College Diversity and Investment Incentives*

Thomas Gall[†], Patrick Legros[‡], Andrew F. Newman[§]

This version, 22 May 2014

Abstract

This paper studies the general equilibrium economic effects of affirmative action and other diversity policies in college admission. In our model, all benefits derived from a college education accrue through peer effects, but cannot be costlessly monetized and are therefore not perfectly transferable among students. Given this constraint, the free market allocation displays excessive segregation relative to the first-best (the hypothetical situation in which there is perfect transferability). Affirmative action policies can restore diversity within colleges but also affect incentives to invest in pre-college scholastic achievement. However, excessive segregation creates its own investment distortions, presentaing a subtle tradeoff that policy evaluation must take into account. Our results indicate that affirmative action policies conditioned on achievement can increase aggregate investment and income, reduce inequality, and increase aggregate welfare relative to the free market outcome. They may also be more effective than decentralized policies such as cross-subsidization of students by colleges, because these subsidies go from privileged (high wealth) students to underprivileged (low wealth) students, while enhancing diversity would require transfers that go in the reverse direction.

Keywords: Matching, misallocation, nontransferable utility, multidimensional attributes, affirmative action, segregation, education.
JEL: C78, I28, J78.

*Some of the material in this paper was circulated in an earlier paper "Mis-match, Rematch and Investment" which this paper now supersedes. We are grateful for comments from John Moore, Glenn Loury, Andy Postlewaite, and seminar participants at Amsterdam, Brown, Budapest, UC Davis, Essex, Frankfurt, Ottawa, Penn, ThReD 2009, and Yale. Gall thanks DFG for financial support (grant GA-1499), Legros thanks financial support of the Communauté Française de Belgique (ARC 00/05-252, PAI Network P6-09) and the FNRS (crédit aux chercheurs 2011).

[†]University of Southampton, UK

[‡]Université Libre de Bruxelles (ECARES) and CEPR

[§]Boston University and CEPR

1 Introduction

In many countries, policy makers and college administrators have been keenly interested in fostering diversity in the composition of students accessing tertiary education. How to achieve diversity is the subject of debate however. Discussions center around the pros and cons of keeping, removing or changing affirmative action policies, and on whether these policies should be focused on creating diversity on the basis of race, gender or social class. This paper contributes to the discussion by examining the effects of diversity policies on aggregate economic variables like the level and distribution of earnings, the composition of colleges, investments prior to college, and overall surplus in the economy. In particular it takes account of the responses of *all* agents, whether beneficiaries of these policies or not.

A salient feature of the college marketplace, and one that makes the resulting free-market allocation of students into colleges potentially problematic, is that the benefits students accrue from attending college cannot easily be transferred among them via a price system; students simply do not engage in transfer payments among themselves to generate the desired level of diversity. Because college benefits are lifetime earnings, even if a price system could be put in place individuals may not have the financial ability to make transfers based on lifetime earnings. In such a world, policies of tax-subsidies may be ineffective or imperfect instruments for achieving the desired goals set by a planner or the college officials. A discussion of affirmative action for college education must therefore be cast in the context of non-transferabilities (NTU).

The NTU characteristic of college education leads by itself to novel positive and normative insights, and as such complements other analyses of diversity policy based on imperfections like search frictions or statistical discrimination. Indeed, the theoretical literature on matching has illustrated that the composition of groups may be significantly effected by non-transferabilities: while groups may have a diverse composition when a full price system exists, they will be segregated when such a price system is lacking. Hence, if there is a surplus-maximizing level of diversity, it is unlikely to be achieved by a free market, opening the door for the possibility that diversity policies may actually lead to improvement.

There is an obvious policy response: just impose it! But such a solution must confront an equally obvious criticism: forcing diversity may distort the incentives to invest in education prior to entering the college, both for those students who are favored by the policy and perhaps more importantly, those who are not. In other words, policy makers seem to be facing a classic equity-efficiency tradeoff: diversity may be desirable from socio or political objectives (equity, diversity or righting past wrongs) but it comes at an economic cost.

As we show, this tradeoff may be misconstrued: affirmative action for college can be beneficial. Indeed, because of non-transferabilities, the free market policy generates too much segregation, already implying a distortion in the incentives to invest in education. Though rematch policies cannot directly address the market imperfections, they may provide an instrument for correcting inefficiency of the match as well as distortions in investment incentives: properly designed, they can raise aggregate output and investment, reduce inequality, and increase welfare.

Our analysis of various forms of rematch builds on the following environment. Colleges are arenas for the acquisition of human capital via a process that is driven entirely by peer effects. At the time they are admitted to college, agents have attributes that reflect their *background* (privileged or underprivileged) and their early education *achievement* (high or low). Privilege and high achievement increase both one's own and one's peers' payoffs to attending college. While background is exogenous, achievement is the result of an earlier investment. We assume that aggregate peer effects are strongest when there is diversity of backgrounds within colleges.

Our emphasis is on contrasting free market outcomes with ones constrained by policy. The free market equilibrium is characterized by full segregation in achievement and background. This implies that returns to college for the underprivileged will be low, giving them minimal incentives to invest. The privileged face the reverse situation, with returns and high investment correspondingly high investment incentives. Compared to a "first best" situation, which could be achieved if every agent had unlimited amounts of wealth with which to make side payments, there is *overinvestment at the top and underinvestment at the bottom* (OTUB).

As in the real world, the rematch policies we consider all aim to match college compositions to the population frequencies of backgrounds, but differ in the extent to which they condition on students' achievements. We first consider "achievement blind" policies that only focused on replicating the diversity of backgrounds in the population, a typical example being "busing". While this type of policy may generate higher aggregate surplus than free market, it guarantees low achievers a "good" match, and high achievers a "bad" one, with sufficient probability as to significantly depress investment incentives.

We then consider an "affirmative action" policy, which is defined as one that conditions the priority given to an underprivileged on achievement: among the underprivileged, only the high achievers are considered candidates for a match with the high achieving privileged. Affirmative action rewards underprivileged high achievers with access to privileged high achievers, encouraging the underprivileged; at the same time, the privileged are discouraged. The former effect dominates the latter, so that affirmative action generates higher aggregate investment and human capital, and less inequality, than the free market. In fact, aggregate investment under affirmative action tends to exceed that in the first best. Numerical simulations indicate that our affirmative action policy can come very close to the optimal rematching policy.

Finally, the results are robust to admitting wealth transfers into the college marketplace, as long as the underprivileged have limited ability to pay. Thinking of colleges as intermediating these payments, this shows that scholarships for the underprivileged alone may be insufficient for achieving college diversity.

Literature

If the characteristics of matched partners are exogenous, and partners can make non-distortionary side payments to each other (transferable utility or TU); there is symmetric information about characteristics; and there are no widespread externalities, stable matching outcomes maximize social surplus: no other assignment of individuals can raise the economy's aggregate payoff. Even if characteristics are endogenous, under the above assumptions re-matching the market outcome is unlikely to be desirable (Cole et al., 2001; Felli and Roberts, 2002). Though it is understood that NTU can distort matching patterns relative to the TU case,[1] there has been little work

[1] Economists are well aware, at least since Becker (1973), that under NTU the equilibrium matching pattern will differ from the one under TU, and need not maximize aggregate surplus (see also Legros and Newman, 2007). This is because a type that receives a large

characterizing those patterns, much less their implications for investment.

The literature on college and neighborhood choice (see among others Bénabou, 1993, 1996; Epple and Romano, 1998) typically finds too much segregation in types, often because of widespread externalities (see also Durlauf, 1996*b*; Fernández and Rogerson, 2001), thereby providing a possible rationale for rematch (called "assocational redistribution" in Durlauf, 1996*a*).

When attributes are fixed, aggregate surplus may be raised by bribing some individuals to migrate (de Bartolome, 1990). Fernández and Galí (1999) compare market allocations of college choice with those generated by tournaments: the latter may dominate in terms of aggregate surplus when capital market frictions lead to non-transferability. They do not consider investments before the match. Peters and Siow (2002) and Booth and Coles (2010) let agents invest in order to increase their attribute before matching in a marriage market with strict NTU. The former finds that allocations are constrained Pareto optimal (with the production technology they study, aggregate surplus is also maximized), and does not discuss policy. The latter compares different marriage institutions in terms of their impact on matching and investments. Gall et al. (2006) analyzes the impact of timing of investment on allocative efficiency. Several studies consider investments before matching under asymmetric information (see e.g., Bidner, 2008; Hopkins, 2012; Hoppe et al., 2009), *mainly* focusing on wasteful signaling, but not considering rematch policies. Finally, that literature assumes that matching depends only on realized attributes from investment, ignoring therefore the fact that both the initial background as well as the realized attribute may matter for sorting.

Rematch has occasionally been supported on efficiency grounds when there is a problem of statistical discrimination. Lang and Lehman (2011) survey the theoretical and empirical literature. Coate and Loury (1993) provide a formalization of the argument that equilibria where under-investment is supported by "wrong" expectations may be eliminated by affirmative action policies (an "encouragement effect"), but importantly also points out a possible downside ("stigma effect").

In their model, affirmative action is indeed consistent with two types of

share of the pie generated in an (efficient) match under TU may be left with a smaller share due to rigidities in dividing that pie if she stays with the same type of partner under NTU. She may then prefer to match with another type with whom she can obtain higher payoffs. If individuals' preferences over matches agree, this can cause excessive segregation.

equilibria. In the good equilibrium, affirmative action provides an incentive for the underprivileged to invest because they believe they will actually get a job; meanwhile employers observe that they are productive, so beliefs are consistent. One would expect after such a policy had been in place for that these benefits would be persistent. This finding appears to be inconsistent with empirical observations for colleges: removing affirmative action policies that have been in place for a while often triggers a reversion to the pre-policy status quo, for instance in case of the end of high college desegregation.[2] In the bad equilibrium, although employment of the underprivileged may increase, beliefs do not change during affirmative action and provide little investment incentive; in this case upon removal, the market equilibrium reverts to what it was before. In the unique equilibrium of our NTU framework, we obtain the same lack of persistence of affirmative action, but the investment effects resemble those of Coate and Loury's good equilibrium.

Existing work tends to evaluate the performance of policies with respect to the objective of colleges, for instance, as in Fryer et al. (2008) who evaluate whether a color-blind policy is a better instrument for increasing enrolment of students from a certain background than a color-sighted policy, or the effect of investment of the target group, but do not evaluate the general equilibrium effects of these policies, e.g., rarely discuss the effects on the group that is not targeted, the privileged, which is a necessary step towards evaluating the effects on inequality or aggregate variables like output or earnings, which are among the questions we analyze in this paper.

Finally, there is a connection between this paper and the recent "misallocation" literature in macroeconomics (e.g., Hsieh et al., 2013) arguing that substantial parts of output gaps between countries are due to inefficient sorting. Though our focus is on higher education rather than the economy at large, we do provide a specific underpinning for misallocation, as well as analysis of policies that might be applied assuming the underlying source of misallocations is intractable. Moreover, similar rigidities are likely to apply in many markets besides education, and so some of the lessons learned here may have wider applicability.

The paper proceeds as follows. Section 2 lays out the model framework.

[2]Orfield and Eaton (1996) report an increase in segregation in the South of the U.S. in districts where court-ordered high college desegregation ended, see also Clotfelter et al. (2006) and Lutz (2011). Weinstein (2011) finds increased residential segregation as a consequence of the mandated desegregation.

In Section 3 we derive the over-investment at the top/under-investment at the bottom (OTUB) result by contrasting the free market outcome, when there is no possibility of transferability, to the idealized situation of perfect transferability. The OTUB result opens the door for rematching policies to be surplus and welfare enhancing, and we indeed show that this is the case in section 4. In fact, we show that when the benefits from diversity are high in terms of total surplus and welfare, an affirmative action policy is close to the second-best policy. We allow in subsection 5 some transferability among students but limited by some of them are wealth constrained and have difficulties borrowing; we comfort the benefits of using affirmative action policies in case. We conclude in section 6. All proofs and calculations not in the text can be found in the appendix.

2 Model

Consider a "market" for colleges populated by a continuum of students with unit measure. Students may differ in their educational *achievement* $a \in \{h, \ell\}$ (for high and low) and their *background* $b \in \{p, u\}$ (for privileged and underprivileged). While individual background is given exogenously, achievement is a consequence of a student's investment in education before entering the market for colleges. Achieving h with probability e requires an investment in education of e at individual cost $e^2/2$.

In the market agents are fully characterized by their *attributes*, a pair ab, and match into colleges as a function of these attributes. A college $(ab, a'b')$ consists of two students. The payoffs are the life time earnings students expect to obtain as a function of their human capital they acquire in college, which depends on the attribute composition $(ab, a'b')$ in the college.

A student with attribute ab attending college $(ab, a'b')$ has output:

$$y(ab, a'b') = f(a, a')g(b, b').$$

The output y is the combined market value of human capital $f(a, a')$, taking as inputs individual cognitive skills acquired before the match, and network capital $g(b, b')$, capturing peer effects such as social networks, role models, or access to resources: the marketability of one's human capital depends on the social connections formed at college; or that the cost of acquiring human capital at college depends on one's own as well as one's peers' background attributes; or that the social environment at college amplifies or depresses the value of individual human capital, or its perception by the market.

Though human capital accumulation obviously depends on one's own characteristics directly as well as through interactions with other students, we will focus on the later aspect. Letting individual payoffs depend also on the student's attribute, as in the specification $y(ab, a'b') = h(ab) + \hat{f}(a, a')\hat{g}(b, b')$ for some function $h(ab)$, would not alter our main results.

For expositional purposes assume that:

$$f(h, h) = 1,\ f(h, \ell) = f(\ell, h) = 1/2,\ f(\ell, \ell) = \alpha, \tag{1}$$

$$g(p, p) = 1,\ g(p, u) = g(u, p) = \delta,\ g(u, u) = \beta, \tag{2}$$

with $\delta < 1$ and $\beta \in [\delta/2, \delta]$. The assumption that the achievement effects $f(\cdot, \cdot)$ have constant differences ($f(a, h) - f(a, \ell)$ is independent of a), is mainly for convenience; at the cost of some notational complexity we could impose other conditions (e.g., increasing differences) without significant effect on the results.

The network effects $g(\cdot, \cdot)$ have strictly decreasing differences on the domain $\{u, p\}$ (that is, $g(u, p) - g(u, u) > g(p, p) - g(p, u)$) if $\delta - \beta > 1 - \delta$, or

$$2\delta > 1 + \beta. \tag{DD}$$

That is, δ captures the desirability of diversity in education at colleges: the higher δ is, the more likely that (DD) is satisfied, hence that integration in colleges is total surplus enhancing. The parameter β reflects the "background gap" $g(p, p) - g(u, u)$ between the privileged and underprivileged, the lower β the higher the gap.

