac{dU_i(\theta)}{d\theta} = y_\theta(\theta, e_i(\theta), E) \qquad i = 1, \ldots, n \quad \theta \in \left[\underline{\theta}, \overline{\theta}\right]. \tag{7}$$

The government must of course satisfy a budget constraint, ensuring that the tuition fees charged are sufficient to pay for the desired level of education:

$$\sum_{i=1}^{n} h_i \int_{\underline{\theta}}^{\overline{\theta}} f_i(\theta)\, \phi_i(\theta)\, d\theta + T - \sum_{i=1}^{n} h_i \int_{\underline{\theta}}^{\overline{\theta}} k e_i(\theta)\, \phi_i(\theta)\, d\theta \geqslant 0,$$

where $T \geqslant 0$ is the amount of resources transferred to the education sector from general taxation, which is assumed to be exogenous.[15] Deriving $f_i(\theta)$ from (5) and re-arranging, the above becomes:

$$u(c^*) + Y - c^* + \sum_{i=1}^{n} h_i \int_{\underline{\theta}}^{\overline{\theta}} \left[y(\theta, e_i(\theta), E) - U_i(\theta) - k e_i(\theta)\right] \phi_i(\theta)\, d\theta + T \geqslant 0. \tag{8}$$

Finally, the overall level of education in the economy, denoted by E, is subject to a definitional constraint:

$$\sum_{i=1}^{n} h_i \int_{\underline{\theta}}^{\overline{\theta}} e_i(\theta)\, \phi_i(\theta)\, d\theta = E. \tag{9}$$

[14] (7) follows from a straightforward application of the revelation principle: the government must ensure that each household prefer "its" combination of education level and associated payment to all other combinations available to households in the same group. If a household where the daughter has ability θ reports instead ability $\widehat{\theta}$, its utility is given by:

$$\psi\left(\theta, \widehat{\theta}\right) = u(c^*) + y\left(\theta, e_i\left(\widehat{\theta}\right), E\right) + Y - f_i\left(\widehat{\theta}\right) - c^*,$$

and so the first order condition for choice of $\widehat{\theta}$ is

$$\frac{\partial \psi(\theta, \widehat{\theta})}{\partial \widehat{\theta}} = y_e\left(\theta, e_i\left(\widehat{\theta}\right), E\right) \frac{de_i(\widehat{\theta})}{d\theta} - \frac{df_i(\widehat{\theta})}{d\theta} = 0. \tag{**}$$

Truthful reporting implies that the first order condition is satisfied at $\widehat{\theta} = \theta$. From (5)

$$\frac{dU(\theta)}{d\theta} = y_\theta(\theta, e_i(\theta), E) + y_e(\theta, e_i(\theta), E) \frac{de_i(\theta)}{d\theta} - \frac{df_i(\theta)}{d\theta}.$$

Substitute the above and $\widehat{\theta} = \theta$ into (**) to obtain (7). The derivation of the second order conditions is also standard, and can be found in De Fraja (2002).

[15] We take it as given here to concentrate the analysis on differences among groups: in De Fraja (2002), the amount of tax paid by each household is determined as part of the solution of a more general problem for the government.

We can now state formally the optimisation problem faced by the government:

$$\max_{\substack{\{e_i(\theta),U_i(\theta)\}_{i=1}^n \\ E}} \sum_{i=1}^{n} h_i \int_{\underline{\theta}}^{\overline{\theta}} U_i(\theta)\,\phi_i(\theta)\,d\theta, \quad \text{subject to (6) (7) (8) and (9).} \tag{10}$$

Optimality also requires $\frac{de_i(\theta)}{d\theta} \geqslant 0$, which will be shown to hold at the solution of problem (10). Towards solving (10), let $e_i^G(\theta;\beta,x)$ be defined as the solution in e of:

$$y_e(\theta,e,E) - x = \frac{\beta-1}{\beta}\frac{1-\Phi_i(\theta)}{\phi_i(\theta)} y_{e\theta}(\theta,e,E), \tag{11}$$

and, for given $\beta > 0$, $\sigma \geqslant 0$, and k, let

$$k^* = k - \frac{\sigma}{\beta}.$$

Finally, let $\theta_i^*(\beta,x)$ be defined as the solution in θ of:

$$e^S(\theta;k) = e_i^G(\theta;\beta,k^*),$$

if a solution exists, and by $\theta_i^*(\beta,x) = \underline{\theta}$ if $e^S(\underline{\theta};k) < e_i^G(\underline{\theta};\beta,k^*)$. We can now state the optimal policy for the government.

Proposition 2 *Let the budget constraint (8) be binding at the solution of problem (10).*[16] *The education policy solving (10), $e_i^*(\theta)$, is given by*

$$e_i^*(\theta) = \begin{cases} e^S(\theta;k) & \text{for} \quad \theta < \theta_i^*(\beta,x) \\ e_i^G(\theta;\beta,k^*) & \text{for} \quad \theta \geqslant \theta_i^*(\beta,x) \end{cases}.$$

Moreover, $\sigma = \sum_{i=1}^n h_i \int_{\theta_i^*(\beta,k^*)}^{\overline{\theta}} y_E\left(\theta, e_i^G(\theta;\beta,k^*), E\right)\phi_i(\theta)\,d\theta > 0.$

Proof. See Appendix ■

Notice first of all that, since individuals' income depends positively on the overall education in the economy, $\sigma > 1$, and therefore $k^* < k$. This implies that the highest ability individuals (those with ability $\theta = \overline{\theta}$) receive strictly more education under public provision than they would purchase privately: by continuity, this is also true for individuals whose ability is "close" to $\overline{\theta}$. On the other hand, individuals whose ability is below a certain level receive the same education they would receive if they purchased it privately.

[16] That is let the government not have sufficient external resources fully to exhaust all the external benefits of education. This seems a plausible assumption; moreover, in De Fraja (2002), it is shown that when taxes (and therefore T in (8)) are set endogenously, the solution to this integrated problem is indeed such that the budget constraint is binding.

The intuition for the particular shape for the relationship between ability and education for individuals in a given group is described in De Fraja (2002): "optimality requires that the education level decrease sharply as ability decreases. This is to provide a sufficient disincentive for able children to pretend to be dim: since they benefit more from education than less bright children (as $y_{e\theta}(\cdot) > 0$), it is more expensive for them to give up education for a given reduction of the contribution to the budget [...]. This process, however, cannot be pushed too far: for low ability children, the education level $e_i^G(\theta; \beta, k^*)$ becomes too low, and the mother of these less bright individuals would opt out of public provision: private education creates a lower bound to the level of education that can be provided by the government (De Fraja 2002, p 19)". In De Fraja (2002), ability and the observable group characteristic (income in that paper) are independently distributed, and therefore $e_i^G(\theta; \beta, k^*)$ is the same for all i's: all individuals of sufficiently high ability receive the same education, irrespective of their household group. The consequences of allowing for differences in ability distribution among groups differ are investigated in the next Proposition, which is the main result of this paper.

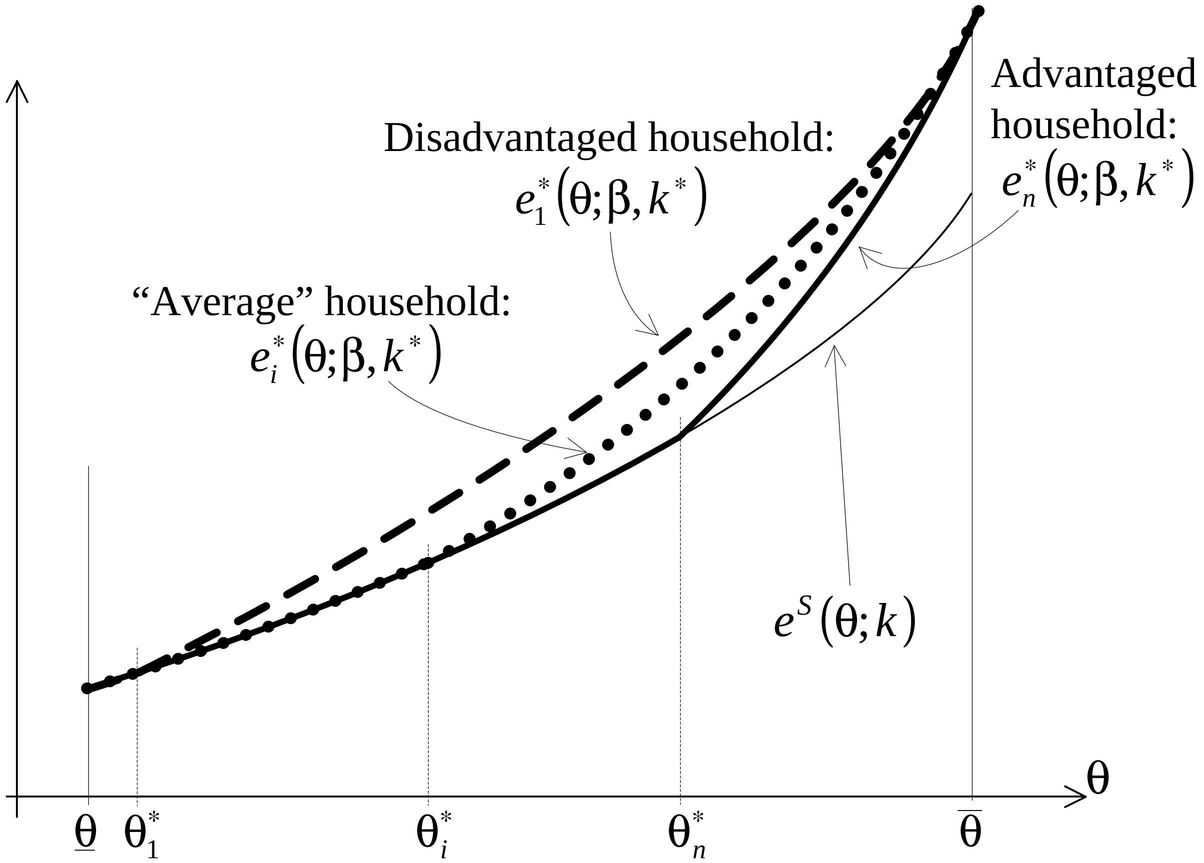
Proposition 3 *Let Assumption 2 hold, then* $e_{i-1}^*(\theta) > e_i^*(\theta)$, *for every* $i = 2, ..., n$, *for every* $\theta \in \left(\theta_{i-1}^*(\beta, k^*), \overline{\theta}\right)$.

