HCEO
hceconomics.org

Human Capital and Economic Opportunity
Global Working Group

Working Paper Series

Working Paper No. 2015-001

College Diversity and Investment Incentives

Thomas Gall
Patrick Legros
Andrew F. Newman

January, 2015

Human Capital and Economic Opportunity Global Working Group
Economics Research Center
University of Chicago
1126 E. 59th Street
Chicago IL 60637
www.hceconomics.org

College Diversity and Investment Incentives*

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

This version, December 2014

Abstract

We study the aggregate economic effects of diversity policies such as affirmative action in college admission. If agents are constrained in the side payments they can make, the free market allocation displays excessive segregation relative to the first-best. Affirmative action policies can restore diversity within colleges but also affect incentives to invest in pre-college scholastic achievement. Affirmative action policies that are achievement-based 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.

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

1 Introduction

Student diversity in higher education is both a goal embraced by college administrators and policy makers, and a subject of much controversy. Few would doubt that the college experience is heavily influenced by the nature of

*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 Steve Durlauf, Glenn Loury, John Moore, Andy Postlewaite, and seminar participants at Amsterdam, Boston University, Brown, Budapest, UC Davis, Essex, Frankfurt, the Measuring and Interpreting Inequality Working Group at the University of Chicago, Ottawa, Penn, ThReD 2009, and Yale. Gall thanks DFG for financial support (grant GA-1499), Legros thanks the Communauté Française de Belgique for financial support of(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

one's peers, both in terms of what one learns from them and what networks one forms with them. A student who graduates having encountered peers from a broad range of backgrounds will likely have different life opportunities and earning potential than one who has not. Yet there is still little consensus on how to design diversity policies to make them effective, or even on whether they are desirable. This paper contributes to the debate 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 direct 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. For a number of reasons, including but not limited to moral hazard, social norms, or regulations, students seldom engage in significant transfer payments among themselves. Because college benefits are lifetime earnings, even if a price system could be put in place, individuals might not have the financial ability to make transfers based on lifetime earnings. In such a world, policies of tax-subsidies might be ineffective or imperfect instruments for achieving the desired goals set by a planner or the college officials. Any discussion of diversity policies for college education ought to take account of the implications of and limitations imposed by these "non-transferabilities" (NTU).

Modeling the college marketplace with NTU leads to novel positive and normative insights, and as such complements other analyses of diversity policy based on imperfections such as search frictions or statistical discrimination. Indeed, the theoretical literature on matching has illustrated that the composition of groups may be significantly affected 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, thus 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 "rematch" 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 trade-off: diversity may be desirable from social or political objectives (equity, diversity or righting past wrongs) but it comes at an economic cost.

As we show, this trade-off may be misconstrued. Indeed, because of non-transferabilities, the free market policy generates too much segregation, which generally implies that free market investment incentives are also distorted. Though rematch policies cannot directly address the market imperfections, they may provide an instrument for correcting the 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. Affirmative action may be beneficial both for equity and efficincy.

Our analysis of various forms of rematch builds on the following environment. Colleges are arenas for the acquisition of human capital. To make our points starkly, we assumpe that the process is driven entirely by peer effects.[1]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 a diversity of backgrounds within colleges.

Our emphasis is on contrasting free market outcomes, represented by stable matches, with ones constrained by policy. This modeling strategy frees the analysis from the confounding effects of informational constraints or search frictions. Under NTU, the free market equilibrium is characterized by full segregation in achievement and background, implying that incentives to invest are distorted with respect to a "first best" situation, which could be achieved if every agent had unlimited amounts of wealth with which to make side payments. In general, returns to college for the underprivileged will be low, giving them minimal incentives to invest. The privileged could also have lower incentives to invest than in the first-best situation but could

[1]It would be straightforward to extend our analysis to the case in which colleges vary in the inherent quality of their faculty or facilities, with little change in results.

also face the reverse situation, with returns and high investment creating high investment incentives, in which case, the free market situation might be characterized by *over-investment at the top and under-investment at the bottom* (OTUB),

Similar to actual practice, 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 focus 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 re-matching policy.

Finally, we show that our results are robust when privileged agents have sufficient wealth to make transfers into the college marketplace, but underprivileged have limited ability to pay. Naturally, the free market outcome changes; instead of global segregation, privileged *low* achievers match with underprivileged high achievers. This still fails to be welfare maximizing, and affirmative action policies help improve aggregate performance. What is also new in the case of limited transfers is the possibility that under affirmative action, there are incentives for privileged high achievers to pay underprivileged high achievers *not* to exercise the option afforded them by the policy. If these incentives are effectuated, affirmative action could appear ineffective, because the matching pattern would be that of the free market. In fact, however, the policy still effectively redistributes surplus contingently on achievement to the underprivileged, and therefore generates similar investment incentives and aggregate effects on income and welfare. The analysis also shows shows that scholarships only targeted to the underprivileged 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,[2] there has been little work 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. Peters and Siow (2002) find that allocations

[2] 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 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.

are constrained Pareto optimal (with the production technology they study, aggregate surplus is also maximized), and do not discuss policy. The result of Peters and Siow (2002) has been recently challenged by Bhaskar and Hopkins (2014) who show that, except for special cases, investments are not first best in a model of where individuals on both sides of the market invest and the surplus is not perfectly transferable. We obtain a similar result in our model but our focus is on the static (matching) and dynamic (investment) effects of affirmative action policies play in environments with non-transferabilities. Booth and Coles (2010) compare different marriage institutions in terms of their impact on matching and investments. Gall et al. (2006) analyze 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 (see Lang and Lehman, 2011, for a survey of the theoretical and empirical literature). Coate and Loury (1993) provide a formalization of the argument that equilibria, when underinvestment 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 consistent with two types of equilibria. In the "bad" affirmative action equilibrium, although employment of the underprivileged may increase, beliefs do not change, leaving investment incentives and wages unchanged or reduced. But in the "good" equilibrium, as in our (unique) 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 a while 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.[3] In our NTU framework, affirmative action is not persistent, but the investment effects function as in 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 enrollment 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.

The paper proceeds as follows. Section 2 lays out the model framework. In Section 3 we show that segregation obtains when agents have no wealth and how it leads to distorted investment incentives with respect to the ideal situation where agents have large initial wealth. This opens the door for rematching policies to be surplus and welfare enhancing and we show that this is the case in Section 4. In fact, 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 since the underprivileged are wealth constrained and have difficulties borrowing; we comfort the benefits of using affirmative action policies in this 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). Students may also have a wealth endowment ω_i. In the NTU case, ω_i is "small" for all agents, implying that transfers are insufficient to change the matching outcome obtained when $\omega_i = 0$. We will also consider the idealized first best case where ω_i is "large" for all agents, as well as the

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

case where only privileged agents have wealth sufficient for making transfers.

Individual background is given exogenously, while 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 their wealth. They match into colleges, which we model as pairs of students, as a function of these attributes. 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.

We assume that:

$$f(h, h) = 1,\ f(h, \ell) = f(\ell, h) = 1/2,\ f(\ell, \ell) = \alpha,$$
$$g(p, p) = 1,\ g(p, u) = g(u, p) = \delta,\ g(u, u) = \beta,$$

with

$$\alpha \geq 0,\ \delta < 1, \beta \in [\delta/2, \delta]. \tag{1}$$

As α is non-negative, $f(\cdot, \cdot)$ has increasing differences, consistent with usual complementarity assumption for production functions. By contrast, the network effects function $g(\cdot,\cdot)$ has strictly decreasing differences on the domain $\{u,p\}$ (that is, $g(u,p)-g(u,u)>g(p,p)-g(p,u)$) whenever $\delta-\beta>1-\delta$, 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. There are many reasons to suspect that diversity in backgrounds is indeed desirable. 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 privileged since there may be replication rather than complementarity of information. 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 B 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 choose a non-contractible investment e. 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 colleges 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 colleges; attributes ab are determined by individuals' optimal choice of education acquisition e under rational expectations.

For a measurable set of agents S, an allocation consists of a partition of this set into colleges as well as transfer among these individuals consistent with their initial wealth endowment. That is if w_i is the wealth of individual i, each agent can obtain a transfer t_i, where $t(a_ib_i) \geq -w_i$ and $\int_{i\in S} t(a_ib_i) \leq \int_{i\in S} w_i$.

A *college choice equilibrium* is defined as a measure preserving matching function between individuals such that the following conditions are satisfied.

- (Payoff Feasibility) Within a pair (i, j), the payoffs are respectively $y(a_ib_i, a_jb_j) + t(a_ib_i)$ and $y(a_jb_j, a_ib_i) + t(a_jb_j)$, where $t(a_ib_i) \geq -w_i$ and $\int_i t(a_ib_i)di \leq \int_i w_idi$.

- (Finite Stability) There does not exist a match and feasible transfers among a finite set of individuals that will make all the individuals strictly better off with respect to the equilibrium payoffs.

If there is no possibility of transfer among agents, either within a pair or across pairs, and the equilibrium condition reduces to the usual stability condition that a pair cannot destabilize the equilibrium. In general if there are no other constraints on matching, the stability condition reduces indeed to deviations of a pair, ignoring the possibility of transfers across pairs. Things will be different when we consider affirmative action policies since some agents cannot prevent other agents from joining them in a group, making transfers across pairs potentially useful for improving payoffs. Existence of such an equilibrium is standard, see, e.g., Kaneko and Wooders (1986, 1989), but our proofs will be constructive.

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. Without loss of generality, 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 with Non-Transferabilities and Investment Distortions

Before discussing the positive and normative effects of re-matching policies, it is useful to contrast the matching pattern and the investment levels obtained in the free market situation where agents have no wealth (or "little" wealth as we will see) 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.