We will assume throughout the paper that diversity is desirable, that is that (DD) holds.[3]

There are many reasons to suspect that diversity in backgrounds is indeed desirable, that is (DD) holds. For instance, when the privileged have preferential access to resources, distribution channels, or information, the benefit of having a peer with a privileged background will be lower for a student who is

[3]The submodularity of g is sufficient for having a benefit of diversity but is not a necessary condition. Indeed, even if g is supermodular, that is $2\delta < 1 + \beta$, the interaction between $f(a, a')$ and $g(b, b')$ imply that the function $f(a, a')g(b, b')$ is not supermodular when considering the order $\ell u < \ell p < hu < hp$.

privileged. Furthermore, exposure to peers of a different background enables a student later to cater to customers of different socio-economic characteristics, for instance through language skills and knowledge of cultural norms. Finally, meeting peers of different backgrounds will expose students to methods of problem-solving, equipping them with a broader portfolio of heuristics they can draw on when employed in firms (following the argument by Hong and Page, 2001). Appendix A discusses alternate assumptions on the output functions.

2.1 Timing

The timing in the model economy is as follows.

1. Policies, if any, are put in place.
2. Agents of background b choose investment e_b. Given an investment e, the probability of achievement h is e and of achievement ℓ is $1 - e$.
3. Achievement is realized and is publicly observed.
4. Agents form colleges of size two in a matching market without search frictions, though it may be constrained by policies.
5. Once groups are formed, payoffs are realized and accrue to the agents.

2.2 Equilibrium

The matching market outcome (absent a policy intervention) is determined by a stable assignment of individuals into groups of size two given attributes ab, which are in turn determined by individuals' optimal choice of education acquisition e under rational expectations. A *college choice equilibrium* is therefore defined as a measure preserving matching function between individuals such that the following conditions are satisfied.

- (Payoff Feasibility) Within a group (i, j), the payoffs are respectively $y(a_ib_i, a_jb_j)$ and $y(a_jb_j, a_ib_i))$.
- (Stability) There do not exist two individuals who can be strictly better off by matching and choosing a feasible share of output given their equilibrium payoff.

Existence of such an equilibrium is standard, see, e.g., Kaneko and Wooders (1986). A college choice equilibrium determines individual payoffs for each attribute ab. Equilibrium payoffs will generally depend on the distribution of attributes, which is determined by education choices and the initial distribution of backgrounds. An *investment equilibrium* is defined as individual education choices $\{e_i\}$ such that:

- (Individual Optimality) Given investments $\{e_j, j \in [0,1]\}$, i's investment e_i maximizes his expected utility.

The fact that attributes in the college match are determined by stochastic achievement realizations of a continuum of agents simplifies matters. Indeed, let individuals be indexed by $i \in [0,1]$, with Lebesgue measure on the unit interval. W.l.o.g. assume that all agents $i \in [0,\pi)$ have background p and all agents in $i \in (\pi, 1]$ have background u. If the aggregate investment level of agents with background b is e_b, then, by a law of large numbers, the measures of the different attributes ℓu, ℓp, hu, and hp are respectively $(1-\pi)(1-e_u)$, $\pi(1-e_p)$, $(1-\pi)e_u$, and πe_p. Hence, given education choices e_b the distribution of attributes in the college match is unique.

This implies that college choice equilibrium payoffs only depend on aggregates e_u and e_p. Therefore in any investment equilibrium all u individuals face the same optimization problem, and all p individuals face the same optimization problem. Hence, in all investment equilibria all agents of the same background b choose the same education investment e_b.

Our analysis will describe the matching patterns in terms of attributes; because there may be 'unbalanced' measures of different attributes, the equilibrium matches of a given attribute may specify different attributes. For instance, both (hp, hu) and $(hp, \ell u)$ matches may be part of an equilibrium. This can be consistent with our definition of equilibrium matches only if the matches between attributes are measure-preserving.

3 Free Market and Investment Distortions

Beyond modifying the nature of matching, and generating segregation instead of diversity, NTU creates distortions on ex-ante investments. Before discussing the positive and normative effects of rematching policies, it is useful to contrast the matching pattern and the investment levels obtained in the free market situation with an ideal situation in which agents have no financial constraints and a price system exists for transferring utility at the college level. We consider this idealized situation below and then derive our "OTUB" result. By considering an economy where total surplus in a match is perfectly transferable, it becomes clear what these transfers should be, and how sensitive they will be to the relative scarcity of different types of individuals, e.g., the relative proportion of privileged in the economy.

3.1 Free Market Equilibrium

In the free market environment a student in a college $(ab, a'b')$ obtains payoff $y = f(a, a')g(b, b')$, yielding a joint surplus of $z(ab, a'b') = 2f(a, a')g(b, b')$; the Pareto frontier for a match $(ab, a'b')$ consists therefore of a single point. Our assumptions imply that the payoffs to each student in a match are given by the following matrix. Note that there is a natural order on attributes

Attributes	hp	hu	ℓp	ℓu
hp	1	δ	1/2	$\delta/2$
hu	δ	β	$\delta/2$	$\beta/2$
ℓp	1/2	$\delta/2$	α	$\alpha\delta$
ℓu	$\delta/2$	$\beta/2$	$\alpha\delta$	$\alpha\beta$

Table 1: Individual payoffs from matching into college $(ab, a'b')$

$hp > hu > \ell p > \ell u$ since a student's payoff is increasing in the attribute of his partner in this order. The free market equilibrium allocation without side payments has full segregation in attributes, because monotonicity implies that $\max\{y(ab, ab), y(a'b', a'b')\} > y(ab, a'b')$. This in turn precludes having in equilibrium a positive measure of $(ab, a'b')$ colleges, with $ab \neq a'b'$ because this would violate stability. Equilibrium payoffs are therefore:

$$v^F(hp) = 1, v^F(\ell p) = \alpha, v^F(hu) = \beta, v^F(\ell u) = \alpha\beta.$$

Therefore an agent of background b chooses e_b to maximize $e_b v^F(hb) + (1 - e_b)v^F(\ell b) - \frac{e_b^2}{2}$ implying that $e_b = v^F(hb) - v^F(\ell b)$, and therefore the equilibrium investment levels are:

$$e_p^F = 1 - \alpha \text{ and } e_u^F = \beta(1 - \alpha). \tag{3}$$

In the free market market equilibrium segregation by background is accompanied by differences between individuals of different backgrounds in outcomes such as investments e_b made before the match or payoffs $y_b \equiv e_b v^F(hb) + (1 - e_b) v^F(\ell b)$, which can be interpreted as individual education acquisition at college. We use background outcome gaps e_p/e_u and y_p/y_u to quantify investment and payoff inequality.

3.2 Full Transferability

Utility is fully transferable between partners in a match $(ab, a'b')$ when they can share the total output

$$z(ab, a'b') = 2f(a, a')g(b, b').$$

in a 1-1 fashion, that is when the Pareto frontier for a match $(ab, a'b')$ is obtained by sharing rules in the set

$$\{s : v(ab) = s, v(a'b') = z(ab, a'b') - s\}.$$

In our definition of equilibrium, the payoff feasibility condition in section 2.2 must be replaced by the condition that payoffs for i, j are bounded by this frontier.

The maximum transfer an individual is willing to make is equal to $y(ab, a'b')$, which corresponds to his life time earnings, which is of a degree of magnitude higher than the fees requested for attending the university. Hence, the case of perfect transferability is an ideal rather than a realistic case.

It is well known that under full transferability agents with the same attribute must obtain the same payoff.[4] Because of equal treatment there is no loss of generality in defining *the* equilibrium payoff of an attribute $v(ab)$. It is also well-known that the college choice equilibrium under fully transferable utility maximizes total surplus given realized attributes. The structure of payoffs and the stability conditions lead to the following observations.

Fact 1. (i) $(hp, \ell u)$ matches cannot be part of a first best allocation.

(ii) Conditional on agents of a given background matching together, segregation in achievement maximizes aggregate surplus.

[4] Otherwise, if one agent obtains strictly less than another this violates stability, as the first agent and the partner of the second agent could share the payoff difference.

(iii) Conditional on agents of a given achievement matching together, segregation by background is surplus inefficient when (DD) holds.

(iv) A first best allocation exhausts all possible (hp, hu) matches if $\delta > (1+\alpha\beta)/(1+2\alpha)$, and then all $(hu, \ell p)$ matches if $\delta > \beta(1-\alpha)(1-2\alpha)$.

The first statement follows because in a $(hp, \ell u)$ college hp agents lose more compared to their segregation payoff than ℓu agents gain. Indeed, when hp segregate, the average surplus in matches (hp, hp) is 1; when lu segregate, the average surplus is 0, and because the total surplus in a match (hp, lu) is equal to $\delta/2 < 1/2$, it is inferior to what hp agents obtain under segregation. The second statement (ii) holds whenever $f(a, a')$ has weakly increasing returns, hence in our constant return case. For (iii) note that condition (DD), is equivalent to $2z(hp, hu) > z(hp, hp) + z(hu, hu)$,implying that segregation in background is inefficient. Observation (iv) is perhaps a little surprising: even when both $f(\cdot,\cdot)$ and $g(\cdot,\cdot)$ have increasing differences, which tends to favor segregation, some integration of hu and ℓp is efficient.[5] The reason for this is that achievement and background are complements in a college. Therefore in a match of a privileged low achiever and an underprivileged high achiever, who were previously segregated, the increase in surplus $z(\cdot,\cdot)$ due to network effects $g(\cdot,\cdot)$ is sufficient to offset the possible loss of surplus due to the change of inputs to production $f(\cdot,\cdot)$:

$$
\begin{aligned}
&f(h,\ell)[g(u,p) - g(u,u)] - f(h,\ell)[g(p,p) - g(u,p)] \\
&> [f(\ell,\ell) - f(h,\ell)]g(p,p) + [f(h,h) - f(h,\ell)]g(u,u).
\end{aligned}
$$

Since one (hp, hp) and two $(\ell p, hu)$ colleges achieve higher aggregate payoffs than one $(\ell p, \ell p)$ and two (hu, hp) colleges, $(\ell p, hu)$ matches are exhausted, and only any remaining hu students match with hp students. Figure 1 shows the possible equilibrium matching patterns under full transferability depending the desirability of diversity. The plain arc indicates the first priority matching, the dashed arc indicates the second priority potential match, once the first priority matches are exhausted, and the ellipsis matches when these second matches are exhausted.[6]

[5]This extends to cases when both $f(a, a')$ and $g(b, b')$ have strictly increasing differences. Hence, the condition to have segregation as the surplus maximizing allocation, i.e., supermodularity of the surplus function $z(ab, a'b')$, is substantially more demanding in a world with multidimensional attributes than in a one-dimensional world.

[6]When the function f exhibits sufficient supermodularity, that is when $f(\ell, h)$ is smaller

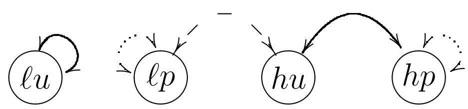

Figure 1: TU equilibrium matchings

As above investments depend on the market premium for high achievement $v^*(hb) - v^*(\ell b)$. Since ℓu students segregate under TU, $v^*(\ell u) = 0$. Payoffs for other attributes will depend on relative scarcity, which in turn will depend on the initial measure of privileged π and achievable surplus $z(ab, a'b')$. The following statement summarizes the properties of TU equilibrium investment levels when there is a high diversity benefit.

Fact 2. Suppose (DD) holds. Under full TU investment levels e_p^* and e_u^* increase in π. $\delta \leq e_p^* < 1$ for $\pi < 1$ and $e_p^* = 1$ for $\pi = 1$. $\beta \leq e_u^* < \delta$ for $\pi < 1$ and $e_u^* = \delta$ for $\pi = 1$.

Investment in education increases in the share of the privileged, because $v^*(\ell u) = 0$, hu students' payoffs determine e_u^*, which increases in the measure of available privileged matches and approaches δ as ℓp agents become abundant. The payoff of hp agents increases in the measure of hp agents that will segregate, while ℓp agents' payoffs decrease as hu agents become scarcer.

If one thinks of the first best outcome as the matching pattern that maximizes total surplus, the following lemma states that the equilibrium of the TU environment indeed leads to a first best allocation. In the proof we show that the payoff difference $v^*(hb) - v^*(\ell b)$ coincides with the social marginal benefit of investment by an individual of background b.

Lemma 1. *The equilibria of the TU environment lead to first best allocations: matching is surplus efficient given the realized attributes, and investment levels maximize ex-ante total surplus net of investment costs.*

3.3 OTUB

Comparing free market market equilibrium investments e_b^F to the first-best ones given in Fact 2, shown in Figure 2, yields the following proposition.

Proposition 1 (OTUB). *The privileged over-invest for $\pi < 1$. The underprivileged never over-invest and under-invest if $\pi > \frac{\beta}{1+\beta}$, in which case there is both over-investment at the top and under-investment at the bottom of the background distribution.*

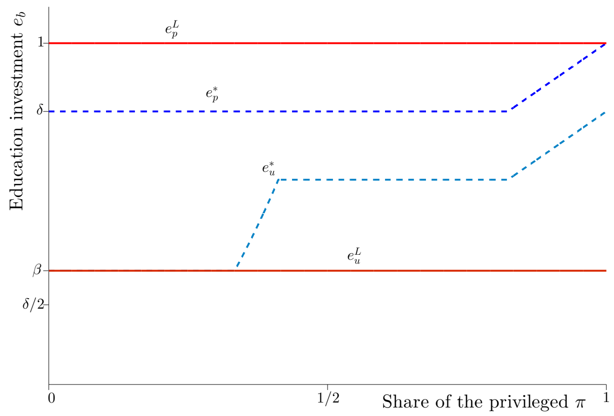

Figure 2: Education investments under free market and TU

The presence of simultaneous under-investment by the underprivileged and over-investment by the privileged reflects two properties of the surplus function. First, diversity in backgrounds is beneficial holding constant the composition of achievements (implied by $2\delta > 1 + \beta$). The second property is complementarity of diversity and returns to investments (here implied by separability of achievement and background in $z(.)$ and the fact that $g(u, p) > g(u, u)$).[7] Both properties guarantee that there will be over-investment at the top and under-investment at the bottom for π high enough. As we show in Appendix A this observation extends to more general settings.

The extent to which the privileged overinvest is determined by the desirability of diversity δ, which captures the marginal social benefit of privileged investment in (u, p) colleges. The extent of underinvestment at the bottom depends on both δ, determining the marginal social benefit, and the disadvantage β, which pins down the incentive to invest in the free market.

This result is interesting for several reasons. First, it formalizes the idea that an imperfect price system can generate excessive segregation, a static inefficiency, but is also accompanied by insufficient investment of underprivileged and excessive investment for privileged, a dynamic inefficiency. This suggest that the discouragement effect that rematching policies arguably have on those not favored, i.e., the privileged, can be desirable. Second, the result connects well with empirical findings. Interpreting background as race,

than 1/2, the first priority match is (hu, hp) and the second priority match is $(\ell p, hu)$.

[7] Desirability of background diversity is not a necessary condition for OTUB in general. For instance, OTUB occurs in this setting also when $1/2 < \delta < 2/3$ and $\beta = \delta/2$, for $\pi \in (1/2, 1)$. It is necessary, however, that some background integration occurs in the benchmark allocation.

a black-white test score gap already in place at early ages (Heckman, 2008) would be amplified by background segregation. Recent evidence for this is provided in Card and Rothstein (2007) and Hanushek et al. (2009).

Third, excessive segregation also has implications for inequality and polarization. Indeed, computing background gaps as a measure of inequality yields the following corollary.