In words, *the optimal education policy is such that individuals from disadvantages groups receive more education than individuals with the same ability from more advantaged groups.* This is a striking result: it provides a rationale for affirmative action purely on efficiency ground, with no appeal to equity or redistributive reasons whatsoever. Moreover, note that, since in the absence of public intervention the amount of education received by an individual is independent of her group (Proposition 1), the affirmative action policy implied by Proposition 3 is not the redress of (inefficient) market discrimination against disadvantaged groups (as for instance in Coate and Loury 1993). Rather, it is an instance of "positive discrimination": people from a disadvantaged group are treated, other things equal, more favourably, precisely because of their appurtenance to the group.

Figure 1 illustrates Proposition 3; it depicts the education received as a function of an individual's ability for three groups. The solid (respectively, dotted, respectively, dashed) line depicts the education level received by individuals coming from the most advantaged (respectively, an averagely advantaged, respectively, the most disadvantaged) households. The three lines meet

Figure 1

Education and ability for three groups of households



at $\theta = \overline{\theta}$: the highest ability individuals all receive the same education, irrespective of their group: see (11), which shows that $y_e\left(\theta, e_i^*\left(\overline{\theta}\right), E\right) = k^*$ for every $i = 1, ..., n$; this is an instance of the "efficiency at the top" principle. The education received by less able individuals declines as their ability decreases, but, as illustrated in Figure 1, it declines less rapidly for disadvantaged groups: therefore, for sufficiently high ability (above the intersection of the higher of the two curve with the private education level) individuals from disadvantaged groups receive more education than individuals from advantaged groups with the same ability.

A different interpretative angle for the result in Proposition 3 is obtained if we draw a horizontal line from a given point on the vertical axis. This gives the individuals who are enrolled at a certain education level: the same "school class" or "degree course". According to Proposition 3, the optimal policy is such that in a given class (below the very top, and above the lower levels of education), individuals from disadvantaged group have lower ability than their classmates from more advantaged groups. This is in line with the data reported by Herrnstein and Murray (1994, pp 472).[17] Given our definition of ability, an empirical analysis of the optimal education policy would show that the (post-education) labour market income of individuals from disadvantaged groups is lower than the labour market income of their classmates from more advantaged groups: controlling for education and not for ability, there would be a gap in earnings between advantaged and disadvantaged groups. Again, this tallies with Herrnstein and Murray's findings (1994, p 323), although their analysis has been criticised on methodological grounds (Neal and Johnson 1996, Cavallo et al. 1997, Cawley et al. 1997).

We can now turn to the proof of Proposition 3. This hinges on the following Lemma, which also has independent interest.

Lemma 3 *At the solution of problem (10),*

$$e_i^G\left(\theta; \beta, k^*\right) \gtreqless e_j^G\left(\theta; \beta, k^*\right) \text{ according to } \frac{1-\Phi_i\left(\theta\right)}{\phi_i\left(\theta\right)} \lesseqgtr \frac{1-\Phi_j\left(\theta\right)}{\phi_j\left(\theta\right)}. \qquad (12)$$

[17] If there are utility costs associated with this (such as the perpetuation of racial stereotypes implied by this observation (Murray 1994), or the sense of despair felt by minority students who find themselves in environments where they are unable to compete (D'Souza 1991)), then they could be explicitly taken into account in the utility function. The wide-ranging investigation by Bowen and Bok (1998, especially pp 191-217) fails to unearth convincing evidence of these costs. Even if they were present, however, it should be apparent that they would dampen, but not eliminate altogether, the differences in provision for individuals in different groups.

Proof. Let $i > j$. To lighten notation, write $e_i = e_i^G(\theta;\beta,k^*)$ and $e_j = e_j^G(\theta;\beta,k^*)$. Relationship (11) for the two groups considered becomes

$$y_e(\theta,e_i,E) - k^* = \frac{\beta-1}{\beta}\frac{1-\Phi_i(\theta)}{\phi_i(\theta)}y_{e\theta}(\theta,e_i,E),$$
$$y_e(\theta,e_j,E) - k^* = \frac{\beta-1}{\beta}\frac{1-\Phi_j(\theta)}{\phi_j(\theta)}y_{e\theta}(\theta,e_j,E).$$

Subtract the second from the first, and add and subtract $\frac{\beta-1}{\beta}\frac{1-\Phi_i(\theta)}{\phi_i(\theta)}y_{e\theta}(\theta,e_j,E)$:

$$\begin{aligned} & y_e(\theta,e_i,E) - y_e(\theta,e_j,E) = \\ & = \tfrac{\beta-1}{\beta}\left[\tfrac{1-\Phi_i(\theta)}{\phi_i(\theta)}\left[y_{e\theta}(\theta,e_i,E) - y_{e\theta}(\theta,e_j,E)\right] + \left[\tfrac{1-\Phi_i(\theta)}{\phi_i(\theta)} - \tfrac{1-\Phi_j(\theta)}{\phi_j(\theta)}\right]y_{e\theta}(\theta,e_j,E)\right]. \end{aligned}$$

Applying the mean value theorem, there exist $\hat{e}$ and $\tilde{e}$ such that the above can be written as:

$$\begin{aligned} & y_{ee}(\theta,\hat{e},E)(e_i - e_j) = \\ & = \tfrac{\beta-1}{\beta}\left[\tfrac{1-\Phi_i(\theta)}{\phi_i(\theta)}y_{ee\theta}(\theta,\tilde{e},E)(e_i-e_j) + \left[\tfrac{1-\Phi_i(\theta)}{\phi_i(\theta)} - \tfrac{1-\Phi_j(\theta)}{\phi_j(\theta)}\right]y_{e\theta}(\theta,e_j,E)\right], \end{aligned}$$

or:

$$\tfrac{\beta}{\beta-1}\left[y_{ee}(\theta,\hat{e},E) - \tfrac{1-\Phi_i(\theta)}{\phi_i(\theta)}y_{ee\theta}(\theta,\tilde{e},E)\right](e_i-e_j) = \left\{\tfrac{1-\Phi_i(\theta)}{\phi_i(\theta)} - \tfrac{1-\Phi_j(\theta)}{\phi_j(\theta)}\right\}y_{e\theta}(\theta,e_j,E). \tag{13}$$

The term in the square brackets in (13) is negative, since $y_{ee}(\theta,\hat{e},E) < 0$ and $y_{ee\theta}(\theta,\tilde{e},E) > 0$,[18] and therefore the sign of $(e_i - e_j)$ is the opposite of the sign of the term in the curly brackets. This establishes the Lemma. ■

To complete the proof of Proposition simply invoke Assumption 2.

Lemma 3 has independent interest because it asserts that given two individuals of the same ability belonging to different groups, which of the two should receive more education depends exclusively on the ability distribution of the groups to which these individuals belong. Since it is in principle simple to determine the ability distribution of different groups, Proposition 3 can be applied even to situations where the distributions of abilities do not conform to one of the functional forms given in Lemmata 1 and 2.

[18] Lest be thought that the result depend crucially on the sign of the third cross derivative $y_{ee\theta}(\cdot)$, note that, for the Lemma to hold, it is necessary and sufficient that the term in the square brackets on the LHS of (13) is negative. But as can be seen from the proof of Proposition 2, the negativity of the term in the square brackets on the LHS of (13) is a necessary condition for $e_i^*(\theta)$ to be increasing, which is a constraint to problem (10). In other words, in order for problem (10) to have the solution identified in Proposition 2, the term in the square bracket in (13) must be negative.

While the result in Proposition 3 may appear surprising, the intuition underlying it is relatively natural. To present it as clearly as possible, it is useful to investigate some further features of the optimal policy. The first result illustrates that, with symmetric information, all individuals of the same ability receive the sale education.

Corollary 1 *Suppose the government can costlessly observe* θ*. Then the optimal policy satisfies:*

$$e_i^*(\theta) = e^S(\theta; k^*) \qquad \textit{for every } i = 1, \ldots, n \quad \theta \in [\underline{\theta}, \overline{\theta}] .$$

Proof. The government problem in this case is obtained from problem (10) by eliminating constraint (7), and the results follows immediately from the proof of Proposition 2, by setting the Pontagryin multiplier $\mu_i(\theta)$ identically to 0. ■

The next result derives the rent obtained by each individual. Denote by $U_i^*(\theta)$ the utility obtained by an individual in group i whose ability is θ at the solution of problem (10).