3.1 Free Market with Non-Transferabilities

In such an environment where transfers are not possible, a student in a college $(ab, a'b')$ obtains payoff $y = f(a, a')g(b, b')$; 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. The

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 $(ab, a'b')$

free market equilibrium allocation without side payments has full segregation in attributes. Indeed, hp cannot obtain more than 1 in any match and will segregate; since $\beta > \delta/2$, hu will also segregate since they cannot attract hp in a match. Now, because $\delta < 1$, ℓp segregate. 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:[4]

$$v^0(hp) = 1, v^0(\ell p) = \alpha, v^0(hu) = \beta, v^0(\ell u) = \alpha\beta.$$

Therefore an agent of background b chooses e_b to maximize $e_b v^0(hb) + (1 - e_b)v^0(\ell b) - \frac{e_b^2}{2}$ implying that $e_b = v^0(hb) - v^0(\ell b)$, and therefore the equilibrium investment levels are:

$$e_p^0 = 1 - \alpha \text{ and } e_u^0 = \beta(1 - \alpha). \tag{2}$$

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 $y_b \equiv e_b v^0(hb) + (1 - e_b)v^0(\ell b)$, 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.

[4] Note that this outcome will be the case whenever the underprivileged have wealth $\omega_u < \alpha(1 - \delta)$ and the privileged have wealth $\omega_p < \beta - \delta/2$. If a college (hu, hp) forms, hp obtains at most payoff $\delta + \omega_u < 1$, and if a college $(hu, \ell p)$ forms, hu obtains at most $\delta/2 + \omega_p < \beta$; $(\ell u, \ell p)$ cannot form either as the maximum payoff to ℓp would be $\alpha\delta + \omega_u < \alpha$.

3.2 First-Best with 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 college. 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.[5] 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.

(i) $(hp, \ell u)$ matches cannot occur in a first best allocation. Indeed, in an $(hp, \ell u)$ college hp agents lose more compared to their segregation payoff than ℓu students gain: the average surplus in matches (hp, hp) is 1, and $\alpha\beta$ in $(\ell u, \ell u)$ matches. The total surplus in a match $(hp, \ell u)$ is $\delta/2 < 1/2$, which is less than what hp students obtain in segregation.

(ii) If an equilibrium match has agents of the same background, they also have the same achievement. That is, matches $(hp, \ell p)$ or $(hu, \ell u)$ cannot occur. This follows from increasing returns of $f(a, a')$.

(iii) If agents of a given achievement match together, surplus is higher if

[5] 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.

backgrounds are diverse. Indeed, note that condition (DD), is equivalent to $2z(hp, hu) > z(hp, hp) + z(hu, hu)$, implying that segregation in background is surplus inefficient.

(iv) If $\alpha > \delta - \beta$, $(hu, \ell p)$ matches are not stable, since the sum of segregation payoffs, $1 + \alpha$, is greater than the total surplus in an $(hu, \ell p)$ match, δ.

(v) If $\alpha < 1 - \delta$, then surplus is higher when matching $(hu, \ell p)$ and segregating hp, than matching (hp, hu) and segregating ℓp: in the former case, total surplus is $2\delta + 2$, compared to $4\delta + 2\alpha$ in the latter case. Hence, in any equilibrium, all $(hu, \ell p)$ matches will be exhausted and matches (hu, hp) will form only if there is an excess supply of hu agents.

The policy discussion will be the most relevant when (hu, hp) are the most desirable but do not arise in the free market. At the same time we would like to allow for $(hu, \ell p)$ matches. For these reasons, we will restrict attention in the following to the set of α satisfying the following condition:

$$1 - \delta < \alpha < \delta - \beta. \tag{3}$$

Lemma 1. *Under* (3)*, a first best allocation exhausts all possible* (hp, hu) *matches, then all* $(hu, \ell p)$ *matches, and then all* $(\ell p, \ell u)$ *matches, while all other remaining attributes segregate.*

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.

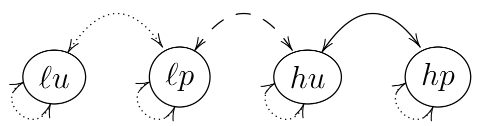

Figure 1: TU equilibrium matchings

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.

Lemma 2. *Suppose (DD) holds. Under full TU, investment levels e_u^* and e_p^* are non-monotonic in π and vary in opposite directions; e_p^* being U shaped and e_u^* being an inverted U shape.*

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 3. *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 Distortions in Investment

With a price system and unconstrained transfers among agents, college returns reflect scarcity: scarce agents in the matching market can claim a high share of the total college return. For this reason the scarcity of privileged as measure by π will affect the returns from college and therefore the incentives to invest in education. By contrast, when there is no possibility of transfer because of wealth constraints, the college returns will not reflect scarcity: there will be segregation and therefore the college return of an attribute is independent of the distribution of attributes, hence of π. This explains why privileged agents may have lower or higher incentives to invest in the NTU case than in the ideal first-best situation. And indeed, comparing the equilibrium investments e_b^0 under non-transferability to the first-best investment levels e_b^* given in Lemma 2, there is an interval of π for which *privileged agents will over-invest and the underprivileged under-invest* with respect to the first-best. This "over-investment at the top, under-investment at the bottom" (OTUB) outcome starkly illustrates the possible investment distortions that can be brought about by non-transferabilities.

A more precise characterization of the investment outcomes is offered in the following proposition that is illustrated in Figure 2.[6]

Proposition 1. *The underprivileged never over-invest and under-invest if $\pi > \frac{(1-\alpha)\beta}{1+(1-\alpha)\beta}$. The privileged over-invest for $\frac{2\alpha+(1-\alpha)(2\delta-1)}{2\alpha+(1-\alpha)2\delta} < \pi < \frac{1-(1-\alpha)(2\delta-1)}{2-(1-\alpha)(2\delta)}$,*

[6]In this figure as well as others in the paper we use the parametrization $\delta = 0.85$, $\beta = 0.5$, $\alpha = 0.18$.

in which case there is both over-investment at the top and under-investment at the bottom of the background distribution.

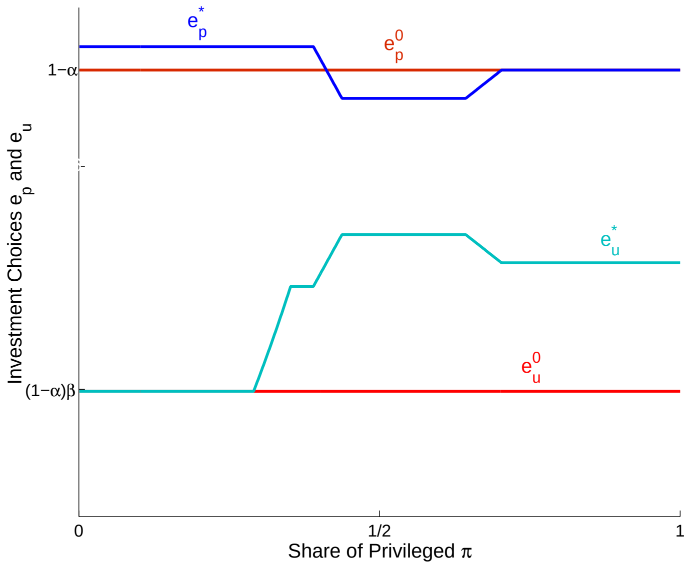

Figure 2: Education investments: NTU vs TU

This result formalizes the idea that an imperfect price system not only can generate excessive segregation, a static inefficiency, but also generates investment distortions, a dynamic inefficiency. Specifically, there will tend to be insufficient investment by the under-privileged; as for the privileged, their investment will be insufficient or excessive depending on whether they are a small enough minority. As we shall see, this suggests that the possible discouragement effects on the privileged that diversity policies introduce can be desirable.

Excessive segregation also has implications for inequality and polarization, but not necessarily in the "obvious" way. Indeed, computing background gaps as a measure of inequality yields the following corollary.

Corollary 1. *For intermediate and high π, inequality in investments e and in payoffs y is higher under NTU 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

Real world policies aim at replicating population measures of backgrounds in colleges, but vary in the degree to which they allow colleges to condition admission on achievement. We will focus on two extreme policies. First, we will consider an "achievement-blind" policy, which re-matches students by background without regard to achievements: each college's expected background composition equals that of the population. Second, we study "affirmative action," which gives priority to the under-privileged 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 element of the attributes, an achievement-blind policy tends often to perform 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 Policy

Several real-world policies are essentially achievement blind. Post 1968, public European universities often did not condition admission on achievement beyond the basic requirement of finishing high school; formally, this is akin to an assignment rule that randomly integrates colleges in background, ignoring achievement. In the U.S., this type of policy has been mainly restricted to primary and secondary education. Possibly the most prominent example is the use of "busing" to achieve high school integration, which operated mainly by redesigning school districts to reflect aggregate population measures. Other examples are the integration of school catchment areas in Brighton and Hove, U.K.; reservation in India to improve representation of scheduled 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).

Definition 1. An *achievement Blind policy* (denoted B policy) exhausts all possible matches of underprivileged and privileged backgrounds, using uniform rationing conditioned 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 3 below. Ellipses indicate matches subject to availability of agents after exhausting matches denoted by solid arrows.

Lemma 4. *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 $(1-e_p)\max\{\pi/(1-\pi); 1\}$. 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.*

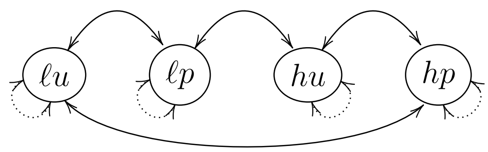

Figure 3: Equilibrium matching under a B policy.

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 be depressed compared to the free market in general.[7] 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 4 to verify this intuition; details are in the appendix:

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

Proposition 2. *Investments under a B policy are lower than in the free market outcome for both backgrounds if $\beta > (2-\sqrt{2})\delta$, as are aggregate investment and payoffs. This 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.*

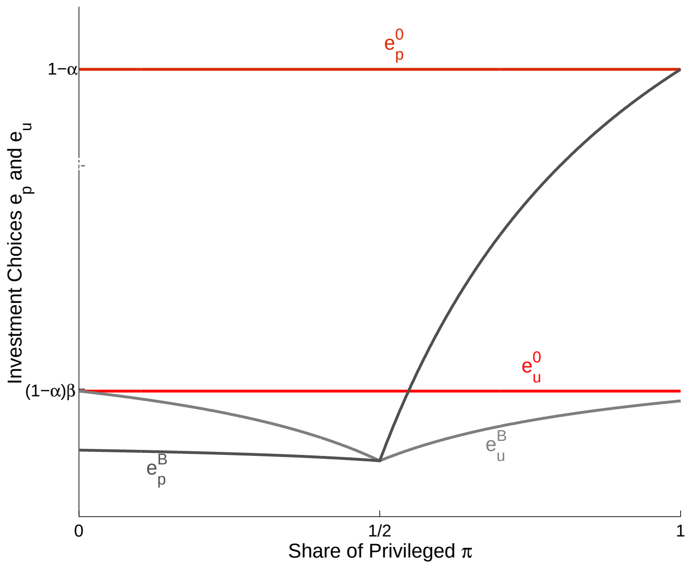

Figure 4: Education investments using a B policy.