Corollary 1. *Inequality in investments e is higher in the free market outcome than in the first best. Inequality in payoffs y is higher for intermediate and high π in the free market outcome than in the first best.*

Hence, if backgrounds are distributed relatively equally, excessive segregation is accompanied by excessive income inequality. In other instances however, income inequality may be greater in the first best benchmark as scarce attributes are paid their full market price (for instance when π is close to 0, hp agents obtain $2\delta - \beta$ in the first best, but only 1 under free market).

4 The Positive and Normative Effects of Diversity Policies

We now introduce some policies that are widely used and examine them in terms of their effects on aggregate outcomes such as investments, payoffs, and inequality, as well on welfare.

Real world policies aim at replicating population measures of backgrounds in colleges, but vary in the degree to which they allow to condition admission on achievement. We will focus on two extreme policies. First, we will consider a policy of background integration that re-matches students without regard to achievements until each college's expected background composition equals that in the population. Second, a policy of affirmative action that gives priority to the underprivileged only over privileged students who have at most the same achievement level. Because a large part of the efficiency of the match is linked to the achievement part of the attributes, a busing policy tends to perform often worse than an affirmative action policy. Studying these polar cases allows some inference on intermediate ones, e.g., scoring policies where a score reflecting both achievement and background determines priority.

4.1 Achievement Blind Policies

The first policy we consider replicates the population distribution of backgrounds in each college, unconditional on achievements. The most prominent example of such a policy is the use of "busing" in the U.S. to achieve high school integration.[8] Public European colleges often do not condition admission to achievement beyond the basic requirement of having finished high school. Formally, the policy is an assignment rule that randomly integrates colleges in background, ignoring achievement, until no further integration is possible.

Definition 1. An *achievement Blind policy* (denoted B policy) exhausts all possible matches of underprivileged and privileged background, using uniform rationing conditioning on background.

Uniform rationing means for instance that when u students outnumber p students, a u student is matched to a p student with probability $\pi/(1-\pi)$. The rule is silent on the matching of any remaining students from the larger background group, who may segregate in achievements. Note that the expected background composition at colleges equals the one in the population. Such a policy is thus best understood as one that departs from the free market outcome of full segregation and randomly reassigns agents to match the expected share of privileged students at each college to their population measure π.

The definition of the policy and the fact that high achievers of both backgrounds strictly prefer to segregate in achievements if they are not subject to a random re-match implies the following equilibrium matching pattern, characterized in the lemma and Figure 3below. Dotted arrows indicate matches subject to availability of agents after exhausting matches denoted by solid arrows.

Lemma 2. *Under a B policy a u agent obtains an hp match with probability* $e_p \max\{\pi/(1-\pi); 1\}$ *and an ℓp match with probability* $(1-e_p)\max\{\pi/(1-$

[8] Busing policies operated mainly by redesigning college districts to reflect aggregate population measures. A public college match then would assign students to colleges. A more recent example is the integration of college catchment areas in Brighton and Hove, U.K. Other examples include reservation in India to improve representation of schedule castes and tribes, the Employment Equality Act in South Africa, under which some industries such as construction and financial services used employment or representation quotas, or the SAMEN law in the Netherlands (until 2003).

π); 1}. If $\pi > 1/2$*, a measure* $(2\pi - 1)$ *of privileged segregate in achievements; if* $\pi < 1/2$*, a measure* $1 - 2\pi$ *of underprivileged segregate in achievements.*

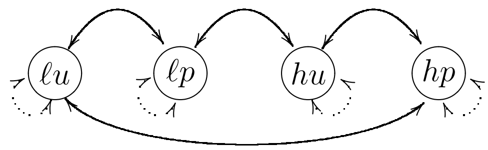

Figure 3: Equilibrium matching under a B policy.

Because this pattern allows both (hu, hp) matches and $(\ell p, hu)$ matches, this policy may be beneficial for increasing surplus *if investment in achievement is not important*, e.g., if the distribution of types is given. However, because the assignment rule does not depend on achievement investment incentives are likely to b depressed compared to the free market in general.[9] This may explain why these policies have been mainly used at the primary or secondary levels rather than at the university level where prior investment in human capital is more important.

The following statement uses Lemma 2 to verify this intuition; details are in the appendix:

Proposition 2. *Investments under a B policy are lower than in the free market outcome for both backgrounds, as are aggregate investment and payoffs. A B policy induces lower payoff inequality than free market for* $\pi \in (0, 1)$*, and lower investment inequality if* π *is not too large. For* $\pi < 1/2$ *the investment gap between backgrounds reverses and aggregate investment by the underprivileged exceeds that of the privileged.*

That is, a B policy is indeed subject to the classic equity-efficiency trade-off that seems to guide much of the policy discussion. Reducing outcome inequality in the economy comes at the cost of undesirable incentive effects depressing levels of investment and output: both the privileged and the underprivileged are discouraged relative to the free market outcome, because higher investment does not increase the probability of obtaining a better match, see Figure 4 (the parameters used to generate this and all other figures are $\beta = .5$ and $\delta = .85$). In fact, when the privileged are a minority a B policy can reverse the background gap in investment, so that $e_p^B < e_u^B$.This comparative statics exercise assumes that when π varies, both δ, β stay constant, which may be a strong assumption in general.

[9]Both ℓ and h agents of background b have the same chance of being matched to an h agent of background b'

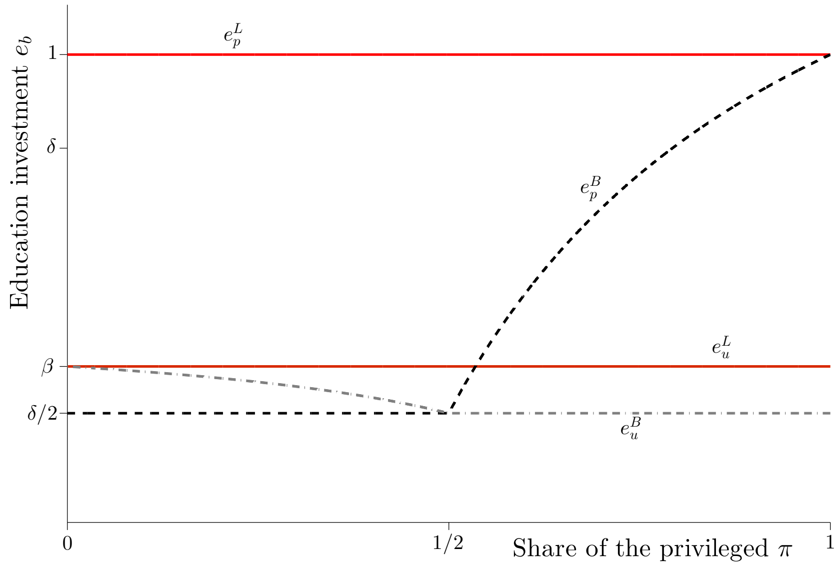

Figure 4: Education investments using a B policy.

4.2 Affirmative Action Policy

We examine now the case where precedence is given for an underprivileged candidate over a privileged competitor of the *same* achievement level only. Formally, affirmative action is a priority for the underprivileged for positions at a given level of achievement. It is widely used in colleges (for instance the "positive equality bill" in the U.K., *Gleichstellung* in the German public service, or reservation of places for highly qualified minority students at some *grandes écoles* in France, like Sciences Po Paris).

Definition 2. Consider an equilibrium and a match $(ab, a'b)$. An *affirmative action policy* (denoted A policy) requires that an agent with attribute au must not strictly prefer to join $a'b$ to staying in his current assignment.

For instance, if a school wants to attract high achievement students, priority should be given to hu students, hence a school $(hp, \ell p)$ can form only if there is no hu student who would like to be matched in a school with a ℓp student.

Lemma 3. *Under an A policy, low achievers do not match with high achievers, and all (hp, hu) matches are exhausted, that is the measure of such integrated matches is* $\min\{(1-\pi)e_u, \pi e_p\}$.

Proof. While hp agents would prefer to segregate, since hu agents strictly prefer a match with an hp agent to one with any other agent, (hp, hp) can occur only if there are no hu agents who are not already matched with hp agents. Hence, all possible (hp, hu) matches must be exhausted, and the measure of such matches is $\min\{(1-\pi)e_u, \pi e_p\}$. The other high achievers segregate. The matches of the low achievers are indeterminate, as any match between them gives zero payoff. □

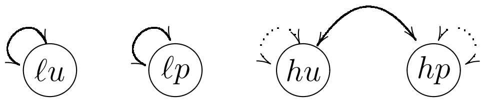

Figure 5: Equilibrium matching under an A policy.

The equilibrium matching pattern under an A policy is shown in Figure 5. As under the B policy optimal individual investment levels will depend on the match an agent expects to obtain, and thus on relative scarcities. Since the privileged only have to accept underprivileged matches if they have the same achievement level, privileged investments will be less depressed than under the B policy. The following proposition states this and other properties of aggregate outcomes under an A policy; details are in the appendix.

Proposition 3. *Under an A policy the underprivileged invest more than under free market ($e_u^A > e_v^F > e_u^B$), and the privileged less ($e_p^F > e_p^A > e_p^B$). Inequality of both investment and payoffs between backgrounds is smaller under the A policy than under free market. Aggregate investment and payoffs are higher, if diversity is desirable enough, or if backgrounds are distributed asymmetrically (π is small or large).*

Not only does an A policy crowd out privileged investment by less than a B policy, but also underprivileged investment is boosted compared to free market, see Figure 6. This is because under an A policy an underprivileged student's expected return from investment is given by the difference of being matched into an (hu, hp) to an $(\ell u, \ell p)$ college, not insuring the agent against low achievement as did the B policy. That is, expected returns to investment are now conditional on integrating in backgrounds. This encourages the underprivileged and discourages the privileged, and, if diversity is desirable – that is condition (DD) holds – the aggregate effect on investment is positive. If diversity is desirable or backgrounds are distributed unevenly also aggregate output is higher.

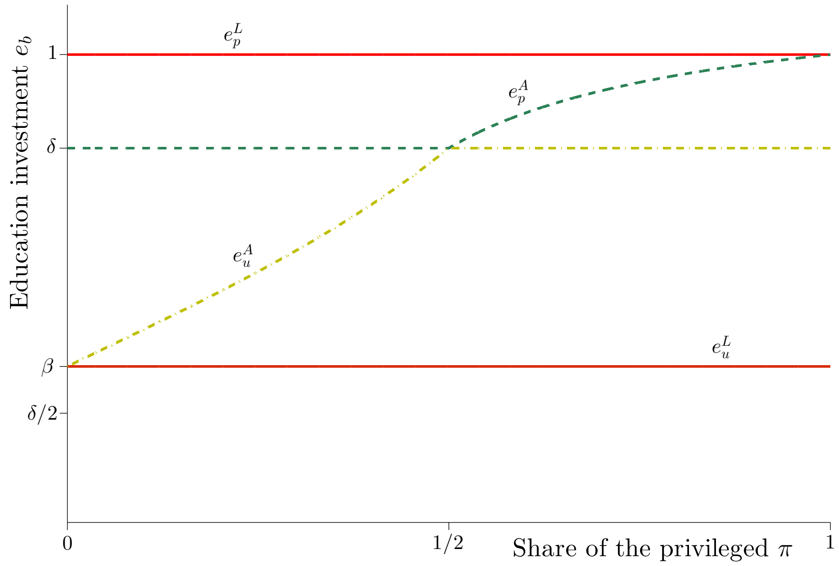

Figure 6: Education investments under an A policy.

4.3 Discussion

The two policies of re-match considered above differ substantially in terms of their position in the trade-off between static and dynamic concerns, i.e., between achieving more efficient sorting ex post, when attributes have realized, and maintaining investment incentives by rewarding investments adequately through the match. Policies that emphasize replicating population frequencies of backgrounds in each college (B policies) may do well in terms of the first but will in general fail in terms of the second. Policies that implement integration only between students that have similar achievement levels forego some benefits of improving the sorting ex post, since for instance matches $(\ell p, hu)$ will not be realized, but induce high investment incentives, mainly by providing access to mixed firms for the underprivileged. Figure 7 illustrates the differences in aggregate performance.[10]

Both types of policy tend to decrease inequality in the economy compared to free market: they decrease the privileged's investment incentives substantially, while the underprivileged's incentives increase with access to better matches. Here investment inequality is also an indicator of social mobility,

[10]Even if the proportions of attributes is given, that is even if one is not concerned about investment incentives, an affirmative action policy dominates an achievement blind policy, and also free market: the A policy foregoes $(hu, \ell p)$ matches but avoids many other surplus decreasing matches, like $(hp, \ell u)$ that arise under a B policy. Obviously, if incentives are ignored, the "naive" policy that replicates the first best match distribution under TU performs even better than the A policy.

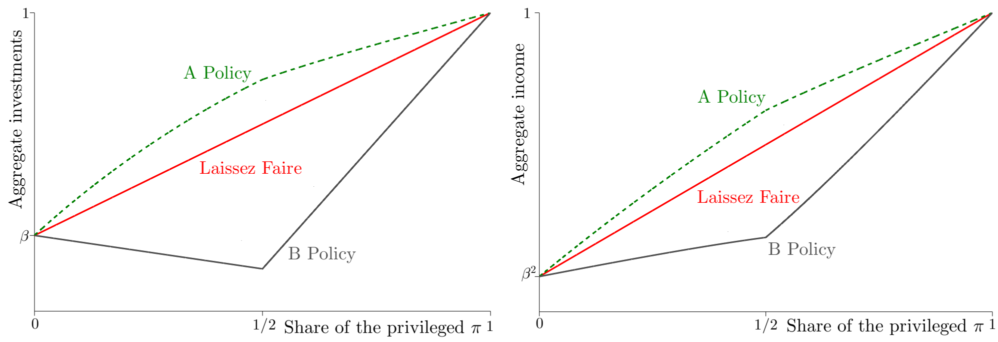

Figure 7: Aggregate investments (left) and aggregate payoff (right).

in terms of the predictive power of parental background on own achievement and payoffs. Figure 8 shows the investment and payoff ratios of privileged to underprivileged.

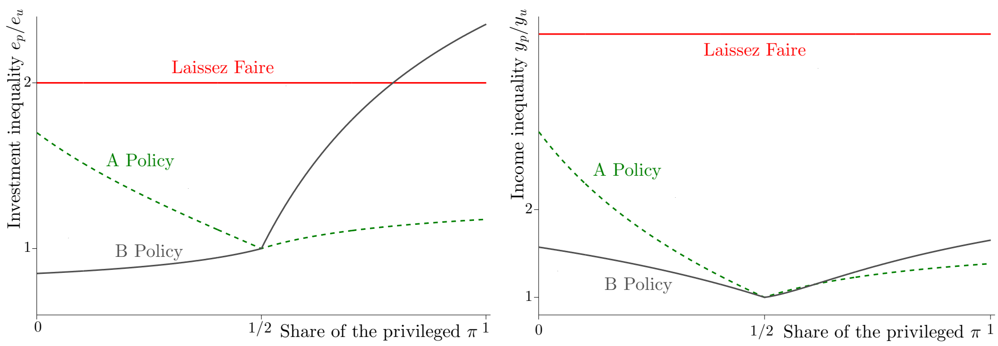

Figure 8: Background gap in investments (left) and payoff (right).