Corollary 2 *Let Assumption 2 hold, then, for every* $i = 2, ..., n$*, for every* $\theta \in (\theta_{i-1}^*(\beta, k^*), \overline{\theta}]$*,* $U_{i-1}^*(\theta) > U_i^*(\theta)$*.*

Proof. Note that $\theta_{i-1}^* < \theta_i^*$ (omitting the argument (β, k^*)), and, therefore, for $\theta \in (\theta_{i-1}^*, \theta_i^*]$ we have $U_{i-1}^*(\theta) > U_i^*(\theta)$, and, in particular, $U_{i-1}^*(\theta_i^*) > U_i^*(\theta_i^*)$. Next consider $\theta \in (\theta_i^*, \overline{\theta})$. Here we have: $\frac{dU_{i-1}^*(\theta)}{d\theta} = y_\theta(\theta, e_{i-1}^*(\theta), E) > y_\theta(\theta, e_i^*(\theta), E) = \frac{dU_i^*(\theta)}{d\theta}$: the inequality follows from the fact that $e_{i-1}^*(\theta) > e_i^*(\theta)$ and $y_{e\theta}(\theta, e, E) > 0$. But clearly, $U_{i-1}^*(\theta_i^*) > U_i^*(\theta_i^*)$ and $\frac{dU_{i-1}^*(\theta)}{d\theta} > \frac{dU_i^*(\theta)}{d\theta}$ imply the corollary. ■

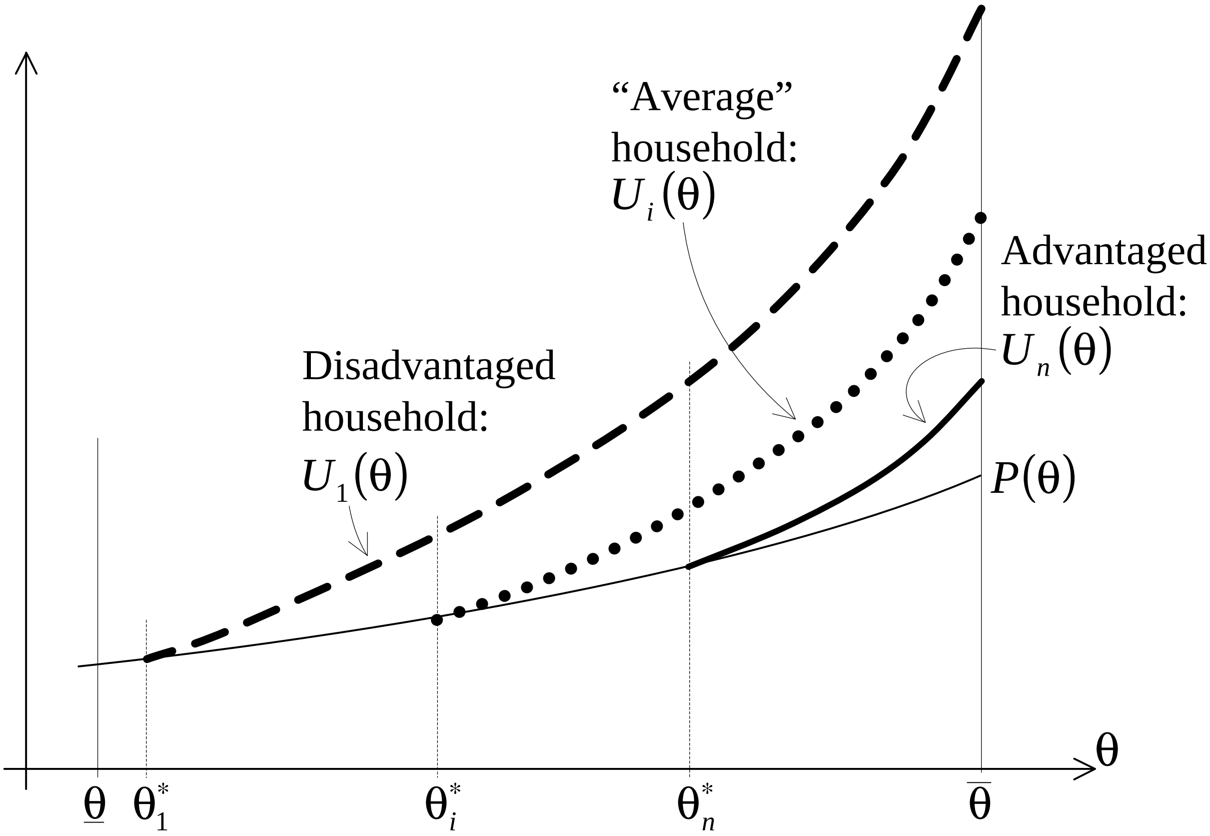
That is, individuals from disadvantaged groups receive, *ceteris paribus*, more utility.

Figure 2 illustrates the corollary: it depicts the utility received by individuals of various ability for three groups of households: as before, the solid (respectively, dotted, respectively, dashed) line depicts the utility achieved by individuals coming from the most advantaged (respectively, with an average advantage, respectively, the most disadvantaged) households, indexed by n (respectively, i, respectively, 1).

We can now describe the intuition for Proposition 3. Figure 2 also depicts the utility households would achieve if they obtained their education privately. Note that the households where the daughter has high ability receive more utility than they would from private provision. This utility, and the incentive to acquire more education than the private level, is provided in the form

Figure 2

Utility for three groups of households



of subsidised education (see De Fraja (2002), Proposition 4): the tuition fee charged to an individual with ability θ from group i, $f_i(\theta)$, is less than the cost sustained by the government for her education, $ke_i(\theta)$. But, obviously, to the extent that the government budget constraint is binding, this subsidy has a social cost, and therefore the government attempts to reduce it, by reducing sharply the education offered to individuals progressively less able, as described in the discussion following Proposition 3. Note that this does not happen in conditions of symmetric information, (Corollary 1), or when the shadow cost of public funds is 1 (set $\beta = 1$ in (11)): in both these cases, rent extraction is not a concern, and the education received by each individual is independent of her group. In general, however, there is a trade-off between reducing the rent received by high ability individuals and distorting the education received by lower ability individuals, both of which have a social cost. But now notice that the relative weight of these costs in the social welfare function is determined by the relative number of individuals of high ability and of lower ability. Roughly speaking, take a given ability level, θ_0: if in one group there are fewer individuals with ability above the given level θ_0 for every individual of ability θ_0 than in a second group, then it is relatively less costly, in aggregate, to give more rent to the high ability individuals in the first group in order to reduce the social loss given by the reduction in education below the social optimum for ability θ_0 individuals. And this is where Assumption 2 is crucial for the result to hold: the hazard ratio measures the ratio of the number of individuals with ability higher that θ_0 and the number of individuals of ability in a small interval $(\theta_0 - \varepsilon, \theta_0 + \varepsilon)$. When a group has a lower hazard rate than another, it therefore has fewer individuals of ability above θ_0 (relative to the number of individuals of ability in $(\theta_0 - \varepsilon, \theta_0 + \varepsilon)$), which makes it worthwhile to pay the extra rent in order to increase the education level provided to individuals with ability in $(\theta_0 - \varepsilon, \theta_0 + \varepsilon)$.

4 Conclusion

The paper derives the optimal education policy in the presence of groups which differ according to the distributions of individuals' ability to benefit from education. The optimal policy is an instance of affirmative action: it favours individuals from disadvantaged background, who need lower ability and pay a lower fee to receive a given education level than individuals from more advantaged backgrounds. As we note in the paper, this result holds irrespectively of the reason why one group is disadvantaged: indeed, paradoxically, if, as argued by Herrnstein and Murray in their controversial book (1994, pp 269-315), differences in ability between groups (be they genetically or environmentally determined) are unavoidable and will not be changed easily, if at all, by conscious policy intervention ("for the foreseeable future, the problems of low [...] ability are not going to be solved by outside interventions", p 389), then the bias in education policy illustrated in Proposition 3 should also be persistent. If, on the other hand, differences in the distribution of ability can be reduced by other forms of intervention, the optimal education policy would tend to become "group blind": in terms of Figures 1 and 2, the three curves would tend to approach a single curve.

In conclusion, it is worth noting the point of view of the paper, which derives the optimality of affirmative action policies not from a sense of justice or fairness, or from the desire of righting past wrongs, but from a dispassionate calculation of society's costs and benefits, using the viewpoint (standard in normative analysis in public economics) of a utilitarian welfare function. The liberal benchmark that identical individuals should be treated identically is not an end in itself: society can to discriminate among individuals when the "collective good" improve if such discrimination is practiced.[19]

[19] In our stylised set-up the "collective good" is the value of a utilitarian welfare function; in the debate on affirmative action in university admission, Loury (2002, pp 154) argues that the social benefits to be obtained by "constructing a racially integrated elite in America" far outweigh the costs of excluding a small number of well qualified white applicants in order to consider race in admission procedures for elite colleges. Bowen and Bok (1998, p 36) estimate that these admission procedures imply that an additional 1.5% of the white applicants are rejected.

References

Bagnoli, M. and T. Bergstrom (1989). Log-concave probability and its applications. University of Michigan.