That is, a B policy is indeed subject to the classic equity-efficiency tradeoff 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.

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 in segregated universities. It is widely used (for instance, the reservation of places for highly qualified minority students at some *grandes écoles* in France, like Sciences Po Paris, the "positive equality bill" and *Gleichstellung* in the public sectors in the U.K. and Germany).

Definition 2. Consider an equilibrium and a match $(ap, a'p)$. An *affirmative action policy* (denoted A policy) requires that an agent with attribute au must not strictly prefer joining $a'p$ 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 5. *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. □

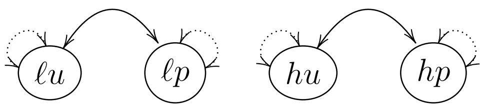

Figure 5: Equilibrium matching under an A policy.

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 ($e_u^A > e_u^0 > e_u^B$), and the privileged less ($e_p^0 > e_p^A > e_p^B$). 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.*

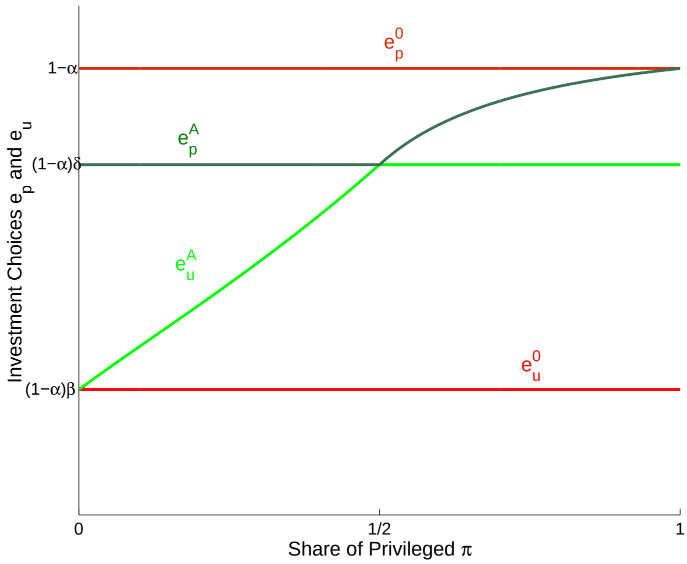

Figure 6: Education investments under an A policy.

Not only does an A policy crowd out privileged investment by less than a B policy, but also underprivileged investment is boosted compared to the 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.

4.3 Aggregate Effects

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 been 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.[8]

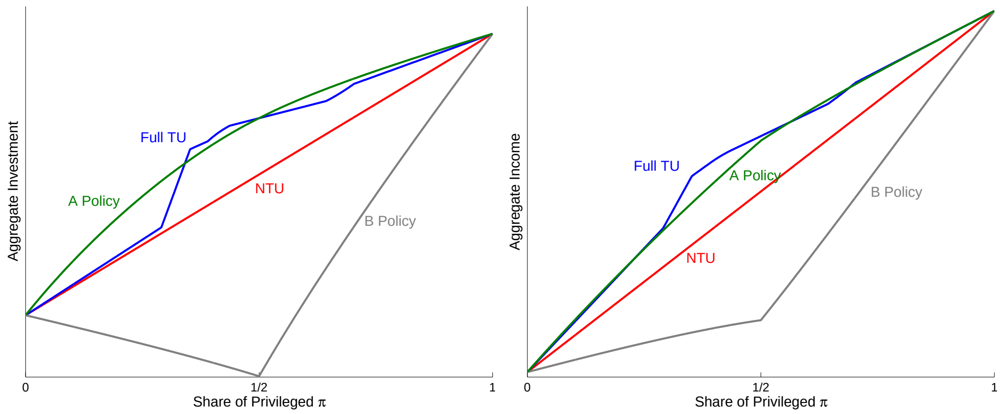

Figure 7: Aggregate investments (left) and aggregate payoff (right).

Both types of policy tend to decrease inequality in the economy compared to a 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, 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.

Our results may be summarized to suggest that policies that ignore achievement, focusing only on background, are likely to be far less effective in improving various aggregate outcome measures, and some of them will do more harm than good. Properly designed achievement based policies,

[8]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.

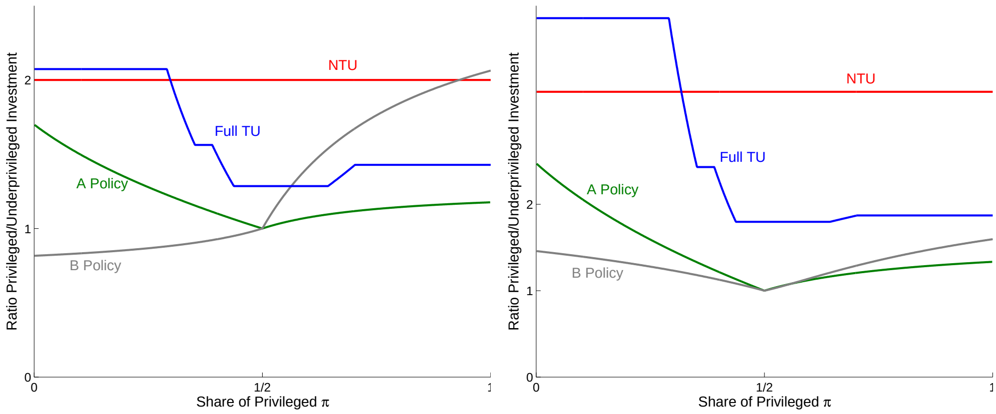

Figure 8: Inequality of investments (left) and payoffs (right).

for instance in the form of scoring rules that assign high weight to high attainments, are preferable to those that simply mix in terms of backgrounds, and can be quite effective in imporoving both aggregate efficiency and equity.

The same conclusions apply if we focus not on outcomes such as output, inequality and investment, but on terms of welfare, measured in aggregate surplus, that is, expected payoff net of investment cost. See Figure 9.

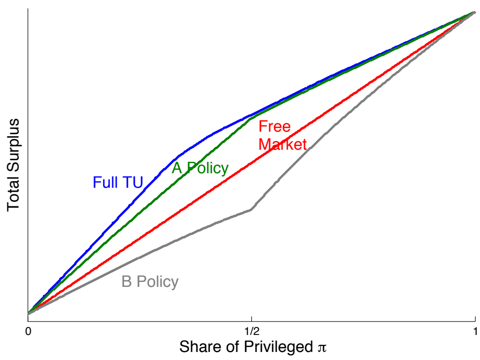

Figure 9: Total Surplus

In this figure the A policy clearly dominates the free market under NTU and the B policy. The dominance of A over B in terms of welfare is a general property, but that of A with respect to NTU requires that δ be large enough (as in the figure where $\delta = 0.85$).

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

(ii) For each $\pi \in (0,1)$, there is $\hat{\delta}(\pi) < 1$ such that an A policy induces strictly higher total surplus than the free market with NTU if $\delta > \hat{\delta}(\pi)$.

4.4 Second-Best Surplus Maximizing Policy

While figure 9 suggests that the A policy is in fact close to the surplus maximizing policy for high values of δ, it may be of independent interest to compute the second-best optimal policy, that is when a planner has full control over the way agents will match, hence 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 4.

The problem above has six 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 our parameters. See Appendix A for details. In fact the A policy realizes more than 97% of the gains in surplus that the second best policy achieves (for $\delta = .85$, $\beta = .5$, and $\alpha = .18$, used for all figures).

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 university of Michigan until 2003) or on whether a student attended a public high school (used in college admission in Brazil), or comes from a disadvantaged neighborhood.

Another policy is one that would replicate the first best matching, that is as in Figure 1. This "naive" policy faces a similar trade-off as the B policy: while maximizing the static gains from re-matching ex post, it falls short of optimizing the incentives, because the *payoffs*, which are still constrained by NTU, cannot replicate the TU outcome.[9] Instead it may be better for the planner to approximate the TU investment incentives by generating convex combinations of NTU payoffs that differ from those that would be accomplished by the naive policy – the second best policy takes full advantage of this possibility. Indeed, an A policy can sometimes outperform the naive policy, and the naive policy is the A policy when $\pi > 1/2$, see Appendix.

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 to public colleges, at all levels of education there are private colleges that charge tuition fees that may reflect 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 lifetime 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 with higher wealth. As

[9]Calling this policy "naive" is a bit of misnomer, as it has a serious practical drawback: it would require considerable sophistication on the part of the policy maker to compute the (counterfactual) TU outcome!

mentioned in footnote 4, for $\omega_u < \alpha(1-\delta)$ and $\omega_p < \beta - \delta/2$ our previous analysis goes through unchanged, because hu students cannot compensate hp students enough to depart from the segregated outcome; neither can ℓp's compensate hu's, nor can ℓu's attract ℓp's. Suppose for simplicity that

$$\omega_p > \delta/2, \text{ and } \omega_u = 0. \tag{4}$$

This implies that the privileged can compensate the underprivileged, but not vice versa; Figure 10 shows the resulting possible payoffs for some attribute combinations.