For practical purposes our results suggest that policies that condition on achievement, for instance in form of scoring rules that assign high weight on attainments, are preferable to those that simply mix in terms of backgrounds. Second, background in our model refers to socio-economic circumstances, which may or may not correlate highly with other markers such as race or gender. Our results then suggest that socio-economic characteristics should take precedence over race the less correlated both become, since an A policy conditioning on race will increasingly fail to ensure access of the underprivileged to (hu, hp) colleges. This seems interesting in light of the public interest aroused by the imminent U.S. Supreme Court decision on the case Fisher v. college of Texas in whether social background based will and should replace race based affirmative action in college admission.

Other types of rematch can be imagined and are used. One type could condition the match on achievement, integrating low and high achievers as much as possible. A move from the free market outcome to using such a policy would, for instance, correspond to abolishing tracking at colleges, that is, assortative sorting of pupils based on past grade achievements. Achievement based policies tend to fare worse than background based policies in terms of output and investment, since incentives tend to be depressed by the guarantee of a good match in case of low achievement, and we have assumed that there are no benefits from integration in achievement ($f(\ell,\ell)+f(h,h)=2f(h,\ell)$).

4.4 Welfare

Until now we have compared the policies to free market in terms of aggregates like output, inequality, investment in human capital, which are aggregates that may matter when thinking of growth for instance. A policy may perform well in these aggregates but perform poorly in terms of welfare, for instance, output and investment may grow but the net surplus may decrease. As we show now, affirmative action policies may also improve welfare, measured in aggregate surplus, that is, payoff net of investment cost. Note first that total surplus can be written as:

$$S = \pi\frac{e_p^2}{2} + \pi v(\ell p) + (1-\pi)\frac{e_u^2}{2} + (1-\pi)y(\ell u).$$

Since $v^A(\ell u) = v^F(\ell u) = 0$ and $v^A(\ell p) = v^F(\ell p) = 0$ total surplus under an A policy exceeds that under free market, $S^A > S^F$, if, and only if:

$$(1-\pi)[(e_u^A)^2 - (e_u^F)^2] > \pi[(e_p^F)^2 - (e_p^A)^2].$$

That is, an A policy induces higher welfare than free market if the encouragement of the underprivileged ($e_u^A - e_v^F$) outweighs the discouragement of the privileged ($e_p^F - e_p^A$). Since e_p^A is a convex combination of $e_p^F = 1$ and δ, and e_u^A a convex combination of $e_v^F = \beta$ and δ, an A policy appears desirable if diversity is sufficiently desirable, i.e., δ is sufficiently close to 1. The following proposition states this and compares A and B policies, its proof is in the appendix.

Proposition 4 (Welfare). *(i) An A policy dominates a B policy in terms of total surplus, that is, $S^A > S^B$ for $0 < \pi < 1$.*

(ii) A sufficient condition for an A policy to induce higher total surplus than

free market is $\delta > \sqrt{(1+\beta^2)/2}$.

A B policy that discourages both the privileged and underprivileged alike may induce higher aggregate surplus than free market as it induces a more efficient match than the free market. This static gain in output from rematching may be sufficient to compensate for the dynamic loss by depressing investment, in particular since welfare accounts for investment cost.

There are indeed two different routes to improve in welfare with respect to the free market outcome: one is to induce a lot of mixing in the market, depressing dramatically incentives but inducing static gains from rematch and savings on investment costs; the other one constrains rematch to preserve incentives, possibly forgoing static gains from rematch.

This raises the question of how a policy would perform that naively replicates the first best matching, i.e., exhausts all $(\ell p, hu)$ matches, and then all (hu, hp) matches. In fact this "naive" policy is quite sophisticated, since its informational demands for implementation are substantial. This policy faces a similar trade-off as the B policy: while maximizing the static gains from rematching ex post, it ignores possible incentive effects, resulting in lower investment by both backgrounds than under free market. Nevertheless a naive policy may induce higher surplus than free market if δ is high and β low. An A policy performs better than such a naive policy if δ is high enough, a B policy if δ and β high enough, see Appendix for details.

Finally, one might be interested in how close the A policy is from the second-best optimal. We focus on a policy space such that the social planner controls the measures $\rho(ab, a'b')$ of matches $(ab, a'b')$ subject to feasibility. The optimization problem of a planner is:

$$\max_{\rho} \sum_{ab,a'b'} \rho(ab, a'b')z(ab, a'b') - \pi\frac{e_p^2}{2} - (1-\pi)\frac{e_u^2}{2}$$

subject to incentive constraints: for $b = p, u$:

$$e_b = \sum_{a'b'} \frac{\rho(hb, a'b')}{\pi_b e_b} y(hb, a'b') - \sum_{a'b'} \frac{\rho(\ell b, a'b')}{\pi_b(1-e_b)} y(\ell b, a'b'),$$

and feasibility: for $b = p, u$:

$$\sum_{a'b'} \rho(hb, a'b') + \rho(hb, hb) = \pi_b e_b \text{ and}$$
$$\sum_{a'b'} \rho(\ell b, a'b') + \rho(\ell b, \ell b) = \pi_b (1 - e_b).$$

That is, the set of policies contains all feasible matching patterns ex post, which define the probabilities of being assigned to different attributes, which in turn determine investments. The A and B policies can be defined in terms of the control variables $\rho(ab, a'b')$. For instance, an A policy will require that $\rho(hu, hp)$ is equal to $\min\{\frac{(1-\pi)e_u}{\pi e_p}, \frac{\pi e_p}{(1-\pi)e_u}\}$, and that $\rho(\ell p, hu) = \rho(\ell u, hp) = 0$. The ρ values for the B policies are those in Lemma 2.

The set of policies also includes scoring polices that give priority to students based on scores: convex combinations of achievements and backgrounds. For instance, one could give "grade subsidies" based on ethnicity (as the college of Michigan until 2003) or on whether a student attended a public college (used in college admission in Brazil), or comes from a disadvantaged neighborhood.

The problem above has five control variables and a discontinuous objective function, making the problem hard to solve analytically. Numerical solutions indicate that the second best policy closely resembles an A policy for δ high enough and $\pi \geq 1/2$, see Appendix B for details. In fact the A policy realizes about 95 to 97% of the gains in surplus that the second best policy achieves when $\pi \geq 1/2$ (for $\delta = .85$ and $\beta = .5$, used for all figures), and more when δ becomes larger.

5 Partial Transferability

Another remedy to excessive segregation implied by NTU could consist in "bribing" ex-ante some students to re-match. Indeed, while a complete lack of side payments appears to describe well the assignment of pupils into public colleges, at all levels of education there are private colleges that charge tuition fees that may depend on students' academic achievements, for instance by offering scholarships. This introduces a price system for attributes, potentially affecting both the matching outcome and investment incentives. Often such a price system suffers from imperfections, for instance because individuals differ in the financial means at their disposal that can be used to pay tuition fees and some of them face borrowing constraints. As we already pointed out, since benefits from college are related to life time earnings, it is likely that the financial constraint binds for most students.

We introduce the possibility of transfers among students by assuming that agents differ in their wealth levels ω_b, depending on their background b. Plausibly, privileged background is associated to higher wealth. For $\omega_u < 1 - \delta$ and $\omega_p < \beta - \delta/2$ our previous analysis goes through unchanged, because neither can hu compensate hp students enough, nor ℓp students hu students. Suppose then that $\omega_p > \beta - \delta/2$ and, for simplicity that $\omega_u = 0$, i.e., the underprivileged are poor without access to loans. This means that the privileged can compensate the underprivileged, but not vice versa; Figure 9 shows the resulting possible payoffs for some attribute combinations.

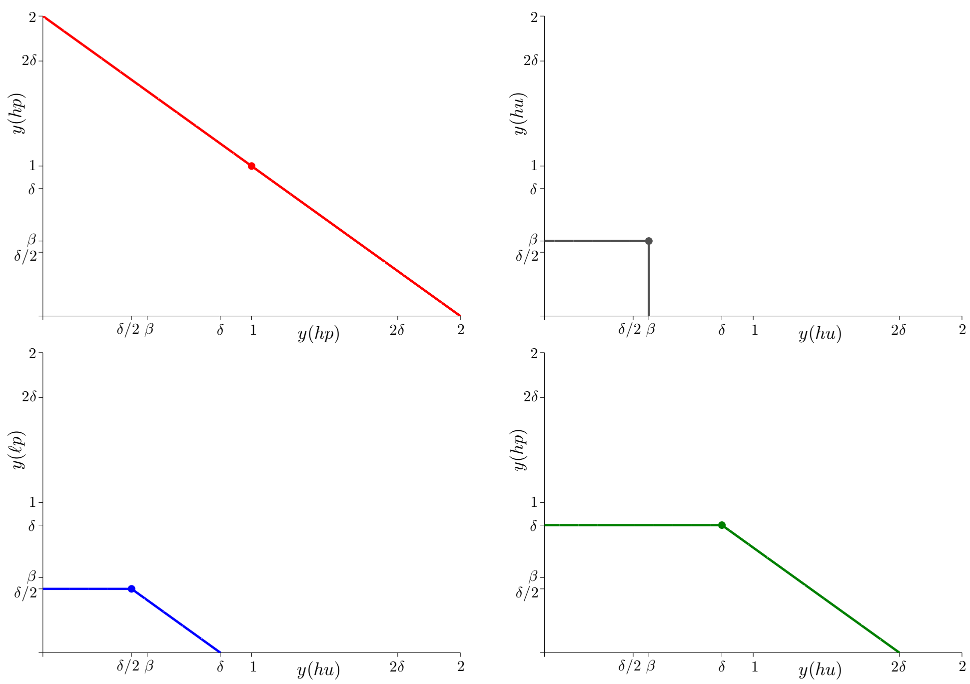

Figure 9: Possible distribution of payoffs in (hp, hp) and (hu, hu) colleges (top) and $(\ell p, hu)$ and (hp, hu) colleges (bottom) when individuals can make lump-sum transfers but the underprivileged face borrowing constraints.

The next statement follows directly from this observation.

Lemma 4. *A free market equilibrium when $\omega_p > \beta - \delta/2, \omega_u = 0$ exhausts all possible $(hu, \ell p)$ matches, ℓu and hp agents segregate.*

Figure 10 shows the resulting equilibrium matching pattern. The underprivileged match with the privileged, but only in $(hu, \ell p)$, not in (hu, hp)

colleges, and the elite (hp, hp) colleges are solely populated by the privileged, which seems to resonate well with the evidence.[11]

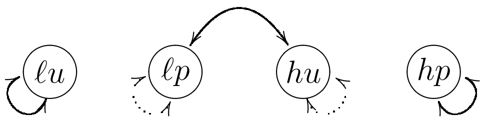

Figure 10: Equilibrium matching with transfers

We consider below the situation where the privileged have large weath, i.e., $\omega_p \gg \delta/2$ and analyze the effect of an A policy as defined above. An equilibrium under an A policy exhausts all (hp, hu) matches, as hu agents obtain δ in an (hu, hp) match, and at most δ in an $(hu, \ell p)$ match.

Proposition 5. *Suppose Condition (DD) holds, $\omega_p \geq \delta/2$, and $\omega_u = 0$. An A policy induces higher investment and payoffs for the underprivileged, and reduces the investment and payoff gap between backgrounds. If π is low the underprivileged invest more than the privileged. If π is intermediate or high an A policy induces higher investment, and, if δ is high enough, also higher payoffs for both backgrounds.*

As in the case without side payments, an A policy encourages investment by the underprivileged, since underprivileged high achievers are rewarded with access to privileged high achievers. By contrast, when side payments are possible an A policy may encourage investments by students of *both* backgrounds. This is because limited wealth limits competition among ℓps, thereby giving rents to privileged low achievers. An A policy depresses these rents for privileged low achievers, forcing them to compete with privileged high achievers for scarce underprivileged high achievers (when π is intermediate or high). This effect outweighs the decrease of the privileged high achievers' payoffs who are forced to match with the underprivileged, so that investment incentives for the privileged increase. This encouragement effect

[11] For instance, Dillon and Smith (2013) find evidence for substantial mismatch in the U.S. higher education system, in the sense that students' abilities do not match that of their peers at a college. This mismatch is driven by students' choices, not by college admission strategies, and financial constraints play the expected role: wealthier students, and good students with close access to a good public college are less likely to match below their own ability. Hoxby and Avery (2012) report that low-income high achievers tend to apply to colleges that are of lower quality than their own and seem less costly, in marked contrast to the behavior of high income high achievers (Table 3). They also find that prices at very selective institutions were not higher for the underprivileged than at non-selective institutions, although this does not account for opportunity cost of, e.g., moving.

is so strong that the expected payoff ex post of a privileged student is higher under an A policy, if diversity is desirable enough (δ sufficiently large). Figure 11 sums up the investment behavior when colleges use tuition fees.

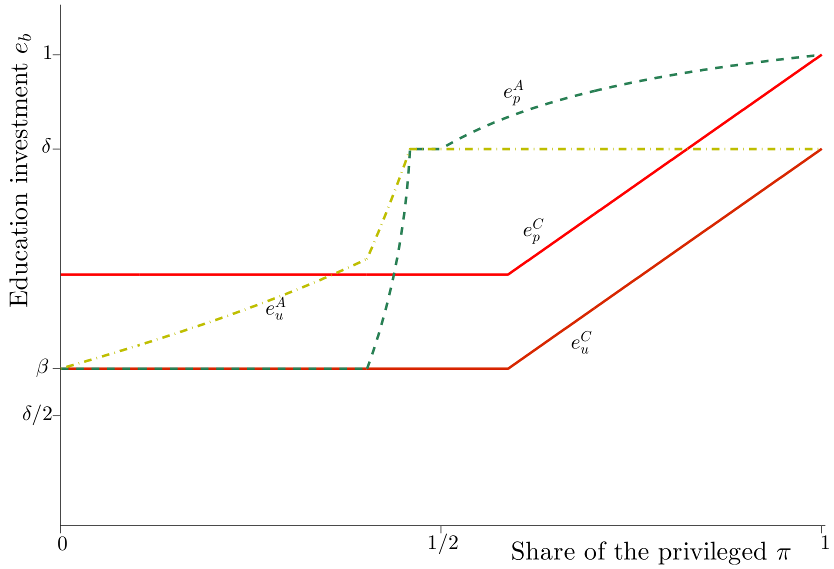

Figure 11: Investments when $\omega_p > \beta - \delta/2, \omega_u = 0$

When borrowing constraints are the only form of market frictions, endowment redistribution may have a similar effect as the re-matching of students given sufficient aggregate wealth. This could be done by giving tax financed grants to underprivileged high achievers to enable them to compensate privileged high achievers in (hu, hp) colleges. This grant could be conditional on the choice of college, like a scholarship, or unconditional, like a stipend. Though both versions encourage the underprivileged by inducing competition between ℓp and hp for hu agents, the privileged will in fact prefer an A policy to both stipends and scholarships from an ex ante point of view, as they will have higher rents in $(hu, \ell p)$ colleges under an A policy.