Becker, G. S. (1971). *The economics of discrimination.* University of Chicago Press, Chicago.

Bowen, W. G. and D. Bok (1998). *The Shape of the River: Long-Term Consequences of Considering Race in College and University Admissions.* Princeton University Press, Princeton, NJ.

Cavallo, A., H. El-Abbadi, and R. Heeb (1997). The hidden gender restriction: The need for proper controls when testing for racial discrimination. In B. Devlin, S. E. Fienberg, D. P. Resnick, and K. Roeder (Eds.), *Intelligence, Genes, and Success*, pp. 193–214. Springer Verlag, New York.

Cawley, J., K. Conneely, J. Heckmann, and E. Vytlacil (1997). Cognitive ability, wages, and meritocracy. In B. Devlin, S. E. Fienberg, D. P. Resnick, and K. Roeder (Eds.), *Intelligence, Genes, and Success*, pp. 179–192. Springer Verlag, New York.

Coate, S. and G. C. Loury (1993). Will affirnmative-action policies eliminate negative stereotypes? *American Economic Review 83*, 1220–40.

De Fraja, G. (2002). The design of optimal education policies. *Review of Economic Studies*, forthcoming.

Dearden, L., S. Machin, and H. Reed (1995). Intergenerational mobility in Britain. *Economic Journal 105*, .

D'Souza, D. (1991). *Illiberal Education.* Free Press, New York.

Herrnstein, R. J. and C. Murray (1994). *The Bell curve: Intelligence and class structure in american life.* The Free Press, New York.

Hirshleifer, J. and J. G. Riley (1992). *The analytics of uncertainty and information.* Cambridge University Press, Cambridge, England.

Holzer, H. and D. Neumark (2000). Assessing affirmative action. *Journal of Economic Literature 38*, 483–568.

Laffont, J.-J. and J. Tirole (1993). *A theory of incentives in procurement and regulation.* MIT Press, Cambridge, Massachusetts.

Leonard, D. and N. van Long (1992). *Optimal conntrol theory and static optimization in economics.* Cambridge University Press, Cambridge, England.

Loury, G. C. (2002). *The Anatomy of Racial Inequality.* Harvard University Press, Cambridge, Mass.

Lundberg, S. J. and R. Starz (1983). Private discrimination and social intervention in competitive labor markets. *American Economic Review 73*, 340–347.

Milgrom, P. and S. Oster (1987). Job discrimination, market forces and the invisibility hypothesis. *Quarterly Journal of Economics 102*, 453–76.

Murray, C. (1994). Affirmative racism. In N. Mills (Ed.), *Debating Affirmative Action: Race Gender Ethnicity and the Policy of Inclusion*, pp. 191–208. Delta, New York.

Neal, D. and W. Johnson (1996). The role of premarket factors in black-white wage differentials. *Journal of Political Economy 104*, 869–895.

Solon, G. (1992). Intergenerational income mobility in the US. *American Economic Review 82*, 393–409.

Stephanopoulos, G. and C. Edley (1995). Review of federal affirmative action programs. White House, Washington DC.

Zimmerman, D. J. (1992). Regression toward mediocrity in economic stature. *American Economic Review 82*, 409–29.

Appendix (not for publication)

Proof of Lemma 1

The strategy of the proof is to show that the assertion is true for each of the distribution functions considered. We begin with some preliminary, and then take each functional form in turn.

To fix ideas, let, in the whole of the proof,

$$i > j$$

Let the following function be constructed. Take a suitable re-scaling of the indices, $\{1, ...n\} \longrightarrow \{k_1, ..., k_n\}$, with $k_i \in \mathbb{R}$, $i = 1, ..., n$, $k_{i-1} < k_i$, $i = 2, ..., n$. Take $\varepsilon > 0$, and let $\underline{g} = k_1 - \varepsilon$ and $\overline{g} = k_n + \varepsilon$. Take any function $\xi : [\underline{\theta}, \overline{\theta}] \times (\underline{g}, \overline{g}) \longrightarrow [0, 1]$ satisfying:

(i) $\xi(\theta, g) = \Phi_i(\theta)$ for every $i = 1, ..., n$, for every $\theta \in [\underline{\theta}, \overline{\theta}]$;

(ii) $\frac{\partial \xi(\theta, g)}{\partial g} < 0$ for every $g \in (\underline{g}, \overline{g})$, for every $\theta \in [\underline{\theta}, \overline{\theta}]$.

That is, $\xi(\theta, g)$ coincides with $\Phi_i(\theta)$ in the (re-scaled) index set $g = k_i \in \{k_1, ..., k_n\}$ (this is (i)), and there is positive correlation between the index i and ability ((ii); g is a mnemonics for group). Notice that the assumption that $\Phi_i(\theta)$ first order stochastically dominates $\Phi_{i-1}(\theta)$ is equivalent to the assumption that θ and g be positively correlated.

The statement in the Lemma can be reformulated, using the function ξ as follows.

$$\frac{1 - \xi(\theta, k_i)}{\frac{\partial \xi(\theta, k_i)}{\partial \theta}} - \frac{\partial \xi(\theta, k_j)}{\frac{\partial \xi(\theta, k_j)}{\partial \theta}} > 0 \tag{A1}$$

if and only if $i > j$. Use the Mean Value Theorem to write (A1) as

$$\frac{\partial}{\partial g}\left(\frac{1 - \xi(\theta, g_0)}{\frac{\partial \xi(\theta, g_0)}{\partial \theta}}\right)(k_i - k_j) > 0$$

for some $g_0 \in [k_j, k_i]$. Expand the first term in the above:

$$\frac{\partial}{\partial g}\left(\frac{1 - \xi(\theta, g_0)}{\frac{\partial \xi(\theta, g_0)}{\partial \theta}}\right) = -\frac{\frac{\partial \xi(\theta, g_0)}{\partial g}}{\frac{\partial \xi(\theta, g_0)}{\partial \theta}} - \frac{(1 - \xi(\theta, g_0))\frac{\partial^2 \xi(\theta, g_0)}{\partial g \partial \theta}}{\left(\frac{\partial \xi(\theta, g_0)}{\partial \theta}\right)^2} \tag{A2}$$

We will now develop the term on the RHS of (A2) for the various distribution functions. We proceed first by identifying the parameter which can be associated to the group, that is a parameter of the distribution which is positively correlated with θ: $\frac{\partial \xi(\theta, g)}{\partial g} < 0$. Once we have identified this parameter, we simply check that the term on the RHS of (A2) is in fact positive.

Beta Normalise the interval $[\underline{\theta}, \overline{\theta}]$ to $[0, 1]$, and let

$$\xi(\theta, g) = \frac{\int_0^\theta x^{g-1}(1-x)^{w-1}\,dx}{\int_0^1 x^{g-1}(1-x)^{w-1}\,dx} = \frac{a(\theta)}{a(1)} \tag{A3}$$

Where, to lighten notation, $\int_0^\theta x^{g-1}(1-x)^{w-1}\,dx = a(\theta)$. We have:

$$\frac{\partial \xi(\theta, g)}{\partial \theta} = \frac{\theta^{g-1}(1-\theta)^{w-1}}{\int_0^1 x^{g-1}(1-x)^{w-1}\,dx} = \frac{\theta^{g-1}(1-\theta)^{w-1}}{a(1)}$$

and therefore $\frac{\partial \xi(\theta,g)}{\partial \theta}$ is the Beta density function with parameters g and w.

Given the distribution (A3), θ and g are positively correlated. To see this, expand $\frac{\partial \xi(\theta,g)}{\partial g}$:

$$\frac{\partial \xi(\theta, g)}{\partial g} = \frac{\int_0^\theta x^{g-1}(\ln x)(1-x)^{w-1}\,dx \int_0^1 x^{g-1}(1-x)^{w-1}\,dx - \int_0^\theta x^{g-1}(1-x)^{w-1}\,dx \int_0^1 x^{g-1}(\ln x)(1-x)^{w-1}\,dx}{\left(\int_0^1 x^{g-1}(1-x)^{w-1}\,dx\right)^2}$$

consider the numerator, and write $\zeta(x)$ for $x^{g-1}(1-x)^{w-1} > 0$.

$$\begin{aligned}
&\int_0^\theta x^{g-1}(\ln x)(1-x)^{w-1}\,dx \int_0^1 x^{g-1}(1-x)^{w-1}\,dx \\
&\quad - \int_0^\theta x^{g-1}(1-x)^{w-1}\,dx \int_0^1 x^{g-1}(\ln x)(1-x)^{w-1}\,dx = \\
&\int_0^\theta \zeta(x)\ln(x)\,dx \int_0^1 \zeta(x)\,dx - \int_0^\theta \zeta(x)\,dx \int_0^1 \zeta(x)\ln(x)\,dx = \\
&\int_0^\theta \zeta(x)\ln(x)\,dx \left(\int_0^\theta \zeta(x)\,dx + \int_\theta^1 \zeta(x)\,dx\right) \\
&\quad - \int_0^\theta \zeta(x)\,dx \left(\int_0^\theta \zeta(x)\ln(x)\,dx + \int_\theta^1 \zeta(x)\ln(x)\,dx\right) = \\
&\int_0^\theta \zeta(x)\ln(x)\,dx \int_\theta^1 \zeta(x)\,dx - \int_0^\theta \zeta(x)\,dx \int_\theta^1 \zeta(x)\ln(x)\,dx = \\
&\int_0^\theta \zeta(x)\left[\ln(x)\int_\theta^1 \zeta(u)\,du - \int_\theta^1 \zeta(u)\ln(u)\,du\right]dx = \\
&\int_0^\theta \zeta(x)\left[\int_\theta^1 [\ln(x) - \ln(u)]\,\zeta(u)\,du\right]dx =
\end{aligned}$$

finally, notice that in the term $[\ln(x) - \ln(u)]$ x varies in the range $[0, \theta]$ and u in the range $[\theta, 1]$ and therefore $x < u$, implying that the term on the last line is negative, and establishing that (ii) is satisfied.