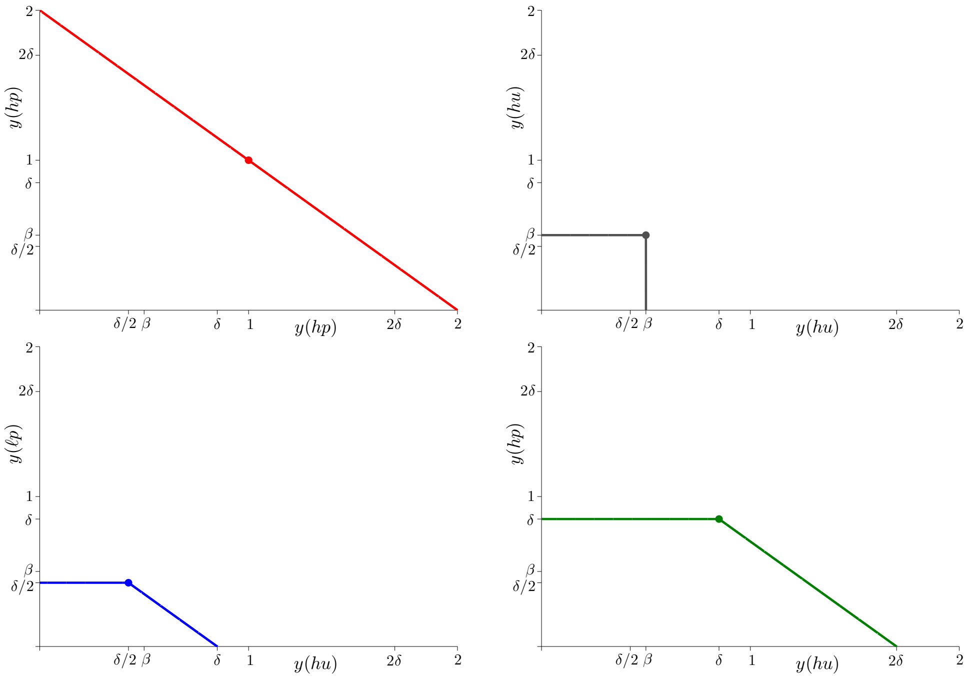

Figure 10: Possible distribution of payoffs in (hp, hp) and (hu, hu) colleges (top) and $(\ell p, hu)$ and (hp, hu) colleges (bottom) when individuals can make lump-sum transfers but the underprivileged face borrowing constraints.

The next statement follows directly from this observation.

Lemma 6. *Under* (4)*, a free market equilibrium exhausts all possible* $(hu, \ell p)$ *matches,* ℓu *and* hp *agents segregate.*

Figure 11 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.[10]

[10]For instance, Dillon and Smith (2013) find evidence for substantial mismatch in the

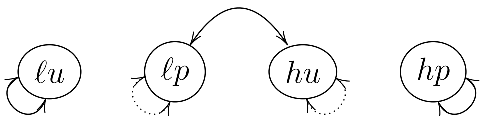

Figure 11: Equilibrium matching with transfers

Note that with (4), an A policy yields the first-best matching pattern, since hu students have priority over hp students in (hp, hp) matches, and ℓu students have priority over ℓp students in $(\ell p, \ell p)$ matches but not in $(hu, \ell p)$ matches, which are possible if ℓp students offer a side payment to hu.

Lemma 7. *Under* (4)*, the matching equilibrium under an A policy is ex post efficient, exhausting first* (hp, hu)*, then* $(hu, \ell p)$*, then* $(\ell p, \ell u)$*.*

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 ℓp's, 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). 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. Indeed for intermediate π this encouragement effect 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).

Proposition 5. *Suppose Conditions (DD) and* (4) *hold. An A policy induces higher investment and payoffs for the underprivileged, and reduces 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 (2013) report that low-income high achievers tend to apply to colleges where the average achievement of students is lower than their own achievement 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.

investment gap between backgrounds. For intermediate π an A policy induces higher investment, and, if δ is high enough, also higher payoffs for both backgrounds.

Figure 12 sums up aggregate outcomes when colleges use tuition fees.

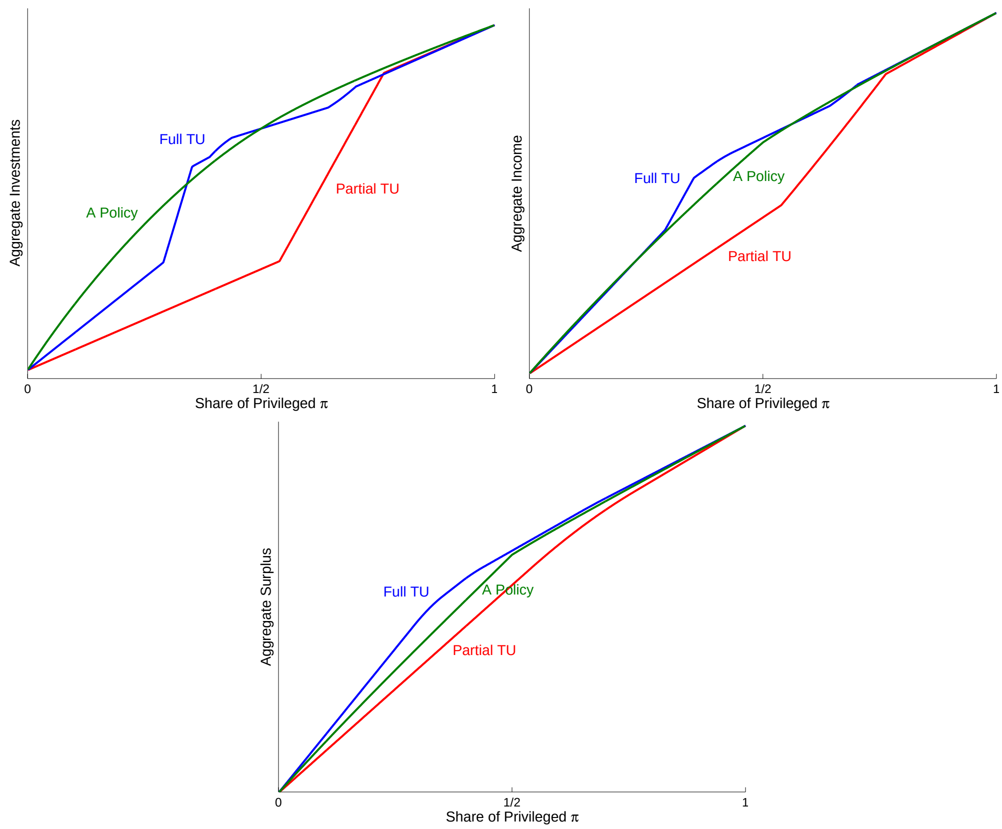

Figure 12: Aggregate investments (left), income (right), and surplus (bottom) when $\omega_p > \delta/2 - \alpha, \omega_u = 0$

Until now, we have considered the possibility of transfers between students who attend the same college, and have shown that an affirmative action policy still has a role to play in generating (hu, hp) colleges, and improving on aggregate variables like output, investment and welfare.

However, because hu students have the right but are not compelled to match with hp students under affirmative action, and because the privileged have wealth with which to make side payments (perhaps intermediated through universities), there may be incentives for hp's to encourage the hu's to match elsewhere, as well as for ℓp's to attract the hu's. This requires some transfers across colleges (from hp's to hu', who would attend (hu, hu) or $(\ell p, hu)$ schools instead of (hu, hp) ones, and the consideration of deviations by coalitions of more than two individuals.[11]

For instance, two hp and ℓp students each, who are matched into (hp, hu) and $(\ell p, \ell u)$ schools, could jointly offer side payments to the two hu students to achieve a rematch of schools $\{(hp, hp), (hu, \ell p), (hu, \ell p), (\ell u, \ell u)\}$. Since ℓu students have no priority in mixed schools nor over h students they do not have to be bought off. hu students would prefer this arrangement, if the side payment exceeds $\delta/2$. An hp student would be prepared to pay at most $1 - \delta$ to obtain an hp match, and ℓp students would pay at most $\delta/2 - \alpha\delta$ to replace their ℓu with an hu match. That is, given an A policy, an outcome that exhausts all (hp, hu) and $(\ell p, \ell u)$ matches *will not* be stable when

$$\delta < \frac{1}{1+\alpha}.$$

Under this condition, an A policy will not lead to (hu, hp) matches but will in fact replicate the free market equilibrium of Figure 11.

But despite the fact that the policy does not seem to have had an effect on *matching*, it still benefits the underprivileged, increasing their incomes, investment incentives and welfare (in fact in our example, the investment incentives of the u's are higher than they would be if the A policy only led to rematch, while the p's have the same investment incentives whether or not the rematch is effected – thus the A policy generates higher aggregate investment than the market outcome whether or not it can be destabilized). Affirmative action may lead to a redistribution of wealth, even if it does not lead to a redistribution of matches.

A second category of diversity policies is the use of scholarships, especially for hu's, financed by private endowments or government funds. These try to generate (hu, hp) matches by giving the hu's sufficient wealth to make the side payment needed to attract an hp (in practice this is a voucher or scholarship, since the wealth given to the hu cannot be spent arbitrarily, and in practice might take the form of reduced or waived tuition along with a living stipend). Observe however, that if the hp with whom the hu is supposed to be paired does not also receive the side payment (perhaps in the form of his own tuition discount), he will not be willing to match with the hu

[11] In practice, such transfers could be effectuated through donations by the hp's (or their parents) to the scholarship funds of the second tier colleges; c.f. the recent controversy over donations by the Koch brothers to the United Negro College fund (http://www.thewire.com/politics/2014/07/major-union-blacklists-united-negro-college-fund-for-koch-brothers-relationship/374264/).

and will instead segregate with another hp. As in the free market outcome, the result is a preponderance of (hp, hp) matches, along with $(\ell p, hu)$. The outcome is the result of market forces among fully informed rational actors, with only borrowing constraints at play.

Some Ivy League universities have expressed consternation at their seeming inability to attract as many underprivileged high achievers as they would like, despite offering generous scholarships to the under-privileged (Hoxby and Avery, 2013). In terms of our model, without transfers to the privileged high achievers, the rational expectation of an hu receiving financial aid to attend such a university is that he will not derive the full benefit of contact with hp's. Insofar as there can be segregation *within* the university, this hu student may prefer a second tier university ((hu, hu) or $(\ell p, hu)$) instead.