6 Conclusion

Though an excess of segregation in the collegiate marketplace has inspired many policy responses as well as much controversy, there has been little attempt to assess their *aggregate* economic consequences, that is taking into account the behavior of parties favored by the policy as well as those who are not. Starting with a model in which the benefits of college are a local public good, and students have limited means with which to bid for peers, we show that the free market will indeed generate excessive segregations, and as a consequence, under-investment by the underprivileged and over investment by the privileged. We study two simple policies, one integrating backgrounds to match population measures without considering achievement and one giving priority to one background only conditional on achievement, and show that these policies may improve on the free market in terms of aggregate investment, output, surplus, and inequality, and can be ranked in terms of aggregate performance. Policies giving priority on the basis of achievement tend to perform better overall.

Though not exhaustive, the set of policies we examine covers the two extremes in terms of conditioning integration on achievement, allowing us to uncover considerable differences in the consequences for investment incentives, suggesting that conditioning on achievement is desirable. Moreover, numerical simulations show that this policy may in fact come close to a second best. While of interest, the question of the "optimal policy" in general settings is best left to future research. This quest will require us to compute complex contingencies, which will raise the issue of its practical implementation. Our focus on policies that are actually used by policymakers yields a convincing economic rationale for the use of such policies, when students' ability to make side payments is constrained and diversity is desirable.

We have introduced the possibility of transfers and as long as underprivileged have limited wealth or difficulties to borrow, vouchers or grants have limited success in generating diversity. Vouchers are feasible but not a market equilibrium. Similarly, need blind policies are feasible but diversity would require that smart underprivileged hu apply for admission while the smart privileged hp are also willing to apply: as we argue this would require that hp actually pay less than hu for otherwise they would segregate.

An extension of the approach would be to consider a dynamic setting in which the background (at least if it is interpreted as socioeconomic status) as well perhaps as the diversity parameters β and δ evolve endogenously. Such a model could provide an efficiency rationale for affirmative action as a remedy for "righting past wrongs," if, for instance, the background gap parameter β could be induced to increase, thereby raising the productivity of the underprivileged. It would also provide an avenue for understanding "segregation traps," in which lineages of underprivileged remain so because they lack the benefits of exposure to the privileged. Finally, such a model would be an appropriate one to revisit the question of the optimal duration of diversity policies.

Another question concerns relaxing the assumption that both backgrounds have the same investment costs. It is straightforward to modify the model to allow, for example, higher marginal costs for the underprivileged than for the privileged. This will tend to mitigate the benefits of an affirmative action policy, both because the underprivileged's investments will be less responsive, and because the privileged, now less likely to match with the underprivileged, will reduce their investment less. A pertinent observation is that investments often happen in environments such as primary and secondary school or neighborhoods, in which there are peer effects and in which the market outcome is characterized by similar imperfections as the one we considered here. Rematching policies can be applied at the school or neighborhood level as well las at college, and this raises questions of how rematching policies in one level impact on the performance of matching policies in another, as well as the complementarity or substitutability of rematch policies on sequential markets. Some progress on these issues has been made in Estevan et al. (2013) and Gall et al. (2014).

Finally, we have focused on how students match into colleges, where rigidities arise naturally from local public goods and borrowing constraints. Our results extend to other settings as well, e.g., the labor market. A firm can be formalized as the set of its members and a contractual arrangement among its members, which is often designed to address agency problems. This typically results in a second best contract, inducing substantial nontransferabilities between firm members. This can be sufficient to generate excessive segregation and opens the door to a similar analysis of the aggregate effects of affirmative action policies in the labor market.

A Appendix: Generalized Surplus Function

Denote attributes by $s \in \{\ell u; \ell p; hu; hp\}$, endowed with a natural order, satisfying $\ell u < \ell p, hu$ and $hp > hu, \ell p$. Let $z(s, s')$ be monotone in its arguments ($z(s, s') > z(s, s'')$ if $s' > s''$).[12] Assume that $z(hp, hp) < 2$ to permit easy interpretation of investments as probabilities. The functional

[12] A weaker form of monotonicity, $z(s, s') < \max\{z(s, s); z(s', s')\} \leq z(hp, hp)$ for all $s \neq s'$ is sufficient.

form $z(s,s') = 2f(a,a')g(b,b')$ satisfies these assumptions.

Diversity is desirable, that is, for $s = ab$ and $s' = a'b'$ with $b \neq b'$

$$2z(s,s') > z(s,s) + z(s',s'). \tag{DD}$$

This corresponds to the case of $2\delta > 1+\beta$ in the functional form used above. Note that this property does not restrict the surplus function with respect to the composition of achievements ℓ and h, in particular decreasing and increasing differences are possible.

$z(.)$ satisfies *complementarity* of diversity and returns to education if

$$2[z(hu,s) - z(\ell u,s)] \geq z(hu,hu) - z(\ell u,\ell u) \text{ for } s \in \{hp, \ell p\}. \tag{C}$$

For this general surplus function, our OTUB result generalizes when (DD) and (C) hold.

Proposition 6. *Suppose properties (DD) and (C) hold.*
(i) There is $\underline{\pi} > 1/2$ *such that for all* $\underline{\pi} < \pi \leq 1$ *under free market privileged agents over-invest* ($e_p^* > e_p^T$)*, and underprivileged agents under-invest* ($e_u^* < e_u^T$)*.*
(ii) If $\pi < \underline{\pi}$ *and* $z(hu,hu) - z(\ell u,\ell u) < 1$ *there is under-investment by the underprivileged* ($e_u^* \leq e_u^T$)*. Under-investment is strict if additionally* $z(hu,hu) - z(\ell u,\ell u) < 2(z(hu,\ell u) - z(\ell u,\ell u))$*.*

The threshold $\underline{\pi}$ is given by $\underline{\pi} = \frac{1}{2(z(hp,hp) - z(hp,\ell p))}$ if $2z(hp,\ell p) > z(hp,hp) + z(\ell p,\ell p)$ and by $\underline{\pi} = \frac{1}{z(hp,hp) + z(\ell u,\ell u) - 2z(\ell p,\ell u)}$ otherwise.

Proof. Because of property (DD) under TU there cannot be positive measures of both matches $(ab, a'b)$ and $(ab', a'b')$. Hence, for any composition of achievements (a,a') the TU allocation exhausts all possible matches with background composition (u,p).

(i) Start by examining the case of $\pi e_p^T > 1/2$, i.e., oversupply of hp agents under TU. In this case $v(hp) = z(hp,hp)/2$ and $v(hu) = z(hp,hu) - z(hp,hp)/2$ by property (DD).

Suppose $(hp,\ell p)$ matches occur in equilibrium then $v(\ell p) = z(hp,\ell p) - z(hp,hp)/2$ and $e_p^T = z(hp,hp) - z(hp,\ell p)$ yielding the condition

$$\pi > 1/2(z(hp,hp) - z(hp,\ell p)).$$

Moreover, $e_p^T = z(hp, hp) - z(hp, \ell p) > (z(hp, hp) - z(\ell p, \ell p))/2 = e_p^*$ since $(hp, \ell p)$ matches occur (and thus are preferred by both hp and ℓp agents to segregation). $v(\ell u) = z(hp, \ell u) - v(hp)$ by property (DD), since $(hp, \ell p)$ matches occur. This means $e_u^T = z(hu, hp) - z(\ell u, hp) > (z(hu, hu) - z(\ell u, \ell u))/2 = e_u^*$ by property C.

Suppose $(hp, \ell p)$ matches do not occur in equilibrium. Then $(\ell p, \ell u)$ matches occur in equilibrium by property (DD). If $\pi(1-e_p) < (1-\pi)(1-e_u)$ then $v(\ell u) = z(\ell u, \ell u)/2$ and $v(\ell p) = z(\ell p, \ell u) - z(\ell u, \ell u)/2 > z(\ell p, \ell p)/2$. Hence, $e_p^T = z(hp, hp)/2 + z(\ell u, \ell u)/2 - z(\ell p, \ell u) < e_p^*$. $e_u^T = v(hu) - z(\ell u, \ell u)/2 > (z(hu, hu) - z(\ell u, \ell u))/2 = e_u^*$. Using these expressions reveals that $\pi e_p^T > 1/2$ implies $\pi(1 - e_p) < (1 - \pi)(1 - e_u)$. Therefore oversupply of hp agents and absence of $(hp, \ell p)$ matches is only consistent with $\pi(1 - e_p) < (1 - \pi)(1 - e_u)$.

(ii) If there are $(\ell u, \ell u)$ matches $v(\ell u) = v(\ell u, \ell u)/2$. If $z(hu, hu) + z(\ell u, \ell u) < 2z(hu, \ell u)$ there cannot be (hu, hu) matches as well. Therefore $w(hu) > z(hu, hu)/2$ and $e_u^T > e_u^*$. Otherwise ℓu agents' payoffs are determined by the equilibrium matches $(\ell u, s)$ yielding $v(\ell u) = z(\ell u, s) - v(s)$ for some skill level $s \in \{hu; \ell p; hp\}$. $w(hu) \geq z(hu, s) - v(s)$ with strict inequality if matches (hu, s) do not occur in equilibrium. Suppose there is $s \in \{hp; \ell p\}$ so that $(\ell u, s)$ matches occur in equilibrium, then by Property (C) $e_p^T = z(hu, s) - z(\ell u, s) > [z(hu, hu) - z(\ell u, \ell u)]/2 = e_u^*$. Otherwise all ℓu agents must be matched to hu, which requires $e_u^T > 1/2$. If $z(hu, hu) - z(\ell u, \ell u) < 1$ this implies $e_u^T > e_u^*$. □

A Policy vs. Free Market

The following proposition provides an analogue to Proposition 4, stating that surplus under an A policy is higher than under free market if δ is close enough to 1.

Proposition 7. *Aggregate surplus under an A policy is higher than under free market if $z(hp, hu)$ is sufficiently close to $z(hp, hp)$.*

Proof. As shown above there is full segregation in an equilibrium under free market with investments:

$$e_p^F = \frac{z(hp, hp) - z(\ell p, \ell p)}{2} \text{ and } e_u^F = \frac{z(hu, hu) - z(\ell u, \ell u)}{2}.$$

Total surplus under free market is

$$S^F = \pi\frac{(z(hp,hp)-z(\ell p,\ell p))^2}{8} + \pi\frac{z(\ell p,\ell p)}{2} + (1-\pi)\frac{(z(hu,hu)-z(\ell u,\ell u))^2}{8} + (1-\pi)\frac{z(\ell u,\ell u)}{2}.$$

Under an A policy both ℓp and ℓu agents segregate, so that $v^A(\ell p) = z(\ell p,\ell p)/2$ and $v^A(\ell u) = z(\ell u,\ell u)/2$. This means total surplus is higher under the A policy if

$$\pi\frac{(e_p^A)^2}{2} + (1-\pi)\frac{(e_u^A)^2}{2} > \pi\frac{(z(hp,hp)-z(\ell p,\ell p))^2}{8} + (1-\pi)\frac{(z(hu,hu)-z(\ell u,\ell u))^2}{8}.$$

Since h types' wages depend on relative scarcity of background two different cases may arise. The first is that $(1-\pi)e_u > \pi e_p$. Then $v^A(hp) = z(hp,hu)/2$ and

$$v^A(hu) = \frac{\pi}{(1-\pi)e_u}\frac{z(hp,hu)-z(\ell p,\ell p)}{2}\frac{z(hp,hu)-z(hu,hu)}{2}.$$

This implies that

$$e_u^A = \frac{z(hu,hu)-z(\ell u,\ell u)}{4} + \frac{1}{2}\sqrt{\frac{(z(hu,hu)-z(\ell u,\ell u))^2}{4} + \frac{\pi}{1-\pi}(z(hp,hu)-z(\ell p,\ell p))(z(hp,hu)-z(hu,hu))}.$$

Using this the condition $(1-\pi)e_u > \pi e_p$ becomes

$$\pi \leq \frac{1}{2}\frac{z(hp,hu)-z(\ell u,\ell u)}{z(hp,hu)-[z(\ell u,\ell u)+z(\ell p,\ell p)]/2}.$$

Comparing surplus, $S^F < S^A$ if

$$\begin{aligned}
&\left(\frac{z(hp,hp)-z(\ell p,\ell p)}{z(hp,hu)-z(\ell p,\ell p)}\right)^2 \\
&< 1+\frac{z(hp,hu)-z(hu,hu)}{z(hp,hu)-z(\ell p,\ell p)}+\frac{1-\pi}{\pi}\left(\frac{z(hu,hu)-z(\ell u,\ell u)}{(z(hp,hu)-z(\ell p,\ell p))}\right)^2 \\
&\qquad \times\sqrt{\frac{1}{4}+\frac{\pi}{1-\pi}\frac{(z(hp,hu)-z(\ell p,\ell p))(z(hp,hu)-z(hu,hu))}{(z(hu,hu)-z(\ell u,\ell u))^2}}.
\end{aligned}$$

A sufficient condition is

$$\left(\frac{z(hp,hp)-z(\ell p,\ell p)}{z(hp,hu)-z(\ell p,\ell p)}\right)^2 < 1+\frac{z(hp,hu)-z(hu,hu)}{z(hp,hu)-z(\ell p,\ell p)},$$

which holds if $z(hp,hu)$ is sufficiently close to $z(hp,hp)$.

The second case arises when $(1-\pi)e_u < \pi e_p$, that is, when

$$\frac{1-\pi}{\pi} < \frac{z(hp,hu)-z(\ell p,\ell p)}{z(hp,hu)-z(\ell u,\ell u)}.$$

Then $v^A(hu) = z(hp,hu)/2$ and

$$\begin{aligned}
v^A(hp) =&\frac{z(hp,hp)-z(\ell p,\ell p)}{2} \\
&-\frac{1-\pi}{\pi e_p}\frac{(z(hp,hu)-z(\ell u,\ell u))(z(z(hp,hp)-z(hp,hu))}{4}.
\end{aligned}$$

This implies that

$$\begin{aligned}
e_p^A =&\frac{z(hp,hp)-z(\ell p,\ell p)}{4} \\
&+\frac{1}{2}\sqrt{\frac{(z(hp,hp)-z(\ell p,\ell p))^2}{4}-\frac{1-\pi}{\pi}(z(hp,hu)-z(\ell u,\ell u))(z(hp,hp)-z(hp,hu))}.
\end{aligned}$$

Comparing surplus, $S^F < S^A$ if

$$\begin{aligned}
&\left(\frac{z(hu,hu)-z(\ell u,\ell u)}{z(hp,hu)-z(\ell u,\ell u)}\right)^2 \\
&< 1-\frac{z(hp,hp)-z(hp,hu)}{z(hp,hu)-z(\ell u,\ell u)}+\frac{\pi}{1-\pi}\left(\frac{z(hp,hp)-z(\ell p,\ell p)}{(z(hp,hu)-z(\ell u,\ell u))}\right)^2 \\
&\qquad \times\sqrt{\frac{1}{4}+\frac{1-\pi}{\pi}\frac{(z(hp,hu)-z(\ell u,\ell u))(z(hp,hp)-z(hp,hu))}{(z(hp,hp)-z(\ell p,\ell p))^2}}.
\end{aligned}$$

Again a sufficient condition is

$$\left(\frac{z(hu,hu)-z(\ell u,\ell u)}{z(hp,hu)-z(\ell u,\ell u)}\right)^2 < 1-\frac{z(hp,hp)-z(hp,hu)}{z(hp,hu)-z(\ell u,\ell u)},$$

which holds if $z(hp, hu)$ is sufficiently close to $z(hp, hp)$. □

B Appendix: Proofs

B.1 Proofs for Section 3

Proof of Fact 1

For (i):

$$z(hp,hp)+z(\ell u,\ell u)=2(1+\alpha\beta)>2\delta=2z(hp,\ell u),$$

implying that any contract for $(hp, \ell u)$ can be competed away by (hp, hp) since $z(\ell u, \ell u) = 2\alpha\beta$. For (ii):

$$z(hp,hp)+z(\ell p,\ell p)=2(1+\alpha)>2=2z(hp,\ell p) \text{ and}$$
$$z(hu,hu)+z(\ell u,\ell u)=2\beta(1+\alpha)>2\beta=2z(hu,\ell u).$$

For (iii):

$$z(hp,hp)+z(hu,hu)=2+2\beta>4\delta=2z(hp,hu)$$

if and only if $\delta < (1+\beta)/2$.