Now define $\int_0^\theta (\ln x)\, x^{g-1}(1-x)^{w-1}\,dx = b(\theta)$, and note that

$$\begin{aligned}
\frac{\partial \xi(\theta, g)}{\partial g} &= \frac{\int_0^\theta x^{g-1}(\ln x)(1-x)^{w-1}\,dx \int_0^1 x^{g-1}(1-x)^{w-1}\,dx - \int_0^\theta x^{g-1}(1-x)^{w-1}\,dx \int_0^1 x^{g-1}(\ln x)(1-x)^{w-1}\,dx}{\left(\int_0^1 x^{g-1}(1-x)^{w-1}\,dx\right)^2} \\
&= \frac{b(\theta)\,a(1) - b(1)\,a(\theta)}{[a(1)]^2}
\end{aligned}$$

and that:

$$\begin{aligned}
\frac{\partial^2 \xi(\theta, g)}{\partial g \partial \theta} &= -\theta^{g-1}(1-\theta)^{w-1}\,\frac{-(\ln\theta)\int_0^1 x^{g-1}(1-x)^{w-1}\,dx + \int_0^1 x^{g-1}(\ln x)(1-x)^{w-1}\,dx}{\left(\int_0^1 x^{g-1}(1-x)^{w-1}\,dx\right)^2} \\
&= -\frac{\theta^{g-1}(1-\theta)^{w-1}\,(b(1) - (\ln\theta)\,a(1))}{a(1)^2}
\end{aligned}$$

(A2) becomes:

$$
\begin{aligned}
& -\frac{\frac{\partial\xi(\theta,g)}{\partial g}}{\frac{\partial\xi(\theta,g)}{\partial\theta}} - \frac{(1-\xi(\theta,g))}{\left(\frac{\partial\xi(\theta,g)}{\partial\theta}\right)^2}\frac{\partial\xi(\theta,g)}{\partial\theta\partial g} = \\
& -\frac{\frac{b(\theta)a(1)-a(\theta)b(1)}{a(1)^2}}{\frac{\theta^{g-1}(1-\theta)^{w-1}}{a(1)}} + \frac{\left(1-\frac{a(\theta)}{a(1)}\right)}{\left(\frac{\theta^{g-1}(1-\theta)^{w-1}}{a(1)}\right)^2}\frac{\theta^{g-1}(1-\theta)^{w-1}(b(1)-(\ln\theta)a(1))}{a(1)^2} = \\
& -\frac{b(\theta)a(1)-a(\theta)b(1)}{a(1)^2}\frac{a(1)}{\theta^{g-1}(1-\theta)^{w-1}} + \\
& \quad \frac{a(1)-a(\theta)}{a(1)}\left(\frac{a(1)}{\theta^{g-1}(1-\theta)^{w-1}}\right)^2\frac{\theta^{g-1}(1-\theta)^{w-1}(b(1)-(\ln\theta)a(1))}{a(1)^2} = \\
& -\frac{b(\theta)a(1)-a(\theta)b(1)}{a(1)\theta^{g-1}(1-\theta)^{w-1}} + \frac{(a(1)-a(\theta))(b(1)-(\ln\theta)a(1))}{a(1)\left(\theta^{g-1}(1-\theta)^{w-1}\right)} = \\
& -\frac{b(\theta)a(1)-a(\theta)b(1)-(a(1)-a(\theta))(b(1)-(\ln\theta)a(1))}{a(1)\theta^{g-1}(1-\theta)^{w-1}} = \\
& -\frac{b(\theta)a(1)-a(1)b(1)+(a(1)-a(\theta))(\ln\theta)a(1)}{a(1)\theta^{g-1}(1-\theta)^{w-1}} = \\
& -\frac{b(\theta)-b(1)+(a(1)-a(\theta))(\ln\theta)}{\theta^{g-1}(1-\theta)^{w-1}} = \\
& \frac{b(1)-b(\theta)-(a(1)-a(\theta))(\ln\theta)}{\theta^{g-1}(1-\theta)^{w-1}} =
\end{aligned}
$$

Finally notice that $b(1) > b(\theta)$ and $a(1) < a(\theta)$ and therefore the term on the last line is positive.

Linear: Again, there is no loss in generality in normalising ability to be in $[0,1]$. Any linear density function can be written as $\frac{w+gx}{\int_0^1(w+gx)dx}$ with $\frac{\partial\xi(\theta,g)}{\partial g} = -2\theta w\frac{1-\theta}{(2w+g)^2} < 0$ and so expression (A2) becomes:

$$
-\frac{-2\theta w\frac{1-\theta}{(2w+g)^2}}{2\frac{w+g\theta}{2w+g}} - \frac{\left(1-\frac{w\theta+\frac{1}{2}g\theta^2}{w+\frac{1}{2}g}\right)}{\left(2\frac{w+g\theta}{2w+g}\right)^2}\left(2w\frac{2\theta-1}{(2w+g)^2}\right) = \frac{w\left(\theta^2-2\theta+1\right)}{2(w+g\theta)^2} > 0
$$

Weibull: $\xi(\theta,g) = 1-e^{-\theta^a g^{-a}}$.

$$
\begin{aligned}
\frac{\partial\xi(\theta,g)}{\partial\theta} &= \theta^{a-1}ag^{-a}e^{-\theta^a g^{-a}} \\
\frac{\partial\xi(\theta,g)}{\partial g} &= -\theta^a g^{-a-1}ae^{-\theta^a g^{-a}} \\
\frac{\partial^2\xi(\theta,g)}{\partial\theta\partial g} &= -\theta^{a-1}a^2g^{-a-1}e^{-\theta^a g^{-a}} + \theta^{2a-1}a^2g^{-2a-1}e^{-\theta^a g^{-a}}
\end{aligned}
$$

and so expression (A2) becomes

$$-\frac{-\theta^a g^{-a-1} a e^{-\theta^a g^{-a}}}{\theta^{a-1} a g^{-a} e^{-\theta^a g^{-a}}}$$
$$-\frac{\left(1-\left(1-e^{-\theta^a g^{-a}}\right)\right)}{\left(\theta^{a-1} a g^{-a} e^{-\theta^a g^{-a}}\right)^2}\left(-\theta^{a-1} a^2 g^{-a-1} e^{-\theta^a g^{-a}}+\theta^{2a-1} a^2 g^{-2a-1} e^{-\theta^a g^{-a}}\right)$$
$$= g^{a-1}\theta^{1-a} > 0$$

Exponential: $\xi(\theta, g) = 1 - e^{-\theta/g}$. This is trivial, since g is the hazard rate.

Normal: Here $\xi(\theta, g) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\theta} e^{\frac{-(u-g)^2}{2\sigma^2}} du$, and

$$\begin{aligned}
\frac{\partial \xi(\theta, g)}{\partial \theta} &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(\theta-g)^2}{\sigma^2}} \\
\frac{\partial \xi(\theta, g)}{\partial g} &= -\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\theta-g)^2}{2\sigma^2}} \\
\frac{\partial^2 \xi(\theta, g)}{\partial \theta \partial g} &= \frac{1}{\sigma^3\sqrt{2\pi}}(\theta-g) e^{-\frac{(\theta-g)^2}{2\sigma^2}}
\end{aligned}$$

and notice that

$$\frac{\partial^2 \xi(\theta, g)}{\partial \theta^2} = -\frac{1}{\sigma^3\sqrt{2\pi}}(\theta-g) e^{-\frac{(\theta-g)^2}{2\sigma^2}}$$

and so expression (A2) becomes

$$\frac{\partial}{\partial g}\left(\frac{1-\xi(\theta, g_0)}{\frac{\partial \xi(\theta, g_0)}{\partial \theta}}\right) = -\frac{\frac{\partial \xi(\theta, g_0)}{\partial g}}{\frac{\partial \xi(\theta, g_0)}{\partial \theta}} - \frac{(1-\xi(\theta, g_0))\frac{\partial^2 \xi(\theta, g_0)}{\partial g \partial \theta}}{\left(\frac{\partial \xi(\theta, g_0)}{\partial \theta}\right)^2} = -\frac{\partial}{\partial \theta}\left(\frac{1-\xi(\theta, g_0)}{\frac{\partial \xi(\theta, g_0)}{\partial \theta}}\right) > 0$$

because the hazard rate is decreasing for the normal distribution.

Lognormal: Clearly the same argument applies.