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 make transfers, we show that the free market will indeed generate excessive segregation, 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 borrowing, 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. re-matching policies can be applied at the school or neighborhood level as well as at college, and this raises questions of how re-matching 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. Contractual arrangement among the members of a firm are 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: Proofs

A.1 Proofs for Section 3

Proof of Lemma 1

Using (i)-(iii), and reversing the argument (iv), noting that under $\delta > \alpha + \beta$ $(hu, \ell p)$ matches induce higher surplus than the sum of partners' segregation payoffs, the possible stable heterogeneous colleges are (hp, hu), $(hu, \ell p)$, and $(\ell p, \ell u)$ (i.e., all three matches will be formed if the alternative is segregation). Reversing the argument in (v), if $\alpha > 1 - \delta$ having matches (hp, hu) and segregating ℓp induces higher surplus than $(hu, \ell p)$ matches and segregating hp students. Hence, under the condition, (hp, hu) matches are exhausted. Comparing matches $(hu, \ell p)$ and segregating ℓu students, yielding surplus $\delta + \alpha\beta$ to matching $(\ell p, \ell u)$ and segregating hu students, yielding surplus $\alpha\delta + \beta$, the former surplus is higher than the latter if $\delta > \beta$, as assumed.

Proof of Lemma 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

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

The intermediate cases where e_p and e_u are determined by $\pi(1-e_p) = (1-\pi)(1-e_u)$, $\pi e_p = (1-\pi)e_u < \pi$, and $e_u = \pi/(1-\pi)$ are omitted. To summarize, for

- $\pi \leq \frac{1-2\alpha}{2(1-\alpha)}$, $e_p = \delta$.
- $\frac{1-2\alpha}{2(1-\alpha)} < \pi < \frac{2\delta-1-\alpha\beta}{2\delta(1-\alpha)}$ e_p^* strictly decreases,
- $\frac{2\delta-1-\alpha\beta}{2\delta(1-\alpha)} \leq \pi \leq \frac{2(1-\delta)+\alpha\beta}{2(1-\delta+\alpha\delta)}$ e_p reaches a minimum at $e_p = 1 - \alpha(2\delta - \beta)$.
- $\frac{2(1-\delta)+\alpha\beta}{2(1-\delta+\alpha\delta)} < \pi < \frac{2(1-\delta(1-\alpha))-\alpha}{2(1-\delta(1-\alpha))}$ e_p^* strictly increases.
- $\pi \geq \frac{2-2\delta(1-\alpha)-\alpha}{2-2\delta(1-\alpha)}$, $e_p^* = 1 - \alpha$.

Similarly, for

- $\pi \leq \frac{(1-\alpha)\beta}{1+(1-\alpha)\beta}$, $e_u = (1-\alpha)\beta$.
- $\frac{(1-\alpha)\beta}{1+(1-\alpha)\beta} < \pi < \frac{(1-2\alpha)\delta}{1-(1-2\alpha)\delta}$ e_u strictly increases,
- $\frac{(1-2\alpha)\delta}{1-(1-2\alpha)\delta} \leq \pi \leq \frac{1-2\alpha}{2-2\alpha}$, $e_u = (1-2\alpha)\delta$,
- $\frac{1-2\alpha}{2-2\alpha} < \pi < \frac{2\delta-1-\alpha\beta}{2\delta(1-\alpha)}$ e_u strictly increases,
- $\frac{2\delta-1-\alpha\beta}{2\delta(1-\alpha)} \leq \pi \leq \frac{2(1-\delta)+\alpha\beta}{2(1-\delta+\alpha\delta)}$ e_u reaches a maximum at $e_u = 2\delta - 1 - \alpha\beta$,
- $\frac{2(1-\delta)+\alpha\beta}{2(1-\delta+\alpha\delta)} < \pi < \frac{2-2\delta(1-\alpha)-\alpha}{2-2\delta(1-\alpha)}$, $e_u = 1-\alpha)(2\delta-1)$ e_u strictly decreases.
- $\pi \geq \frac{2-2\delta(1-\alpha)-\alpha}{2-2\delta(1-\alpha)}$, $e_u = (1-\alpha)(2\delta-1)$.

Proof of Lemma 3

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 $(ab, a'b')$ schoolss 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) schools, 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 student matched to an hu student into an hp student matched to an hu, and from turning an ℓu student matched to an ℓu student into an hu student matched to an hp, who used to be matched to an hp.

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 $(\ell p,\ell p)$ students 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.

A.2 Proofs for Section 4

Proof of Proposition 2

Recall that in the free market investments are given by $e_p^0 = 1-\alpha$ and $e_u^0 = (1-\alpha)\beta$, and expected payoffs by $y_p^0 = 1-\alpha+\alpha^2$ and $y_u^0 = (1-\alpha)^2\beta^2 + \alpha\beta$.

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

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

Therefore $e_p^B = \delta/2 - \alpha\delta(1 - e_u^B) < \delta/2 < e_p^0$. u agents obtain a p match with probability $\pi/(1-\pi)$, and otherwise the policy allows them to segregate in achievement. Hence:

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

Then $e_u^B = (1-\alpha)\beta + \frac{\pi}{1-\pi}(\delta(1/2 - \alpha(1 - e_p^B)) - (1-\alpha)\beta) < (1-\alpha)\beta = e_u^0$.

$$\begin{aligned} e_u^B &= \frac{(1-2\pi)(1-\alpha)\beta + \pi\delta(1/2-\alpha)(1+\alpha\delta)}{1-\pi-\pi\alpha^2\delta^2} \\ e_p^B &= \frac{(1-2\pi)(1-\alpha)\alpha\delta + \delta(1/2-\alpha)(1-\pi+\pi\alpha\delta)}{1-\pi-\pi\alpha^2\delta^2}. \end{aligned}$$

Hence, e_u^B and e_p^B are decreasing in $\pi \in [0, 1/2]$. Investment inequality under the B policy is lower if $e_p^0/e_v^0 > e_p^B/e_u^B$, which must be true for $\pi < 1/2$ since $e_u^B \geq e_p^B$ (to see this note that $e_u^B = e_p^B$ for $\pi = 1/2$ and e_u^B decreases faster in π than e_p^B). Expected payoff can be written as $y_b = e_b^2 + w(\ell b)$, yielding:

$$\begin{aligned} y_p^B &= (e_p^B)^2 + \left(\frac{\delta}{2} - \alpha\delta\right) e_u^B + \alpha\delta \text{ and} \\ y_u^B &= (e_u^B)^2 + \frac{\pi}{1-\pi}\left(\frac{\delta}{2} - \alpha\delta\right) e_p^B + \frac{\pi}{1-\pi}\alpha\delta + \frac{1-2\pi}{1-\pi}\alpha\beta. \end{aligned}$$

Indeed $y_p^B = y_u^B$ if $\pi = 1/2$. Payoff inequality under the B policy is lower if $y_p^0/y_u^0 > y_p^B/y_u^B$. A sufficient condition is:

$$\frac{(e_p^B)^2 + \left(\frac{\delta}{2} - \alpha\delta\right) e_u^B + \alpha\delta}{(e_u^B)^2 + \frac{\pi}{1-\pi}\left(\frac{\delta}{2} - \alpha\delta\right) e_p^B + \frac{\pi}{1-\pi}\alpha\delta + \frac{1-2\pi}{1-\pi}\alpha\beta} < \frac{1}{\beta} < \frac{(1-\alpha)^2 + \alpha}{(1-\alpha)^2\beta^2 + \alpha\beta},$$

which can be shown (by using the facts that the condition slackens in e_u and $e_u^B \geq e_p^B$) to hold for $\pi < 1/2$. As for payoffs, $y_p^B < y_p^0 = 1 - \alpha + \alpha^2$ using that $e_u^B \leq (1-\alpha)\beta$, $e_p^B \leq \delta/2$, and that $\alpha < \delta/2 < \beta < \delta$. Aggregate payoff under a B policy is given by:

$$y^B = \pi(e_p^B)^2 + (1-\pi)(e_u^B)^2 + \pi\left(\frac{\delta}{2} - \alpha\delta\right)(e_u^B + e_p^B) + 2\pi\alpha\delta + (1-2\pi)\alpha\beta.$$

Using that $e_p^B < \delta/2$, $e_u^B \leq (1-\alpha)\beta$, and $\alpha < \delta/2$, it can be verified that $y^B < \pi(1-\alpha)^2(1-\beta^2) + \alpha(1-\beta)] + (1-\alpha)\beta^2 + \alpha\beta = y^0$.

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

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

Therefore $e_u^B = \delta/2 - \alpha\delta(1 - e_p^B) \leq \delta(1/2 - \alpha^2)$. $e_u^B < e_u^0$ if $\beta > \delta(1/2 - \alpha^2)/(1-\alpha)$. A sufficient condition is $\beta > (2-\sqrt{2})\delta$. p agents obtain a p match with probability $(2\pi - 1)/\pi$, in which case the policy allows them to segregate in achievement. Hence:

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

Therefore:

$$e_p^B = \frac{1-\pi}{\pi}\left(\frac{\delta}{2} - \alpha\delta(1 - e_u^B)\right) + \frac{2\pi-1}{\pi}(1-\alpha) < e_p^0 \text{ for } \pi < 1.$$

$$\begin{aligned} e_p^B &= \frac{(1-\pi)\left(\frac{\delta}{2} - \alpha\delta\right)(1 + \alpha\delta) + (2\pi - 1)(1-\alpha)}{\pi - (1-\pi)\alpha^2\delta^2} \text{ and} \\ e_u^B &= \frac{\pi\left(\frac{\delta}{2} - \alpha\delta\right) + (1-\pi)\alpha\delta\left(\frac{\delta}{2} - \alpha\delta\right) + \alpha\delta(1-\alpha)(2\pi - 1)}{\pi - (1-\pi)\alpha^2\delta^2}. \end{aligned}$$

Computing e_p^B and e_u^B reveals that $e_p^B > e_u^B$ for $\pi > 1/2$, and that $\frac{\partial e_p^B}{\partial \pi} > \frac{\partial e_u^B}{\partial \pi}$. Hence, investment inequality e_p^B / e_u^B is 1 for $\pi = 1/2$ from above, and strictly increases to reach its maximum at $\pi = 1$, where $e_p^B = 1 - \alpha$ and $e_u^B = \delta/2 - \alpha^2\delta$. Therefore, if $(1-\alpha)\beta < \delta(1/2 - \alpha^2)$ investment inequality is strictly lower under the B policy, and otherwise there is $\hat{\pi} > 1/2$ such that investment inequality is strictly lower under the B policy for $\pi < \hat{\pi}$ and strictly higher for $\pi > \hat{\pi}$.