For (iv):

$$z(hu,hu)+2z(\ell p,\ell u)=2\beta+2\alpha\delta<2\delta+2\alpha\beta=2z(hu,\ell p)+z(\ell u,\ell u)$$

implying that segregation for $hu, \ell p$ is unstable since $(hu, \ell p)$ matches can offer strictly higher payoffs to both attributes. Since $hu, \ell p$ will not segregate, let us compare now the benefit of hu integrating with ℓp versus integrating with hp. By facts (i), (ii) and (iii), it is sufficient to compare the matching pattern $\{(hp,hp),(hu,\ell p),(hu,\ell p),(\ell u,\ell u)\}$ to the matching pattern $\{(hp,hu),(hp,hu),(\ell p,\ell u),(\ell p,\ell u)\}$. Aggregate payoffs in the first pattern amount to $2+2\delta+2\alpha\beta$, and to $4\delta+4\alpha\delta$ in the second. Hence, aggregate payoffs in the second are greater if $\delta > (1+\alpha\beta)/(1+2\alpha)$. Note that this is implied by $\delta > 1 - \alpha$.

Proof of Fact 2

Depending on relative scarcity of hu, ℓp, and hp agents there are five cases.

Case (1): $\pi e_p > (1-\pi)e_u$ and $\pi(1-e_p) > (1-\pi)(1-e_u)$: Then some hp segregate and $v(hp) = 1$. hu match with hp and obtain $v(hu) = 2\delta - 1$. Likewise, some ℓp remain unmatched and obtain $v(\ell p) = \alpha$, whereas $v(\ell u) = (2\delta - 1)\alpha$. Hence, $e_p = 1 - \alpha$ and $e_u = (2\delta - 1)(1 - \alpha)$. The conditions become

$$\frac{\pi}{1-\pi} > \max\{2\delta - 1; (1 - (1-\alpha)(2\delta - 1))/\alpha\} = \frac{1 - (1-\alpha)(2\delta - 1)}{\alpha}.$$

Case (2): $\pi e_p > (1-\pi)e_u$ and $\pi(1-e_p) < (1-\pi)(1-e_u)$: Then $v(hp) = 1$ and $v(hu) = 2\delta - 1$ as above. But now $v(\ell u) = \alpha\beta$ and $v(\ell p) = \alpha(2\delta - \beta)$. Hence, $e_p = 1 - \alpha(2\delta - \beta)$ and $e_u = 2\delta - 1 - \alpha\beta$. The conditions become

$$\frac{2\delta - 1 - \alpha\beta}{1 - \alpha(2\delta - \beta)} < \frac{\pi}{1-\pi} < \frac{2 - 2\delta + \alpha\beta}{\alpha(2\delta - \beta)}.$$

Case (3): $\pi e_p < (1-\pi)e_u$ and $\pi > 1-\pi$. Then some ℓp segregate, so that $v(\ell p) = \alpha$. Therefore $v(hu) = \delta - \alpha$ and $v(hp) = \delta + \alpha$. $v(\ell u) = \alpha(2\delta - 1)$. Therefore $e_p = \delta$ and $e_u = (1 - 2\alpha)\delta$. The first condition then would imply $\pi/(1-\pi) < 1 - 2\alpha$, which is a contradiction to the second, $\pi/(1-\pi) > 1$.

Case (4): $\pi e_p < (1-\pi)e_u < \pi$ and $\pi < 1 - \pi$. Now some ℓu segregate, so that $v(\ell u) = \alpha\beta$. Therefore $v(\ell p) = \alpha(2\delta - \beta)$ and $v(hu) = \delta - \alpha(2\delta - \beta)$ and $v(hp) = \delta + \alpha(2\delta - \beta)$. This means that $e_p = \delta$ and $e_u = (1 - 2\alpha)\delta$. The conditions become

$$(1 - 2\alpha)\delta < \frac{\pi}{1-\pi} < 1 - 2\alpha.$$

Case (5): $\pi < (1-\pi)e_u$: Now some hu segregate, so that $v(hu) = \beta$ and $v(\ell u) = \alpha\beta$. $v(hp) = 2\delta - \beta$ and $v(\ell p) = \delta - \beta$, so that $e_p = \delta$ and $e_u = (1-\alpha)\beta$. The condition becomes

$$\frac{\pi}{1-\pi} < (1-\alpha)\beta.$$

Case (4): $\pi(1 - e_p) = (1-\pi)e_u < \pi$. Again hp segregate, so that $v(hp) = 1$. $v(\ell p) = \delta - v(hu)$ and $\beta < (2\delta - 1) \leq v(hu) \leq \delta$. $e_u = v(hu)$ and

$e_p = (1-\delta) + v(hu)$. That is:

$$v(hu) = \pi\delta \text{ and } v(\ell p) = (1-\pi)\delta.$$

This case obtains if:

$$0 \leq \frac{1-\pi}{\pi} \leq \frac{1-\delta}{2\delta-1}.$$

Case (5): $\pi(1-e_p) < (1-\pi)e_u = \pi$. Then $v(\ell p) = \delta - v(hu)$ and $v(hp) = 2\delta - v(hu)$, and $\beta \leq v(hu) \leq 2\delta - 1$. $e_p = \delta$ and $e_u = v(hu) = \pi/(1-\pi)$. For this case we need:

$$\frac{1}{\beta} \geq \frac{1-\pi}{\pi} \geq \frac{1}{2\delta-1}.$$

To summarize, $e_p = \delta$ if $\pi \leq \frac{2\delta-1}{\delta}$ and $e_p = 1$ if $\pi \geq 1$, and $\delta < e_p < 1$ otherwise. $e_u = \beta$ if $\pi \leq \frac{\beta}{1+\beta}$, $\beta < e_u < 2\delta - 1$ if $\frac{\beta}{1+\beta} < \pi < \frac{2\delta-1}{2\delta}$, $e_u = 2\delta - 1$ if $\frac{(2\delta-1)}{2\delta} \leq \pi \leq \frac{2\delta-1}{\delta}$, $2\delta - 1 < e_u < \delta$ if $\frac{2\delta-1}{\delta} < \pi < 1$, and $e_u = \delta$ if $\pi \geq 1$.

Proof of Lemma 1

To establish static surplus efficiency, suppose the contrary, i.e., a set of agents can be rematched to increase total payoff of all these agents. Then the increase in total payoff can be distributed among all agents required to rematch, which makes all agents required to re-match also strictly prefer their new matches, a contradiction to stability. Therefore matching is surplus efficient given investments.

The second part of the lemma requires some work. Let $\{ab\}$ denote a distribution of attributes in the economy, and $\mu(ab, a'b')$ the measure of $(ab, a'b')$ firms in a surplus efficient match given $\{ab\}$. Since $\mu(ab, a'b')$ only depends on aggregates πe_p, $\pi(1-e_p)$, $(1-\pi)e_u$, and $(1-\pi)(1-e_u)$ and investment cost is strictly convex, in an allocation maximizing total surplus all p agents invest the same level e_p, and all u agents invest e_u.

An investment profile (e_u, e_p) and the associated surplus efficient match $\mu(.)$ maximize total surplus ex ante if there is no (e'_u, e'_p) and an associated surplus efficient match $\mu(.)$ such that total surplus is higher.

Denote the change in total surplus Δ_b by increasing e_b to $e'_b = e'_b + \epsilon$. If there are positive measures of (hp, hp) and (hp, hu) firms, it is given by:

$$\Delta_p = \epsilon[z(hp, hu) - z(\ell p, hu)] - \epsilon e_p - \epsilon^2/2 \text{ and}$$
$$\Delta_u = \epsilon[z(hp, hu) - z(hp, hp)/2] - \epsilon e_u - \epsilon^2/2,$$

reflecting the gains from turning an ℓp agent matched to an hu agent into an hp agent matched to an hu agent, and from turning an ℓu agent matched to an ℓu agent into an hu agent matched to an hp agent, who used to be matched to an hp agent.

That is, assuming that indeed $\pi > (1-\pi)e_u > \pi(1-e_p)$ the optimal investments are given by $e_p = z(hp, hp)/2$ and $e_u = z(hp, hu) - z(hp, hp)/2$. Recall that TU wages are given in this case by $v(hp) = z(hp, hp)/2 = 1$ and $v(\ell p) = z(hu, \ell p) - v(hu)$, and $v(hu) = z(hp, hu) - z(hp, hp)/2 = 2\delta - 1$ and $y(\ell u) = 0$. Hence, TU investments are $e_p^T = z(hp, hu) - z(hu, \ell p)$ and $e_u^T = z(hp, hu) - z(hp, hp)/2$. That is, TU investments are optimal with respect to marginal deviations.

Checking for larger deviations suppose only e_u increases by ϵ, such that the measure of (hu, hu) firms becomes positive after the increase. The change in total surplus is now:

$$\Delta = \epsilon_1[z(hp, hu) - z(\ell p, hu)] + \epsilon_2[z(hu, hu)/2 - z(\ell u, \ell u)/2] - \epsilon e_p - \epsilon^2/2,$$

for $\epsilon_1 + \epsilon_2 = \epsilon$ such that the measure of (hp, hp) under e_u was $\epsilon_1/2$. Clearly, $\Delta < 0$ for $e_u = z(hp, hu) - z(\ell p, hu)$, since cost is convex and surplus has decreasing returns in an efficient matching. Suppose now that e_p decreases by ϵ large enough to have a positive measure of $(\ell p, \ell p)$ firms after the decrease (a decrease in e_u would have the same effect). The change in total surplus is:

$$\Delta = -\epsilon_1[z(hp, hu) - z(\ell p, hu)] - \epsilon_2[z(hp, hp)/2 - z(\ell p, \ell p)/2] + \epsilon e_p - \epsilon^2/2,$$

which is negative for $e_p = z(hp, hu) - z(hu, \ell p)$ since cost is convex and surplus has decreasing returns in an efficient matching. Finally, an increase of e_p will not affect the condition $\pi > (1-\pi)e_u > \pi(1-e_p)$.

A similar argument holds in all the five cases present in the proof of Fact 2.

B.2 Proofs for Section 4

Proof of Proposition 2

Recall that under free market investments are given by $e_p^F = 1$ and $e_u^F = \beta$, and expected payoffs of $v_p^F = 1$ and $v_u^F = \beta^2$.

Suppose first that $\pi < 1/2$. Then:

$$v^B(hp) = \frac{\delta}{2}\left(1 + e_u^B\right) \text{ and } y^B(\ell p) = \frac{\delta}{2}e_u^B.$$

Therefore $e_p^B = \delta/2 < e_p^F$. u agents obtain a p match with probability $\pi/(1-\pi)$, and otherwise the policy allows them to segregate in achievement. Hence:

$$v^B(hu) = \frac{\pi}{1-\pi}\frac{\delta}{2}\left(1 + e_p^B\right) + \frac{1-2\pi}{1-\pi}\beta \text{ and } v^B(\ell u) = \frac{\pi}{1-\pi}\frac{\delta}{2}e_p^B.$$

Then $e_u^B = \beta + \frac{\pi}{1-\pi}(\delta/2 - \beta) < e_u^F$ for $0 < \pi < 1/2$. Investment inequality under the B policy is lower if $e_p^F/e_v^F > e_p^B/e_u^B$, which must be true for $\pi < 1/2$ since $e_u^B \geq e_p^B$. Payoff can be written as $y_b = e_b^2 + w(\ell b)$, yielding:

$$v_p^B = \frac{\delta^2}{4} + \frac{\delta}{2}e_u^B \text{ and } v_p^B = (e_u^B)^2 + \frac{\pi}{1-\pi}\frac{\delta^2}{4}.$$

Indeed $v_p^B < v_u^B$ is possible for π small. Payoff inequality under the B policy is lower if $y_p^F/y_u^F > y_p^B/y_u^B$, that is, if:

$$(e_u^B)^2 - \beta^2\frac{\delta}{2}e_u^B + \frac{\pi}{1-\pi}\frac{\delta^2}{4} - \beta^2\frac{\delta^2}{4} > 0,$$

which can be shown (by computing the minimum of the LHS) to hold for $\pi < 1/2$. Aggregate payoff is:

$$v^B = \pi v_p^B + (1-\pi)y_u^B = \pi\frac{\delta^2}{2} + \pi\frac{\delta}{2}e_u^B + (1-\pi)(e_u^B)^2.$$

The fact that $e_u \leq \beta < \delta$ then immediately implies $v^F > y^B$.

If $\pi \geq 1/2$ on the other hand:

$$v^B(hu) = \frac{\delta}{2}\left(1 + e_p^B\right) \text{ and } v^B(\ell u) = \frac{\delta}{2}e_p^B.$$

Therefore $e_u^B = \delta/2 < e_u^F$. p agents obtain a p match with probability $(2\pi-1)/\pi$, in which case the policy allows them to segregate in achievement. Hence:

$$v^B(hp) = \frac{1-\pi}{\pi}\frac{\delta}{2}\left(1 + e_u^B\right) + \frac{2\pi-1}{\pi} \text{ and } v^B(\ell p) = \frac{1-\pi}{\pi}\frac{\delta}{2}e_u^B.$$

Therefore:

$$e_p^B = \frac{1-\pi}{\pi}\frac{\delta}{2} + \frac{2\pi - 1}{\pi} < e_p^F \text{ for } \pi < 1.$$

Investment inequality under the B policy is lower if $e_p^F/e_v^F > e_p^B/e_u^B$, that is, if:

$$\pi < \frac{\beta - \delta/2}{\beta - \delta/2 + \beta(1-\delta/2)} \in (1/2, 1).$$

Payoffs are given by:

$$v_p^B = (e_p^B)^2 + \frac{1-\pi}{\pi}\frac{\delta^2}{4} \text{ and } v_u^B = \frac{\delta^2}{4} + \frac{\delta}{2}e_p^B.$$

$v_p^F/v_u^F > v_p^B/v_u^B$ if:

$$(\delta - \beta^2 e_p^B)e_p^B + \frac{\delta^2}{4}\left(1 - \beta^2\frac{1-\pi}{\pi}\right) > 0,$$

which must be true, since $e_p^B \leq 1$ and $\delta > \beta$. Aggregate payoff is:

$$y^B = \pi(e_p^B)^2 + (1-\pi)\frac{\delta}{2}e_p^B + (1-\pi)\frac{\delta^2}{2}.$$

Using the expression for e_p^B above, $v^B < v^F$ for $\pi \geq 1/2$.