Gamma: In order to have positive correlation between θ and g, let

$$\xi(\theta, Y) = \frac{1}{Y^a \Gamma(a)}\int_0^{\theta} u^{a-1} e^{-u/Y} du$$

where $\Gamma(y) = \int_0^{\infty} e^{-x} x^{y-1} dx$. So we have:

$$\begin{aligned}
\frac{\partial \xi(\theta, g)}{\partial \theta} &= \frac{d\left(\frac{1}{g^a \Gamma(a)}\int_0^{\theta} u^{a-1} e^{-u/g} du\right)}{d\theta} = \frac{g^{-a}}{\Gamma(a)}\theta^{a-1} e^{-\frac{\theta}{g}} \\
\frac{\partial \xi(\theta, g)}{\partial g} &= -\frac{g^{-a}\int_0^{\theta}(ga-u) u^{a-1} e^{-\frac{u}{g}} du}{g^2 \Gamma(a)} = -\frac{\theta^a g^{1-a} e^{-\frac{\theta}{g}}}{g^2 \Gamma(a)} < 0 \\
\frac{\partial^2 \xi(\theta, g)}{\partial \theta \partial g} &= g^{-a-2}\theta^{a-1} e^{-\frac{\theta}{g}} \frac{-ga+\theta}{\Gamma(a)} \\
\frac{\partial^2 \xi(\theta, g)}{\partial \theta^2} &= g^{-a-1}\theta^{a-2} e^{-\frac{\theta}{g}} \frac{ga-g-\theta}{\Gamma(a)}
\end{aligned}$$

Now write (A2) as:

$$-\frac{\frac{\partial\xi(\theta,g)}{\partial g}}{\frac{\partial\xi(\theta,g)}{\partial\theta}}-\frac{(1-\xi(\theta,g))}{\left(\frac{\partial\xi(\theta,g)}{\partial\theta}\right)^2}\frac{\partial\xi(\theta,g)}{\partial\theta\partial g}=$$

$$-\frac{-\frac{\theta^a g^{1-a}e^{-\frac{\theta}{g}}}{g^2\Gamma(a)}}{\frac{g^{-a}}{\Gamma(a)}\theta^{a-1}e^{-\frac{\theta}{g}}}-\frac{\left(1-\left(\frac{1}{g^a\Gamma(a)}\int_0^\theta u^{a-1}e^{-u/g}du\right)\right)}{\left(\frac{g^{-a}}{\Gamma(a)}\theta^{a-1}e^{-\frac{\theta}{g}}\right)^2}\left(g^{-a-2}\theta^{a-1}e^{-\frac{\theta}{g}}\frac{-ga+\theta}{\Gamma(a)}\right)=$$

$$\frac{\theta}{g}+\frac{\left(1-\left(\frac{1}{g^a\Gamma(a)}\int_0^\theta u^{a-1}e^{-u/g}du\right)\right)}{\left(\frac{g^{-a}}{\Gamma(a)}\theta^{a-1}e^{-\frac{\theta}{g}}\right)}\frac{ga-\theta}{g^2}=$$

$$\frac{\theta}{g}\left(1+\frac{1-\left(\frac{1}{g^a\Gamma(a)}\int_0^\theta u^{a-1}e^{-u/g}du\right)}{\left(\frac{g^{-a}}{\Gamma(a)}\theta^{a-1}e^{-\frac{\theta}{g}}\right)}\frac{\frac{g}{\theta}a-1}{g}\right)=$$

$$\frac{\theta}{g}\left(1+\frac{(1-\xi(\theta,g))}{\frac{\partial\xi(\theta,g)}{\partial\theta}}\frac{\frac{g}{\theta}a-1}{g}\right) \tag{A4}$$

We want this to be positive, and so, if $\frac{g}{\theta}a-1>0$. the Lemma holds. Let therefore $\frac{g}{\theta}a<1$. Consider now the derivative with respect to θ of the hazard rate:

$$\frac{\partial}{\partial\theta}\left(\frac{(1-\xi(\theta,g))}{\frac{\partial\xi(\theta,g)}{\partial\theta}}\right)=-1-\frac{(1-\xi(\theta,g))}{\frac{\partial\xi(\theta,g)}{\partial\theta}}\frac{\frac{\partial^2\xi(\theta,g)}{\partial\theta^2}}{\frac{\partial\xi(\theta,g)}{\partial\theta}}<0$$

because the Gamma distribution has a decreasing hazard rate. The above is:

$$-1-\frac{(1-\xi(\theta,g))}{\frac{\partial\xi(\theta,g)}{\partial\theta}}\frac{(a-1)\frac{g}{\theta}-1}{g}<0$$

$$1-\frac{(1-\xi(\theta,g))}{\frac{\partial\xi(\theta,g)}{\partial\theta}}\frac{1-(a-1)\frac{g}{\theta}}{g}>0$$

Now notice that

$$1-\frac{(1-\xi(\theta,g))}{\frac{\partial\xi(\theta,g)}{\partial\theta}}\frac{1-a\frac{g}{\theta}}{g}>1-\frac{(1-\xi(\theta,g))}{\frac{\partial\xi(\theta,g)}{\partial\theta}}\frac{1-(a-1)\frac{g}{\theta}}{g}>0$$

Thus establishing that the hazard rate is decreasing in g.

Power Let $\xi(\theta,Y)=\theta^Y$ with $Y>0$.

$$\xi(\theta,Y)=\theta^Y$$

$$\frac{\partial\xi(\theta,Y)}{\partial\theta}=\frac{d\left(\theta^Y\right)}{d\theta}=\theta^{Y-1}Y$$

$$\frac{\partial\xi(\theta,Y)}{\partial Y}=\frac{d\left(\theta^Y\right)}{dY}=\theta^Y\ln\theta<0$$

$$\frac{\partial^2\xi(\theta,Y)}{\partial\theta\partial Y}=\frac{d\left(\frac{d\left(\theta^Y\right)}{d\theta}\right)}{dY}=\theta^{Y-1}(Y\ln\theta+1)$$

$$\frac{\partial^2\xi(\theta,Y)}{\partial\theta^2}=\frac{d\left(\frac{d\left(\theta^Y\right)}{d\theta}\right)}{d\theta}=\theta^{Y-2}(Y-1)Y$$

and expression (A2) becomes:

$$-\frac{\frac{\partial \xi(\theta,Y)}{\partial Y}}{\frac{\partial \xi(\theta,Y)}{\partial \theta}} - \frac{(1-\xi(\theta,Y))}{\left(\frac{\partial \xi(\theta,Y)}{\partial \theta}\right)^2} \frac{\partial \xi(\theta,Y)}{\partial \theta \partial Y} =$$

$$-\frac{\theta^Y \ln \theta}{\theta^{Y-1} Y} - \frac{\left(1-\theta^Y\right)}{\left(\theta^{Y-1} Y\right)^2} \theta^{Y-1} (Y \ln \theta + 1) =$$

$$\frac{\theta}{Y^2} \left(1 - \theta^{-Y} (Y \ln \theta + 1)\right)$$

consider the term $\left(1 - \theta^{-Y} (Y \ln \theta + 1)\right)$: this is increasing in Y ($\frac{d\left(-\theta^{-Y}(Y \ln \theta + 1) + 1\right)}{dY} = \theta^{-Y} \left(\ln^2 \theta\right) Y$) and is 0 at $Y = 0$, and therefore it is always positive, establishing the assert for the power distribution.

Proof of Proposition 2

Rewrite the integral constraints (7) and (6) as (see Leonard and van Long (1992, p 191) for details):

$$\dot{S}(\theta) = \sum_{i=1}^{n} h_i e_i(\theta)\phi_i(\theta) \qquad S(\underline{\theta}) = 0 \quad S(\overline{\theta}) = E$$
$$\dot{B}(\theta) = \sum_{i=1}^{n} h_i [y(\theta, e_i(\theta), E) - U_i(\theta) - ke_i(\theta)]\phi_i(\theta) \quad B(\underline{\theta}) = 0 \quad B(\overline{\theta}) = T + Y + u(c^*) - c^*$$

Where a dot above a variable denotes its derivative with respect to θ. Let σ and β be the Pontragyin multipliers associated to the above constraints; these are independent of θ (Leonard and van Long 1992, p 191). Let $\mu_i(\theta)$ be the Pontragyin multiplier associated to constraint (4); let $\rho_i(\theta)$ be the Lagrange multiplier associated to constraint (5). The lagrangean for problem (8) can now be written as:

$$\mathcal{L} = \sum_{i=1}^{n} h_i [(1-\beta)U_i(\theta) + \beta y(\theta, e_i(\theta), E) - (k\beta - \sigma)e_i(\theta)]\phi_i(\theta) + \tag{A5}$$
$$\mu_i(\theta) y_\theta(\theta, e_i(\theta), E) + \rho_i(\theta)[U_i(\theta) - P(\theta)]$$

the first order conditions can be written as:

$$\frac{\partial \mathcal{L}}{\partial e_i(\theta)} = h_i(\beta y_e(\theta, e_i(\theta), E) - (k\beta - \sigma))\phi_i(\theta) + \mu_i(\theta) y_{e\theta}(\theta, e_i(\theta), E) = 0 \tag{A6}$$
$$\frac{\partial \mathcal{L}}{\partial U_i(\theta)} = h_i(1-\beta)\phi_i(\theta) + \rho_i(\theta) = -\dot{\mu}_i(\theta) \tag{A7}$$

It will be shown later that the assumption that the budget constraint is binding implies $\beta > 1$. Let therefore be assumed, in what follows, that $\beta > 1$. Moreover, as discussed in De Fraja (2002), the interesting case is $\beta > 1$. The rest of the proof goes through a number of steps, which are organised in a series of lemmata.