Payoffs are given by:

$$y_p^B = (e_p^B)^2 + \frac{1-\pi}{\pi}\left(\left(\frac{\delta}{2} - \alpha\delta\right) e_u^B + \alpha\delta\right) + \frac{2\pi - 1}{\pi}\alpha \text{ and}$$
$$y_u^B = (e_u^B)^2 + \alpha\delta + \left(\frac{\delta}{2} - \alpha\delta\right) e_p^B.$$

Indeed $y_u^B > y_u^0$ for $\pi \in [1/2, 1]$, and if α sufficiently close to $\delta/2$ then $y_p^B < y_p^0$ for $\pi \in [1/2, 1)$, implying that indeed $y_p^0/v_u^0 > v_p^B/v_u^B$. Aggregate payoff is greater under a B policy if:

$$\pi(e_p^B)^2 + (1-\pi)(e_u^B)^2 + (1-\pi)\left(\left(\frac{\delta}{2} - \alpha\delta\right)(e_p^B + e_u^B) + 2\alpha\delta\right)$$
$$> \pi(1-\alpha)^2 + (1-\pi)(1-\alpha)\beta^2 + \pi\alpha + (1-\pi)\alpha\beta.$$

It can be verified that in the neighborhood of $\pi = 1$ the LHS increases faster in π than does the RHS. Since $y^B = y^0$ for $\pi = 1$, $y^0 > y^B$ for $\pi < 1$.

Proof of Proposition 3

Students' payoffs depend on relative scarcity:

$$v(hp) = \begin{cases} \delta & \text{if } \pi e_p \leq (1-\pi)e_u \\ 1 - \frac{1-\pi}{\pi}\frac{e_u}{e_p}(1-\delta) & \text{otherwise.} \end{cases}$$
$$v(\ell p) = \begin{cases} \alpha\delta & \text{if } \pi(1-e_p) \leq (1-\pi)(1-e_u) \\ \alpha - \frac{1-\pi}{\pi}\frac{1-e_u}{1-e_p}\alpha(1-\delta) & \text{otherwise.} \end{cases}$$
$$v(hu) = \begin{cases} \delta & \text{if } \pi e_p \geq (1-\pi)e_u \\ \beta + \frac{\pi}{1-\pi}\frac{e_p}{e_u}(\delta-\beta) & \text{otherwise.} \end{cases}$$
$$v(\ell u) = \begin{cases} \alpha\delta & \text{if } \pi(1-e_p) \geq (1-\pi)(1-e_u) \\ \alpha\beta + \frac{\pi}{1-\pi}\frac{1-e_p}{1-e_u}\alpha(\delta-\beta) & \text{otherwise.} \end{cases}$$

That is, four distinct cases may potentially arise. Turn to the case $\pi e_p < (1-\pi)e_u$ and $\pi(1-e_p) > (1-\pi)(1-e_u)$ first. Investments are

$$e_p = \delta - \alpha + \frac{1-\pi}{\pi}\frac{1-e_u}{1-e_p}(1-\delta)\alpha \text{ and}$$
$$e_u = \beta - \alpha\delta + \frac{\pi}{1-\pi}\frac{e_p}{e_u}(\delta-\beta).$$

This case cannot occur, however. To see this note first that for $\pi = 1/2$ $e_p = e_u = (1-\alpha)\delta$. Using the total differential reveals that e_p decreases in π while e_u increases in π. This would imply that $e_p > e_u$ for $\pi < 1/2$, a contradiction to $\pi(1-e_p) > (1-\pi)(1-e_u)$ for $\pi < 1/2$. Computing the total differential in the neighborhood of $\pi = 1/2$ yields a contradiction to $\pi e_p < (1-\pi)e_u$ for $\pi > 1/2$, as e_p given above does not decrease fast enough to keep the condition satisfied. The case $\pi e_p > (1-\pi)e_u$ and $\pi(1-e_p) < (1-\pi)(1-e_u)$ can be discarded using an analogous argument.

This leaves us with the 'symmetric' cases. Consider first $\pi e_p \leq (1-\pi)e_u$ and $\pi(1-e_p) \leq (1-\pi)(1-e_u)$, yielding:

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

This implies that $e_p^A \geq e_u^A$, so that this case requires $\pi \leq 1/2$. Clearly, $e_p^A \leq 1-\alpha = e_p^0$ and $e_u^A \geq (1-\alpha)\beta = e_u^0$. Therefore investment inequality is lower under an A policy. Turning to payoffs, $y_p^A = (1-\alpha)^2\delta^2 + \alpha\delta$ and $y_u^A \geq (1-\alpha)^2\beta^2 + \alpha\beta$. That is, $y_p^A < (1-\alpha)^2 + \alpha = y_p^0$ and $y_u^A \geq (1-\alpha)^2\beta^2 + \alpha\beta = y_u^0$ with strict inequality for $\pi > 0$. Hence, payoff inequality is lower under the A policy as well. The difference in aggregate payoff between an A policy and the free market is:

$$\begin{aligned}\Delta =& \pi[(e_p^A)^2 - (1-\alpha)^2] + (1-\pi)[(e_u^A)^2 - (1-\alpha)^2\beta^2] \\ &+ \pi[v(\ell p) - \alpha] + (1-\pi)[v(\ell u) - \alpha\beta].\end{aligned}$$

Both e_u^A and e_p^A strictly increase in δ for $\pi \in (0, 1/2)$. The latter follows because $1 - e_p^A < 1 - e_u^A$ and $2\delta > \beta + 1 > \beta + 2\alpha\delta$. Since also $\pi v(\ell p) + (1-\pi)v(\ell u)$ increases in δ, this implies that also the difference Δ increases in δ. Therefore, for any $0 \leq \pi < 1/2$ there is $\hat{\delta}(\pi) < 1$ such that $\delta > \hat{\delta}(\pi)$ implies $y^A > y^0$.

For $\pi e_p > (1-\pi)e_u$ and $\pi(1-e_p) > (1-\pi)(1-e_u)$:

$$e_p^A = 1 - \alpha - (1-\delta)\frac{1-\pi}{\pi}\left(\frac{e_u}{e_p} - \alpha\frac{1-e_u}{1-e_p}\right) \text{ and } e_u^A = (1-\alpha)\delta$$

Again $e_p^A \geq e_u^A$ is implied, so that this case requires $\pi \geq 1/2$. Again $e_p^A \leq e_p^0$ and $e_u^A \geq e_u^0$ and investment inequality is lower under an A policy. Moreover, $e_u^A > \delta(1/2 - \alpha^2) \geq e_u^B$. Comparing aggregate investment and using the fact that $2\delta > 1+\beta$ yields $\pi e_p^A + (1-\pi)e_u^A > \pi e_p^0 + (1-\pi)e_u^0$ for $0 < \pi \leq 1/2$. Payoffs are $y_u^A = (1-\alpha)^2\delta^2 + \alpha\delta > y_u^0$ and $y_p^A \leq (1-\alpha)^2 + \alpha = y_p^0$, so that payoff inequality is lower under the A policy. The difference of aggregate payoffs $y^A - y^0$ is now given by:

$$\begin{aligned}\Delta =&\pi[(e_p^A)^2 - (1-\alpha)^2] + (1-\pi)[(e_u^a)^2 - (1-\alpha)^2\beta^2] \\ &+ \pi[v(\ell p) - \alpha] + (1-\pi)[v(\ell u) - \alpha\beta].\end{aligned}$$

Both e_u^A and e_p^A strictly increase in δ for $\pi \in (1/2, 1)$, the latter follows because $(1-\alpha)\delta \leq e_p^A \leq 1-\alpha$ and $\delta \geq 1-\alpha$. This implies that also $(e_p^A)^2 + v(\ell p)$ increase in δ and thus the difference Δ increases in δ. Therefore, for any $1/2 \leq \pi < 1$ there is $\hat{\delta}(\pi) < 1$ such that $\delta > \hat{\delta}(\pi)$ implies $y^A > y^0$.

For instance, $\hat{\delta}(1/2)$ is implicitly defined by

$$(1-\alpha^2)(2\hat{\delta}(1/2)^2 - 1 - \beta^2) + \alpha(2\hat{\delta}(1/2) - 1 - \beta) = 0.$$

Proof of Proposition 4

Recall that $e_p^0 = 1$ and $e_u^0 = \beta$. The difference in surplus between an A policy and the free market is therefore:

$$\begin{aligned}\Delta S =&\pi\frac{(e_p^A)^2 - (1-\alpha)^2}{2} + (1-\pi)\frac{(e_u^A)^2 - (1-\alpha)^2\beta^2}{2} \\ &+ \pi[v(\ell p) - \alpha] + (1-\pi)[v(\ell u) - \alpha\beta].\end{aligned}$$

As shown in the proof of Proposition 3, both e_p^A and e_u^A weakly increase in δ, and one of them strictly. Similar to above both $(e_p^A)^2/2 + v(\ell p)$ and $\pi v(\ell p) + (1-\pi)v(\ell u)$ increase in δ, which implies that ΔS strictly increases in δ. $\Delta S > 0$ for $\delta = 1$. Hence, there is $\hat{\delta}(\pi) < 1$ such that $\delta > \hat{\delta}(\pi)$ implies $S^A > S^0$.

For instance, $\hat{\delta}(1/2)$ is implicitly defined by

$$(1-\alpha^2)(2\hat{\delta}(1/2)^2 - 1 - \beta^2) + 2\alpha(2\hat{\delta}(1/2) - 1 - \beta) = 0.$$

Hence, for $\pi = 1/2$ $y^A > y^0$ implies $S^A > S^0$.