Proof of Proposition 3

Since low achievers match with low achievers $v^A(\ell p) = v^A(\ell u) = 0$. High achievers' payoffs depend on relative scarcity, however.

Case 1: $(1-\pi)e_u^A \geq \pi e_p^A$. Then hu agents outnumber hp agents and $v(hp) = \delta$. The expected payoff of an hu agent is:

$$Ev^A(hu) = \frac{\pi e_p}{(1-\pi)e_u}\delta + \left(1 - \frac{\pi e_p}{(1-\pi)e_u}\right)\beta.$$

Since $e_p^A = v^A(hp) - v^A(\ell p) = \delta$, with $e_p^B < \delta/2 < e_p^A < 1 = e_p^F$, and $v^A(\ell u) = 0$ this becomes a quadratic equation in e_u^A. It has a solution $\beta \leq e_u^A \leq \delta$ if $\pi \leq 1/2$, which is given by:

$$e_u^A = \frac{\beta}{2} + \frac{1}{2}\sqrt{\beta^2 + 4\frac{\pi}{1-\pi}\delta(\delta - \beta)} > e_u^F.$$

This implies that $e_p^A/e_u^A < e_p^F/e_u^F$, and since $y_b^A = (e_b^A)^2$ and $y_b^F = (e_b^F)^2$ also $y_p^A/y_u^A < v_p^F/v_u^F$. Comparing aggregate investment, $\pi e_p^A + (1-\pi)e_u^A >$

$\pi + (1-\pi)\beta$ if $\delta > (1+\beta)/2$, which holds by assumption. Aggregate payoff is greater under the A policy if:

$$\beta^2(3\delta^2 - 2\beta\delta - 1) > \frac{\pi}{1-\pi}(1 + \beta\delta - 2\delta).$$

There is $\pi^A > 0$ sufficiently small, such that the condition holds for $\pi < \pi^A$. $\pi^A = 1/2$ if $2\delta > 1 + \beta^2$.

Case 2: $(1-\pi)e_u^A < \pi e_p^A$. Then hp agents outnumber hu agents and $v^A(hu) = \delta$. The expected payoff of an hp agent is:

$$Ev^A(hu) = \frac{(1-\pi)e_u^A}{\pi e_p^A}\delta + \left(1 - \frac{(1-\pi)e_u^A}{\pi e_p^A}\right).$$

Since $e_u^A = v^A(hu) - v^A(\ell u) = \delta > \beta = e_u^F$, and $v^A(\ell p) = 0$ this becomes a quadratic equation in e_p^A. It has a solution $\delta \leq e_p^A \leq 1$ if $\pi \geq 1/2$, which is given by:

$$e_p^A = \frac{1}{2}\left(1 + \sqrt{1 - 4\frac{1-\pi}{\pi}\delta(1-\delta)}\right) < 1 = e_p^F.$$

Comparing this to the expression for $e_p^B = ((1-\pi)\delta/2 + 2\pi - 1)/\pi$ and noting that $\delta > 2/3$ yields $e_u^B < e_p^A$. This means that $e_p^A/e_u^A < e_p^F/e_u^F$, and as above also $y_p^A/y_u^A < v_p^F/v_u^F$. For aggregate investment, as above $\pi e_p^A + (1-\pi)e_u^A > \pi + (1-\pi)\beta$ if $\delta > (1+\beta)/2$. Comparing aggregate payoffs, $v^A > v^F$ if:

$$3\delta^2 - 2\delta - \beta^2 > \frac{1-\pi}{\pi}(2\delta^2 - \delta - \beta^2)^2.$$

Both sides are positive, the conditions slackens in π. There is $\pi^A < 1$ such that $\pi \geq \pi^A$ implies $v^A > v^F$. $\pi^A = 1/2$ if $2\delta^2 > 1 + \beta^2$.

Proof of Proposition 4

Recall that $e_p^F = 1$ and $e_v^F = \beta$. $S^A > S^F$ holds if, and only if:

$$\pi\frac{\delta^2 - 1}{2} + \frac{1-\pi}{2}\left(\frac{\beta^2}{4}\left(1 + \sqrt{1 + 4\frac{\pi}{1-\pi}\frac{\delta}{\beta}(\frac{\delta}{\beta} - 1)}\right)^2 - \beta^2\right) > 0 \text{ if } \pi \leq 1/2$$

$$\pi\frac{(1 + \sqrt{1 - 4\frac{1-\pi}{\pi}\delta(1-\delta)})^2 - 4}{8} + (1-\pi)\frac{\delta^2 - \beta^2}{2} > 0 \text{ if } \pi > 1/2.$$

For $\pi \leq 1/2$ the condition becomes:

$$\frac{\pi}{1-\pi}\left(1-2\delta^2+\delta\beta\right) < \frac{\beta^2}{2}\left(\sqrt{1+4\frac{\pi}{1-\pi}\frac{\delta}{\beta}\left(\frac{\delta}{\beta}-1\right)}-1\right). \tag{B.1}$$

This condition holds for all $0 < \pi < 1/2$ if $1 \leq (2\delta-\beta)\delta$. It is quickly verified that the condition holds with equality when $\delta = \sqrt{(1+\beta^2)/2}$. Otherwise, if $1 > (2\delta-\beta)\delta$, the condition can be rewritten as:

$$\frac{\pi}{1-\pi} < \beta^2\frac{3\delta^2-2\delta\beta-1}{(1-2\delta^2+\delta\beta)^2}.$$

The LHS strictly increases in π and the RHS strictly increases in δ if $1 > (2\delta-\beta)\delta$, as assumed. Therefore condition (B.1) defines $\delta^*(\pi,\beta)$ such that $S^A > S^*$ if and only if $\delta > \delta^*(.)$, and $\delta^*(.)$ increases in π. Inspecting (B.1) reveals that the condition is satisfied for all positive β and δ at $\pi = 0$. Moreover, $\delta^*(1/2,\beta) = \sqrt{(1+\beta^2)/2}$ as argued above.

For $\pi \geq 1/2$ the respective condition becomes:

$$\frac{1-\pi}{\pi}(\delta^2-\beta^2-\delta(1-\delta)) > \frac{1}{2}\left(1-\sqrt{1-4\frac{1-\pi}{\pi}\delta(1-\delta)}\right). \tag{B.2}$$

Note that $\delta \geq (1+\beta)/2$ implies that the LHS is strictly positive and the condition can be rewritten as:

$$\frac{1-\pi}{\pi} < \frac{\delta^2-\beta^2-2\delta(1-\delta)}{(\delta^2-\beta^2-\delta(1-\delta))^2}.$$

That is, $\delta > (1+\sqrt{1+3\beta^2})/3$ is a necessary (and at $\pi = 1$ also sufficient) condition for the above expression. The LHS strictly decreases in π and the RHS strictly decreases in β and strictly increases in δ for $\delta > 1/2$. Hence, condition (B.2) defines $\delta^*(\pi,\beta)$ such that $S^A > S^*$ if and only if $\delta > \delta^*(.)$, and $\delta^*(.)$ decreases in π and increases in β.

Turning to a B policy, total surplus is given by:

$$S^B = \pi\frac{(e_p^B)^2}{2} + (1-\pi)\frac{(e_u^B)^2}{2} + \begin{cases} \pi\frac{\delta}{2}(e_p^B+e_u^B) & \text{if } \pi \leq 1/2 \\ (1-\pi)\frac{\delta}{2}(e_p^B+e_u^B) & \text{if } \pi \geq 1/2 \end{cases}$$

Comparing this expression to total surplus under free market, $S^B > S^F$ if, and only if:

$$\begin{aligned}&\frac{3}{4}\delta^2 + 2\beta\delta - 2\beta^2 - 1 + \frac{\pi}{1-\pi}\left(\frac{3}{4}\delta^2 - 2\beta\delta + \beta^2\right) > 0 \text{ if } \pi \leq 1/2 \text{ and}\\ &\frac{3}{4}\delta^2 + 2\delta - \beta^2 - 2 + \frac{1-\pi}{\pi}\left(\frac{3}{4}\delta^2 - 2\delta + 1\right) > 0 \text{ if } \pi \geq 1/2.\end{aligned} \tag{B.3}$$

Because both conditions strictly increase in δ for $\pi \neq 1/2$ and coincide for $\pi = 1/2$, they define a unique threshold $\delta^B(\pi, \beta)$ such that $S^B > S^F$ if and only if $\delta > \delta^B(\pi, \beta)$.

Indeed δ^B increases in π for $\pi < 1/2$, since the first line in (B.3) strictly decreases in π as $(1+\beta)/2 < \delta < 2\beta$, and $\delta^B(0, \beta) = \frac{2}{3}\left(\sqrt{3 + 10\beta^2} - 2\beta\right) > 2/3$. The second condition increases in π for $\delta > 2/3$. Since $\delta^B(1/2, \beta) = \sqrt{\frac{2}{3}(1 + \beta^2)} > 2/3$, δ^B necessarily decreases in π and to reach $\delta^B(1, \beta) = \frac{2}{3}\left(\sqrt{10 + 3\beta^2} - 2\right) > 2/3$. Since the LHS of both conditions in (B.3) strictly decrease in β, $\delta^B(\pi, \beta)$ strictly increases in β.

Finally, $S^A > S^B$ holds if, and only if:

$$\pi\left((e_p^A)^2 - (e_p^B)^2\right) + (1-\pi)\left((e_u^A)^2 - (e_u^B)^2\right) + \min\{pi; 1-\pi\}(e_u^B + e_p^B).$$

For $\pi \geq 1/2$, $S^A > S^B$ if, and only if:

$$\begin{aligned}\frac{1-\pi}{\pi}&\left(\frac{\delta^2}{4} + 2(1-\delta) + \frac{1-\pi}{\pi}\left(2\delta - \frac{3}{4}\delta^2 - 1\right)\right)\\ &+\frac{1}{4}\left(1 + \sqrt{1 - 4\frac{1-\pi}{\pi}\delta(1-\delta)}\right)^2 > 1.\end{aligned}$$

This becomes:

$$\begin{aligned}&9\delta^2 - 16\delta + 8 >\\ &\quad\frac{1-\pi}{\pi}\left(4 - 8\delta - 3\delta^2 + \left(\frac{5}{2}\delta^2 - 6\delta + 4 + \frac{1-\pi}{\pi}\left(4\delta - \frac{3}{2}\delta^2 - 2\right)\right)^2\right),\end{aligned}$$

Computations verify that the condition holds at $\pi = 1/2$ for all $\delta \geq 2/3$, which is sufficient for the condition to hold for $\pi \geq 1/2$ for all $\delta \geq 2/3$ (since either the RHS decreases in π or the RHS is negative anyway).

For $\pi \geq 1/2$, $S^A > S^B$ if, and only if:

$$\frac{\pi}{1-\pi}\left(\frac{\delta^2}{4} - \beta\delta + 2\beta - \delta\right) + (e_u^A)^2 - \beta^2 > 0.$$

Using the expression for e_u^A this becomes:

$$\frac{\pi}{1-\pi}\left(\frac{\delta^2}{4} + (1-\delta)(2\beta - \delta)\right) + \frac{\beta}{2}\left(\sqrt{\beta^2 + 4\frac{\pi}{1-\pi}\delta(\delta-\beta)} - \beta\right) > 0,$$

which holds for $\pi \in (0, 1/2)$ and $\delta \in [2/3, 1]$ and $\beta \in [\delta/2, 2\delta - 1]$.

Second Best Policy

Given a policy $\rho(ab, ab')$ the payoffs of the different attributes are given by:

$$v(hp) = (2\rho(hp,hp) + \rho(hp,hu)\delta + \rho(hp,\ell p)/2 + \rho(hp,\ell u)\delta/2)/(\pi e_p),$$
$$v(\ell p) = (\rho(hp,\ell p)/2 + \rho(hu,\ell p)\delta/2)/(\pi(1-e_p)),$$
$$v(hu) = (2\rho(hu,hu) + \rho(hp,hu)\delta + \rho(hu,\ell p)\delta/2 + \rho(hu,\ell u)\beta/2)/((1-\pi)e_u),$$
$$y(\ell u) = (\rho(\ell u,hp)\delta/2 + \rho(\ell u,hu)\beta/2)/((1-\pi)(1-e_u)).$$

Total surplus can be written as:

$$\begin{aligned} S =& \pi e_p + (1-\pi)e_u\beta - \frac{\pi e_p^2 + (1-\pi)e_u^2}{2} \\ &+ \rho(hp,hu)(2\delta - \beta - 1) + \rho(hu,\ell p)(\delta - \beta) - \rho(hp,\ell u)(1-\delta). \end{aligned}$$

Since $e_b = y(hb) - y(\ell b)$ and given the constraints, the optimization problem reduces to choosing the measures of (hp, hu), $(hp, \ell p)$, $(hp, \ell u)$, $(hu, \ell p)$, and $(hu, \ell u)$ matches. Unfortunately, all these could be positive in optimum. It is possible to show that $(hp, \ell p)$ matches cannot occur in an optimal policy, when $e_p \geq 1/2$.[13] We solved the problem numerically and Figure 12 shows the second best optimal matching for the parametrization used to generate all the figures ($\delta = .85$, $\beta = .5$). The broken lines correspond to matching probabilities under an A policy for comparison. That is, an A policy is indeed

[13]Note that replacing (hb, hb) and $(\ell b, \ell b)$ matches by $(hb, \ell b)$ only has an effect on e_b, lowering surplus if $e_p > 1$ and $e_u > \beta$. Hence, if e_p decreases in $\rho(hp, \ell p)$ necessarily $\rho(hp, \ell p) = 0$ in optimum. e_b decreases in $\rho(hb, \ell b)$ if, and only if, $1 - 3e_b + 2e_b^2 - y(\ell b) \leq 0$, which holds if $e_b \geq 1/2$.

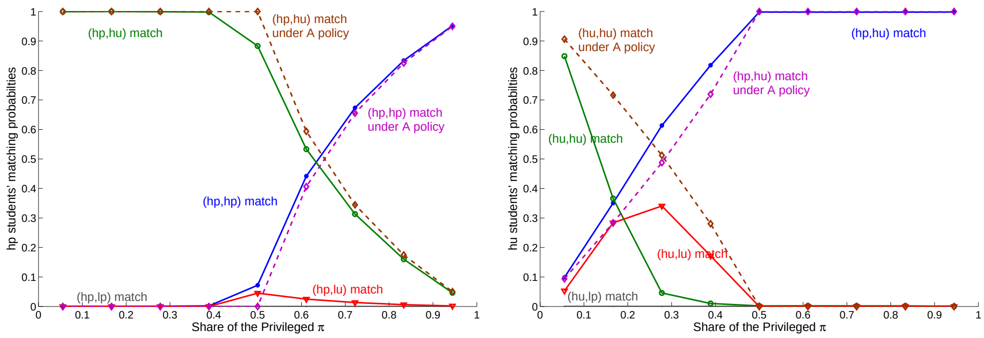

Figure 12: hp (left) and hu (right) students' matching probabilities in the second best.

very close to second best for this particular parametrization when $\pi \geq 1/2$.[14] Comparing surplus values to those under an A policy and free market yields the numbers in the text.