According to the first Lemma, the "participation constraint" must be binding for some households: someone is indifferent between state and private provision.

Lemma A1 *For every $i = 1, \ldots, n$, it cannot be that $U_i(\theta) > P(\theta)$ for every $\theta \in [\underline{\theta}, \overline{\theta}]$*

Proof. By contradiction: let there exist $i \in \{1, ...n\}$ such that $U_i(\theta) > P(\theta)$ for every $\theta \in [\underline{\theta}, \overline{\theta}]$. Then $\rho_i(\theta) = 0$ for every $\theta \in [\underline{\theta}, \overline{\theta}]$, and, from (A7), $\dot{\mu}_i(\theta) = h_i(\beta - 1)\phi_i(\theta)$, implying, $\mu_i(\theta) = h_i(\beta - 1)\Phi_i(\theta) + M_i$. But we have $U_i(\underline{\theta})$ and $U_i(\overline{\theta})$ free, implying $\mu_i(\underline{\theta}) = \mu_i(\overline{\theta})$, a contradiction. ■

According to the second lemma, if there are households whose utility is strictly above its reservation level, then all households with the same income and higher ability also enjoy utility strictly above their reservation level.

Lemma A2 *For every $i = 1, \ldots, n$, let there exist $\hat{\theta}_i \in [\underline{\theta}, \overline{\theta})$ and $\delta > 0$ with $U_i(\theta) > P(\theta)$ for $\theta \in (\hat{\theta}_i, \hat{\theta}_i + \delta)$. Then $U_i(\theta) > P(\theta)$ for (almost) every $\theta \in (\hat{\theta}_i, \overline{\theta}]$.*

Proof. Assume by contradiction that there exists $\theta_i^{**} > \theta_i^*$ with $U_i(\theta)$ continuous in θ_i^{**}, and $U_i(\theta_i^{**}) = P(\theta_i^{**})$. Without loss of generality, assume that $U_i(\theta_i) > P(\theta_i)$ for $\theta \in (\theta_i^*, \theta_i^{**})$ (thus $U_i(\theta)$ goes above $P(\theta)$ at θ_i^* and goes back to it at θ_i^{**}; note that we are not implying $U_i(\theta) = P(\theta)$ for $\theta < \theta_i^*$or for $\theta > \theta_i^{**}$). Consider the following problem

$$\max_{e_i(\theta),U_i(\theta)} V = h_i \int_{\theta_i^*}^{\theta_i^{**}} U_i(\theta)\,\phi_i(\theta)\,d\theta \tag{A8}$$

$$\begin{aligned}
\text{s.t.: } & \dot{U}_i(\theta) = y_\theta(\theta, e_i(\theta), E) \qquad \theta \in (\theta_i^*, \theta_i^{**}) \\
& U_i(\theta_i) > P(\theta_i) \qquad \theta \in (\theta_i^*, \theta_i^{**}) \\
& U_i(\theta_i^*) = P(\theta_i^*) \\
& U_i(\theta_i^{**}) = P(\theta_i^{**}) \\
& \dot{S}(\theta) = h_i e_i(\theta)\,\phi_i(\theta) \qquad S(\theta_i^*) = S_i^* \quad S(\theta_i^{**}) = S_i^{**} \\
& \dot{B}(\theta) = h_i\left[y(\theta, e_i(\theta), E) - U_i(\theta) - k e_i(\theta)\right]\phi_i(\theta) \qquad B(\theta_i^*) = B_i^* \quad B(\theta_i^{**}) = B_i^{**}
\end{aligned}$$

This is a truncated version of the original problem, in the sense that, for appropriately chosen constants S_i^*, S_i^{**}, B_i^*, and B_i^{**}, the optimum of problem (A8) coincides with the optimum of the original problem, (8), for the household with in group i where the daughter has ability $\theta \in [\theta_i^*, \theta_i^{**}]$. Note that the FOC of problem (A8) determine

$$\mu_i(\theta) = h_i(\beta - 1)\,\Phi_i(\theta) + \underline{M}_i \tag{A9}$$

where $\mu_i(\theta)$ is continuous. Note that $\frac{\partial V}{\partial P(\theta_i^*)} > 0$ and $\frac{\partial V}{\partial P(\theta_i^{**})} > 0$, and hence $\mu(\theta_i^*) > 0$ and $\mu(\theta_i^{**}) > 0$ (see Leonard and van Long 1992, p 153). This however cannot happen if $\mu_i(\theta)$ satisfies (A9). This contradiction establishes the Lemma. ■

Lemma A2 implies that in each income class there is a cut-off ability level, $\hat{\theta}_i \in [\underline{\theta}, \overline{\theta}]$ such that households with ability below this cut-off only receive their reservation utility, households with higher ability receive strictly more. The next lemma is the core of the proof: it draws the implication of Lemma A2 for the education received by the daughters in these two groups of households.

Lemma A3 *$e_i^*(\theta) = e^S(\theta; k)$ for $\theta < \hat{\theta}_i$, $e_i^*(\theta) = e_i^G(\theta; \beta, k^*)$ for $\theta \geqslant \hat{\theta}_i$.*

Proof. $U_i(\theta) > P(\theta, Y_i, E)$ for $\theta \in \left(\hat{\theta}_i, \overline{\theta}\right]$ implies $\int_{\underline{\theta}}^{\hat{\theta}_i} \rho_i\left(\tilde{\theta}\right) d\tilde{\theta} = \int_{\hat{\theta}_i}^{\overline{\theta}} \rho_i\left(\tilde{\theta}\right) d\tilde{\theta}$, for every $\theta > \hat{\theta}_i$. Solving the differential equation (A7), we obtain: $\mu_i(\theta) = h_i(\beta - 1)\,\Phi_i(\theta) - \int_{\underline{\theta}}^{\theta} \rho_i\left(\tilde{\theta}\right) d\tilde{\theta} + M_i$. Together with the transversality condition $\mu_i(\underline{\theta}) = 0$, this implies $M_i = -h_i(\beta - 1)\,\Phi_i(\theta) + \int_{\underline{\theta}}^{\theta} \rho_i\left(\tilde{\theta}\right) d\tilde{\theta}$. That is

$$\mu_i(\theta) = -h_i(\beta - 1)\left(1 - \Phi_i(\theta)\right) \qquad \text{for } \theta \in \left(\hat{\theta}_i, \overline{\theta}\right] \tag{A10}$$

Substitute into (A6) to obtain $e_i^*(\theta) = e_i^G(\theta; \beta, k^*)$ for $\theta > \hat{\theta}_i$. Consider now $\theta \leqslant \hat{\theta}_i$. We have $U_i(\theta) = P(\theta)$, which implies $\dot{U}_i(\theta) = P_\theta(\theta)$, that is $y_\theta(\theta, e_i^*(\theta), E) = y_\theta\left(\theta, e^S(\theta; k), E\right)$, that is $e_i^*(\theta) = e^S(\theta; k)$. ■

We now show that for (at least) some income groups, there is a positive measure of households who are in the group above the cut-off ability level obtained in Lemma 2, that is, who receive strictly more utility that they obtain from private provision.

Lemma A4 *There exists $i \in \{1, ..., n\}$ such that $\hat{\theta}_i < \overline{\theta}$.*

Proof. Suppose not: then $U_i(\theta) = P(\theta)$ and $e_i^*(\theta) = e^S(\theta; k)$ for every $\theta \in [\underline{\theta}, \overline{\theta}]$ and $i = 1, ..., n$. Consider the government budget constraint in this case is clearly slack if $T > 0$ (the government could duplicate exactly the private sector outcome and still be left with T). Let therefore $T = 0$. Consider a policy where each household is offered education $e^S(\theta; k) + \varepsilon$. For ε "small", this raises additional revenues of

$$\sum_{i=1}^{n} h_i \int_{\underline{\theta}}^{\overline{\theta}} y_E\left(\theta, e^S(\theta; k) + \varepsilon, E\right) \phi_i(\theta)\, d\theta > 0$$

This is a first order effect. The additional cost, net of any increase in the tuition fee is

$$\sum_{i=1}^{n} h_i \int_{\underline{\theta}}^{\overline{\theta}} \left[y(\theta, e_i(\theta) + \varepsilon, E) - k e_i(\theta) - k\varepsilon\right] \phi_i(\theta)\, d\theta < 0$$

which is a second order effect: for ε sufficiently close to 0, the positive first order effect prevails. ■

We next show that $e_i^*(\theta)$ is continuous, and therefore that θ_i^* is in fact equal to $\hat{\theta}_i$ and given by the condition $e^S(\theta_i^*; k) = e_i^G(\theta_i^*; \beta, k^*)$

Lemma A5 *For every $i = 1, ..., n$, $e_i^*(\theta)$ is continuous.*

Proof. To establish the Lemma, we need to show that $\mu_i(\theta)$ is continuous. From Leonard and van Long (1992, p 335, (10.68e)) we get the condition:

$$\mu_i\left(\theta_i^{*-}\right) - \mu_i\left(\theta_i^{*+}\right) \geqslant 0 \tag{A11}$$

where θ_i^* is determined by the condition $\int_{\theta_i^*}^{\overline{\theta}} y_\theta(\theta, e_i^*(\theta), E)\, d\theta = P(\theta_i^*)$, and θ_i^{*-} (θ_i^{*+}) is a shorthand for the limit of a function as θ tends to θ_i^* from below (above). From the first order condition for $e_i(\theta)$, (A6), we obtain:

$$\mu_i\left(\theta_i^{*-}\right) = -h_i \frac{\beta\left[y_e\left(\theta_i^*, e^S(\theta_i^*; k), E\right) - k^*\right]}{y_{e\theta}\left(\theta_i^*, e^S(\theta_i^*; k), E\right)} \phi_i(\theta_i^*)$$