To compare A and B policies, $S^A > S^B$ holds if, and only if:

$$\pi\frac{(e_p^A)^2 - (e_p^B)^2}{2} + (1-\pi)\frac{(e_u^A)^2 - (e_u^B)^2}{2}$$
$$+ \pi(v^A(\ell p) - v^B(\ell p)) + (1-\pi)(v^A(\ell u) - v^B(\ell u)) > 0. \tag{A.1}$$

If $\pi \leq 1/2$, $e_u^A - e_u^B \geq \frac{\pi}{1-\pi}[\delta(1-\alpha) - \delta(1/2 + \alpha\delta/2 - \alpha)]$, and $e_u^A + e_u^B \geq (1-\alpha)\beta + \delta\frac{1/2-\alpha}{1-\alpha\delta}$, and $e_p^A - e_p^B \geq (1-\alpha)\delta - \delta/2$ and $e_p^A + e_p^B \geq (1-\alpha)\delta + \delta\frac{1/2-\alpha}{1-\alpha\delta}$. Moreover, $v^A(\ell p) - v^B(\ell p) \geq \alpha(1-\delta) - \delta(1/2-\alpha)(1-\alpha)\beta$ and $v^A(\ell u) - v^B(\ell u) = \frac{\pi}{1-\pi}\left(\left(\frac{1-e_p^A}{1-e_u^A} - 1\right)\alpha(\delta - \beta) - \delta(1/2-\alpha)e_p^B\right) \geq -\frac{\pi}{1-\pi}\delta^2(1/4 - \alpha/2)$. All these observations imply that inequality (A.1) holds indeed, and thus $S^A > S^B$ for $\pi \in [0, 1/2]$.

If $\pi \geq 1/2$, for the underprivileged $e_u^A - e_u^B \geq \delta(1 - \alpha - 1/2 + \alpha^2)$ and $e_u^A + e_u^B \geq (1-\alpha)\delta + \delta(1/2-\alpha) \geq 2\delta/3 \geq 1$, and $v^A(\ell u) - v^B(\ell u) \geq -(\delta/2 - \alpha\delta)(1-\alpha)$. For the privileged, $e_p^A - e_p^B \geq \frac{1-\pi}{\pi}\left(\delta(1-\alpha) - -\delta/2 + \alpha\delta - \alpha\delta^2(1/2 - e_u^B)\right) \geq \frac{1-\pi}{\pi}\left(\delta/2 - \alpha\delta^2(1/2 - \alpha^2)\right)$, and $e_p^A + e_p^B \geq (1-\alpha)\delta + \delta(1/2-\alpha)/(1-\alpha\delta)$, and wages are $v^A(\ell p) - v^B(\ell p) \geq -\frac{1-\pi}{\pi}(1/2-\alpha)(1/2-\alpha^2)\delta^2$. Again all these observations imply that inequality (A.1) holds, and thus $S^A > S^B$ for $\pi \in [1/2, 1]$.

Second Best Policy

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

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

Since $\sum \rho(hp,.) = \pi e_p$ and similarly for the other attributes this leaves six choice variables.

We solved the problem numerically and Figure 13 shows the second best optimal matching for the parametrization used to generate all the figures ($\delta = .85$, $\beta = .5$, $\alpha = .18$). The broken lines correspond to matching probabilities under an A policy for comparison. That is, an A policy is indeed very close to second best for this particular parametrization when

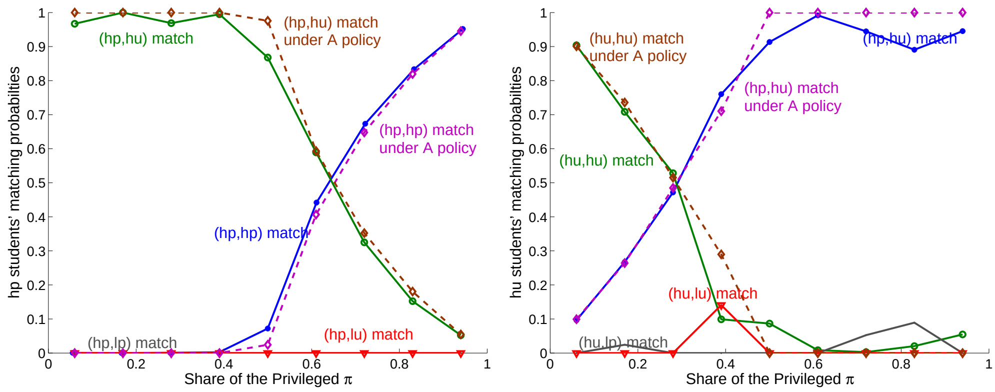

Figure 13: hp (left) and hu (right) students' matching probabilities in the second best.

$\pi \geq 1/2$.[12] Comparing surplus values to those under an A policy and free market yields the numbers in the text.

Naive Policy

A naive policy replicates the TU outcome, exhausting all possible (hu, hp) matches first, then all $(hu, \ell p)$ and then all $(\ell p, \ell u)$ matches, while all remaining students segregate. Again several cases may occur.

(i) $\pi e_p \geq (1-\pi)e_u$: in this case only (hp, hu) and $(\ell p, \ell u)$ form, and the outcome coincides with that of that an A policy. This regime occurs for $\pi \geq 1/2$

(ii) $\pi e_p < (1-\pi)e_u < 1-\pi < \pi$: in this case (hp, hu), $(hu, \ell p)$, and $(\ell p, \ell u)$ matches occurs and some ℓp students segregate. $v(hp) = \delta$ and $v(\ell u) = \alpha\delta$. $v(hu) = \frac{\pi e_p}{(1-\pi)e_u}\frac{\delta}{2} + \frac{\delta}{2}$, and

$$e_u = \frac{\pi e_p}{(1-\pi)e_u}\frac{\delta}{2} + \delta\left(\frac{1}{2} - \alpha\right) \leq (1-\alpha)\delta.$$

ℓp students obtain

$$v(\ell p) = \frac{\delta}{2} - \frac{2\pi - 1}{\pi(1-e_p)}\left(\frac{\delta}{2} - \alpha\right) - \frac{(1-\pi)(1-e_u)}{\pi(1-e_p)}\delta\left(\frac{1}{2} - \alpha\right),$$

[12]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.

and thus

$$e_p = \frac{\delta}{2} + \frac{2\pi - 1}{\pi(1-e_p)}\left(\frac{\delta}{2} - \alpha\right) - \frac{(1-\pi)(1-e_u)}{\pi(1-e_p)}\delta\left(\frac{1}{2} - \alpha\right).$$

For $\pi = 1/2$, $e_u = e_p = \delta(1-\alpha)$. As π increases beyond $1/2$, e_p decreases less than e_u, implying $e_p > e_u$, a contradiction to $\pi > 1-\pi$ and $\pi e_p < (1-\pi)e_u$.

(iii) $\pi e_p < (1-\pi)e_u < \pi < 1-\pi$: then (hp, hu), $(hu, \ell p)$, and $(\ell p, \ell u)$ matches form and some ℓu students segregate. $v(hp) = \delta$ and $v(\ell u) = \alpha\delta - \frac{1-2\pi}{(1-\pi)(1-e_u)}\alpha(\delta-\beta)$. $v(hu) = \frac{\pi e_p}{(1-\pi)e_u}\frac{\delta}{2}+\frac{\delta}{2}$, and $v(\ell p) = \alpha\delta + \frac{(1-\pi)e_u - \pi e_p}{\pi(1-e_p)}\left(\frac{\delta}{2} - \alpha\delta\right)$. Investments are

$$e_p = (1-\alpha)\delta - \frac{(1-\pi)e_u - \pi e_p}{\pi(1-e_p)}\left(\frac{\delta}{2} - \alpha\delta\right) < (1-\alpha)\delta = e_p^A,$$

and

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

$e_u^A > e_u^N$ for $\pi < 1/2$ follows, since assuming the contrary and computing the difference would yield $e_u^A - e_u^N > 0$, contradiction.

(iv) $\pi < (1-\pi)e_u$: in this case (hp, hu) and $(hu, \ell p)$ matches occur and some hu as well as all ℓu students segregate. Thus $v(hp) = \delta$, $v(\ell p) = \delta/2$, $v(\ell u) = \alpha\beta$, and $v(hu) = \frac{\pi e_p \delta/2 + \pi\delta/2}{(1-\pi)e_u} + (1 - \frac{\pi}{(1-\pi)e_u})\beta$. Therefore, $e_p = \delta/2$ and

$$e_u = (1-\alpha)\beta + \frac{\pi}{1-\pi}\left(\frac{e_p+1}{e_u}\frac{\delta}{2} - \frac{\beta}{e_u}\right) < (1-\alpha)\beta + \frac{\delta^2}{4} + \frac{\delta}{2} - \beta.$$

A.3 Proofs for Section 5

Proof of Lemma 6

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 ℓu 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, β and α, respectively. This establishes the lemma.

Proof of Lemma 7

hp students cannot be compensated by a side payment from any ℓ student. Hence, they match with hu, who have priority, or segregate, which they prefer. An hu obtains δ when matched with an hp and β when segregated. An ℓp gets α when segregated and $\alpha\delta$ when matched to an ℓu. Hence, an ℓp student would be prepared to pay at most $\delta/2 - \alpha\delta$ to an hu to avoid an ℓu, and $\delta/2 - \alpha$ to avoid an ℓp student. Hence, as $\delta > \delta - \alpha\delta$, an hu student will always prefer an hp match to an ℓp match. Any hu not matched to an hp will prefer an ℓp over an hu, since $\beta < \delta - \alpha$. Finally, under an A policy ℓu students cannot block $(hu, \ell p)$, but only $(\ell p, \ell p)$ pairs, which they prefer since $\alpha\delta > \alpha\beta$. The statement follows.

Proof of Proposition 5

We first derive the competitive equilibrium. Schools compete for students and earn zero profits, therefore $v(\ell u) = \alpha\beta$, $v(hp) = 1$, and $t(\ell p, hu) = -t(hu, \ell p) \in [\beta - \delta/2, \delta/2 - \alpha]$ is determined by the relative scarcity of attributes hu and ℓp. Agents' investments are given by $e_u^C = \delta/2 + t(\ell p, hu) - \alpha\beta$ and $e_p^C = 1 - \delta/2 + t(\ell p, hu)$.

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

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

Finally, if $\pi(1 - e_p^C) > (1 - \pi)e_u^C$, $t(\ell p, hu) = \delta/2 - \alpha$. Then $e_u^C = \delta - (1 + \beta)\alpha$ and $e_p^C = 1 - \alpha$. This regime occurs if $\pi > 1 - \frac{\alpha}{\delta - \alpha\beta}$. $v(\ell p) = \alpha$.

Note that $e_p^C / e_u^C \geq (1 - \alpha)/(\delta - (1 + \beta)\alpha)$, since both e_u^C and e_p^C increase in π at the same rate $\delta(1 - \alpha)$.