B.3 Proofs for Section 5

Proof of Lemma 4

In an equilibrium allocation attributes hp and ℓu segregate with tuition fees $t(ab, ab) = 0$. Since hp students cannot be adequately compensated by any other attribute and ℓp cannot adequately compensate any other attribute, no new college can make positive profit and attract either hp or ℓu students. hu and ℓp agents cannot both segregate (with zero fees), since a college offering $t(\ell p, hu) = \beta - \delta/2 + 2\epsilon$ and $t(hu, \ell p) = -\beta + \delta/2 - \epsilon$ would attract both ℓp and hu students and make strictly positive profit. It is easily verified that neither ℓp nor hu agents obtain more than their segregation payoff. This establishes the lemma. Note the possibility that hp match with ℓp giving both attributes their segregation payoffs, if ℓp agents outnumber hu agents, hp agents will obtain their segregation payoff in any equilibrium, however.

Proof of Proposition 5

We first derive the competitive equilibrium. Schools compete for students and earn zero profits, therefore $v(\ell u) = 0$, $v(hp) = 1$, and $t(\ell p, hu) = -t(hu, \ell p) \geq \beta - \delta/2$ is determined by the relative scarcity of attributes

[14]This result becomes more pronounced when δ is closer to 1. For low δ the second best may take the form of a naive policy, details are available from the authors.

hu and ℓp. Agents' investments are given by $e_u^C = \delta/2 + t(\ell p, hu)$ and $e_p^C = 1 - \delta/2 + t(\ell p, hu)$.

Suppose $\pi(1 - e_p^C) < (1 - \pi)e_u^C$ first. Then $t(\ell p, hu) = \beta - \delta/2$, $e_u^C = \beta$ and $e_p^C = 1 + \beta - \delta$. This regime occurs for $\pi < \frac{\beta}{\delta}$.

Second, suppose that $\pi(1-e_p^C) = (1-\pi)e_u^C$. This implies that $t(\ell p, hu) = (2\pi-1)\delta/2$, and $e_u^C = \pi\delta$ and $e_p^C = 1-(1-\pi)\delta$. This may hold for $\frac{\beta}{\delta} \leq \pi \leq 1$.

*

A Policy Under an A policy ℓu students segregate, so that $v^A(\ell u) = 0$. Payoffs of all other attributes will depend on relative scarcities.
(i) Suppose that $\pi < (1 - \pi)e_u^A$. Then $v^A(hp) = \delta$ and $v^A(\ell p) = \delta - \beta$, so that $e_p^A = \beta < 1 + \beta - \delta = e_p^C$, and:

$$v^A(hu) = \frac{\pi e_p^A}{(1-\pi)e_u^A}\delta + \left(1 - \frac{\pi e_p^A}{(1-\pi)e_u^A}\right)\beta.$$

Since $v^A(\ell u) = 0$ this implies:

$$e_u^A = \frac{\beta}{2} + \frac{1}{2}\sqrt{\beta^2 + 4\frac{\pi}{1-\pi}\beta(\delta - \beta)} > \beta = e_u^C.$$

This regime occurs for $\pi < \frac{\beta(1+\delta-\beta)}{1+\beta(1+\delta-\beta)}$. Since $v^A(\ell p) = y^C(\ell p) = \delta - \beta$, $e_p^C > e_p^A$ implies $y_p^C > y_p^A$, and similarly $y_u^C < y_u^A$. Therefore $e_p^C/e_u^C > e_p^A/e_u^A$ and $y_p^C/y_u^C > v_p^A/v_u^A$. Comparing aggregate investments yields an ambiguous result: $\pi e_p^A + (1 - \pi)e_u^A > \pi e_p^C + (1 - \pi)e_u^C$ for π, or for δ, sufficiently close to 1, and similarly for aggregate output y_b.
(ii) Let $\pi = (1 - \pi)e_u^A$. Then $v^A(hp) = \delta$ and $v^A(\ell p) = \delta/2 - t^A(\ell p, hu)$, $e_p^A = \delta/2 + t^A(\ell p, hu)$, and:

$$e_u^A = \frac{\pi}{1-\pi} = \frac{\pi e_p^A}{(1-\pi)e_u^A}\delta + \left(1 - \frac{\pi e_p^A}{(1-\pi)e_u^A}\right)(\delta/2 + t^A(\ell p, hu)).$$

Using $e_p^A = \delta/2 + t^A(\ell p, hu)$ and $e_u^A = \pi/(1 - \pi)$ and solving for e_p^A yields:

$$e_p^A = \frac{1+\delta}{2} - \frac{1}{2}\sqrt{(1+\delta)^2 - 4\frac{\pi}{1-\pi}}.$$

$t(\ell p, hu) \leq \delta/2$ then implies that this regime requires $\frac{\beta(1+\delta-\beta)}{1+\beta(1+\delta-\beta)} \leq \pi < \frac{\delta}{1+\delta}$.

This in turn implies that $e_u^A > e_u^C$ and therefore $y_u^A > y_u^C$. e_p^A strictly increases in π, and $e_p^C > e_p^A$ for $\pi/(1-\pi) = \beta(1+\delta-\beta)$, but $\delta > e_p^C$. This implies that $e_p^C > e_p^A$ for small π but the reverse holds for π large given this regime. Note that $e_p^A < e_u^A$ and therefore the investment gap reverses under the A policy. Also $v_p^A/v_u^A < y_p^C/y_u^C$. Since $v_p^A = (e_p^A)^2 + \delta - e_p^A$ in this regime, v_p^A has a minimum at either $e_p^A = \beta$ or $e_p = 1/2$. Hence, v^A strictly increases for $e_p^A > \max\{\beta; 1/2\}$. For $e_p^A = \delta$ corresponding to $\pi = \delta/(1+\delta)$, $v^A > v^C$.

(iii) Let now $\pi e_p^A \leq (1-\pi)e_u^A < \pi$. Then $v^A(hp) = \delta$ and $v^A(\ell p) = 0$, so that $e_p^A = \delta$. Similarly, $w^A(hu) = \delta$ as $t^A(\ell p, hu) = \delta$, and therefore $e_u^A = \delta$. This regime occurs for $\frac{\delta}{1+\delta} \leq \pi \leq 1/2$. Therefore $e_b^C < e_b^A$ for $b = u, p$. Since $e_u^A = e_p^A$ and $v_u^A = v_b^A$, there is perfect equality under an A policy. Comparing the privileged students' output, $y_p^A > y_p^C$ if and only if:

$$\delta > \frac{1+\beta+\beta^2}{1+2\beta} =: \delta^A.$$

Note that $(1+\beta)/2 < \delta^A < 1$. As for aggregate output, $v^A \geq v^C$ if and only if:

$$\delta^2 - \beta^2 \geq \pi(1 - \delta\beta - (1-\delta)(\delta-\beta)).$$

This condition gets tighter as β and π increase, and holds for $\beta = 2\delta - 1$ and $\pi = 1/2$.

(iv) Finally, suppose that $\pi e_p > (1-\pi)e_u$. Since all hu agents match with hp agents this regime coincides with the one analyzed above in the proof of Proposition 3, yielding $e_u = \delta > \pi\delta$, $v(\ell p) = 0$, and:

$$e_p = \frac{1}{2} + \frac{1}{2}\sqrt{1 - 4\frac{1-\pi}{\pi}\delta(1-\delta)} > 1 - (1-\pi)\delta.$$

This regime occurs whenever $\pi > 1/2$. Therefore $e_b^A > e_b^C$ for $b = u, p$. $e_p^A/e_u^A < e_p^C/e_u^C$ since $e_p^C - e_u^C = 1 - \delta$ and $e_p^A - e_u^A < 1 - \delta$ and $e_u^A > e_u^C$. Clearly, $v_u^A > v_u^C$, and as shown above $v^A > v^C$ for $1/2 \leq \pi \leq \beta/\delta$, and for $\pi > \beta/\delta$ $v^A > v^C$ is implied by $\pi e_p^A \geq \pi - \pi(1-\pi)\delta - (1-\pi)^2\delta^2$, which must hold. Again, $v_p^A > v_p^C$ can be ensured by δ sufficiently close to 1, for all $\pi \geq 1/2$.

References

Arcidiacono, P., Aucejo, E. M., Fang, H. and Spenner, K. I. (2011), 'Does affirmative action lead to mismatch? A new test and evidence', *Quantitative Economics* **2**(3), 303–333.

Becker, G. S. (1973), 'A theory of marriage: Part I', *Journal of Political Economy* **81**(4), 813–846.

Bénabou, R. (1993), 'Workings of a city: Location, education, and production', *Quarterly Journal of Economics* **108**(3), 619–652.

Bénabou, R. (1996), 'Equity and efficiency in human capital investment: The local connection', *Review of Economic Studies* **63**, 237–264.

Bidner, C. (2008), 'A spillover-based theory of credentialism', *mimeo University of New South Wales*.

Booth, A. and Coles, M. (2010), 'Education, matching, and the allocative value of romance', *Journal of the European Economic Association* **8**(4), 744–775.

Card, D. and Rothstein, J. (2007), 'Racial segregation and the black-white test score gap', *Journal of Public Economics* **91**(11-12), 2158–2184.

Clotfelter, C., Vigdor, J. and Ladd, H. (2006), 'Federal oversight, local control, and the spectre of "resegregation" in southern schools', *American Law and Economics Review* **8**(3), 347–389.

Coate, S. and Loury, G. C. (1993), 'Will affirmative-action policies eliminate negative stereotypes?', *American Economic Review* **5**, 1220–1240.

Cole, H. L., Mailath, G. J. and Postlewaite, A. (2001), 'Efficient non-contractible investments in large economies', *Journal of Economic Theory* **101**, 333–373.

de Bartolome, C. A. (1990), 'Equilibrium and inefficiency in a community model with peer group effects', *Journal of Political Economy* **98**(1), 110–133.

Dillon, E. W. and Smith, J. A. (2013), 'The determinants of mismatch between students and Colleges', *NBER Working Paper Series* Nr. 19286.

Durlauf, S. N. (1996*a*), 'Associational redistribution: A defense', *Politics & Society* **24**(2), 391–410.

Durlauf, S. N. (1996*b*), 'A theory of persistent income inequality', *Journal of Economic Growth* **1**(1), 75–93.

Epple, D. and Romano, R. E. (1998), 'Competition between private and public universities, vouchers, and peer-group effects', *American Economic Review* **88**(1), 33–62.

Estevan, F., Gall, T., Legros, P. and Newman, A. (2013), 'College Admission and High School Integration', mimeo.

Felli, L. and Roberts, K. (2002), 'Does competition solve the hold-up problem?', *CEPR Discussion Paper Series* Nr. 3535.

Fernández, R. and Galí, J. (1999), 'To each according to...? markets, tournaments, and the matching problem with borrowing constraints', *Review of Economic Studies* **66**(4), 799–824.

Fernández, R. and Rogerson, R. (2001), 'Sorting and long-run inequality', *Quarterly Journal of Economics* **116**(4), 1305–1341.

Fryer, R. G. and Loury, G. C. (2005), 'Affirmative action and its mythology', *Journal of Economic Perspectives* **19**(3), 147–162.

Fryer, R. G. and Loury, G. C. (2007), 'Valuing identity: The simple economics of affirmative action policies', *Working Paper Brown University*.

Fryer, R. G., Loury, G. C. and Yuret, T. (2008), 'An Economic Analysis of Color-Blind Affirmative Action', *Journal of Law, Economics, and Organization* **24**(2), 319–355.

Gall, T., Legros, P. and Newman, A. F. (2006), 'The timing of education', *Journal of the European Economic Association* **4**(2-3), 427–435.

Gall, T., Legros, P. and Newman, A. F. (2014), 'The Timing of Affirmative Action', mimeo.

Hanushek, E. A., Kain, J. F. and Rivkin, S. G. (2009), 'New evidence about brown v. board of education: The complex effects of university racial composition on achievement', *Journal of Labor Economics* **27**(3), 349–383.

Harsanyi, J. C. (1953), 'Cardinal utility in welfare economics and in the theory of risk-taking', *Journal of Political Economy* **61**(5), 434–435.

Heckman, J. J. (2008), 'Schools, skills, and synapses', *Economic Inquiry* **46**(3), 289–324.

Holmström, B. and Myerson, R. B. (1983), 'Efficient and durable decision rules with incomplete information', *Econometrica* **51**(6), 1799–1819.

Hong, L. and Page, S. E. (2001), 'Problem solving by heterogeneous agents', *Journal of Economic Theory* **97**, 123–163.

Hopkins, E. (2012), 'Job market signalling of relative position, or becker married to spence', *Journal of the European Economic Association* (forthcoming).

Hoppe, H., Moldovanu, B. and Sela, A. (2009), 'The theory of assortative matching based on costly signals', *Review of Economic Studies* **76**(1), 253–281.

Hoxby, C. M. and Avery, C. (2012), 'The missing "One-offs": The hidden supply of high-achieving, low-income students', *NBER Working Paper Series* Nr. 18586; also forthcoming, Brookings Papers on Economic Activity.

Hsieh, C., Hurst, E., Jones, C. I. and Klenow, P. J. (2013), 'The allocation of talent and U.S. economic growth', *NBER Working Paper Series* Nr. 18693.

Kaneko, M. and Wooders, M. H. (1986), 'The core of a game with a continuum of players and finite coalitions: the model and some results', *Mathematical Social Sciences* **12**, 105–137.

Lang, K. and Lehmann, J. (2011), 'Racial discrimination in the labor market: Theory and empirics', *Journal of Economic Literature*, **50**(4), 9591006.

Legros, P. and Newman, A. F. (2007), 'Beauty is a beast, frog is a prince: Assortative matching with nontransferabilities', *Econometrica* **75**(4), 1073–1102.

Lutz, B. (2011), 'The end of court-ordered desegregation', *American Economic Journal: Economic Policy* **3**(2), 130–168.

Orfield, G. and Eaton, S. E. (1996), *Dismantling Desegregation: The Quiet Reversal of Brown v. Board of Education*, The New Press, New York, NY.

Peters, M. and Siow, A. (2002), 'Competing pre-marital investments', *Journal of Political Economy* **110**, 592–608.

Roth, A. E. and Sotomayor, M. (1990), *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*, Econometric Society Monographs No.18, Cambridge University Press, Cambridge.

Sander, R. (2004), 'A systemic analysis of affirmative action in American law universities', *Stanford Law Review* pp. 367–483.

Shimer, R. and Smith, L. (2000), 'Assortative matching and search', *Econometrica* **68**(2), 343–369.

Smith, L. (2006), 'The marriage model with search frictions', *Journal of Political Economy* **114**(6), 1124–1144.

Weinstein, J. (2011), 'The impact of university racial compositions on neighborhood racial compositions: Evidence from university redistricting', *mimeo*.