Conversely, we saw in (A10) that $\mu_i\left(\theta_i^{*+}\right) = -(\beta - 1) h_i (1 - \Phi_i(\theta_i^*))$, and therefore (A11) becomes:

$$-h_i \frac{\beta\left[y_e\left(\theta_i^*, e^S(\theta_i^*; k), E\right) - k^*\right]}{y_{e\theta}\left(\theta_i^*, e^S(\theta_i^*; k), E\right)} \phi_i(\theta_i^*) + (\beta - 1) h_i (1 - \Phi_i(\theta_i^*)) \geqslant 0$$

or

$$y_e\left(\theta_i^*, e^S(\theta_i^*; k), E\right) - k^* - \frac{\beta - 1}{\beta} \frac{1 - \Phi_i(\theta_i^*)}{\phi_i(\theta_i^*)} y_{e\theta}\left(\theta_i^*, e^S(\theta_i^*; k), E\right) \leqslant 0$$

from (A6) we get substituting $\mu_i\left(\theta_i^{*+}\right) = -(\beta-1)\,h_i\,(1-\Phi_i(\theta_i^*))$:

$$y_e\left(\theta_i^*, e^S(\theta_i^*;k), E\right) - k^* - \frac{\beta-1}{\beta}\frac{1-\Phi_i(\theta_i^*)}{\phi_i(\theta_i^*)} y_{e\theta}\left(\theta_i^*, e^S(\theta_i^*;k), E\right) = 0$$

The LHS of the above is a non-decreasing function of e, because $y_{ee} \leqslant 0$ and $y_{ee\theta} \geqslant 0$, and therefore,

$$e^S(\theta_i^*;k) \geqslant e_i^G(\theta_i^*;\beta,k^*) \tag{A12}$$

We also have $U_i(\theta_i^*) = P(\theta_i^*)$ and $U_i(\theta) > P(\theta)$ for $\theta \in \left(\theta_i^*, \overline{\theta}\right]$. These imply: $\dot{U}_i(\theta_i^*) \geqslant P_\theta(\theta_i^*)$, that is, $y_\theta\left(\theta_i^*, e_i^G(\theta_i^*;\beta,k^*), E\right) \geqslant y_\theta\left(\theta_i^*, e^S(\theta_i^*;k), E\right)$, which in turn implies, given $y_{e\theta}(\cdot) \geqslant 0$:

$$e_i^G(\theta_i^*;\beta,k^*) \geqslant e^S(\theta_i^*;k) \tag{A13}$$

(A12) and (A13) imply the Lemma. ■

The next lemma shows that there are individuals above the cut-off point.

Lemma A6 *For every $i = 1, ..., n$, there exists $a > 0$ such that $e_i^*(\theta) = e_i^G(\theta;\beta,k^*)$ for (almost) every $\theta \in \left[\overline{\theta}-a, \overline{\theta}\right]$*

Proof. Suppose not. Let there exist i_1, i_2, and $a > 0$ such that for $\theta \in \left[\overline{\theta}-a, \overline{\theta}\right]$ it is: $e_{i_1}^*(\theta) = e^S(\theta;k)$ and $e_{i_2}^*(\theta) = e_{i_2}^G(\theta;\beta,k^*)$. Take a sufficiently small, so that $e_{i_2}^G\left(\overline{\theta}-a;\beta,k^*\right) > e^S\left(\overline{\theta}-a;k\right)$. Without altering any household's utility, reduce $e_{i_2}^*(\theta)$ by $\frac{\varepsilon}{h_{i_2}}$ for children with ability $\theta \in \left[\overline{\theta}-a, \overline{\theta}\right]$. Take ε so that $e_{i_2}^*(\theta)$ remains above $e^S(\theta;k)$. This releases revenues $\left(\frac{\varepsilon}{h_{i_2}}\right) k^*$ per child, and hence a total of $\left(1-\Phi_{i_2}\left(\overline{\theta}-a\right)\right) k^*\varepsilon$. It also reduces the aggregate future income of the individuals in group i_2 by:

$$h_{i_2}\int_{\overline{\theta}-a}^{\overline{\theta}} \left[y\left(\theta, e_{i_2}^G(\theta;\beta,k^*), E\right) - y\left(\theta, e_{i_2}^G(\theta;\beta,k^*) - \frac{\varepsilon}{h_{i_2}}, E\right)\right]\phi_{i_2}(\theta)\,d\theta \cong$$
$$\int_{\overline{\theta}-a}^{\overline{\theta}} y_e\left(\theta, e_{i_2}^G(\theta;\beta,k^*), E\right)\phi_{i_2}(\theta)\,d\theta \equiv V_{i_2}$$

In order to maintain the same utility for these households, it is necessary to compensate them for the loss in future income, via an equivalent reduction in their contribution to the education budget.

Consider individuals in group i_1 with ability $\theta \in \left[\overline{\theta}-a, \overline{\theta}\right]$. Let their education be increased by $\left(\frac{\varepsilon}{h_{i_2}}\right)\frac{k^*}{k}\frac{1-\Phi_{i_1}\left(\overline{\theta}-a\right)}{1-\Phi_{i_2}\left(\overline{\theta}-a\right)}$. This has a total cost of $\left(1-\Phi_{i_2}\left(\overline{\theta}-a\right)\right)k^*\varepsilon$, and can therefore be financed with the savings from the reduction in the education of the individuals from group i_2. The increase in future earnings determined by the fact that these individuals are better educated is

$$h_{i_1}\int_{\overline{\theta}-a}^{\overline{\theta}} \left[y\left(\theta, e^S(\theta;k) + \frac{\varepsilon}{h_{i_2}}\frac{k^*}{k}\frac{1-\Phi_{i_1}\left(\overline{\theta}-a\right)}{1-\Phi_{i_2}\left(\overline{\theta}-a\right)}, E\right) - y\left(\theta, e^S(\theta;k), E\right)\right]\phi_{i_1}(\theta)\,d\theta \cong$$
$$\int_{\overline{\theta}-a}^{\overline{\theta}} y_e\left(\theta, e^S(\theta;k), E\right)\phi_{i_2}(\theta)\,d\theta \equiv V_{i_1}$$

Since $y_{ee}(\cdot) < 0$, and $e^G_{i_2}\left(\overline{\theta} - a; \beta, k^*\right) > e^S\left(\overline{\theta} - a; k\right)$, we have $V_{i_1} > V_{i_2}$, and, therefore, for sufficiently low ε, the reduction in the contribution of the households in group i_2 can be financed by the increase in the contribution from the households in group i_1, with something left over. This left over can be used to increase somebody's utility, hence the situation hypothesised cannot be an optimum. This ends the proof of Lemma A6. ■

We have shown (Lemma A3) that households are divided into two subsets, those who receive the same education they would acquire privately, and those who receive more, and that the latter set has positive measure (Lemma A6). For individuals in the latter subset, the first order condition (A6) and the value of $\mu_i(\theta)$ given in (A10) determine $e^*_i(\theta)$.

This establishes the main body of the proposition. There are a few "loose ends". Begin with the value of σ: this follows immediately, using (7.120) in Leonard and van Long (1992), p 255, and differentiating the Lagrangean (A5) with respect to E.

We also need to show that $e^*_i(\theta)$ is increasing in θ. Clearly, $e^S(\theta^*_i; k)$ is increasing in θ; when $\theta \geqslant \theta^*_i$, totally differentiate (9) and rearrange:

$$\frac{de^G_i(\theta; \beta, k^*)}{d\theta} = \frac{\frac{1-\Phi_i(\theta)}{\phi_i(\theta)} y_{e\theta\theta}(\cdot) + \left(\frac{d}{d\theta}\left(\frac{1-\Phi_i(\theta)}{\phi_i(\theta)}\right) - \frac{\beta}{\beta-1}\right) y_{e\theta}(\cdot)}{\frac{\beta}{\beta-1} y_{ee}(\cdot) - \frac{1-\Phi_i(\theta)}{\phi_i(\theta)} y_{ee\theta}(\cdot)}$$

In view of the technical assumptions on the third derivative, the above is positive, establishing that $e^*_i(\theta)$ is increasing.

The entire proof is based on the assumption that $\beta > 1$. We now show that this must hold. We do this by showing that $\beta = 1$ implies that the budget constraint is slack, against the assumption that it should be binding. If $\beta = 1$, then,

$$y_e(\theta, e^*_i(\theta), E) = k - \sigma \qquad i = 1, \ldots, n \qquad \theta \in \left[\underline{\theta}, \overline{\theta}\right]$$

implying that the education level is independent of the group. Moreover, it is easy to see that an increase in T does not affect the education level, implying that the budget constraint is not binding.

The final thing to establish is that $f_i(\theta)$ is set, for all households at a level which allows everybody the optimal rate of consumption c^*. Clearly this is the case for the households who receive the same education as they would purchase privately. But just as clearly, for the other households, who have more utility than they would receive privately: if it were not the case, it would be possible to increase the value of the objective function by reducing the education provided to these households, and increasing their current consumption.