A **Policy**

Start with the case $\pi < (1 - \pi)e_u$, i.e., all p students are matched to hu students. This means all ℓu students segregate. Hence, $v(\ell u) = \alpha\beta$, $v(hp) = \delta$, and $v(\ell p) = \delta/2 - (\beta - \delta/2)$, which implies $e_p = \beta$. $v(hu) = \frac{\pi e_p}{(1-\pi)e_u}(\delta - \beta) + \beta$, so that

$$e_u = (1 - \alpha)\beta + (\delta - \beta)\frac{\pi\beta}{(1 - \pi)e_u}.$$

Solving for e_u yields $e_u = (1-\alpha)\beta/2 + \sqrt{(1-\alpha)^2\beta^2 + 4\frac{\pi}{1-\pi}\beta(\delta-\beta)}/2$, that is $e_u \in [(1-\alpha)\beta, (1-\alpha+\delta-\beta)\beta]$ and strictly increases in π. This regime occurs if $\pi \leq (1-\alpha+\delta-\beta)\beta/(1+(1-\alpha+\delta-\beta)\beta) < (1-\alpha)\beta/(\delta-\alpha\beta)$. Clearly $e_u^A > e_u^C = (1-\alpha)\delta$ for $\pi > 0$. Since $e_p^A = \beta < 1+\beta-\delta = e_p^C$, investment inequality is lower under an A policy, too.

For $\pi = (1-\pi)e_u$, since $e_p = \delta - v(\ell p)$, $v(\ell p)$ has to adjust by adjusting the transfer an ℓp pays to an hu. Since $e_u = v(hp) - \alpha\beta$,

$$\begin{aligned} e_u &= \frac{\pi e_p}{(1-\pi)e_u}(\delta - \delta + v(\ell p)) + \delta - v(\ell p) \\ &= \frac{\pi e_p}{(1-\pi)e_u}(\delta - e_p) + e_p = \frac{\pi}{1-\pi}. \end{aligned}$$

This last equation pins down e_p. Solving for e_p and noting that both $v(\ell p) \geq \alpha\delta$ and $e_p \leq (1-\alpha)\delta$ have to hold, this regime will occur if $(1-\alpha+\delta-\beta)\beta/(1+(1-\alpha+\delta-\beta)\beta) < \pi \leq \frac{\frac{(1+\delta)^2}{4}-\alpha\beta-\frac{(1-\delta+2\alpha\delta)^2}{4}}{1+\frac{(1+\delta)^2}{4}-\alpha\beta-\frac{(1-\delta+2\alpha\delta)^2}{4}}$. In this case $e_u^A = \pi/(1-\pi) > e_u^C$. $e_p/e_u \leq (1-\alpha)\delta(1-\pi)/\pi \leq (1-\alpha)\delta/(\beta(1+\delta-\alpha-\beta) < (1-\alpha)/(\delta-(1+\beta)\alpha) \leq e_p^C/e_u^C$.

Otherwise, $\pi > (1-\pi)e_u > \pi e_p$, then $v(hp) = \delta$. Since the case $\pi > 1/2$ can quickly be brought to a contradiction, suppose that $\pi < 1/2$. Then all ℓp are matched, to ℓu or hu. Hence, $v(\ell p) = \alpha\delta$ and therefore $e_p = (1-\alpha)\delta$. $v(hu) = \frac{\pi e_p}{(1-\pi)e_u}\alpha\delta + (1-\alpha)\delta$ and $v(\ell u) = \alpha\delta - \frac{1-2\pi}{(1-\pi)(1-e_u)}\alpha(\delta-\beta)$. Therefore

$$e_u = (1-\alpha)\delta - \alpha\delta\left(1 - \frac{\pi(1-\alpha)\delta}{(1-\pi)e_u}\right) + \alpha(\delta-\beta)\frac{1-2\pi}{(1-\pi)(1-e_u)}.$$

Now $e_u \in [(1-\alpha)\delta, \frac{(1+\delta)^2}{4} - \alpha\beta - \frac{(1-\delta+2\alpha\delta)^2}{4}]$. Since both boundaries exceed $\delta-(1+\beta)\alpha$, $e_u^A > e_u^C$. $e_p^A/e_u^A \leq \max\{1; ((1-\alpha)\delta)/(\frac{(1+\delta)^2}{4}-\alpha\beta-\frac{(1-\delta+2\alpha\delta)^2}{4})\} < (1-\alpha)/(\delta-(1+\beta)\alpha) \leq e_p^C/e_u^C$.

For $\pi > 1/2$ there are more p students than u students. If $\pi e_p < (1-\pi)e_u$, $v(hp) = \delta$, all remaining hu match with ℓp students, all ℓu match with ℓp students, and some ℓp segregate. Therefore $v(\ell u) = \alpha\delta$. Since ℓp need to obtain α in a $(hu, \ell p)$ match, $v(hu) = \frac{\pi e_p}{(1-\pi)e_u}\alpha + \delta - \alpha$ and:

$$e_u = \frac{\pi e_p}{(1-\pi)e_u}\alpha + (1-\alpha)\delta - \alpha.$$

On the other hand, $v(\ell p) = \alpha - \frac{(1-\pi)(1-e_u)}{\pi(1-e_p)}(1-\delta)\alpha$, so that

$$e_p = \delta - \alpha + \frac{(1-\pi)(1-e_u)}{\pi(1-e_p)}(1-\delta)\alpha.$$

These two equations are not compatible with $e_p < e_u$, implied by our assumption, $\pi e_p \leq (1-\pi)e_u$, a contradiction. Suppose therefore that $\pi e_p > (1-\pi)e_u$. Then all (hu, hp) and all $(\ell p, \ell u)$ matches are exhausted, while both some hp and some ℓu or ℓp students segregate. No side payments are made. Hence, using the proof of Proposition 3:

$$e_p = 1 - \alpha - (1-\delta)\frac{1-\pi}{\pi}\left(\frac{e_u}{e_p} - \alpha\frac{1-e_u}{1-e_p}\right) \text{ and } e_u = (1-\alpha)\delta.$$

Using the results above, for $\pi = 1/2$, $e_u = (1-\alpha)\delta = e_u$. $e_u = (1-\alpha)\delta > \delta - (1+\beta)\alpha$ implies also that $e_u > e_u^C$ for $\pi \geq 1/2$. $e_p < 1-\alpha$ implies $e_p/e_u < 1/\delta$, and $1/\delta < (1-\alpha)/(\delta-(1+\beta)\alpha) \leq e_p^C/e_u^C$.

Since ℓu students have at least payoff $\alpha\beta$ under the A policy and exactly $\alpha\beta$ in the market, $e_u^A > e_u^C$ implies also that income and surplus of the underprivileged is higher under an A policy, $S_u^A > S_u^C$ and $Y_u^A > Y_u^C$. A sufficient condition for $e_p^A > e_p^C$ is that $\frac{\frac{(1+\delta)^2}{4} - \alpha\beta - \frac{(1-\delta+2\alpha\delta)^2}{4}}{1+\frac{(1+\delta)^2}{4} - \alpha\beta - \frac{(1-\delta+2\alpha\delta)^2}{4}} \leq \pi \leq 2 - 1/(\delta(1-\alpha))$ (the upper bound strictly exceeds the lower bound under our assumptions).

For the last last statement suppose that $\delta \geq \frac{2}{3}\frac{1}{1-\alpha}$, then $2 - 1/(\delta(1-\alpha)) \geq 1/2$. Suppose $\pi = 1/2$, then

$$\begin{aligned} S^A - S^C =& \frac{1}{2}\left(\frac{(1-\alpha)^2\delta^2 - (1-(1-\alpha)\delta)^2}{2} + \frac{(1-\alpha)^2\delta^2 - (1-\alpha)^2\delta^2/4}{2}\right) \\ &+ \frac{1}{2}\left(2\alpha\delta - \alpha\beta - \frac{1}{2}(\delta - \alpha\beta)\right). \end{aligned}$$

This difference is indeed strictly positive for δ close enough to 1. Since investments and payoffs are continuous in π, and $S_u^A \geq S_u^C$, indeed an A policy yields higher aggregate surplus for intermediate π.

Comparing expected payoffs for the privileged, an A policy induces higher payoffs for $\pi = 1/2$ if:

$$Y_p^A - Y_p^C = \frac{1}{2}\left((1-\alpha)^2\delta^2 - (1-(1-\alpha)\delta)^2 + \alpha\delta - \frac{1}{2}(\delta - \alpha\beta)\right) > 0.$$

Using $\delta \geq \frac{2}{3}\frac{1}{1-\alpha}$, the condition will hold and income among the privileged will be strictly higher under an A policy for $\pi = 1/2$, and by continuity also in a neighborhood.

B 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 $s' > s''$).[13] Assume that $z(hp, hp) < 2$ to permit easy interpretation of investments as probabilities. The functional 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.

[13] 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.

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 $e_u^T > e_u^*$. $\square$

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:

$$e_p^0 = \frac{z(hp, hp) - z(\ell p, \ell p)}{2} \text{ and } e_u^0 = \frac{z(hu, hu) - z(\ell u, \ell u)}{2}.$$

Total surplus under free market is

$$\begin{aligned} S^0 = \pi \frac{(z(hp, hp) - z(\ell p, \ell p))^2}{8} + \pi \frac{z(\ell p, \ell p)}{2} \\ + (1-\pi)\frac{(z(hu, hu) - z(\ell u, \ell u))^2}{8} + (1-\pi)\frac{z(\ell u, \ell u)}{2}. \end{aligned}$$

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

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

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

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

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, $S^0 < S^A$ 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, $S^0 < S^A$ if

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

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 $z(hp,hp)$. $\square$

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.

Bhaskar, V. and Hopkins, E. (2014), 'Marriage as a Rat Race: Noisy Pre-Marital Investments with Assortative Matching', mimeo, University of Edimburgh.

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., 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.(2013), "Low-Income high-achieving students miss out on attending selective colleges," *Brookings Papers on Economic Activity*, Spring 2013.

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.

Kaneko, M. and Wooders, M. H. (1989)The core of a continuum economy with widespread externalities and finite coalitions: from finite to continuum economies. *Journal of Economic Theory* **49**, 135-168.

Lang, K. and Lehmann, J. (2011), 'Racial discrimination in the labor market: theory and empirics', *Journal of Economic Literature*, **50**(4), 959-1006.

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.

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

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